Department of Electronic & Computer Engineering of

Technical University of Crete, Greece

Identification of the EGFR Signaling Pathway using an Asymptotic Observer based on the Rise Approach (Robust Integral of the Sign Error)

Diploma Thesis in Systems Biology

Chatzidamianou Nikolina

Panagiotidou Agapi

Supervisor:

Prof. E. Christodoulou

Examining Committee:

Prof. P. Stavroulakis

Prof. G. Mpoutalis

Acknowledgements

We would like to express our deep appreciation and sincere gratitude to Professor M.Christodoulou, our supervisor, for his wisdom, invaluable guidance and professionalism from the beginning to the end in the course of our Diploma Thesis. We also wish to thank Thomas Giotis for his patience and his invaluable help in every step of the process of this thesis.

Contents

Chapter 1

Introduction

Systems biology is a new and fast-developing field of biology aimed at system-level understanding of biological systems and provides information useful for molecular biology and medicine. System-level understanding is essential to have a set of principles and methodologies that connects the behavior of molecules with the system features and functions. In this thesis, we will simulate and identify the EGFR (Epidermal Growth Factor Receptor) signaling pathway using an asymptotic observer based on the rise approach.

Starting this thesis, in Chapter 2, we make an introduction to molecular biology and its most basic concepts. Cell, DNA and proteins are presented in order to introduce biological issues that will help the reader to comprehend better our work. We also make an introduction to the field of Systems Biology, relevant to the application of systems theory to biology.

In Chapter 3, we introduce the term of protein tyrosine kinase and protein tyrosine kinase receptor where EGFR (Epidermal Growth Factor Receptor) belongs in this family. We analyze the specific signaling pathway called EGFR PATHWAY which is composed of twenty three proteins and their interactions. We also present a complete set of chemical kinetic equations describing the reactions of EGFR pathway proposed by Kholodenko.

Artificial neural networks have been analyzed in Chapter 4, as well as, the training of them. We present the construction of a computational model which predicts the dynamic behavior of each protein in the EGFR PATHWAY. The mathematical tool used is an Asymptotic Observer based on the Rise Approach (Robust Integral of the Sign Error).

In Chapter 5, we analyze the implementation of the Rise Observer in the EGFR model, we introduce the Matlab tool Simulink and we present the results of the identification of the EGFR signaling pathway.

We complete, in Chapter 6, with the conclusions and possible future work in the field of Systems Biology.

Chapter 2

A GENERAL OVERVIEW IN SYSTEMS AND MOLECULAR BIOLOGY

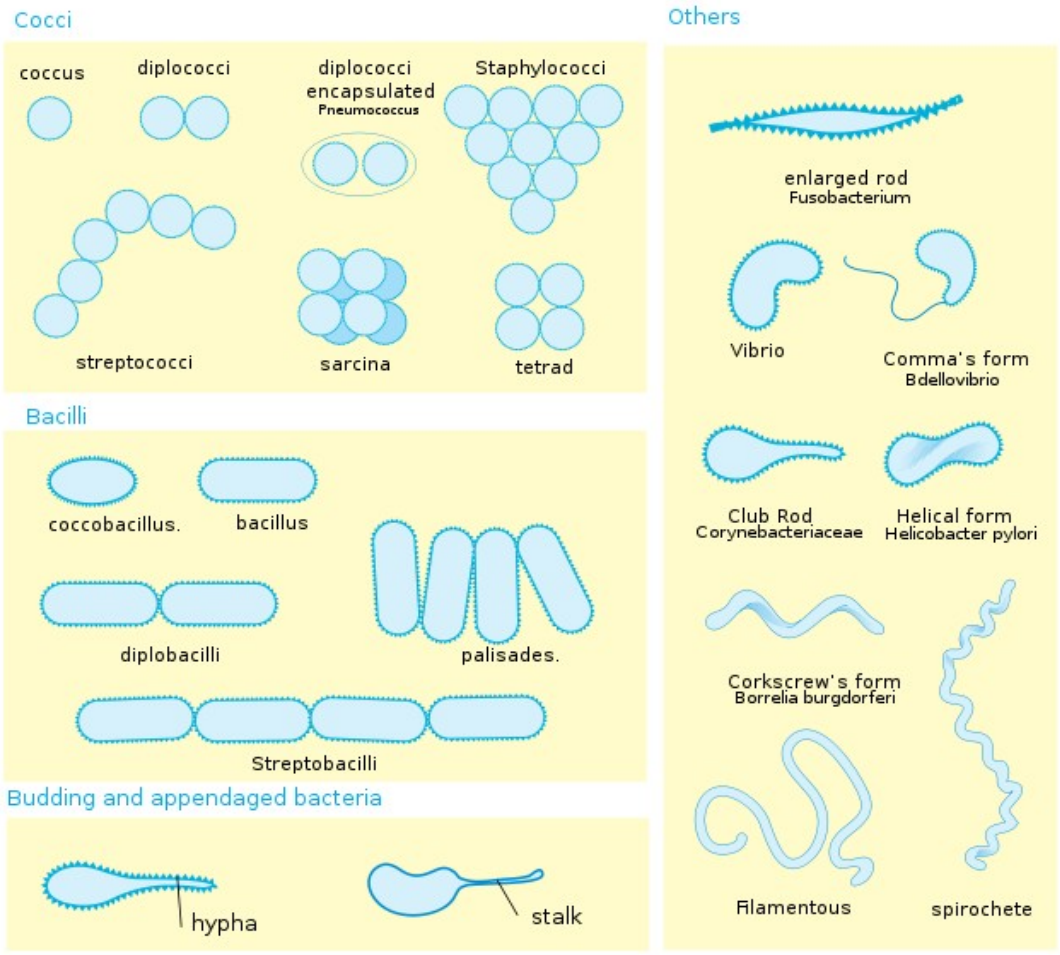
2.1 MOLECULAR BIOLOGY

2.1.1 THE CELL

Cells are often called the building block of life. They are the smallest structures capable of basic life processes, such as growth, taking in nutrients, getting rid of waste and reproducing. All known living organisms are composed of cells. Some microscopic organisms, such as bacteria and protozoa, are unicellular (consist of a single cell).Humans and other organisms are multicellular (consist of more than one cell).Cells implement thousands of biochemical reactions per minute and continuously reproduce new cells.

Cells vary in size and shape. The cell size broadly ranges from 0.2m to 200mm.The smallest cell a type of bacterium, mycoplasma, measures 0.0001mm in diameter and the largest know cell is an undifertilised ostrich egg cell. Human cells have also a variety of sizes from small red blood cells that measure 0.00076mm to liver cells that may be ten time larger.

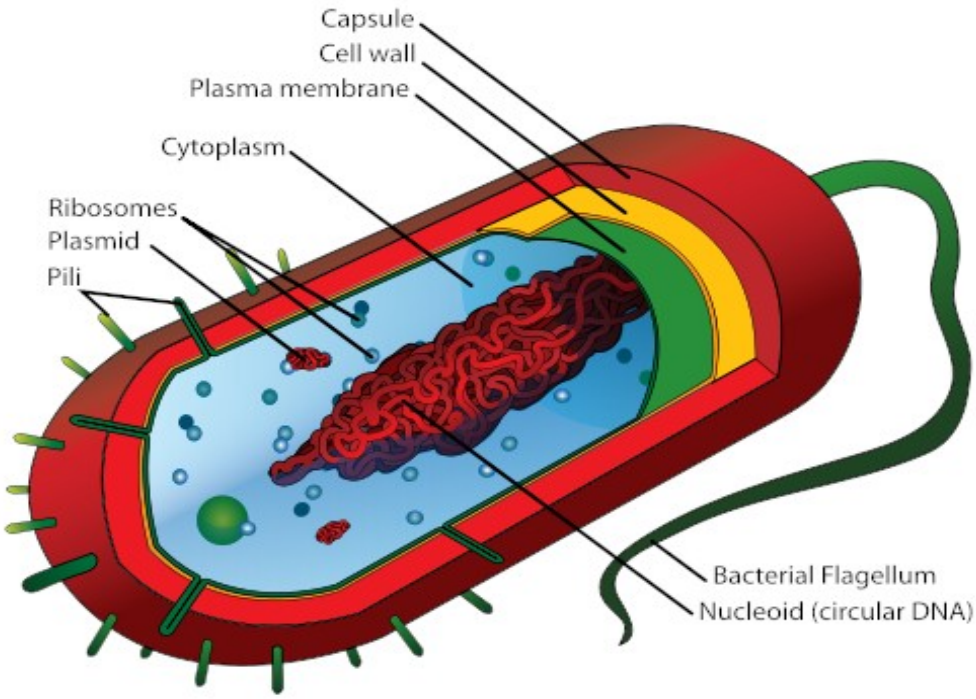
The basic shape of the cell is spherical, but variety of shapes exists among unicellular and multicellular organisms. The bacterium Bacillus resembles rods but the Campylobacter fetus is S-shaped. Certain cells have no fixed shape; hence they keep on changing their shape each minute you observe them under the microscope, such as Amoeba. In most multicellular organisms the location and association of cells can become unimportant factor in determining the shape. So in many cases, the original shape of cell can be altered by changing their environment and position. [1]



[Figure 2.1: Basic morphological differences between bacteria. The most often found forms and their associations.]

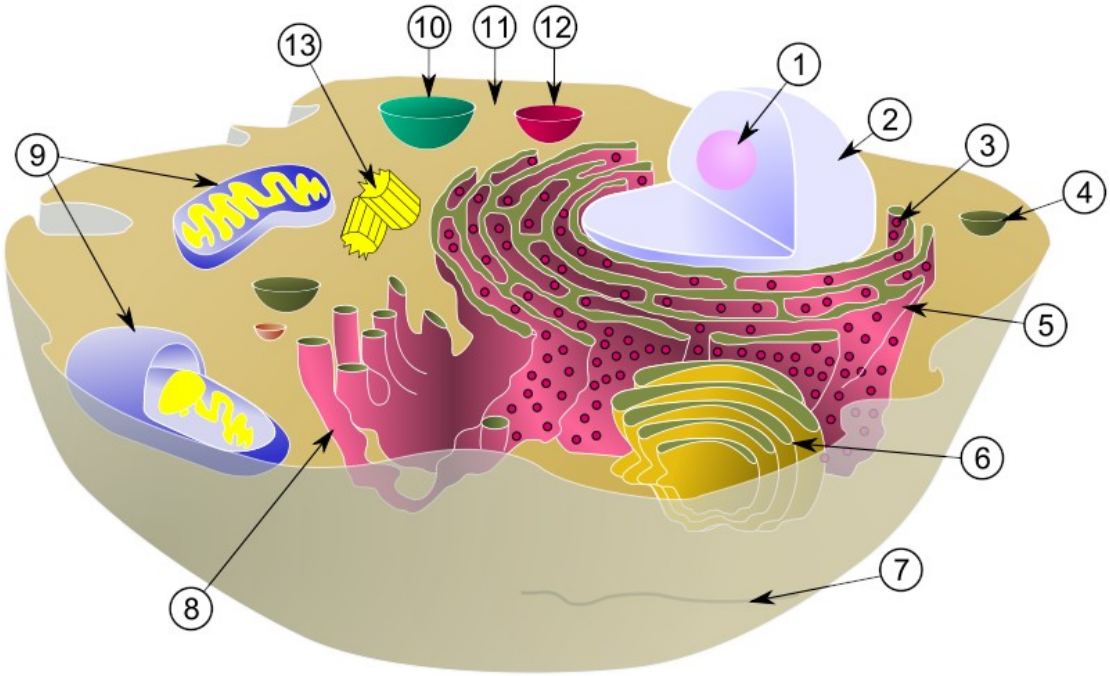
Most of the cells in multicellular organisms are organized into cooperated assemblies called tissues, such as the nervous, muscle, epithelial and connective tissues found in vertebrates. All the tissues components have to be coordinated correctly, and many of the require continual maintenance and renewal. Different tissue types are assembled into organs specialized to perform particular functions. All together, these assembled organ systems form the human body. [2]

Given these differences, it is perhaps surprising that there are only two types of cell: prokaryotic or eukaryotic. Prokaryotes are the cells that lack a membrane-bound nucleus. These cells have few internal structures that are distinguishable under a microscope such as Bacteria and cyanobacteria (also known as blue-green algae). Prokaryotic cells differ significantly from eukaryotic cells. Instead of having chromosomal DNA, their genetic information is in a circular loop called a plasmid. Bacterial cells are very small, roughly the size of an animal mitochondrion (about 1-2µm in diameter and 10µm long). Prokaryotic cells feature three major shapes: rod shaped, spherical and spiral. Instead of going through elaborate replication processes like eukaryotes, bacterial cells divide by binary fission.[3], [4]



[Figure 2.2: Diagram of a typical prokaryotic cell.]

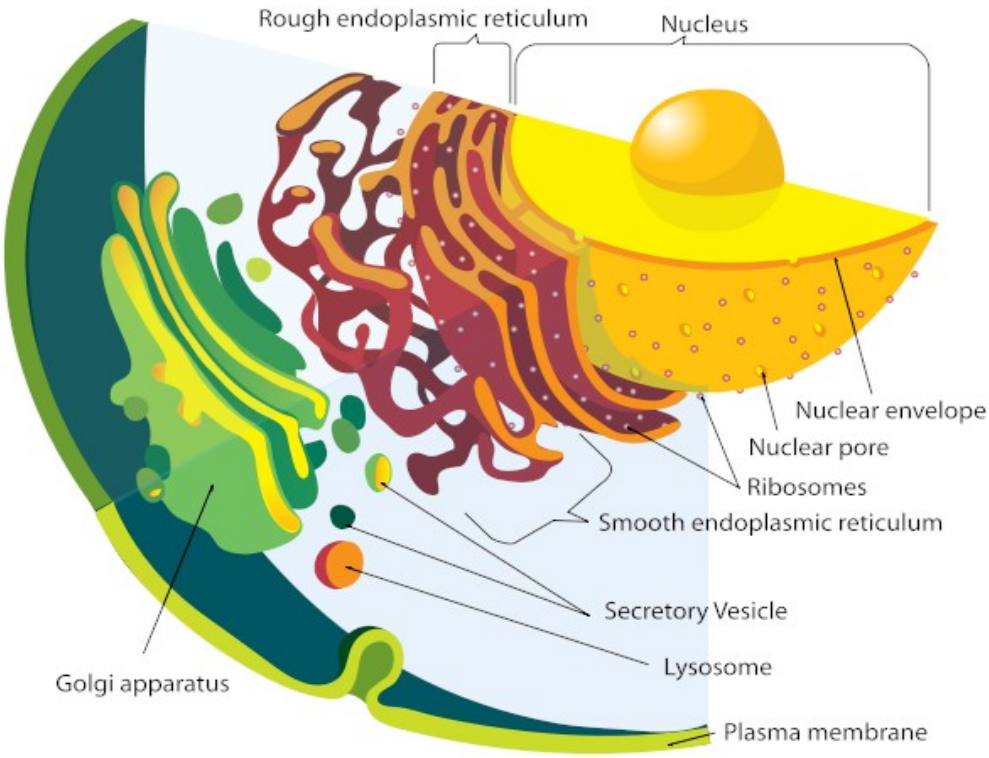
Eukaryotic cells can be easily distinguished through a membrane-bound nucleus. They are the cells of all other organisms, from protists to mammals to fungi to plants. These are generally larger (5-100µm, although some eukaryotic cells are large enough to be seen with the naked eye) and structurally more complex. Eukaryotic cells contain a variety of specialized structures known collectively as organelles, surrounded by a viscous substance called cytosol. The largest organelle, the nucleus contains the genetic information stored in the molecule deoxyribonucleic acid (DNA).



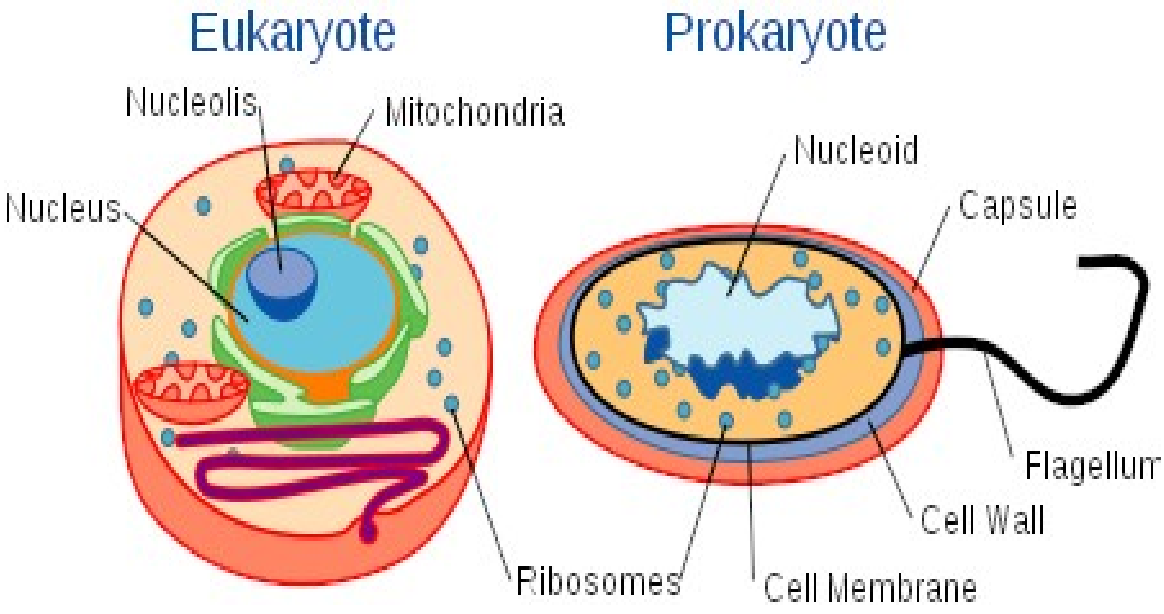
[Figure 2.3: Diagram of a typical animal (eukaryotic) cell, showing subcellular components. Organelles: (1) nucleosite, (2) nucleus, (3) ribosome, (4) vesicle, (5) rough endoplasmic reticulum (ER), (6) Golgi apparatus, (7) Cytoskeleton, (8) smooth endoplasmic reticulum, (9) mitochondria, (10) vacuole, (11) cytoplasm, (12) lysosome, (13) centrioles within centrosome.]

The major difference between prokaryotes and eukaryotes is that eukaryotic cells always have nucleus and other membrane-bound organelles; they often are multicellular and larger cells. In contrary, prokaryotic cells always are unicellular and smaller cells and have no nucleus or any membrane-bound organelles. Unlike the circular prokaryotic DNA, long sessions of eukaryotic DNA pack into the nucleus by wrapping around proteins.

The nucleus, that protects the DNA, is surrounded by a double membrane with pores providing channels through the nucleus that allow material to move in and out. A signal comes into the nucleus and instructions go out to the cytoplasm. In this way, potentially damaging chemical reactions that occur in the cytoplasm are protected. Endoplasmic Reticulum(ER) is called the collection of nuclear membrane, allowing molecules to move from one place to another. It is connected to the outer membrane of the nucleus and plays an important part in protein synthesis. It comes in two distinct forms: the rough ER and the smooth ER. The rough ER has ribosomes on its surface, giving it a "rough" appearance, and secretes proteins into the cytoplasm. The smooth ER synthesizes lipids and steroids and plays a role in calcium sequestration and release. [5]



[Figure 2.4: Diagram of an endomembrane system.]



[Figure 2.5: The cells of eukaryotes (left) and prokaryotes (right)]

2.1.2 DNA

DNA or deoxyribonucleic acid is the hereditary material in humans and almost all other organisms. Most DNA is located in the cell nucleus, but a small amount of DNA can also be found in the mitochondria. DNA is the chemical responsible for preserving coping and transmitting information within cells and from generation to generation. The information in DNA is stored as a code made up of four chemical bases: adenine (A), guanine (G), cytosine(C) and thymine (T). Human DNA consists of about 3 billion bases, and more than 99 percent of those bases are the same in all people.

These bases, according to the order or sequence they appear, sets the information available for building and maintaining an organism. The combined base and sugar is known as a nucleoside to distinguish it from the phosphorylated form, which is called a nucleotide. Four

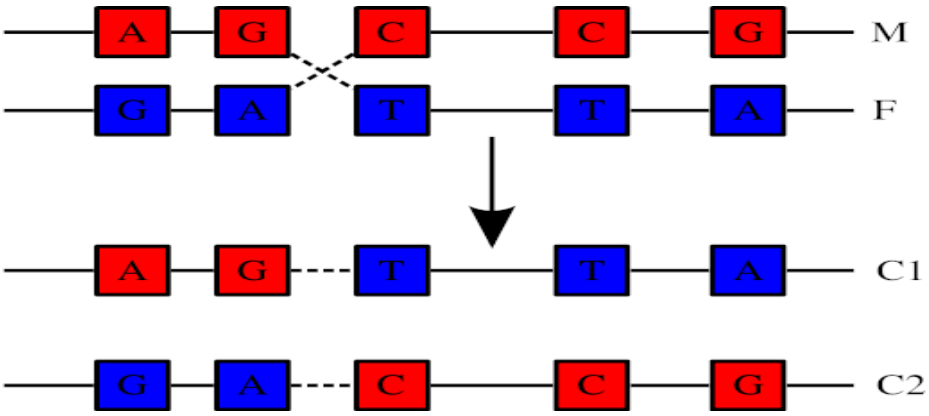
different nucleotides assemble DNA. Nucleotides entwine like vines in two long strands that

form a double helix. Chemical interaction create the phosphate-deoxribose backbone of the double helix within one strand opposite to their direction in the other strand: the strands are ant parallel and the bases pointing inward. The ends of these two DNA strands are called the 5’ (five prime) end, which have a terminal phosphate group, and 3’ (three prime) end, which

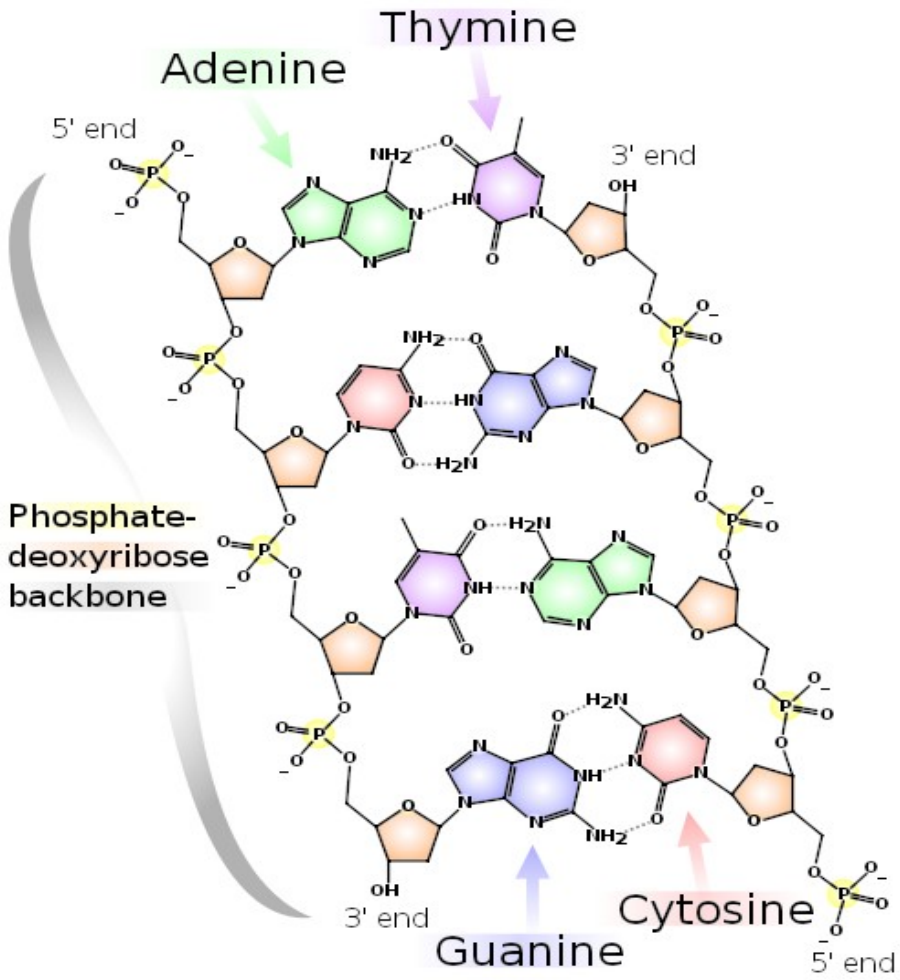
have a terminal hydroxyl group where nucleotides added in this direction synthesizing DNA.

[6]

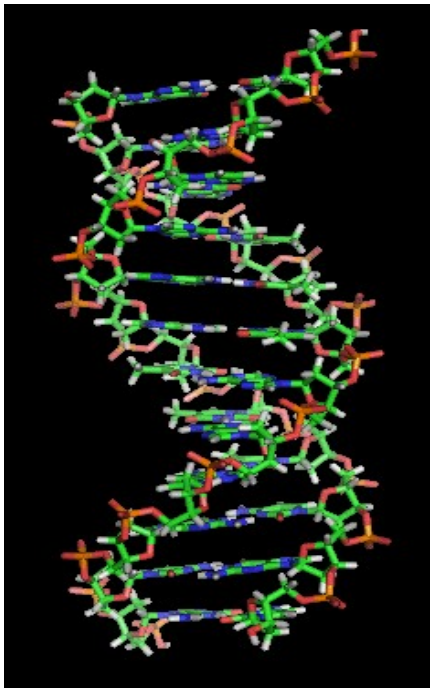
A gene is the basic physical and functional unit of heredity and is a distinct portion of a cell’s DNA. All living organisms depend on genes. They hold the information to build and maintain their cells and pass genetic traits to offspring. They act as instructions to make molecules called proteins. In humans, genes vary in size from a few hundred DNA bases to more than 2 million bases. The Human Genome Project has estimated that humans have 20000 and 25000 genes. Genes are packaged in bundles called chromosomes. Humans have 23 pairs of chromosomes of those 1 pair is the sex chromosomes and the other 22 pairs are autosomal chromosomes. [7], [8]



[Figure 2.6: Recombination involves the breakage and rejoining of two chromosomes (M and F) to produce two re-arranged chromosomes (C1 and C2).]



[Figure 2.7: Chemical structure of DNA. Hydrogen bonds shown as dotted lines.]



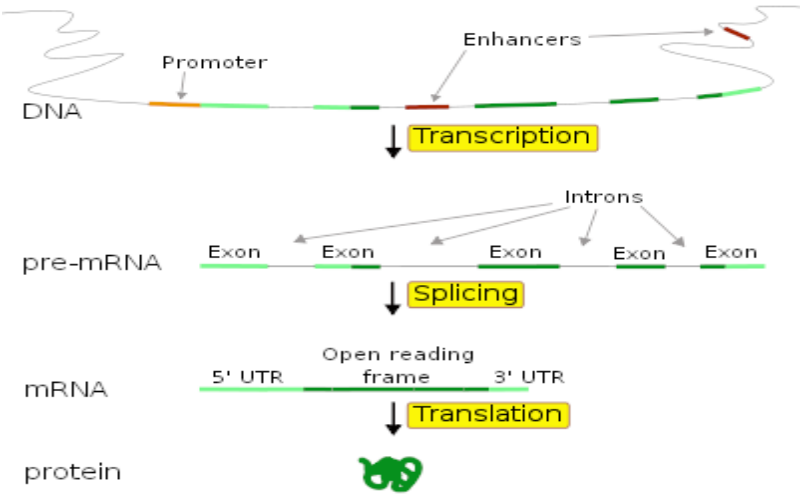
[Figure 2.8: A section of DNA. The bases lie horizontally between the two spiralling strands]

2.1.3 PROTEINS

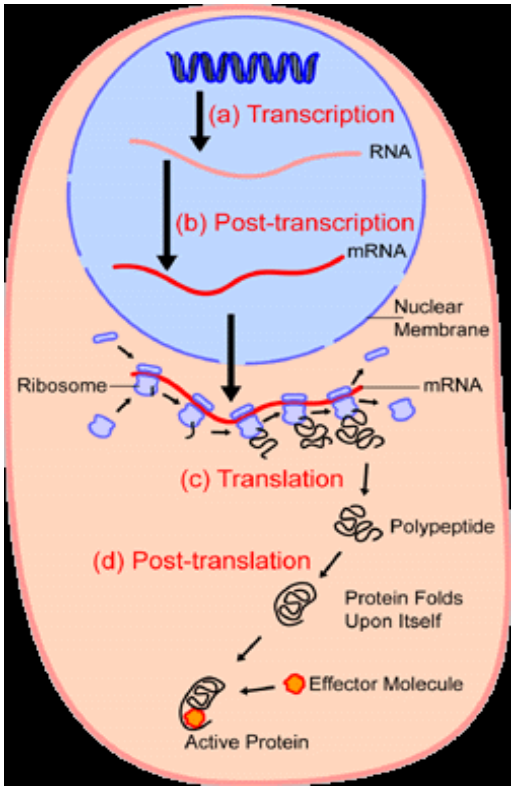
Proteins, DNA and RNA, are the molecules that support life on Earth. DNA is the template of genetic information for protein synthesis. RNA is the main responsible for the protein synthesis. The Messenger (mRNA) is an intermediate copy of genetic information that guides protein synthesis in the cell's protein synthesis machinery, called the ribosome. The Machinery (rRNA) is an essential component of the ribosome and the carrier (tRNA) carries amino acids to the ribosome for protein synthesis.

Proteins are polymers made of amino acids. There are 20 different amino acids commonly found in proteins. To make a protein, the ribosome strings amino acids together into a chain. This chain, called a polypeptide, folds into a specific three-dimensional shape. This process can be divided into two parts, transcription and translation. Before the synthesis of a protein begins, the corresponding RNA molecule is produced by RNA transcription. One strand of the DNA double helix is used as a template by the RNA polymerase to synthesize a messenger RNA (mRNA). This mRNA migrates from the nucleus to the cytoplasm. During this step, mRNA goes through different types of maturation including one called splicing when the non-coding sequence can be described as a unit of three nucleotides called a codon. At the translation, the ribosome binds to the mRNA at the start codon (AUG) that is recognized only by the initiator tRNA. [9]

The ribosomes contribute to the translation of protein synthesis. In translation, a ribosome starts to moves from codon to codon along the RNA, and the activation of amino acid linked to tRNA. As the ribosome moves from codon to codon along the mRNA , tRNA brings the corresponding amino acid to each. At the end, a release factor binds to the stop codon, the final mRNA codon, terminating translation which ends the synthesis of the peptide chain and releases it. Protein molecules perform functions as different as digesting sugars or moving muscles. The protein function is completely determined by its structure. [10]



*[Figure 2.10: Diagram of the "typical" eukaryotic protein-coding **gene**. Promoters and enhancers determine what portions of the DNA will be transcribed into the precursor mRNA (pre-mRNA). The pre-mRNA is then spliced into messenger RNA (mRNA) which is later translated into protein.]*



[Figure 2.9: An overview of protein synthesis. Within the nucleus of the cell (light blue), genes (DNA, dark blue) are transcribed into RNA. This RNA is then subject to post-transcriptional modification and control, resulting in a mature mRNA (red) that is then transported out of the nucleus and into the cytoplasm (peach), where it undergoes translation into a protein. mRNA is translated by ribosomes (purple) that match the three-base codons of the mRNA to the three-base anti-codons of the appropriate tRNA. Newly synthesized proteins (black) are often further modified, such as by binding to an effector molecule (orange), to become fully active.]

[Figure 2.11: The DNA sequence of a gene encodes the amino acid sequence of a protein.]

2.2 Systems biology

Systems biology describes the study of biological systems as a biology-based inter-disciplinary study, by investigating the components of cellular networks and their interactions. Systems biology is the modelling and discovery of a system by applying experimental high throughput, requiring an integration of experimental and computational methods, and whole –genome techniques. In systems biology the network of genes in an organism and all the components and the interactions among them, consist an integrated system. Thus instead of analysing individual components or aspects of an organism, we study the holist complex interactions. The form and functions of an organism are defined by these interactions. As an example, the immune system which fights infections and diseases with immune responses is the integrated system of interactions of numerous genes, proteins and the organism's external environment. [14]

Systems biology emerged with a growing understanding of how genes and their resulting proteins are responsible of biological form and function. As a result, the large increase in data from the genomics and the accompanying advances in high-throughput experiments and bioinformatics, today there are several Institutes of Systems Biology in all over the world. In advance, with the help of the internet, researchers are allowed to store and distribute massive amounts of information which make the study of Systems biology more evolving.

Systems biology can be consider as a holistic approach to understanding biology. It aims at system-level understanding of biology, namely a cycle composed of theory, analytic or computational modelling to propose specific testable hypotheses about a biological system and experimental validation. An examination of a structure and dynamics of cellular and organismal function, using the acquired quantitative description of cells or cell processes to consummate the computational model or theory. System biology refers to a cluster of peripherally overlapping concepts, taking a different approach and tries to integrate the biological knowledge and to understand how the molecules act together within the network of interaction that makes up life. [11]

Systems biology approach mathematical models of biological units.Scientists (computational biologists, statisticians, mathematicians, computer scientists, engineers, and physicists) are working to improve the quality of these approaches in an effort to co-operate on an interdisciplinary basis. These models, such as the reconstruction of dynamic systems from the quantitative properties of their elementary building blocks, require appropriate methods: first, for characterising the biological system of interest, second for systematically modifying the system, and third, for system modelling. Other aspects of computer science and informatics are also used in systems biology. System biology investigates the functioning and function of inter- and intra- cellular dynamic networks, using signal- and systems-

oriented approaches. To understand the functioning and function of cells, systems biology addresses the following central questions:

-how do the components within a cell interact to bring about its structure and function? (intra-cellular dynamics)

-How do cells interact to bring about coherent cell populations? (Inter-cellular dynamics)

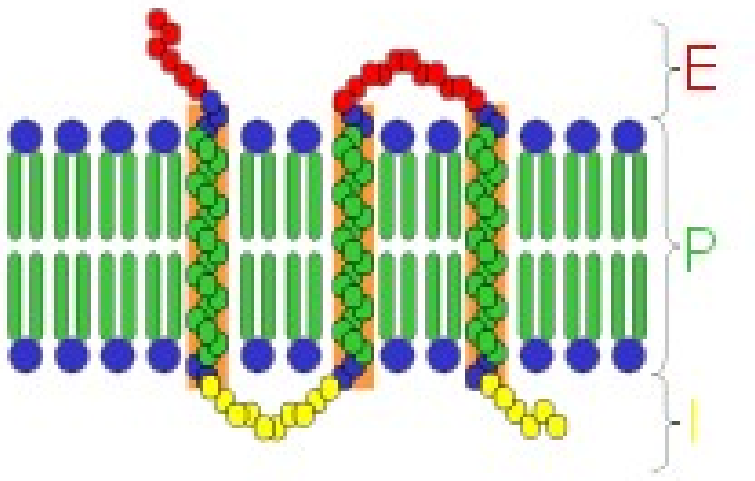
Systems biology signals a shift, away from molecular characterization and cataloguing of the components in the cell, towards an understanding of functional activity. [11], [12], [13], [14]

2.2.1 Cells as I/O systems

Cell life may compare as a collection of "wireless network" of interactions among proteins, RNA, DNA, and smaller molecules involved in signalling and energy transfer. These networks process environmental signals, induce appropriate cellular responses, and sequence internal events such as gene expression, thus allowing cells and entire organisms to perform their basic functions.

Research in molecular biology, genomics and proteomics provide plenty of data describing the basic constituents of such networks, and the mapping of intra- and inter-cellular signalling networks. The genome encodes the list of functional proteins that are potentially present in every cell of a given organism. This is achieved from the particular ordering of the four possible (A, T, C, G) bases in its DNA sequence. Genomics research has as its objective the complete decoding of this information, both to determine the entire DNA sequence of organisms as well as the cataloguing of differences among individual members. The full set of proteins in a cell type or tissue, the shape and the modifications made to a particular set of proteins is what largely determines their function and it's up to proteomics research to determine all the sets of proteins and their three dimensional structure. Proteins, which interact with each other, are the primary components of living things.

Besides proteins are essential parts of organisms, among other roles, they form receptors that endow the cell with sensing capabilities; actuators that make muscles move (myosin, actin) which form a system of scaffolding that maintains cell shape. Detectors form the immune response, enzymes that catalyse chemical reactions and are vital to metabolism, and switch that turn genes on or off. They are also important in cell signalling and they provide structural support and help in the transport of smaller molecules. Notably in cell adhesion and the cell cycle, and in directing the breakdown and reassembly of other cellular and elements. Large scale data sets are being generated by genomics and proteomics projects, using advanced genetic engineering tools and measurement technologies and there is a widely recognized need to organise and interpret these data [15].



[Figure 2.12 : Transmembrane receptor: E=extracellular space, I=intracellular space, P=plasma membrane]

Cells are the most natural example of input-output systems theory and control systems, built from simple components that are linked together according to certain rules. Cells receive information from external inputs, which can be physical (e.g. radiation or temperature) or chemical (e.g. drugs or hormones), and measurable output with chemical signals to other cells, as the activation of transcription factors. Each cell can be regarded as being composed of many subsystems involved in various processes such as cell growth and maintenance, division, and death. The comprehension of cells behaviour in terms of cascades and feedback interconnections of elementary “modules” is one of the challenging topics in molecular biology literature.

Some of the questions arising from these natural systems are: How can someone characterize the input/output behaviours in these networks, and why is it so important the information of processing capabilities? What subsystems appear most often repeated than others? What are the interactions between different signal pathways and how they interact? How is the calculation of the forms of reactions and values of parameters (identification, reverse engineering)? Why do we use cascades and feedback loops? After calculating these forms of reactions, must find how estimate time-varying internal states, such as the concentrations of proteins and other chemical substances, from input/output experiments (observer problem)? Where is the main reason that affects the robustness of the system? What about stability, oscillations, and other dynamical properties of such complex systems? Further deepening to analysis questions, we have to deal with the control of cellular systems through drugs or genetic modifications.

2.2.2 Cell Chemistry

The cell is the basic building block of which higher organizational levels such as tissues and organs and entire organism are composed. This chapter is to review some basic concepts from molecular and cell biology. The cell is rather complex environment, consisting of many different components. Because cells are about 70% water, life depends mostly on aqueous chemical reactions. These reactions occur between molecules, where a molecule is a cluster of atom*s*, held together by so called covalent bonds. The weight of a molecule is its mass relative to that of a hydrogen atom. The mass of a molecule is specified in Daltons, 1 Da being an atomic mass unit approximately equal to the mass of a hydrogen atom.

$$\text{moles} = \frac{\text{weight}}{\text{molecular weight (a quantity)}}$$

One mole, 1 M, corresponds to N_A = 6:022 ¢ 10_{23} molecules of a given substance. N_A is referred to as the Avogadro's number. The morality of a solution is defined by a concentration of 1 moles of the substance in 1 litter of solution:

$$1\text{molar} = 1\text{M} = 1\frac{\text{mol}}{\text{L}} \text{ (a concentration)}$$

If molecules are clusters of atoms, held together by bonds, these bonds can be broken by violent collisions amongst molecules. Average thermal motion does not break these bonds and thus the breaking and making of bonds is the fundamental process that determines the concentrations of chemical species in a reaction. This process requires energy to take place and is care-fully controlled by highly specific, catalysts, called *enzymes*. How fast a reaction occurs is a matter of *kinetics*, defined by the rate of a reaction. In general, energy is the ability of a system to perform work. Therefore, whether or not a reaction can proceed is determined by its *energetic*. [16]

The principle types of reactions are: *catabolic pathways*, breaking down organic matter, for example to harvest energy in cellular respiration, and biosynthetic or *anabolic pathways* which use energy to construct components of cells such as proteins and nucleic acids. Both sets of reactions together constitute what is called the *metabolism* of the cell. Metabolism is the set of chemical reactions that happen in living organisms to maintain life. These processes allow organisms to grow and reproduce, maintain their structures, and respond to their environments.

A protein is any chain of amino acid. An amino acid is small molecule that acts as the building block of any protein. Proteins are particularly versatile, having various roles in maintaining the function of a cell and the organism as a whole. Many proteins serve as *enzymes* that are catalysts that control kinetic (bond-breaking and -making) reactions. Other proteins are used to build the structural components that make up the cell, or they act as motors and produce force and movement. Enzymes have extremely interesting properties that make them little chemical-reaction machines. The purpose of any enzyme in a cell is to allow the cell to carry out chemical reactions very quickly. These reactions allow the cell to

build things or take things apart as needed. This is how a cell grows and reproduces. At the most basic level a cell is really a little bank full of chemical reactions that are made possible by enzymes. Enzymes are made from amino acids and their proteins. When an enzyme is formed it is made by stringing together between 100 and 1000 amino acids in a very specific and unique order. The chain of amino acid then folds into a unique shape. That shape allows the enzyme to carry out specific chemical reactions (an enzyme acts as a very efficient catalyst for a specific chemical reaction). The enzyme speeds that reaction up tremendously. [17], [18]

Referring to ligand is a substance that is able to bind to and form a complex with the bio molecule to serve a biological purpose. It is a signal triggering molecule, binding to a site on a target protein. The receptor-ligand signalling is the process where extracellular molecules, called *Ligands* bind to receptors in the cell membrane. Extracellular signalling molecules include hormones, cytokines and growth factors. Usually extracellular signals are found at

very low concentrations, in the order of $10\text{-}8 mol=L$. Binding to receptors is highly specific to

particular ligands. Not all ligands that bind to receptors results in the activation of that receptor. Ligands that bind to receptors and thereby prevent activation are called *antagonists*. Antibodies are proteins that are found in blood or other body fluids of vertebrates, and are used by the immune system to identify and neutralised foreign objects, such as bacteria and viruses. Antibodies can therefore be used in experiments to select and quantitate proteins. When a person gets immunized against something, the thing that is trying to be accomplished is to activate a person's immune system. Therefore, we will make antibodies which can bind to the thing that we are injected with. This response not only leads to an initial production of antibody, but also leads to more cells which recognize the foreign thing and which can make antibodies against it. After immunization or a natural recovery from some infection, we have more cells specifically able to respond to the same foreign thing. [19], [20]

For example, considering a population of antibody molecules which suddenly encounter a population of ligands, diffusing in the fluid surrounding them. The frequent encounters antibody-ligand complexes. The population of such complexes will initially increase but eventually complexes will also break apart (dissociate). Eventually, a chemical equilibrium is reached in which the number of association events per second is equal to the number of dissociation events. From the concentrations of ligands and, antibody and the complex at

equilibrium, one can calculate the equilibrium constant $Keq$of the strength of binding. The

same principle described here for antibodies, applies to any binding of molecules.

Capital letters are used to denote molecular species. Consider two proteins *A* and *B,* the corresponding complex they form *AB* and the reversible reaction:

$$A+B\leftrightarrow(AB)$$

For dissociation the reaction diagram is:

$$(\mathrm{AB})\ \mathrm{kd}\ \mathrm{A}+\mathrm{B}$$

where the *dissociation rate* equals the product of kd and the complex concentration (*AB*).

For the association of molecules:

$$\mathrm{A}+\mathrm{B}\ \mathrm{ka}\ \mathrm{AB},$$

The *association rate* is the product ofka, *A* and *B*.

At equilibrium,

$$\mathrm{kaA}\cdot\mathrm{B}=\mathrm{kd}(\mathrm{AB})$$

which lead to the definition of the *equilibrium constant*

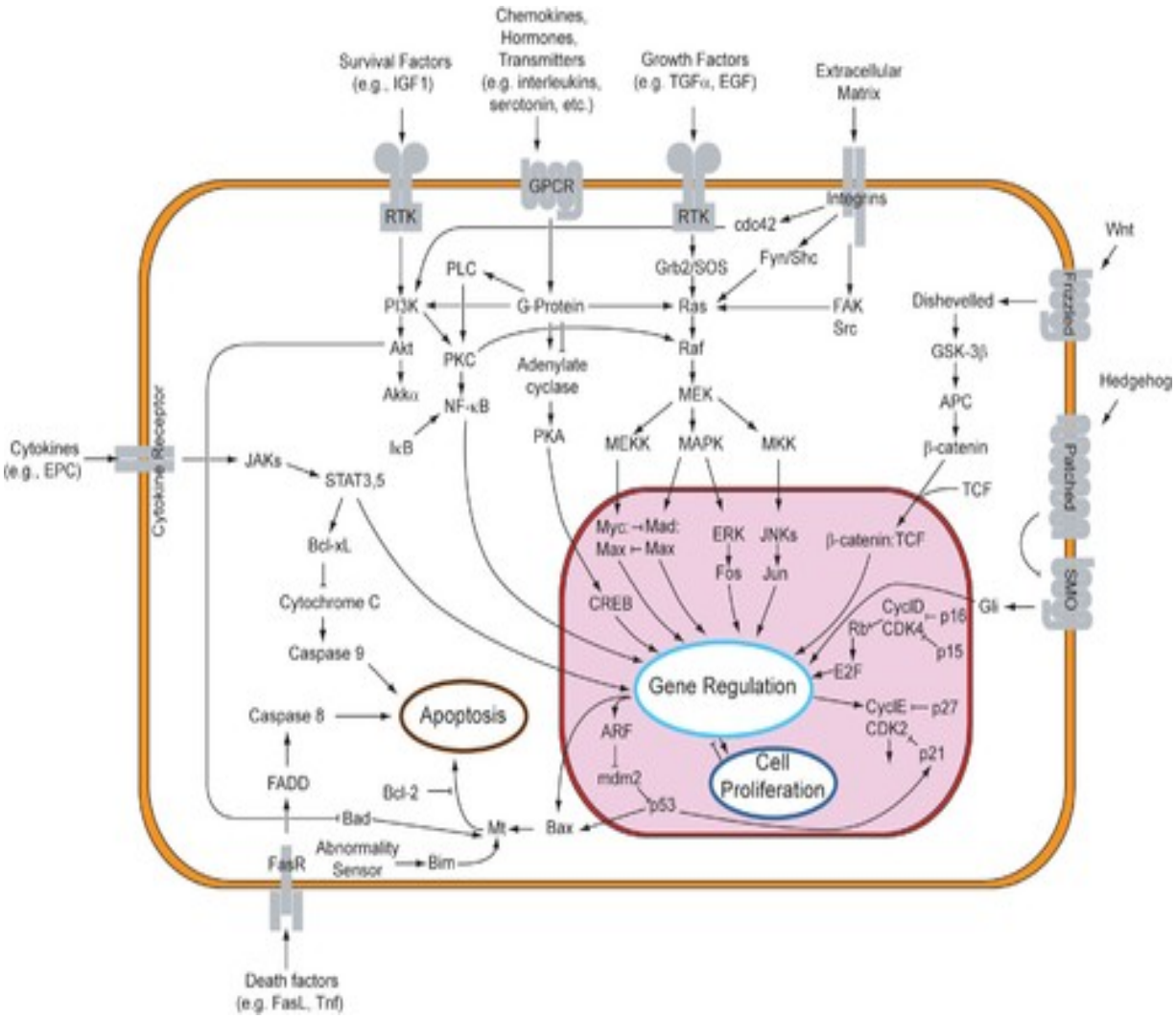
$$\mathrm{Keq}=(\mathrm{AB})\mathrm{A}\cdot\mathrm{B}=\mathrm{kakd}$$

The equilibrium constant has a unit of litters per mole. The larger the equilibrium constant, the stronger the binding between *A* and *B*. Note that for the system to be in equilibrium there is a flow of mass or material. In dynamic systems theory, a steady state is sometimes also referred to as equilibrium so that there is a risk of confusion. For the biochemist a biological system in equilibrium is dead. [21]

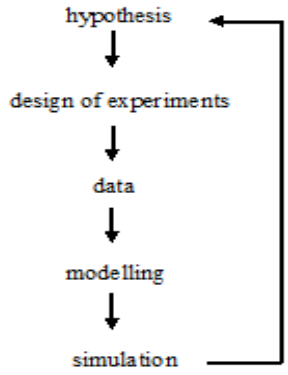
2.2.3 Cell Signaling

For cells to combine into networks that realize higher levels of organization, including for example tissue and organs, it is necessary for them to communicate, exchange information. The basis for this intercellular signalling is the receptors in the cell membrane. The transmission of extracellular information to the genome is referred to intracellular signalling. Inter- and intra- cellular information effects the transcription of information from the genome and the synthesis of proteins.

The transmission of information is realized by chemical reaction networks, called pathways. Signals, passing these networks, are realized through changes in concentrations. The cell membrane and the nucleus in eukaryotic cells form physical barriers. There are principally two ways to pass these barriers- through active transport of molecules passing through the cell surface (e.g. via pores or gap junctions) or nucleus or via signalling transduction, i.e., receptor stimulation and phosphorylation as a means to transmit information without the movement of molecules. The location of a signalling molecule within the cell affects the interaction with other proteins and hence the movement of molecules to different cellular locations, called *translocation*, influences the dynamics of a signalling pathway.



[Figure 2.13: Overview of signal transduction pathways]



[Figure 2.14: Iteration of the modelling loop.]

System biology requires an iteration of the modelling loop as shown. The diagram shows the role of mathematical modelling and simulation in testing hypotheses but also in generating hypotheses through prediction. The dominant mathematical modelling tool for intracellular signalling networks is their description with the help of ordinary differential equations (ODE). As a simple example of the way by which such a model is built, consider the following enzymatic reactions:

$$E_0 + S \underset{k_{-1}}{\overset{k_1}{\square}} E_1 \xrightarrow{k_2} E_0 + P$$

Where:

E_0 the concentration of the unoccupied enzymes

S the concentration of the substrate (the unreacted molecules)

E_1 the concentration of occupied enzymes

P the concentration of product (reacted molecules)

k_1, k_{-1}, k_2 the rate constants.

The reaction scheme is called Michaelis-Menten mechanism. The Michaelis-Menten equation is a differential equation used to model the rate at which enzymatic reactions occur. This model allow scientist to predict how fast the reaction will take place based on the concentrations of the chemicals being reacted. In order to model an enzymatic reaction, some conditions must be maintained:

- Temperature, ionic strength, ph and other physical conditions that might affect the rate must remain constant.
- Each enzyme can act on only one other molecule at a time.
- The enzyme must remain unchanged during the course of the reaction.
- The concentration of substrate must be much higher than the concentration of enzyme.

The rates for the three reactions are:

$E0+S \rightarrow E1 \ : \ Rate1=k1SE0$

$E0+S \leftarrow E1 \ : \ Rate\text{-}1=k\text{-}1E1$

$E1 \rightarrow E0+P \ : \ Rate2=k2E1$

The ordinary differential equation for this system is given by :

$$dSdt=-Rate1+Rate\text{-}1=-k1SE0+k\text{-}1E1$$

$$dE0dt=-Rate1+Rate\text{-}1+Rate2=-k1SE0+k\text{-}1E1+k2E1$$

$$dE1dt = Rate1 - Rate{-}1 - Rate2 = k1SE0 - k{-}1E1 - k2E1$$

$$dPdt = Rate2 = k2E1$$

The rate constants can be difficult if impossible to determine. For the purpose of seeing the behaviour of the system, we give them the values $k_1=10$, $k_{-1}=1$ and $k_2=5$ with initial conditions $S=1.0$ and $E_0=0.08$. By adding equations and doing some algebraic maniplulation:

$$dSdt = -k1SET + k{-}1 + k1SE1$$

$$dE1dt = k1SET - (k1S + k{-}1 + k2)E1$$

As long as $E_T \ll S$ then we can assume that $dE1dt \approx 0$

$$dSdt = -k1SET + (k{-}1 + k1S)E1$$

$$0 \approx k1SET - (k1S + k{-}1 + k2)E1$$

$$E1 = k1SET(k{-}1 + k2 + k1S)$$

By manipulating these equations we can derive the Michaelis-Menten equation fordSdt:

$$\frac{dS}{dt} = -\frac{V_{max}S}{K_M + S}$$

Where $V_{max} = k_2 E_T$ and $K_M = \frac{k_{-1} + k_2}{k_1}$

This one equation replaces the system for modelling the substrate rate equation. There are only two parameters and they can both be determined experimentally. [22], [23], [24], [25]

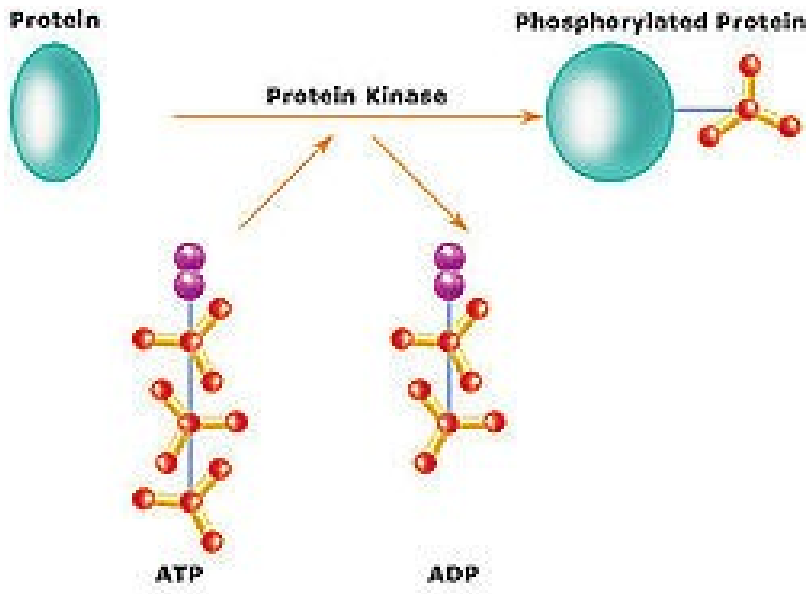
Chapter 3

EGFR Pathway

3.1 Protein Tyrosine Kinase

ATP K)

A large variety of cellular events, such as metabolism, differentiation and growth are modulated by Protein Tyrosine Kinases. The phosphorylation of tyrosine residues in target proteins is vital for maintaining cellular homeostasis and causes a change in the function of the protein. Phosphorylation at tyrosine residues also is responsible for a wide range of properties in proteins such as enzyme activity, subcellular localization, and interaction between molecules, communicating signals within a cell and regulating cellular activity, such as cell division. As a result, Protein Tyrosine Kinases are important targets for both basic research and drug development effort. The are divided into two main families of proteins, including both transmembrane receptor-linked kinases and cytoplasmic proteins, known as non receptor tyrosine kinases. Phosphorylation of proteins by kinases is an important mechanism in signal transduction for regulation of cellular activity. The term kinase describes a large family of enzymes that are responsible for catalyzing the transfer of phosphoryl group from a nucleosite triphosphate donor, such as, to an acceptor molecule and tyrosine kinases also catalyze the phosphorylation of tyrosine residues in proteins. Furthermore, many tyrosine kinases were first identified as the products of activated retroviral or cellular oncogenes in which manifestation they induce neoplastic cell growth. Finally, mutations can trigger some tyrosine kinases to become active, a nonstop state that may provide to initiation or progression of cancer. [26], [27]



[Figure 3.1: Protein phosphorylation]

3.1.1 Protein Tyrosine Kinase Receptors

$(RTK)s$

The Tyrosine Kinases consists are separated in two major groups the transmembrane receptor-linked kinases and the non-receptor tyrosine kinases. Non-receptor tyrosine kinases or non-specific tyrosine kinases can deliver a phosphate group from to a tyrosine remains (residue) in a protein, while receptor tyrosine kinases are the high-affinity cell surface receptors for many polypeptide growth factor, cytokines, and hormones. The receptors are transmembrane proteins that spam the plasma membrane just once. Of the 90 unique tyrosine kinase genes identified in the human genome, 58 encode receptor tyrosine kinases proteins. Receptor tyrosine kinases have not only been key regulators of normal cellular processes but also a vital function in the augmentation and progression of many types of cancer. [28], [29]

$(RTK)s)$

Tyrosine kinases receptors are a family of receptors with a similar structure. They each have a tyrosine kinase domain (which phosphorylate proteins on tyrosine residues), a hormone binding the extracellular domain, and a carboxyl terminal segment with multiple tyrosines for autophosphorylation. Then the tyrosine kinases domains phosphorylate the C terminal tyrosine residues. This phosphorylation generates binding sites for proteins with domains. is one of these proteins. with bound to it then binds to the receptor complex. This causes the activation of, which is a guanyl nucleotide-release protein. When this is activated, it causes certain proteins to release and exchange it for. is one of these proteins. When has bound to it, it becomes active. Activated then causes the activation of a cellular kinase called. kinases then phosphorylate another cellular kinase called. This causes the activation of. Activated then phosphorylates another protein kinase called causing its activation. This series of phosphorylating activations is called a kinase cascade. It results in amplification of the signal. Among the final targets of the kinase cascade are transcriptions factor. Phosphorylation of these proteins causes them to become active and bind to the, causing changes in gene transcription. Different receptor tyrosine kinase classes have been identified. Epidermal growth factor receptor family which is directly related with the development neurodegenerative diseases, such as multiple sclerosis and Alzheimer's Diseases. Fibroblast growth factor receptor family, a mutation in its receptor causes achondroplasia - the most common type of dwarfism. Vascular endothelial growth factor receptor family which is one of the major inducers of endothelial cell reproduction and permeability of arteries. receptor family where is the receptor for members of the glial cell line-derived neurotrophic factor family of extracellular signalling molecules or molecules, as an antibody, hormone, or drug, that binds to a receptor (ligands). receptor family where

and receptors are the largest subfamily of. Insulin receptor family, a transmembrane protein embedded in the plasma membrane of the responding cells. [30], [31], [32], [33], [34], [35]

3.2 Epidermal Growth Factor Receptor

$Elk-1, AP-1, -1), ER1$

The epidermal growth factor receptor (in humans) is the cell surface receptor for members of the epidermal growth factor family of extracellular protein ligands. The epidermal growth factor family of receptor tyrosine kinases contains four receptors, and. Members of the family are composed of a cytoplasmic tyrosine kinase domain, a single transmembrane domain, and an extracellular domain that is involved in ligand binding and receptor dimerization. Activating the has as a result the beginning of a diverse array of cellular pathways. In response to toxic environmental stimuli (e.g. ultraviolet irradiation) or to receptor occupation by, the forms or Homology-2) containing effector proteins. Dimerization occurs autophosphorylation initiating and downstream cascade of events culminating in cellular responses such as cell proliferation or apoptosis. The activate dimer complexes with the adapted protein, coupled to the guanine nucleotide releasing factor,. The complex can either bind directly to phosphotyrosine site in the receptor or indirectly through. These protein interactions bring in close proximity to, allowing for activation. This consequently activates the and signalling pathways that activate transcription factors, such as and, that advance gene expression and contribute to cell proliferation. [36], [37]

$EGFR$

In the similar cell type, signalling throughout the same pathway may have as a consequence completely different outputs relying on the amplitude and persistence of activation of signalling intermediates, i.e. on their kinetic behaviour. Many factors, such as the number of the receptors which are displayed on the surface of the cell, the concentration of the growth factor and targeting proteins and their initial activity states, affect the kinetics of the cellular response to .Furthermore, other signalling pathways that share or combine with one or more components of pathway can affect the kinetic pattern of signalling. Despite a great body of data which describes signalling at the molecular stage, the way in which the complex model of cellular responses to is controlled, remains defectively understood. The major reason for this is the lack of a quantitative description of signalling network. Detailed knowledge of the dynamics of complex cellular responses requires a sequence of experimental and computational approaches.

$EGFR$

The early events of signalling, such as binding and receptor autophosphorylation, binding and activation of , phosphorylation and, and activation of, progress in a time frame of seconds. A significant role in induced signalling plays the receptor internalization and its subsequent degradation in lysosomes, which are slow processes. Activation and binding of ligands causes the recruitment of to clathrin-coated pits and transfer to endosomes. These series of actions are developing over time frames of minutes to hours, (much more slowly

than early signalling events) which develop to a quasi-steady-state level in a time scale of seconds.

$37 \pm C$

Protein phosphorylation that results from stimulation of cellular receptors is a dynamic process, which reflects the receptor protein kinase activation and the intercommunications between distinct signalling parts and the activities of various phosphatases. Following signalling depends on how the phosphorylation of adapter and target molecules builds up over time, and both transient and sustained levels of this phosphorylation/activation ultimately rely on the whole network of signalling reactions. Though, our comprehension of how the cellular response is incorporated to achieve the desired physiological outcome and our knowledge of signalling events at the molecular level are not associated. He has joined experimental analysis with a computational kinetic approach, in order to produce a consolidating framework for studying receptor tyrosine kinase signalling. Experimental analysis of the time course of the response to stimulation in rat hepatocytes displayed a rapid burst in receptor phosphorylation and accumulation of phosphorylated/ activated target proteins, which occurs as early as within 1530s following stimulation. The time resolution of the experiments was not adequate to demonstrate the beginning rate of increment in phosphorylation or activation of its target proteins. A more gradual increase can be seen at temperatures lower than used in these experiments. [38], [39], [40]

3.3 Kinetic Analysis

3.3.1 Schematic representation of Protein-Protein Interactions induced by EGF

$SOS \mid - S$

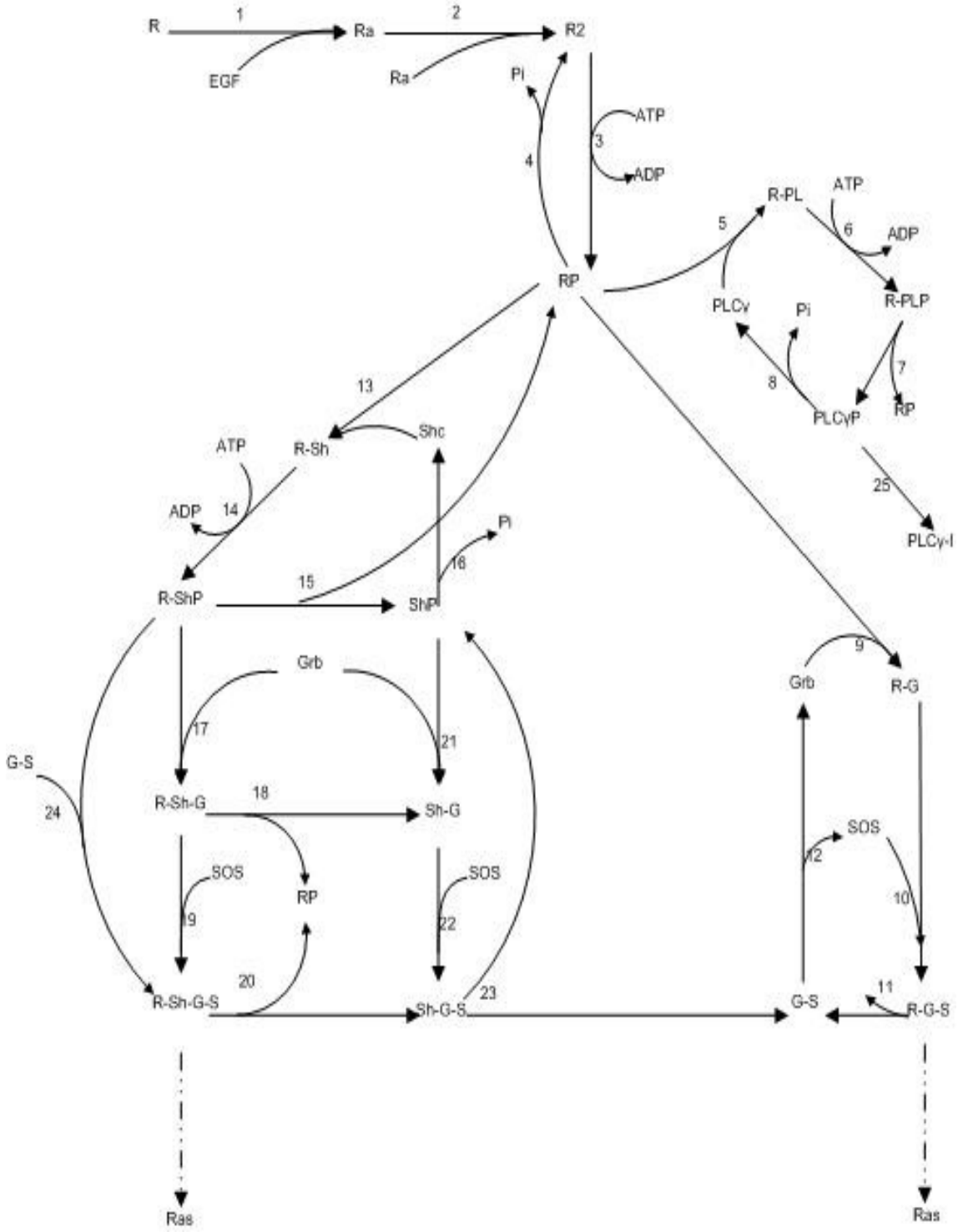
The model considered in the present thesis is based on the signalling network studied by Kholodenko et al. (1999), which considers three coupled cycles of interactions between the phosphotyrosine residues of the receptor and the cytoplasmic proteins , and . As presented in figure 3.2, binds to the extracellular domain of the monomeric to form a receptor- ligand complex, which subsequently dimerizes and is activated autophosphorylation of tyrosine residues. The subset of the cycle considered here is initiated by the binding of to a receptor phosphotyrosine to yield the complex. The binding of the Son of Sevenless homolog protein, to then effects the ternary complex, which subsequently scatters into the phosphorylate receptor,, and the complex , which further dissociates into and .

$Shc \mid P - S - S$

In the cycle, binds to the phosphorylated receptor to form the complex R-Sh, whose subsequent phosphorylation yields , which may then dissociates to yield the phosphorylated receptor, , and phosphorylated . Alternatively, may bind to the complex to produce the

ternary complex, which dissociates to produce and the complex Sh-G. may also bind to to form the four protein complex,, which may also be produced by the association of and. dissociates to produce and the complex, which may in turn dissociate to free and, thereby interacting with cycle. is dephosphorylated by phosphatases to release.

Hence, both theand cycles involve the activation of, whose downstream target is the membrane-bound protein (see figure 3.2). Once activated, may activate other signalling proteins to relate the signal downstream to other cytoplasmic and nuclear targets. -mediated signalling is initiated by the binding of to produce the complex, which is then phosphorylated, yielding, before dissociating into phosphorylated phospholipase Cγ and. may then be dephoshorylated by phosphatases or translocate to cytoskeletal or membrane structures to produce. This subset of network consists of 23 variable which take part in 25 kinetic reactions with 50 associated rate constance. Moreover, the following suite of 23 coupled ordinary differential equations governs the temporal evolution of this group of variables. [41]



[Figure 3.2: Kinetic scheme of EGFR signaling mediated by adapter and target proteins. Numbering of individual steps is arbitrary.]

3.3.2 Kinetic equations comprising the computational model

$$\frac{d[EGF]}{dt} = -u_1$$

$$\frac{d[R]}{dt} = -u_1$$

$$\frac{d[R_a]}{dt} d[R_a]/dt = u_1 - 2u_2$$

$$\frac{d[R_2]}{dt} d[R_2]/dt = u_2 + u_4 - u_3$$

$$\frac{d[RP]}{dt} d[RP]/dt = u_3 + u_7 + u_{11} + u_{15} + u_{18} + u_{20} - u_4 - u_5 - u_9 - u_{13}$$

$$\frac{d[R-PL]}{dt} d[R-PL]/dt = u_5 - u6$$

$$\frac{d[R-PLP]}{dt} d[R-PLP]/dt = u_6 - u_7$$

$$\frac{d[R-G]}{dt} d[R-G]/dt = u_9 - u_{10}$$

$$\frac{d[R-G-S]}{dt} d[R-G-S]/dt = u_{10} - u_{11}$$

$$\frac{d[R-Sh]}{dt} d[R-Sh]/dt = u_{13} - u_{14}$$

$$\frac{d[R-ShP]}{dt} = u_{14} - u_{24} - u_{15} - u_{17}$$

$$\frac{d[R-Sh-G]}{dt} = u_{17} - u_{18} - u_{19}$$

$$\frac{d[R-Sh-G-S]}{dt} = u_{19} - u_{20} + u_{24}$$

$$\frac{d[G-S]}{dt} = u_{11} + u_{23} - u_{12} - u_{24}$$

$$\frac{d[ShP]}{dt} = u_{15} + u_{23} - u_{21} - u_{16}$$

$$\frac{d[PLC\gamma P]}{dt} = u_7 - u_8 - u_{25}$$

$$\frac{d[PLC\gamma P-I]}{dt} = u_{25}$$

$$\frac{d[Grb]}{dt} = u_{12} - u_9 - u_{17} - u_{21}$$

$$\frac{d[Shc]}{dt} = u_{16} - u_{13}$$

$$\frac{d[SOS]}{dt} = u_{12} - u_{10} - u_{19} - u_{22}$$

$$\frac{d[Sh-G-S]}{dt} = -u_{23} + u_{20} + u_{22}$$

[Table 3.1: Kinetic equations comprising the computational model]

3.3.3 Kinetic Equations

RP FR

We transformed the reaction scheme, as we see in figure 3.2, into mathematical equations recognized as chemical kinetics equations to incorporate the experimental research in a description of the dynamic action of the signalling network. For changes with time of the concentration of any part, e.g. the receptor form, one can write the following.

RP

RP -total rate of consumption

$V_m = V_{cw} = 33.3$

The total rate is the sum of the rates that produce or consume in proportion to the kinetic diagram. For example, the total rate of production equals the sum of the (net) rates of six steps (steps 3, 7, 11, 15, 18, and 20; Figure 3.2). A complete set of chemical kinetic equations representing the reactions of (figure 3.2) is specified in Table 3.1. Kinetic equations are occasionally written in terms of concentrations since the reaction rates are functions of concentrations. If the same compound takes part in reactions occurring in different compartments with unequal volumes, having the desired effect in concentration of that compound will be different relying on the volume of the corresponding compartment. The first step (binding to) could be regarded as occurring in the extracellular compartment with a specified first concentration of. The amount of in the extracellular compartment would then be estimated as the number of the receptors on the cell surface separated by the volume of development medium per cell. The second step could be the connection and dissociation of the receptor monomers takes place in the cell membrane. It is considered that all other steps are happening in the cytosolic compartment. As a result, the same mole number of would cause three concentrations which are constituting the different compartments. Though, for computational purposes, it is more suitable to manage only with a single concentration of associated to the cytoplasmic water volume of the cell. This requires rescaling the rate constants of steps 1 and 2. For the reason of this rescaling, the concentration in the representation was also associated with the cytoplasmic water volume (i.e. in the experimental medium was increased by the ratio (see Table 3.2)). Typically, there were in our experiments therefore. Assuming the diameter of a hepatocyte of and a cytoplasmic water volume of about of total intracellular volume, .

$\frac{V_m}{V_{cw}} = 33.3$ 34 05

Concentrations and the Michaelis constants (and) are given in nM. First- and second-order rate constants are expressed in and, respectively. , and are expressed in. =100, ,,,,. Medium concentration was multiplied by the factor to formally rescale it to the cytoplasmic water volume.

Reaction Number	Rate Equation	Parameter values
1	$k_1 * [R] * [EGF] - k_{-1} * [R_a]$	$k_1 = 0.003; k_{-1} = 0.06$
2	$k_2 * [R_a] * [R_a] - k_{-2} * [R_2]$	$k_2 = 0.01; k_{-2} = 0.1$
3	$k_3 * [R_2] - k_{-3} * [RP]$	$k_3 = 1; k_{-3} = 0.01$
4	$\frac{V_4 * [RP]}{(K_4 + [RP])}$	$V_4 = 450; K_4 = 50$
5	$k_5 * [RP] * [PLC\gamma] - k_{-5} * [R-PL]$	$k_5 = 0.06; k_{-5} = 0.2$
6	$k_6 * [R-PL] - k_{-6} * [R-PLP]$	$k_6 = 1; k_{-6} = 0.05$
7	$k_7 * [R-PLP] - k_{-7} * [RP] * [PLC\gamma P]$	$k_7 = 0.3; k_{-7} = 0.006$
8	$\frac{V_8 * [PLC\gamma P]}{(K_8 + [PLC\gamma P])}$	$V_8 = 1; K_8 = 100$
9	$k_9 * [RP] * [Grb] - k_{-9} * [R-G]$	$k_9 = 0.003; k_{-9} = 0.05$
10	$k_{10} * [R-G] * [SOS] - k_{-10} * [R-G-S]$	$k_{10} = 0.01; k_{-10} = 0.06$
11	$k_{11} * [R-G-S] - k_{-11} * [RP] * [G-S]$	$k_{11} = 0.03; k_{-11} = 4.5 * 10^{-3}$
12	$k_{12} * [G-S] - k_{-12} * [Grb] * [SOS]$	$k_{12} = 1.5 * 10^{-3}; k_{-12} = 10^{-4}$
13	$k_{13} * [RP] * [Shc] - k_{-13} * [R-Sh]$	$k_{13} = 0.09; k_{-13} = 0.6$

14	$k_{14}*[R-Sh]-k_{-14}*[R-ShP]$	$k_{14}=6; k_{-14}=0.06$
15	$k_{15}*[R-ShP]-k_{-15}*[ShP]*[RP]$	$k_{15}=0.3; k_{-15}=9*10^{-4}$
16	$\frac{V_{16}*[ShP]}{(K_{16}+[ShP])}$	$V_{16}=1.7; K_{16}=340$
17	$k_{17}*[R-ShP]*[Grb]-k_{17}*[R-Sh-G]$	$k_{17}=0.003; k_{17}=0.1$
18	$k_{18}*[R-Sh-G]-k_{-18}*[RP]*[Sh-G]$	$k_{18}=0.3; k_{-18}=9*10^{-4}$
19	$k_{19}*[R-Sh-G]*[SOS]-k_{-19}*[R-Sh-GS]$	$k_{19}=0.01; k_{-19}=2.14*10^{-2}$
20	$k_{20}*[R-Sh-G-S]-k_{-20}*[Sh-G-S]*[RP]$	$k_{20}=0.12; k_{-20}=2.4*10^{-4}$
21	$k_{21}*[ShP]*[Sh-G]-k_{-21}*[Sh-G]$	$k_{21}=0.003; k_{-21}=0.1$
22	$k_{22}*[Sh-G]*[SOS]-k_{-22}*[Sh-G-S]$	$k_{22}=0.03; k_{-22}=0.064$
23	$k_{23}*[Sh-G-S]-k_{-23}*[ShP]*[G-S]$	$k_{23}=0.1; k_{-23}=0.021$
24	$k_{24}*[R-ShP]*[G-S]-k_{-24}*[R-Sh-G-S]$	$k_{24}=0.009; k_{-24}=4.29*10^{-2}$
25	$k_{25}*[PLC\gamma P]-k_{-25}*[PLC\gamma P-I]$	$k_{25}=1; k_{-25}=0.03$

[Table 3.2: Rate equations and parameter values of the kinetic model]

3.3.4 Conserved Moieties

$[EGFR]_{total}$

In the reaction network described by the equations which are presented in Table3.2, the moiety and the protein moieties are conserved. This hypothesis is excused for the short term responses considered here. Let be the total concentrations of forms. Then the following is true:

$$[EGFR]_{total} = [R] + [R_a] + 2*([R_2] + [RP] + [R-PL] + [R-PLP] + [R-G] +$$

$$+[R-G-S] + [R-Sh] + [R-ShP] + [R-Sh-G] + [R-Sh-G-S])$$

$$[EGF]_{total}$$

Accepting that out of the total of receptors/cell is showed on the cell membrane, the total concentration of surface-expressed, converted to the cytoplasm water volume, is about. Five other moieties sustained in the signalling reactions involve the total concentrations of,,and proteins and,designated below by ,,,, and , orderly *(Table 3.2)*.

$$[EGF]_{total} = [EGF] + [R_a] + 2*([R_2] + [RP] + [R-PLP] + [R-PL] +$$

$$+[R-Sh] + [R-ShP] + [R-G] + [R-G-S] + [R-Sh-G] + [R-Sh-G-S])$$

$$[PLC\gamma]_{total} = [R-PL] + [R-PLP] + [PLC\gamma] + [PLC\gamma P] + [PLC\gamma P-I]$$

$$[Grb2]_{total} = [Grb] + [G-S] + [Sh-G] + [Sh-G-S] + [R-G] + [R-G-S] +$$

$$+[R-Sh-G] + [R-Sh-G-S]$$

$$[Shc]_{total} = [Shc] + [ShP] + [Sh-G] + [Sh-G-S] + [R-Sh] + [R-ShP] +$$

$$+[R-Sh-G] + [R-Sh-G-S]$$

$$[SOS]_{total} = [SOS] + [G-S] + [Sh-G-S] + [R-G-S] + [R-Sh-G-S]$$

3.3.5 Thermodynamic Restrictions along Cyclic Pathways in the Kinetic Scheme

$9-12$

If a kinetic scheme contains true cycles, where the first and last states are the same, the equilibrium parameters of the reactions along any cycle content called detailed balance relationships which depend on the product of the equilibrium constants along a cycle to be equal to 1, since at equilibrium the net flux through any cycle disappears. For that reason, such relations decrease the number of independent rate constants in a kinetic model. The kinetic scheme in figure 3.2 shows that the progression along steps (in the positive direction) completes a cycle without any concomitant transformations and changes in the free energy. Hence, the following restriction exists on the kinetic constants.

$$k_9 * k_{10} * k_{11} * k_{12} / (k_{-9} * k_{-10} * k_{-11} * k_{-12}) = 1$$

Further examination of the kinetic scheme in figure 3.2 shows additional reaction cycles that imply the following constraints.

$$k_{15} * k_{21} * k_{-17} * k_{-18} / (k_{-15} * k_{-21} * k_{17} * k_{18}) = 1$$

$$k_{18} * k_{22} * k_{-19} * k_{-20} / (k_{-18} * k_{-22} * k_{19} * k_{20}) = 1$$

$$k_{12} * k_{22} * k_{21} * k_{23} / (k_{-12} * k_{-22} * k_{-21} * k_{-23}) = 1$$

$$k_{15} * k_{-20} * k_{-23} * k_{-24} / (k_{-15} * k_{20} * k_{23} * k_{24}) = 1$$

3.3.6 EGF Binding Constants

$EGFR\ 6nM$

Reported values for binding to the solubilized extracellular domain of the receptor range from to, whereas full-length in plasma membrane vesicles has a substantially higher affinity

for, with an apparent of. In binding studies carried out on complete hepatocytes, values of and of and for high and low affinity sites, respectively, have been presented. Our recent data also demonstrate that in intact hepatocytes, autophosphorylation saturates at concentrations of, whereas in Triton-solubilized cells maximal activation of tyrosine kinase activity requires . These findings indicate that thefor in intact hepatocytes should be well below the value of measured for the solubilized receptor. We have used for binding to intact liver cells, in our kinetic model. This symbolizes an average value of literature data on binding studies in hepatocytes and is suited with our experimental data on

$$k_d = \frac{k_{-1}}{k_1} = 400nM \; s^{-1}$$

Since the values are valuable for the (quasi)equilibrium conditions, knowledge of the rate constants of the forward and backward reactions is needed to explain the temporal function. The association and dissociation steps are defined by second-order and first-order rate constants. For binding to the recombinant soluble extracellular binding domain of receptor, the ‘‘on’’ (association) and ‘‘off’’ (dissociation) rate constants were reported to be

$$ATP : ADP$$

Binding in a living cell, the concentration is much higher than the Michaelis constant of the receptor kinase for. Hence, the rate of tyrosine phosphorylation of the receptor (step 3) or bound target proteins (steps 6 and 14) can be kinetically characterized by pseudo-first order rate constants. Especially, the regular free energy dissimilarities of the tyrosine phosphorylation reactions are small, so that the equilibrium constants are of the order of unity. Therefore, the phosphorylation steps catalyzed by receptor kinase are regarded changeable (Table 3.2), and the effective rate constants will rely on the ratio, which is believed constant. By contrast, the phosphatase reactions (steps 4, 16, 8) can be considered as kinetically irreversible. The phosphatases are accepted to follow Michaelis-Menten kinetics, and the concentration of inorganic phosphate is considered constant.

$$k_d \; b2$$

Since data for on and off rate constants for the interactions with its target proteins in situ are unavailable, the corresponding rate constants were taken for granted to be in the same range as those reported for the binding of domains to phosphopeptides (Table 3.2). It has been reported that the - complex binds to both and -derived phosphopeptides with higher affinity than alone. These experimental data restrict the values of the corresponding reactions in the kinetic scheme (Figure 3.2) as follows:

$$\frac{k_d^{17}}{k_d^{24}} = 7.$$

Table (3.2) shows the rate expressions and kinetic constants of all of the reactions shown in Figure 3.2. [41]

3.4 Results

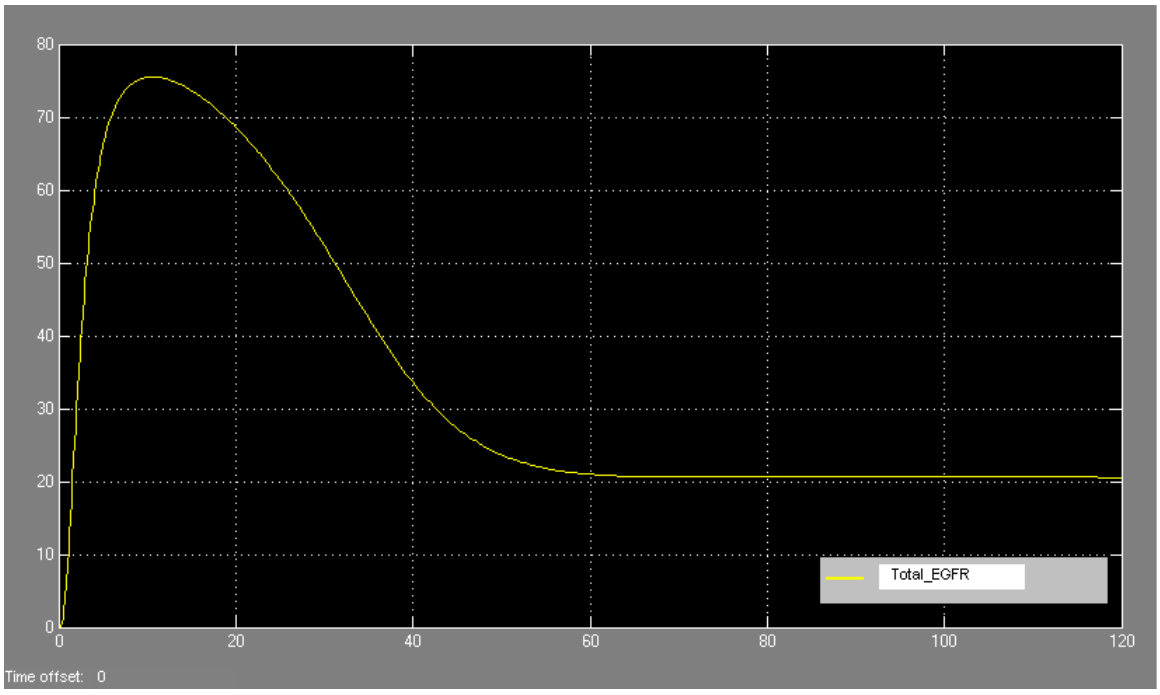
3.4.1 Experimental Analysis of EGFR Signalling

$EGFR$

The time course of the cellular response to was taken place after recently isolated hepatocytes by calculating tyrosine phosphorylation and protein-protein interactions of signalling intermediates showed in Figure 3.2 after stimulation with different concentrations. These experiments provided estimates of the phosphorylated protein as a fraction of total protein in the lysate. Figure 3.3 shows total phosphorylated as a fraction of the total protein in the lysate at different times after stimulation with. The kinetic scheme in figure 3.2 indicates that the following forms contributed to the bands of the phosphorylated receptor.

$$EGFR = 2*([RP]+[R-PL]+[R-PLP]+[R-G]+$$

$$+[R-G-S]+[R-Sh]+[R-ShP]+[R-Sh-G]+[R-Sh-G-S]$$



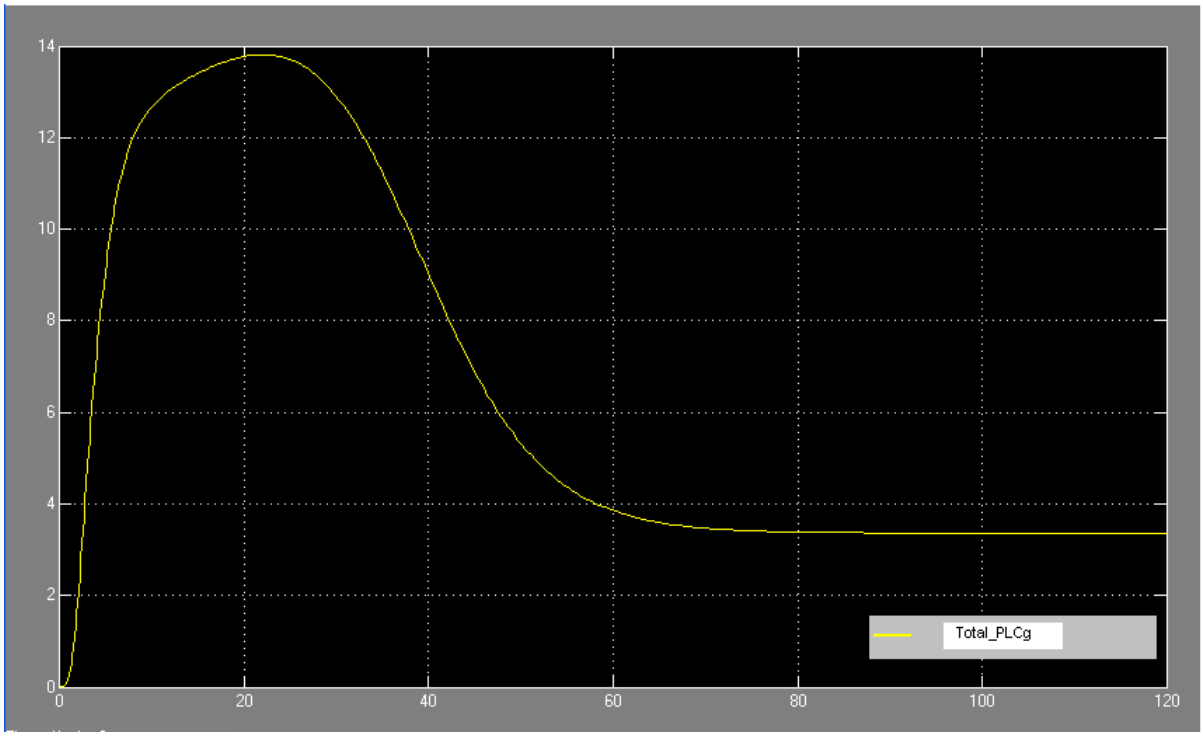
[Figure 3.3: Time course of EGFR autophosphorylation]

Shc R

Figures show the time course of the -induced activation of several downstream signalling events, as detected by tyrosine phosphorylation of (figure 3.4) and proteins (figure 3.5) and by co precipitation with (figure 3.6) and
(figure 3.7).

$$PLC\gamma = [R - PLP] + [PLC\gamma P]$$

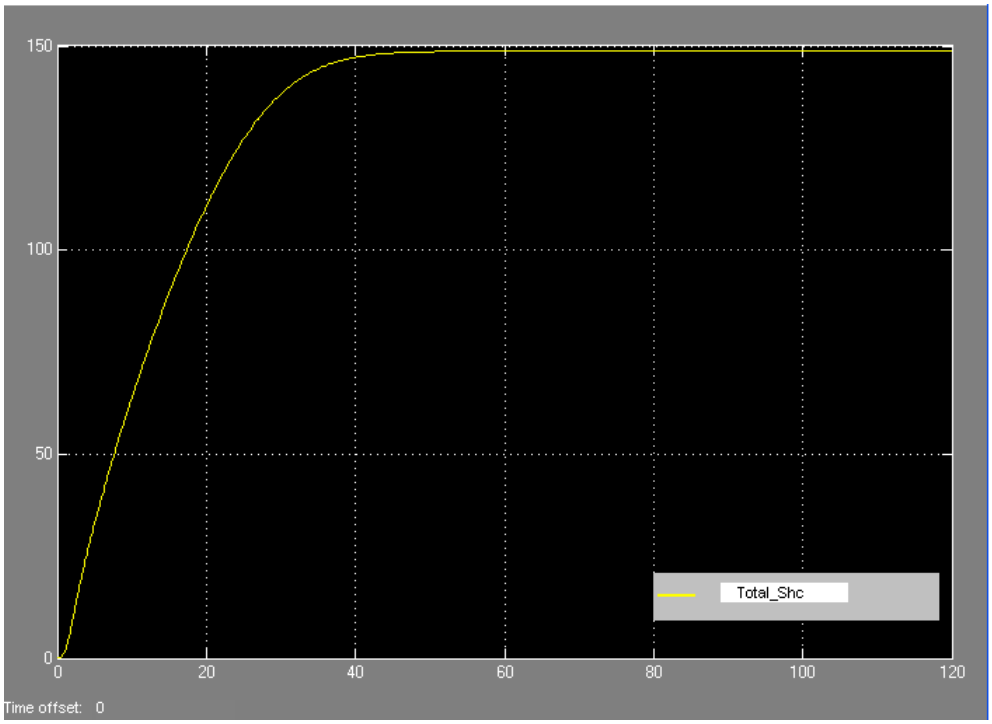
Total phosphorylated



[Figure 3.4: Time course of EGF-induced tyrosine phosphorylation, phosphorylated PLCγ]

$$Shc = [R - ShP] + [R - Sh - G] + [R - Sh - G - S] +$$

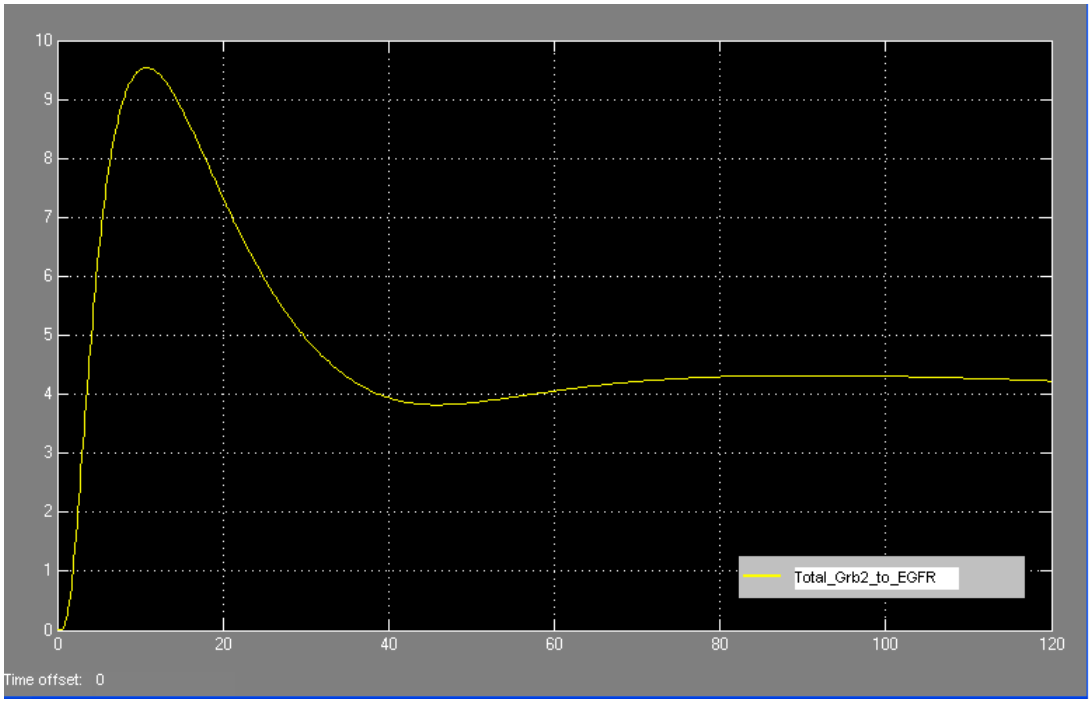
$$+[ShP] + [Sh - G] + [Sh - G - S]$$



[Figure 3.5: Time course of EGF-induced tyrosine phosphorylation, phosphorylated Shc]

$$= [R-G]+[R-G-S]+$$

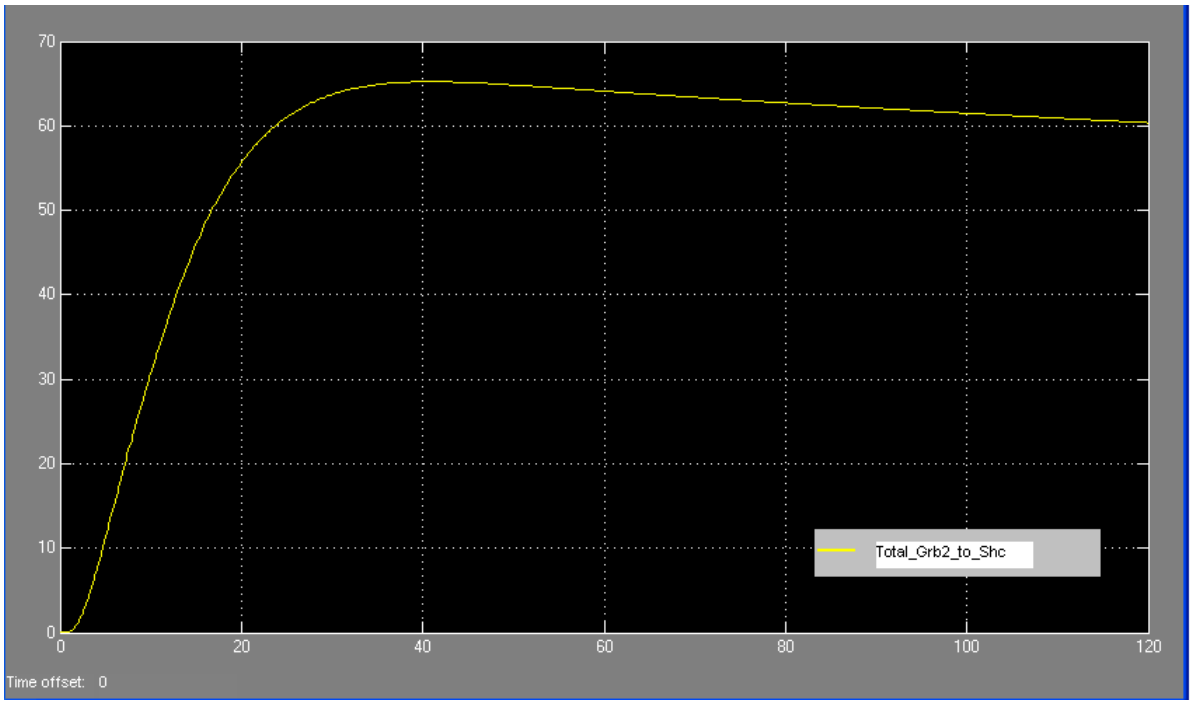
$$+[R-Sh-G]+[R-Sh-G-S]$$



[Figure 3.6: Time course of EGF-induced tyrosine phosphorylation, Grb2 co precipitation with EGFR]

$$= [R - Sh - G] + [Sh - G] +$$

$$+ [R - Sh - G - S] + [Sh - G - S]$$



[Figure 3.7: Time course of EGF-induced tyrosine phosphorylation, Grb2 co precipitation with

$EGFR$

Comparison with the total cell lysates analyzed in parallel on the same gels did quantification of the bands. In agreement with our earlier findings, a larger fraction of was bound to than to (figure 3.6-3.7). [41]

3.4.2 Computational Kinetic Analysis of EGFR Signalling

$Shc\ ShP] + [R - Sh - G] + [R - Sh - G - S]$

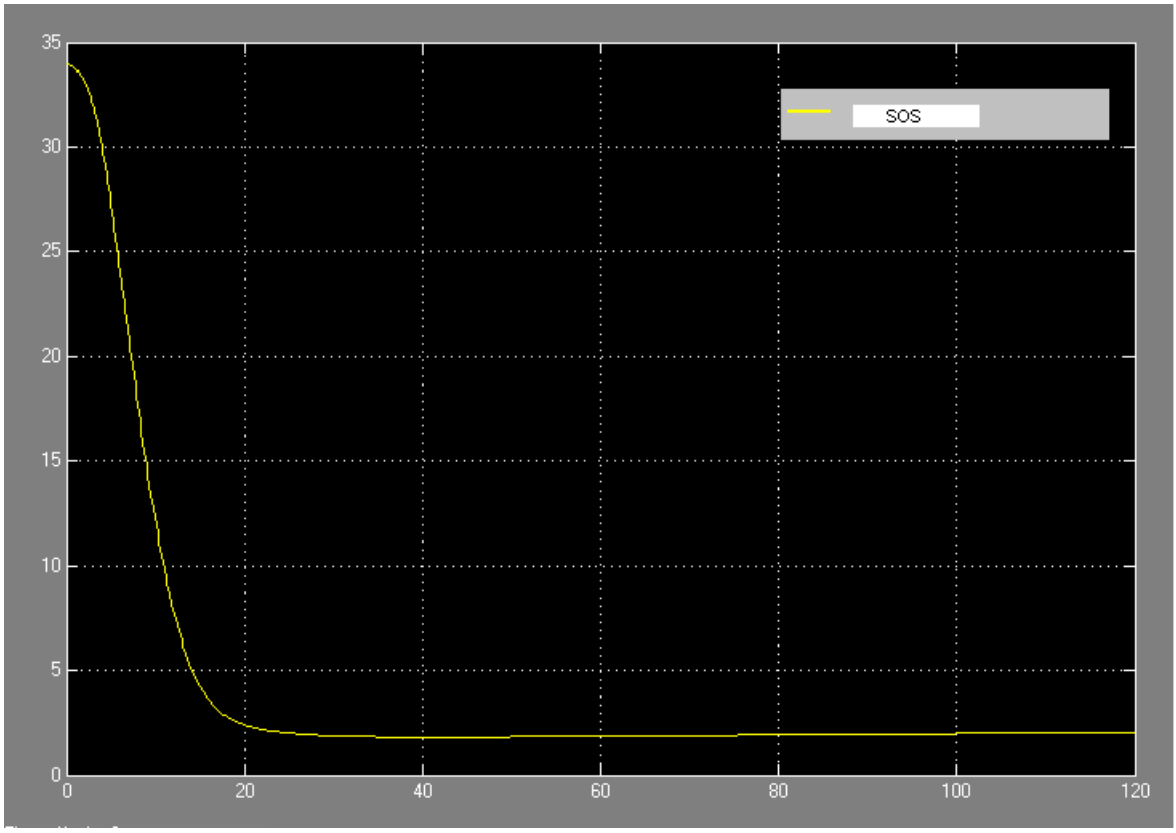
The total concentrations of phosphorylated and of co precipitated with phosphorylated do not exhibit a marked maximum (figure 3.5 and 3.7), and they reach a quasistationary level, in agreement with our experimental observations. Importantly, the kinetic model explains why these transients differ so markedly from the transients of the total phosphorylated (figure 3.3) and co precipitated with (figure 3.6). Computations show that the total phosphorylated bound to (i.e.) exhibit a pronounced peak, descending then to a low sustained level (figure 3.6). We come to the conclusion that the almost monotonic growth in the total phosphorylated is happened by the accumulation of the phosphorylated forms dissociated from the receptor,

$[ShP]+[Sh-G]+[Sh-G-S]$

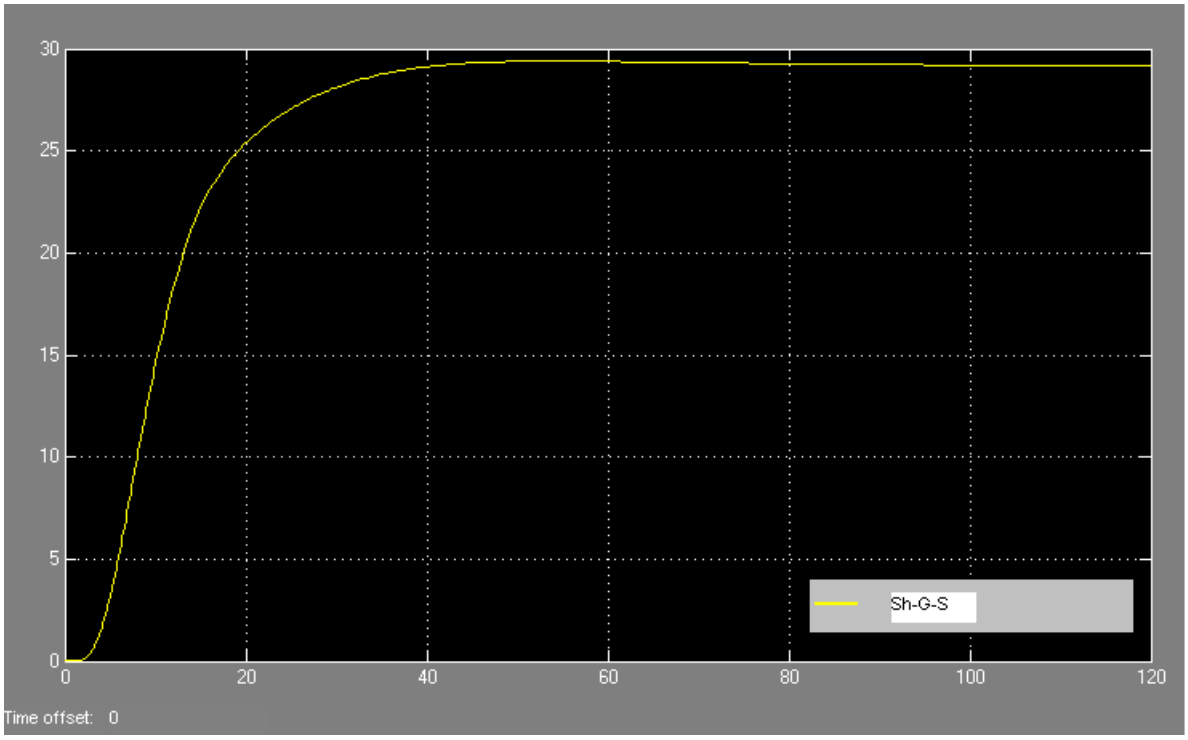
(i.e.).

Shc $Sh-G-S$

The total concentrations of bound to (and) and to the phosphorylated (Sh-G-S) exhibit transient and monotonic increases, respectively (figure 3.8-3.9).



[Figure 3.8: Computation of the time course of downstream EGF signaling, total SOS bound to EGFR]



[Figure 3.9: Computation of the time course of downstream EGF signaling, concentration of Sh-G-S complex]

EGFR

Using the kinetic model, it is helpful to observe how the rates of individual steps alter with time. The rates of steps of the receptor cycles including interaction with and increase to peak values and then shrink to low (less than) or near to zero sustained (stationary) values. Especially, the rates of phosphorylation of the receptor (by intrinsic tyrosine kinase in step 3) and its dephosphorylation by phosphotyrosine phosphatase(s) (step 4) do not decrease to zero with time (figure 3.6) but they reach rather high sustained values.

PLCγP

Figure 3.4 examines the time course of the total phosphorylated when he accepts a translocation (step 25) of to a structural element of the cell (solid lines) and in the absence of such a process.

EGFR

The dynamics of the signalling presents to be robust to significant alterations in the rate constants of the protein interactions included. Usually, a several fold (in many cases 1 or even 2 orders of magnitude) variation of a rate constant does not occur in significant changes of the response to. The kinetic model emphasizes that the dynamic pattern of signal propagation strongly relies on the relative abundance of molecular factors included in the

EGFR

Our results make it possible to predict and interpret the time course of -activated proteins and protein complexes, even those that were not considered explicitly in this paper. [41]

Chapter 4

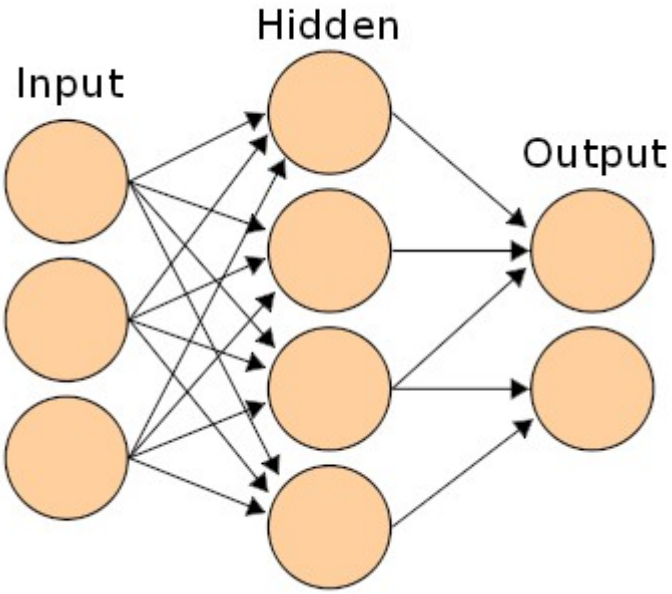
NEURAL NETWORKS AND THE RISE OBSERVER MODEL

4.1 Introduction to Neural Networks

Neural networks are flexible statistical models which are operated like the structure of the brain. They are considered flexible because they can learn to calculate roughly the parameters of some populations as using a small number of its exemplars (one or a few) in sequence. They are not altered typically by the usual statistical models. For example, we can discover neural network architectures associated with discriminate analysis, principal component analysis, logistic regression, and other techniques. In fact, neural networks which are used as statistical tools in a wide range of disciplines, including psychology, statistics, engineering, econometrics, and even the physics are useful for examining the standard statistical models. In the past, the term "neural network" was linked to a network or circuitry of biological neurons, while the modern usage of the term "neural network" is often associated with artificial neural networks, which are formed by artificial neurons or nodes. [42]

McCulloch and Pitts in 1943 introduced the simplified neurons were widely known as artificial neural networks. These neurons are presented as models of biological neurons and as components for circuits that could calculate a variety of tasks. The fundamental type of the neuron is discovered upon the functionality of a biological neuron. Neurons are the most important signalling units of the nervous system and each neuron is a distinct cell whose various processes appear in its cell body. In each structure, the neuron has four main areas. The cell body has two offshoots of this, the dendrites and axis, which results in presynaptic terminals. The cell body is the heart of the cell containing the nucleus and maintaining a good protein synthesis. [43]

An artificial neural network is a system based on the operation of biological neural networks, otherwise we could say that this is a simulation of biological nervous system. Artificial neural network models can perform, among others, non-linear approximation function, data classification, clustering and nonparametric regression or as simulations of group behavior model biological neurons. These are not simulations of real neurons because they do not model the biology, chemistry, or physics of a real neuron. They do, however, model many aspects of information by combining the pattern recognition behaviour of real neurons in a meaningful way. Neural models are very suitable for emulation, analysis, prediction, and association. The major applications of ANNs makes them useful in many ways such as to find patterns in data by memorizing characteristics and features of given data and make associations from new data to the old data, to learn and reproduce rules or operations from given examples, or to function approximation and modelling by analyzing and generalizing from sample facts and make predictions from these. [44]



[Figure 4.1: An artificial neural network is an interconnected group of nodes, akin to the vast network of neurons in the human brain.]

4.1.1 Applications of Neural networks

Artificial neural network models through observations can be used to construct a function. This is usefulness particularly in applications where you cannot design a function by hand because of the complexity of the data or task. A few representative examples of practical problems to which neural network analysis has been applied successfully are:

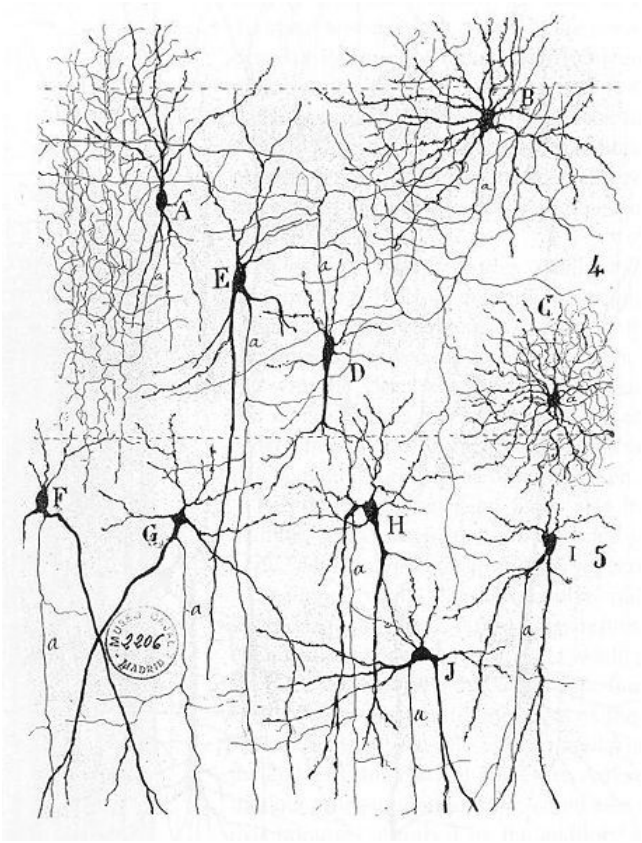
- In financial at stock market prediction and property appraisal : Neural networks can examine a lot of information quickly and sort it all out, and so are being used by many technical analysts to make prediction about stock prices based upon a large number of factors such as past performance of other stocks and various economic indicators.

- Detection and evaluation of medical phenomena, medical diagnosis and treatment cost estimation. Assisting doctors with their diagnosis by analyzing the reported symptoms and monitor many health-related indices. Neural networks have been used to recognize this predictive pattern so that the appropriate treatment can be prescribed.

- Engine management in energy demand forecasting and power control systems. The neural networks controls the various parameters within which the engine functions (the input of sensors from an engine), in order to achieve a particular goal such as minimizing fuel consumption.

- Credit worthiness. Decide whether an applicant for a loan is a good or bad credit risk and a variety of pieces of information are usually known. After training neural network on historical data, neural network analysis can identify the most relevant characteristics and use those to classify applicants as good or bad credit risks.

- Monitoring the condition of machinery and managerial decision making where the best decision option is selected by using the classification capabilities of neural network. Neural networks can be instrumental in cutting costs by bringing additional expertise to scheduling the preventive maintenance of machines. A neural network can be trained to distinguish between the sounds a machine makes when it is running normally (false alarms).[45]



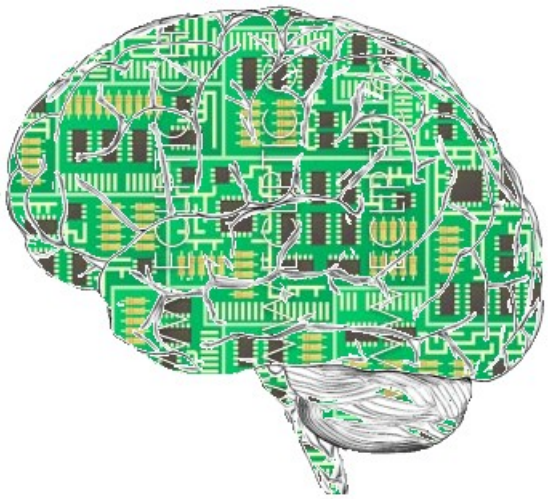
[Figure 4.2: From "Texture of the Nervous System of Man and the Vertebrates" by Santiago Ramón y Cajal. The figure illustrates the diversity of neuronal morphologies in the auditory cortex.]

4.1.2 The basic artificial model

An artificial neuron is a device with a number of inputs. Each input is weighted, and these weights correspond to a biological neuron. How strong the connection between an input and a neuron is is indicated by the value of the weight. Each neuron has also a single threshold value which makes the neuron very flexible and powerful one. In order to activate the neuron, the weighted sum of the inputs is formed, and the threshold is subtracted. Activation functions then, produce the output of the neuron with the activation signal, these describe an individual neuron. If a network has connections between the units and the propagation of data, there must be inputs (which carry the values of variables of interest in the outside world) and outputs (which form predictions, or control signals). A simple network has a feed forward structure where signals flow from inputs strictly feed forward from any hidden units, eventually reaching the output units. Such a structure has stable behaviour. However, if the network is recurrent (contains connections back from later to earlier neurons) it can be unstable, and has very complex dynamical properties. The typical feed forward network has neurons arranged in a distinct layered topology. The data processing can extend over multiple units, but no feedback connections are present. The

input layer is not really neural at all: these units simply shared to introduce the values of the input variables. The hidden and output layered neurons are each connected to all of the units in the presenting layer. Again, it is possible to define networks that are partially-connected to only some units in the presenting layer, however, for most applications fully-connected networks are better.

Executing the network, after the input variable values are placed in the input units, there is a progressively execution of the hidden and output layer units. Each unit calculates each activation value by subtracting the threshold from the weighted sum of the outputs of the units in the presenting layer. The activation value is passed through the activation function to produce the output of the neuron. When the entire network has been executed, the outputs of the output layer act as the output of the entire network. [45]



[Figure 4.3: There is no consensus on how closely the brain should be simulated.]

4.1.3 Neural Network Training

Definite programmed neural networks can perform a task by manually creating the topology and then setting the weight of each link and threshold. Nevertheless, neural networks have a unique strength: the ability to program themselves. For the training of a neural network, the most common method used is trial and error. Thus, by changing the weighting of a random link by a random amount can make the network behave the way it should. If that change cause the decline of the network accuracy, undo the chance and make a different one. Although it takes time, this method does produce result. There is no learning required for the fixed-way networks, so a learning mode is supervised or unsupervised.

At neural model development, supervised learning networks have been established, where at these supervised learning networks the training data consist of many pairs of input/output training patterns. For an unsupervised learning rule, an output unit is trained

to respond to clusters of pattern within the input. The network learns to adapt based on the experiences collected through the previous training pattern. [46]

It is necessary to mention that in the case of supervised learning there are two possible learning's by which the training patterns can be applied during training. Incremental and batch training mode. In incremental training the weights and biases of the network are updating each time a training pair is presented to the network. In batch training the weights and biases are only updated after all of the training pairs are presented.

4.2 An Observer Approach

PNN

An observer-based deterministic learning approach is presented in ''An Observer Approach for Deterministic Learning with Patchy Neural Networks''. The proposed scheme consists of an observer providing asymptotic estimates of the unknown nonlinearly parameterized vector field. The observer's output and the state vector measurements are used subsequently by a localized to extract the general form of the nonlinearity in the regions wherein the state trajectories remain. Detailed analysis demonstrates the capability of learning with limited assumptions. A simulation study was also carried out that verifies our theoretical results .An observer that employs the integral of the sign of the error term, an asymptotic estimation of the respective nonlinear vector field is achieved. Patchy Neural Networks () are then introduced to learn the unknown nonlinearity from the observer's output and the state observations. weights are updated algebraically reducing significantly the computational load of learning.

4.2.1 Introduction

Artificial neural networks have been efficiently used for deterministic learning in the context of adaptive control [48] and for computational or statistical learning in the context of machine learning [49]. A deterministic learning theory [50], [51] was proposed recently and applied to the dynamical pattern recognition problem [52], [53] in which, the authors address the dynamical pattern recognition problem of the temporal patterns generated from the dynamical system

$(1)\,{}_0) = x_0\,(t); p)$

$f_i(x, p) = [f_1(x; p) f_2(x; p) \dots f_n(x; p)]^T$

Where is the state vector, is a vector with system parameters and represents the system dynamics with a smooth, unknown, nonlinear function. Dynamical patterns are defined as general recurrent trajectories from (1) and include among others periodic, quasi-periodic or even chaotic trajectories. As described in [52] the pattern recognition process involves two main tasks: an initial identification task and a recognition task.

NN

For the identification task, a deterministic learning approach based on localised radial basis function neural networks () is adopted, in [50], [51]. Using this learning scheme, information on the dynamical pattern is obtained and stored in the weights. After the identification procedure, a set of dynamical models (the "test set") is constructed. These models are then employed for pattern recognition using comparisons between the actual and the test patterns (generated by the test models) based on some suitable similarity measure. In [52], [53] is given a detailed description of the overall methodology.

$f(x; p)$

The "An Observer Approach for Deterministic Learning with Patchy Neural Networks" paper, focus on the identification task and propose an alternative observer-based approach to deterministic learning. As a first step, an observer is designed based on the robust integral of the sign error (RISE) approach [59], [60], [61]. In this way, an asymptotic estimate of the smooth vector field is obtained. A localized neural network can then be employed to extract and store the information of this estimate.

PNN

A new class of localized neural networks called *patchy neural networks* () is then introduce, with basic functions that are "patches" of the state space. Their universal approximation capability is proved i.e. it is shown that a with a sufficient number of nodes can approximate with desired precision within some compact region a general smooth nonlinear function. Then, a simple is employed to store the information obtained from the observer estimate with an easily implementable algebraic weight update law. The advantages of the proposed

PNN

- There is no need for a persistence of excitation condition for the state vector. Theis capable of learning the unknown nonlinearity in some region of the state space from a single visit of the state trajectories to the patches of the region. This is in contrast to standard deterministic learning schemes [50], [51], [52], wherein a recurrence condition for the state trajectories is necessary for learning in those neighbourhoods

$N_1 \times \dots \times N_n$

- The weight update laws are defined by an algebraic form and are not given in the form of ordinary differential equations () as in [51]. This results in a significant

reduction of the computational cost for learning since we must only solve from the observer, opposed to needed to train the weights (for awith nodes).

4.2.2 Learning Problem Formulation

$x(t; x_0; p_i)$

In pattern recognition of temporal patterns generated by a time-invariant dynamical system in the form of (1), the first step is the -based information of the unknown nonlinearity. Using a test temporal pattern, a localized network is trained to estimate along the trajectory. Then, for each training pattern, N dynamical models are constructed [52]

$$\dot{\bar{x}} = -A(\bar{x} - x) + \bar{W}_i^T P(x), (1 \le i \le N) \quad (2)$$

$$\bar{W}_i := [\bar{W}^i_1, \ldots, \bar{W}^i_n] \in \ll^{kn}$$

with design constants and the weight matrix that corresponds to the -th test pattern.

$$\| x(t) - \bar{x}^i(t) \|$$

In rapid dynamical pattern recognition of temporal pattern, the most similar from the N test

$f(x)(x)$

Thus, from a set of test models, the one describing better a specific temporal pattern is the one with smallest synchronization error. This paper ("An Observer Approach for Deterministic Learning with Patchy Neural Networks"), *propose a novel approach for obtaining an estimator of of* (1). Rapid dynamical pattern recognition can then be carried

c_f, c_f', c_f''

Assumption 1: the state trajectories are bounded within some compact set and there exist

$$c_f' := \max_{1 \le i,j \le n} \sup_{x \in I} | \partial f_i(x) / \partial x_j | < \infty$$

$$c_f'' := \max_{1 \le i,j,k \le n} \sup_{x \in I} | \partial^2 f_i(x) / \partial x_j \partial x_k | < \infty$$

and .

4.2.3 Patchy Neural Network

$I_i = I_1 \times I_2 \times \ldots \times I_n$

Localized are introduced that are relatively simple and easy to train. Let some -dimensional

$a_{i,j} = a_{i,0} + j\delta (1 \le i \le n)$

$A_{1,i_1} \times \ldots \times A_{n,i_n} (1 \le i_j \le N_i, 1 \le i \le n)$

On the sets we can now defined the “patch” functions

$$(3) \quad p_{i_1,i_2,\ldots,i_n}(x) = \begin{cases} 1, x \in A_{1,i_1} \times \ldots \times A_{n,i_n} \\ 0, else \end{cases}$$

.

NN

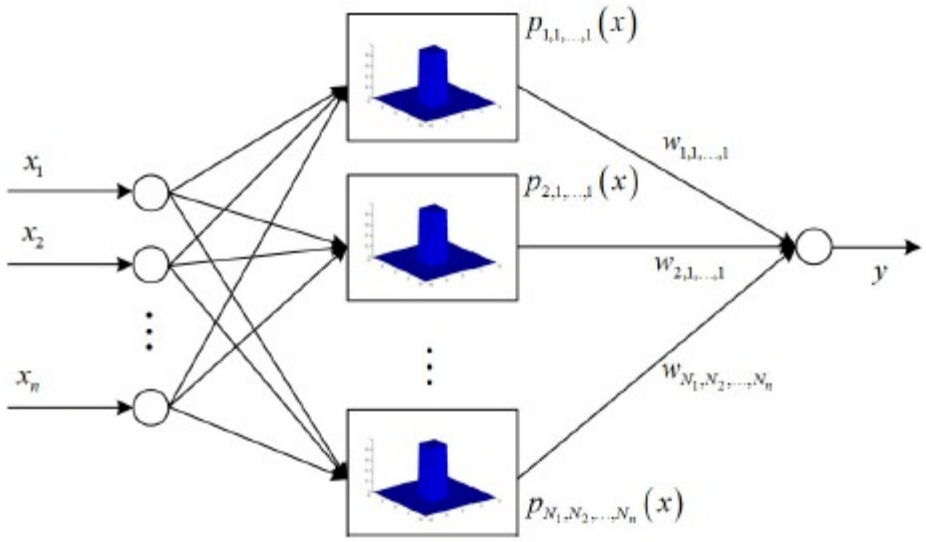
A patchy neural network is a single-hidden layer with basis vector consisting of “patch”

$$(4) \quad y = \sum_{i_1=1}^{N_1} \cdots \sum_{i_n=1}^{N_n} w_{i_1,\ldots,i_n} p_{i_1,\ldots,i_n}(x) = W^T P(x)$$

$P(x) := [p_{1,\ldots,1}(x), \ldots, p_{N_1,\ldots,N_n}(x)]^T$ N_n

PNN

A graphical representation of the is shown in Figure 4.4.



[Figure 4.4: The patchy neural network (PNN).]

$W \in \ll^{N_1 \times \ldots \times N_n} \ll$

The has the ability to approximate general nonlinear functions. Consider a nonlinear function defined on some -dimensional rectangle. Then, for every there exists a with nodes (patch functions) and a weight vector such that

$$(5)\; x) = W^T P(x) + \epsilon_a\ (x)$$

$$R_{i_1 \ldots i_n} := A_{1 i_1} \times \ldots \times A_{n i_n} \; N_1, N_2, \ldots, N_n\ (x)]^T$$

with and approximation error . This is a direct consequence of the patch function definition

$$(6)\; x) = f(x_{i_1 \ldots i_n}) + \frac{\partial f}{\partial x}(\theta_{i_1, \ldots i_n} x + (1 - \theta_{i_1 \ldots i_n}) x_{i_1, \ldots i_n})(x - x_{i_1 \ldots i_n})$$

$$x_{i_1 \ldots i_n} \in R_{i_1, \ldots i_n}$$

$$\| x - x_{i_1, \ldots i_n} \| \leq \prod_{j=1}^{n} (a_{j i_j} - a_{j i_{j-1}}) \forall x, x_{i_1 \ldots i_n} \in R_{i_1 \ldots i_n}$$

$$w_{i_1, \ldots i_n} = f(x_{i_1 \ldots i_n})$$

$$| f(x) - W^T P(x) | \leq \max_{\substack{i_r \in \{1,\ldots,N_r\} \\ 1 \leq r \leq n}} \prod_{j=1}^{n} (a_{j,i_j} - a_{j,i_j-1})$$

. (7)

$$c' := \max_{x \in I} \| \frac{\partial f(x)}{\partial x} \|$$

$$N_1, \ldots, N_n$$

$$\max_{\substack{i_r \in \{1,\ldots,N_r\} \\ 1 \leq r \leq n}} \prod_{j=1}^{n} (a_{j,i_j} - a_{j,i_j-1}) \leq \varepsilon / c'$$

that yields the desired approximation property.

4.2.4 Identification Scheme

$$f(x) \quad f(x(t))$$

The overall scheme consists of two distinct estimation tasks. Initially, an asymptotic observer is proposed based on the robust integral of the sign error (RISE) approach [59], [60], [61] that identifies the nonlinear vector field. Then, a is employed to extract the general form of based on the observer's output and the state vector measurements.

A. RISE Observer

(8)
$$\dot{\hat{x}}(t) = \xi(t) := k(x(t) - \hat{x}(t)) + \lambda k \int_0^t (x(s) - \hat{x}(s))ds + \beta \int_0^t \operatorname{sgn}(x(s) - \hat{x}(s))ds$$

$$k, \lambda, \beta > 0$$

with constants for RISE observer (8) of system (1), the following Lemma can be proved.

$$k > \lambda > 0$$

Lemma 1: let the nonlinear dynamical system (1) satisfying Assumption 1 and the RISE observer described by (8), respectively. Then, for the selection and

(9)
$$\rho \geq n c_f c_f' + (\frac{1}{\lambda}) n^2 c_f (c_f c_f' + c'^2 f)$$

$$w_m := \sup_{1 \le i,j \le N} | w_{i,j} |$$

Where

$f(x)$ $t))$

Remark 1: The most important property of the RISE observer (8) is that it provides an asymptotic time estimate of the respective vector field that can be subsequently used for training a localized to extract the general form of.

PNN

B. Weight Update

PNN $\nu | \nu = x(t), t \in [0, +\infty)\}$

Lemma 1 ensures that the vector signal of the RISE observer converges asymptotically to the vector field. The observer's output and the state vector measurements can now be used to train a to obtain the general form of at a neighbourhood of the region of the state space defined by the state trajectories. The weights are updated algebraically as follows

$$\hat{w}^j_{i_1,\ldots,i_n}(t) = (1 - p_{i_1,\ldots,i_n}(x(t)))\, \hat{w}^j_{i_1,\ldots,i_n}(t^-) + p_{i_1,\ldots,i_n}(x(t))\xi^j_i(t)$$

(10)

$$(11)\quad w^j_{i_1,\ldots,i_n}(t) = \begin{cases} \hat{w}^j_{i_1,\ldots,i_n}(t^-), & p_{i_1,\ldots,i_n}(x(t)) = 0 \\ \xi^j_j(t), & p_{i_1,\ldots,i_n}(x(t)) = 1 \end{cases}$$

$$\hat{w}^j_{i_1,\ldots,i_n}(0) = 0 (1 \le i_1 \le N_1, \ldots 1 \le i_n \le N_n; 1 \le j \le n)$$

$$f(x) \quad \hat{w}_i := [\hat{w}^i_{1,\ldots,1}, \ldots, \hat{w}^j_{N_1,\ldots,N_n}]^T$$

Then, the vector with, can be used to estimate.

4.2.5 PNN Estimation

$t \le T \le 0\}\{v \in R_{i_1,\ldots,i_n} \subset I \mid \exists i_j \in \{1,\ldots,N_j\}, 1 \le j \le n,$

The unknown nonlinear vector field elements can be approximated by in a region of the state trajectories s.t. with s.t.for some -partitioned and .

$s' \in R(x(\tau);\delta(\varepsilon))\}$

$s \in R_{\delta(\varepsilon)}(T(\varepsilon),t) := \{s' \in I \subset \ll^n \mid \exists \tau \in [\mathrm{T}(\varepsilon),t]$

$\sqrt{2c_f n^{\beta/2}}$

Theorem 1: Let the nonlinear dynamical system (1) satisfying Assumption 1 and the RISE observer (8), (9). Then, for every, there exist a time and a with patch length and weight update law (10) such that for all and s.t..

$f(x(t))$

$T(\varepsilon),t)$

Remark 2: practically, the above theorem states, that we can select a sufficiently dense partition of the state space (a small enough) so that the can approximate efficiently the nonlinear function f in the region that is the union of individual patches from which the state trajectories pass during. Time is defined as the time needed for the RISE observer

$T(\varepsilon) \ge 0 \dot{\xi}(t) - f(x(t))) = 0$

$\| \xi(t) - f(x(t)) \| \le \varepsilon / 2 \forall t \ge T(\varepsilon)$

$\tau_{[t_1,t_2]}(s,\delta) := \sup\{\tau \in [t_1,t_2] \mid s \in R(x(\tau);\delta)\}$

.

$\hat{W}^T(t) P(s) = \xi(\tau_{[T(\varepsilon),t]}(s,\delta))$

Then, for every it holds true from (11) that
Equivalently, we have that

(12)

$= \sup_{x \in I} \| \frac{\partial f}{\partial x} \|_F \| s - x(\tau_{[T(\varepsilon),t]}(s,\delta)) \| + \frac{\varepsilon}{2} \le n^{3\beta} c'_f \delta + \frac{\varepsilon}{2} \le \varepsilon$ $x(\tau_{[T(\varepsilon),t]}(s,\delta))) - \xi(\tau_{[T(\varepsilon),t]}(s,\delta)) \|$

.

$I \subset \ll$

Remark 3: From theorem 1 one can see that for a complete learning of some nonlinearity within some, the dynamical system (1) should be mixing, i.e. the state trajectories should visit neighbourhoods of all the points in. This property appears in chaotic systems!

4.2.6 Simulation Study

To verify the effectiveness of the proposed scheme, consider the learning problem for the

$$\dot{x} = y$$

$$\dot{y} = -p_2 x - p_3 x^3 - p_1 y + q\cos(\omega t) \quad (13)$$

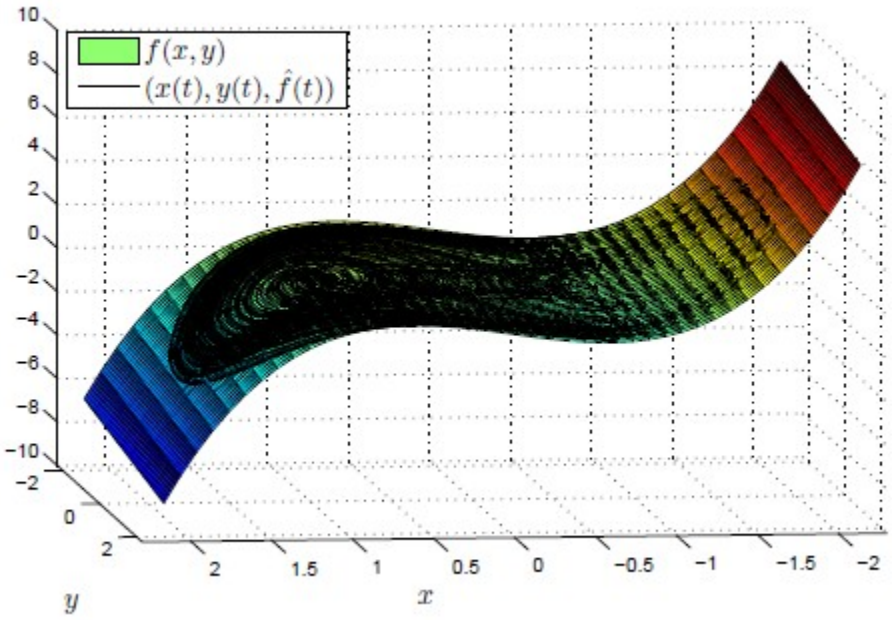
$$f(x,y) := -p_2 x - p_3 x^3 - p_1 y$$

with initial conditions, and parameters ,, , and . address the learning problem of of (13). The RISE observer

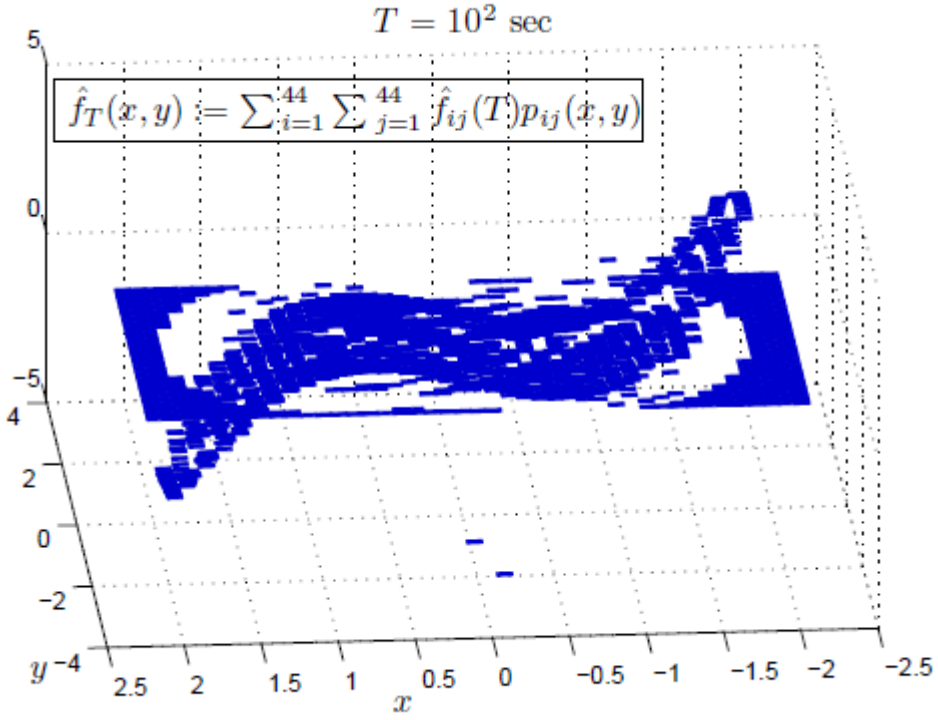
$$\dot{\hat{y}} = \cos(\omega t) + k(y(t) - \hat{y}(t)) + k\lambda \int_0^t (y(s) - \hat{y}(s))ds + \rho \int_0^t \operatorname{sgn}(y(s) - \hat{y}(s))ds \quad (14)$$

$$\hat{f}_t(x,y) = \sum_{i=1}^{44} \sum_{j=1}^{44} \hat{w}_{i,j}(t) p_{i,j}(x,y)$$

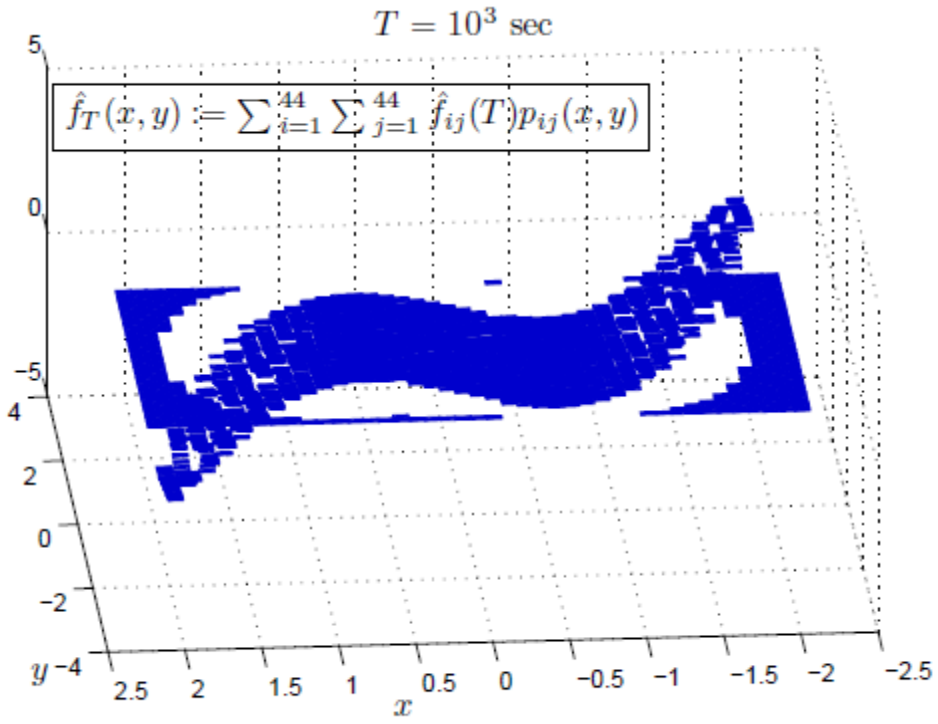
With parameters, and β=20.The square is divided into patches with patch length and define the associated with update law (10) and output. Simulation results are shown in fig.4.5-7.



[Figure 4.5: Graph of function f(x,y) and the trajectories of {xt,yt, ξtT≤t≤103}.]



[Figure 4.6: PNN output ftx,y=i=144j=144wi,jtpi,j(x,y) graph for t=102.]



[Figure 4.7: PNN output ftx,y=i=144j=144wi,jtpi,j(x,y) graph fort=103.]

$$r(t) := \ddot{x} + \lambda \ddot{x}(t) \quad)$$

Proof of Lemma 1: Let as define the observation error and the filtered observation error. Then, for the nonnegative function

$$(15) \quad V(t) := \frac{\lambda^2}{2} \tilde{x}^T(t)\, \ddot{x}(t) + \frac{1}{2} r^T(t) r(t)$$

$$(16) \quad \dot{V} = -\lambda^3 \tilde{x}^T(t)\, \ddot{x}(t) + r^T(t)[\dot{x}(t) - \ddot{x}(t) + \lambda \dot{r}(t)]$$

.

$$\ddot{x}(t) = -k\dot{r}(t) - \beta \operatorname{sgn}(\dddot{x}(t)) \tag{17}$$

.

Using the above identity in (16) we result in

$$\dot{V} = -\lambda^3 \dddot{x}^T(t)\ddot{x}(t) - (k-\lambda)\dot{r}^T(t)r(t) + L(t) \tag{18}$$

$$L(t) := -r^T(t)[\ddot{x}(t) + \beta \operatorname{sgn}(\dddot{x}(t))]$$

with. Following a similar analysis to [60] we obtain

$$+\lambda \sum_{i=1}^{n} \int_0^t |\dddot{x}_i(s)|(|\ddot{x}_i(s)| + \frac{1}{\lambda}|\dddot{x}_i(s)| - \beta)ds \ddot{x}(0) + \dddot{x}(s)ds - \beta \sum_{i=1}^{n} |\dddot{x}_i(s)|\Big|_0^t))ds \tag{19}$$

$$\dot{x}_i(t) = f_i(x)$$

Since we have that

$$\ddot{x}_i(t) = \sum_{j=1}^{n} \frac{\partial f_i}{\partial x_j}(x) f_j(x) \tag{20}$$

$$\sup_{1 \le i \le n} \sup_{0 \le t \le \infty} |\dddot{x}_i(t)| \le n c_f c'_f \tag{21}$$

Further differentiation yields

$$\dddot{x}_i(t) = \sum_{k=1}^{n} \sum_{j=1}^{n} [\frac{\partial^2 f_i}{\partial x_j \partial x_k}(x) f_j(x) f_k(x) + \frac{\partial f_i}{\partial x_j}(x) \frac{\partial f_j}{dx_k}(k) f_k(x)] \tag{22}$$

and the following bound can be obtained directly

$$\sup_{1\le i\le n}\sup_{0\le t\le\infty}|\dddot{x}_i(t)|\le n^2 c_f(c_f c''_f + c'^2_f) \quad (23)$$

β

Thus, for satisfying (9) we have from (22), (23) and (19) that

$$\int_0^t L(s)ds \le \beta\sum_{i=1}^{n}|\ddot{x}_i(0)| + \ddot{x}^T(0)\dot{x}(0) \quad (24)$$

$P(t)$

The above inequality ensures the no negativity of defined by

$$P(t) := \beta\sum_{i=1}^{n}|\ddot{x}_i(0)| + \ddot{x}^T(0)\dot{x}(0) - \int_0^t L(s)ds \ge 0 \quad (25)$$

.

$$\bar{V}(t) := V(t) + P(t)$$

Let us now consider the nonnegative function. Then, from (18), (25) the dynamics of take the following form

$$\dot{V} = \lambda^3 \ddot{x}^T(t)\dot{x}(t) - (k-\lambda)r^T(t)r(t) \le 0 \quad (26)$$

$\bar{V}(t)$ $\bar{V}(t) = \bar{V}_\infty < \infty$

as long as . The fact that function is decreasing and bounded from below ensures

$\|r(t)\| \in L_\infty$

Thus, it holds true that,.

Integrating (26) we obtain

$$\lambda^3\int_0^\infty \ddot{x}^T(s)\dot{x}(s)ds + (k-\lambda)\int_0^\infty r^T(s)r(s)ds = \bar{V}(0) - \bar{V}_\infty < \infty \quad (27)$$

$$\|r(t)\| \in L_\infty \cap L_2$$

i.e. , .

$$\ddot{x}(t) = -\lambda \ddot{x}(t) + r(t) \tag{28}$$

$$\dddot{x}(t) = -(k-\lambda)r(t) - \lambda^2 \ddot{x}(t) - \beta \operatorname{sgn}(\dddot{x}(t)) - \ddot{x}(t) \tag{29}$$

$$\left\| \ddot{x}(t) \right\| \in L_\infty \cap L_2$$

$$\left\| \dot{r}(t) \right\| \in L_\infty$$

$$\lim_{t\to\infty} \ddot{x}(t) = \lim_{t\to\infty} r(t) = 0$$

$$\lim_{t\to\infty} (\xi(t) - f(x(t))) = \lim_{t\to\infty} (\dot{x}(t) - \dot{x}(t)) = \lim_{t\to\infty} \ddot{x}(t) = \lim_{t\to\infty} (r(t) - \lambda \ddot{x}(t)) = 0$$

that concludes the proof of Lemma 1.

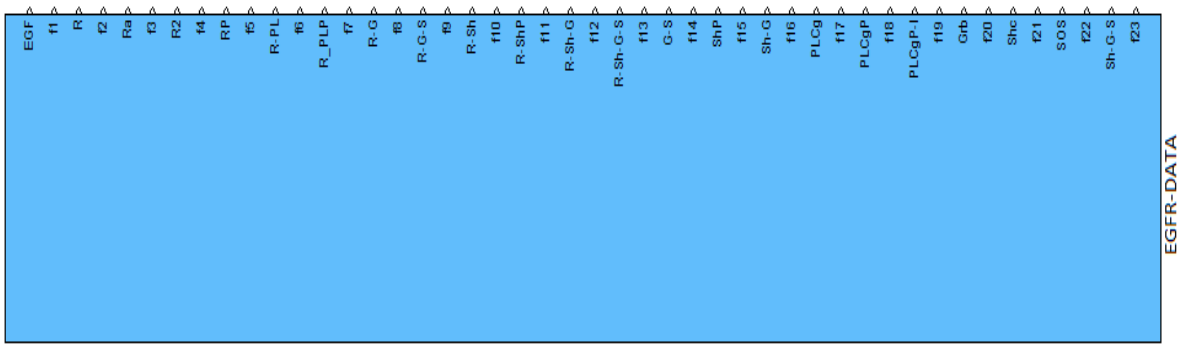
Chapter 5

IDENTIFICATION OF EGFR USING THE RISE OBSERVER

The software tool we use is the Matlab together with Simulink. Simulink is an environment powerful and easy to use for multidomain simulation and Model-Based Design for dynamic and embedded systems. Is an extension of MATLAB by Mathworks Inc. It works with MATLAB to offer modeling, simulation, and analysis of dynamical systems under a graphical user interface (GUI) environment. Its primary interface is a graphical block diagramming tool and a customizable set of block libraries Models are hierarchical, which allow using both top-down and bottom-up approaches. Simulink is widely used in control theory and digital signal processing for multidomain simulation and design.

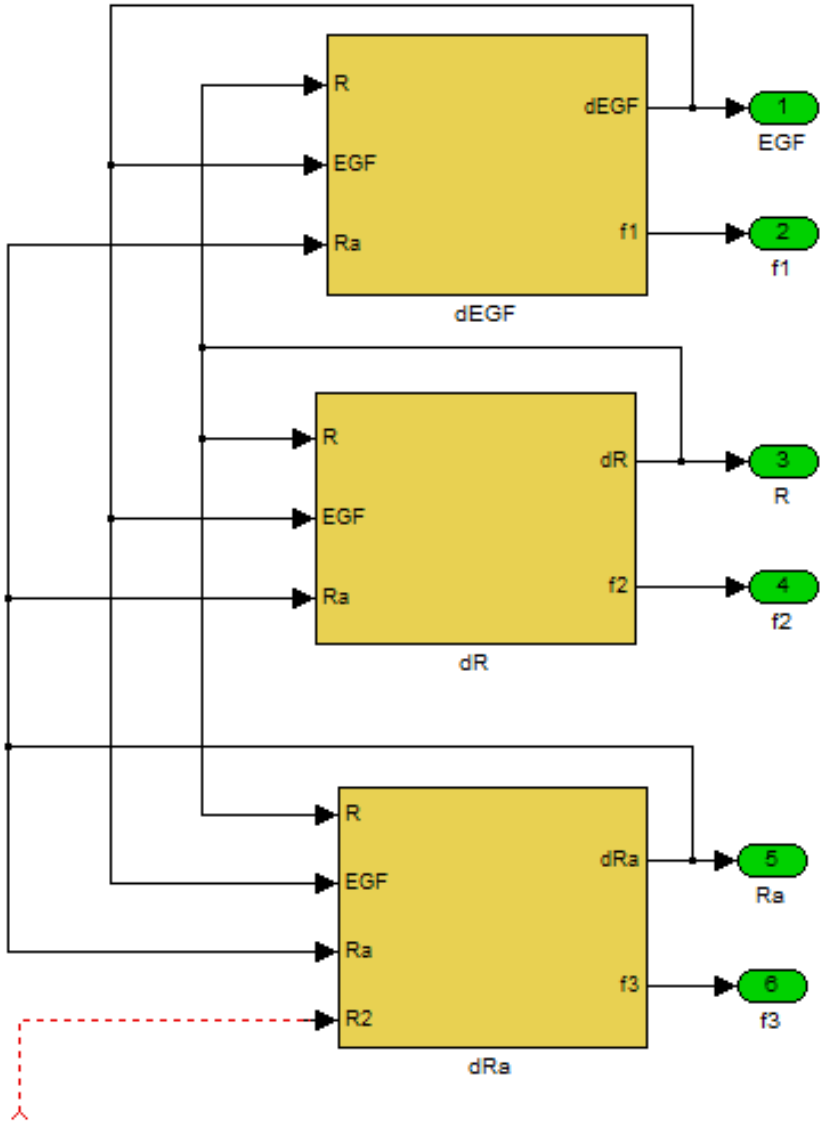
5.1 Implementation of RISE Observer

Originally, data are produced in order to observe the neural network. The stimulation of EGFR pathway is based on the kinetic model as we describe in chapter 3. It is clear that instead of using the numerical model in chapter 3, we could use the real data, if the labs that provided them to Kholodenko, could sent them to our laboratory. Figure 5.1 shows the Simulink block diagram which calculates the dynamic behavior of EGFR pathway.

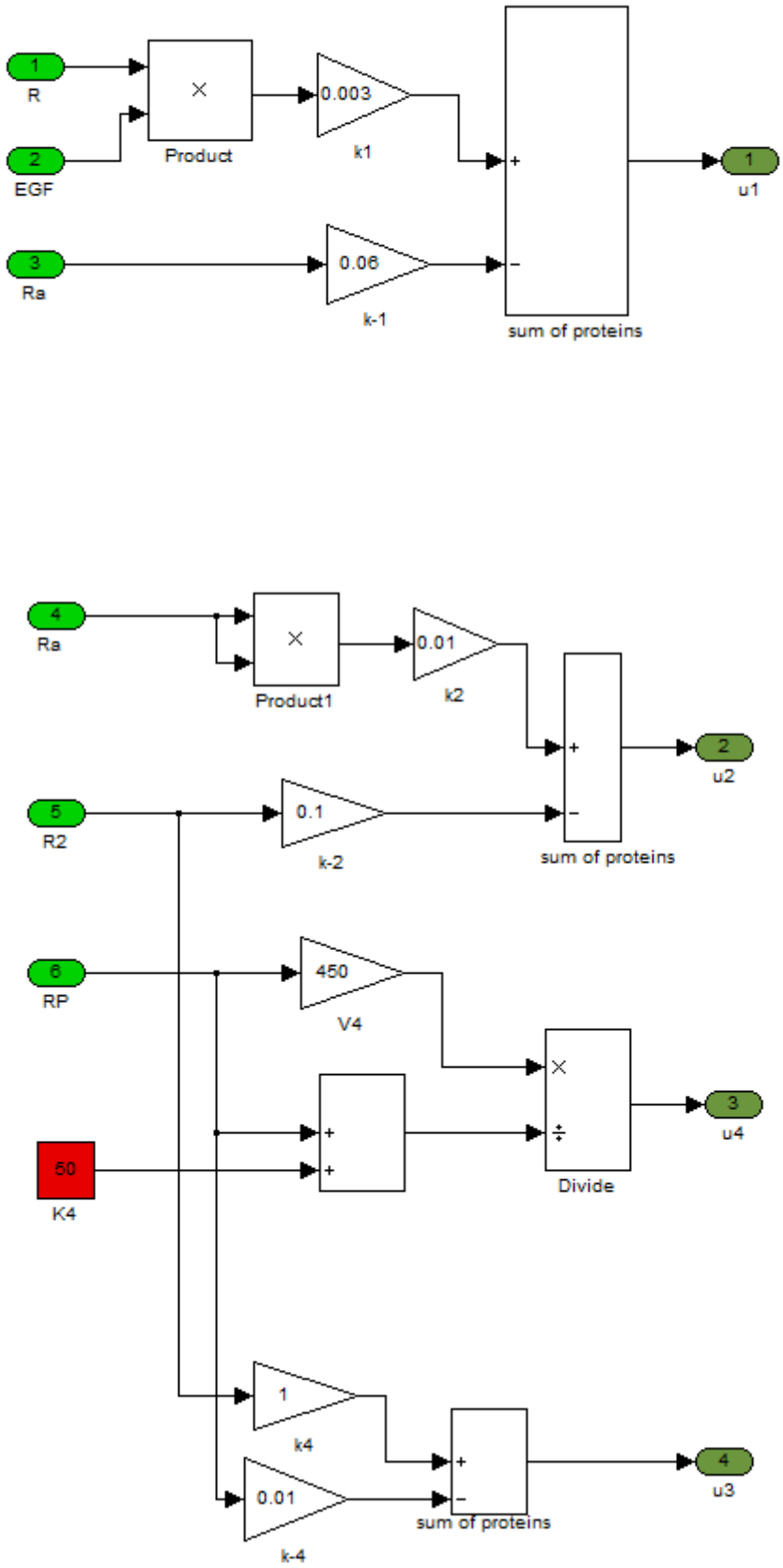


[Figure 5.1: Simulink Block to produce data according to the kinetic model from Kholodenko described in chapter3 for EGFR pathway.]

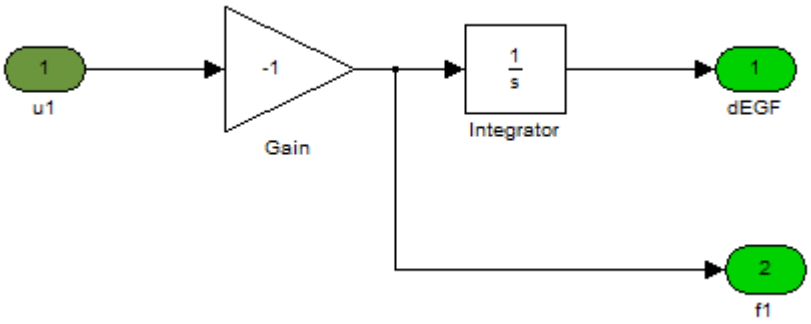
This block is composed of twenty three subsystems each one being divided by other smaller subsystems. Figure 5.2 shows some of those subsystems. The connections are not shown in this figure but they are clear from the names of inputs and outputs. Each subsystem first produces the rates of change for each protein concentration using as inputs the concentrations of the proteins as shown in figure 5.3. These rates are sums of products of appropriate concentrations. Next, each subsystem evaluates the concentration of the protein it represents as shown in figure 5.4. Protein concentrations are integrators outputs.

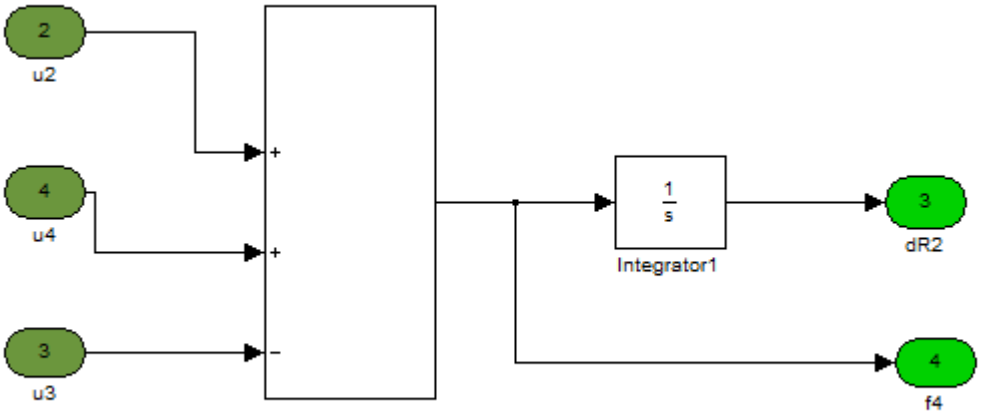


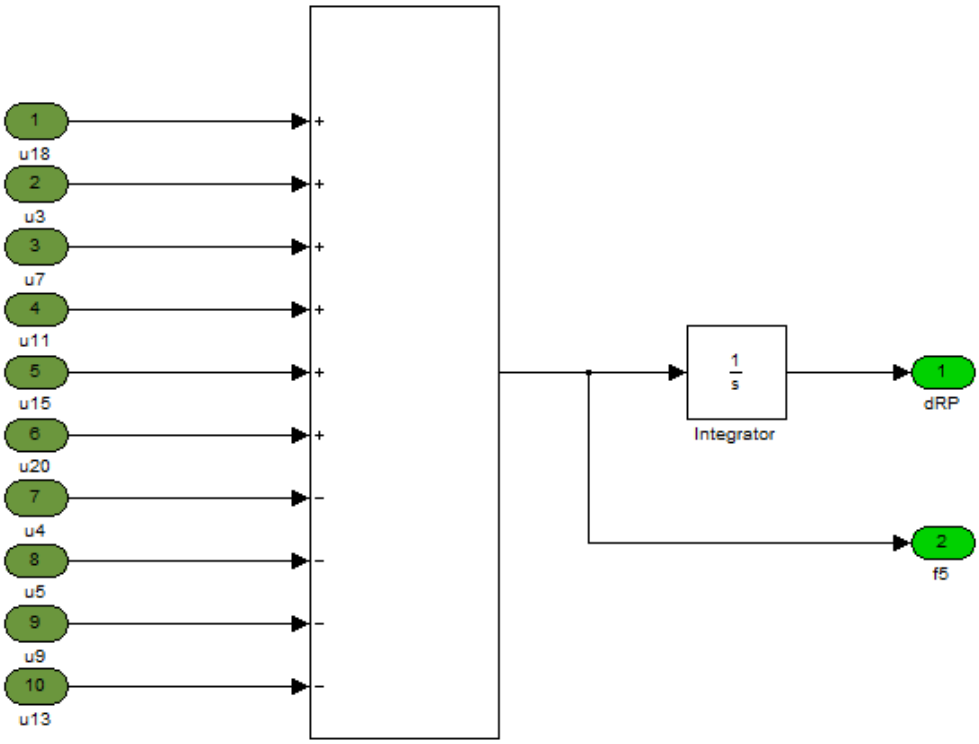
[Figure 5.2: Three of the twenty three subsystems.]



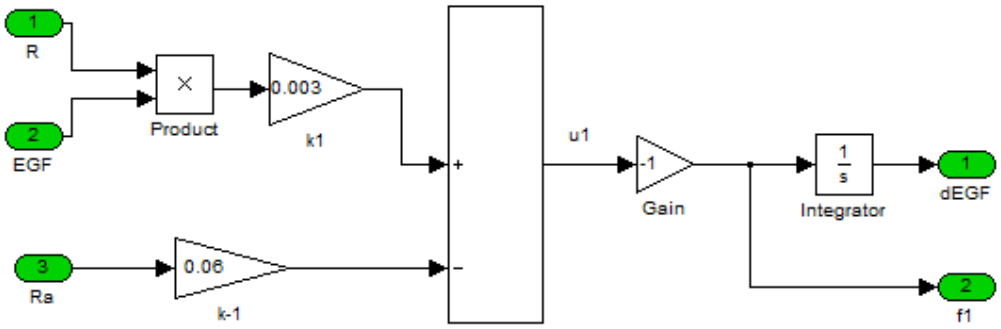
[Figure 5.3: Calculation of the rate concentrations of proteins.]


1
u1
-1
Gain
1
s
Integrator
1
dEGF
2
f1

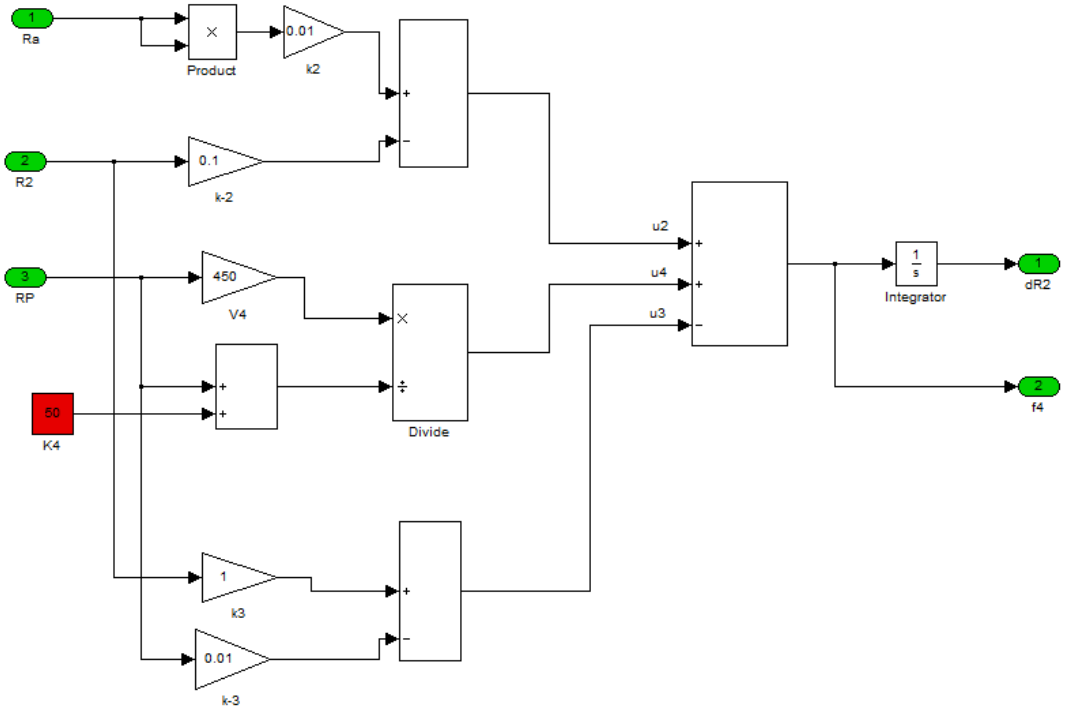




[Figure 5.4: Calculation of the concentrations of the proteins.]

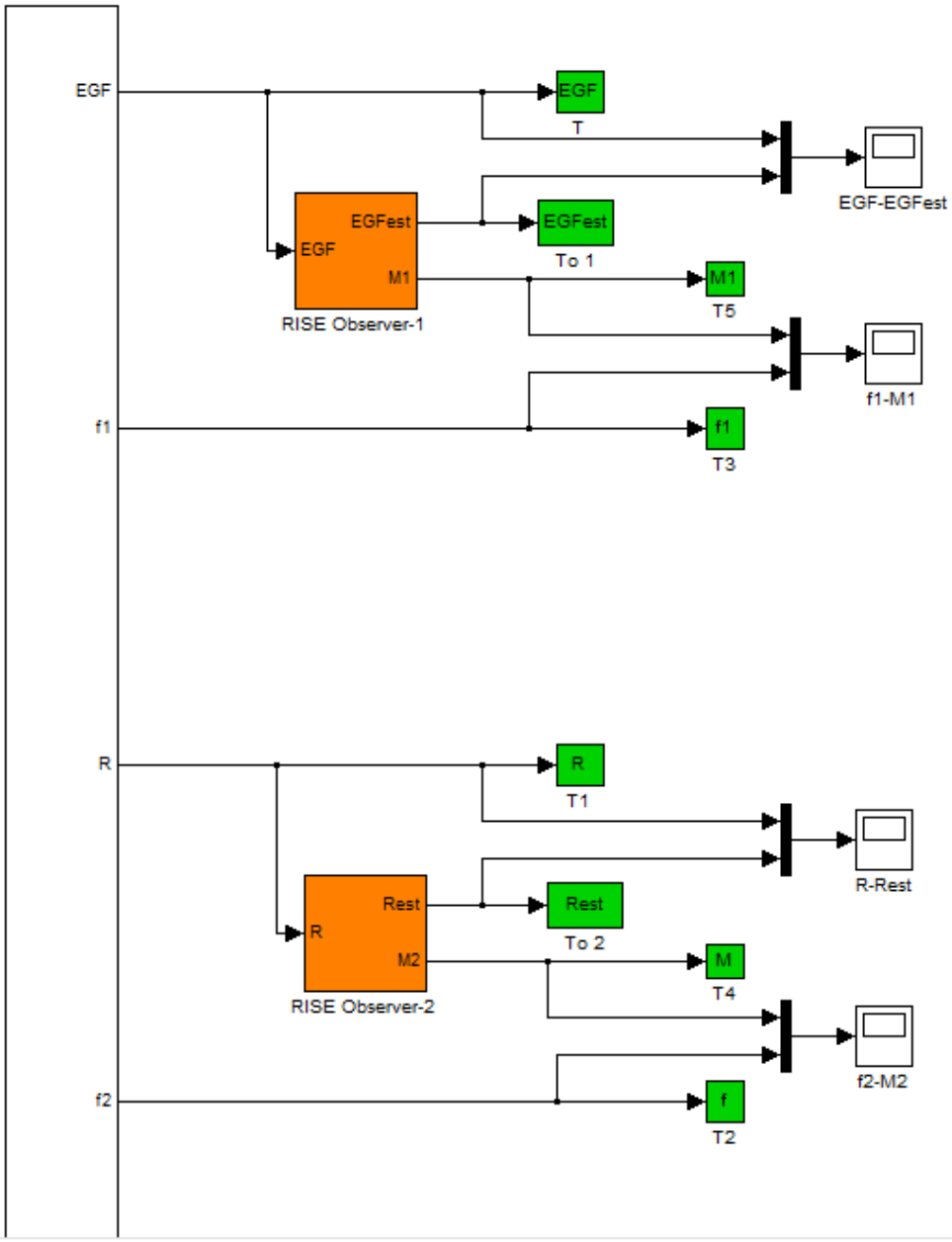


[Figure 5.5: The subsystem of dEGF.]



[Figure 5.6: The subsystem of dR2.]

In order to identify the EGFR pathway we use twenty three asymptotic observers based on the robust integral of the sign error (RISE) approach that identifies the nonlinear vector field $\xi^{*}(t):=f(x(t))$ as described in chapter 4. Each one observer is representing a protein member of the EGFR pathway. In figure 5.7 we show the RISE observer for the proteins EGF and R. This continues for the twenty three proteins.

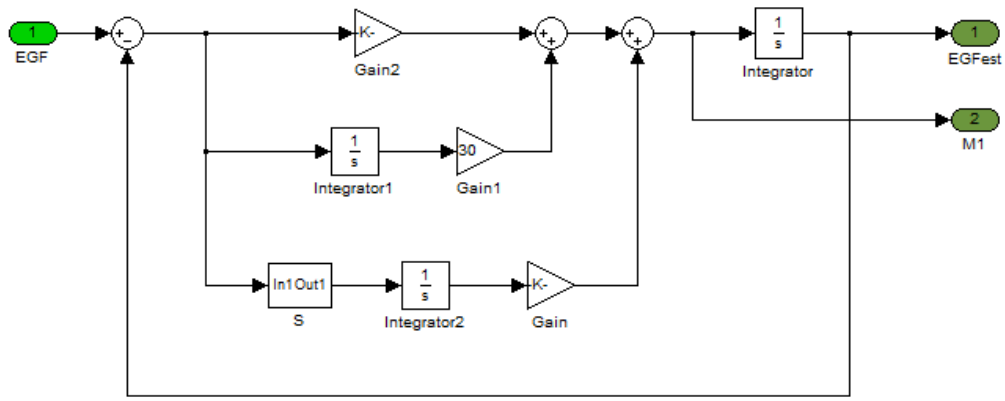


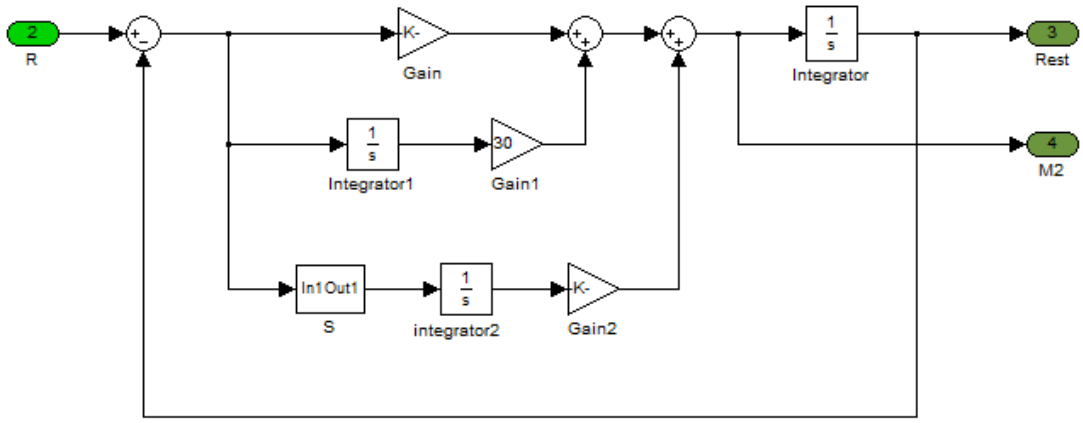
[Figure 5.7: Identification block diagram of EGF and R proteins.]

We have the following full state of RISE observer with constants κ, λ, β > 0:

$$xt=\xi t:=kxt-xt+\lambda\kappa 0txs-xsds+\beta 0tsgn\,(xs-x(s))ds. \quad (1)$$

The most important property of the RISE observer (1) is that it provides an asymptotic time estimate $\xi(t)$ of the respective vector field $f(x(t))$ that can be subsequently used for identification of the EGFR signalling. In figure 5.8 we have the implementation of the RISE observer (1) for the EGF and R protein in simulink.

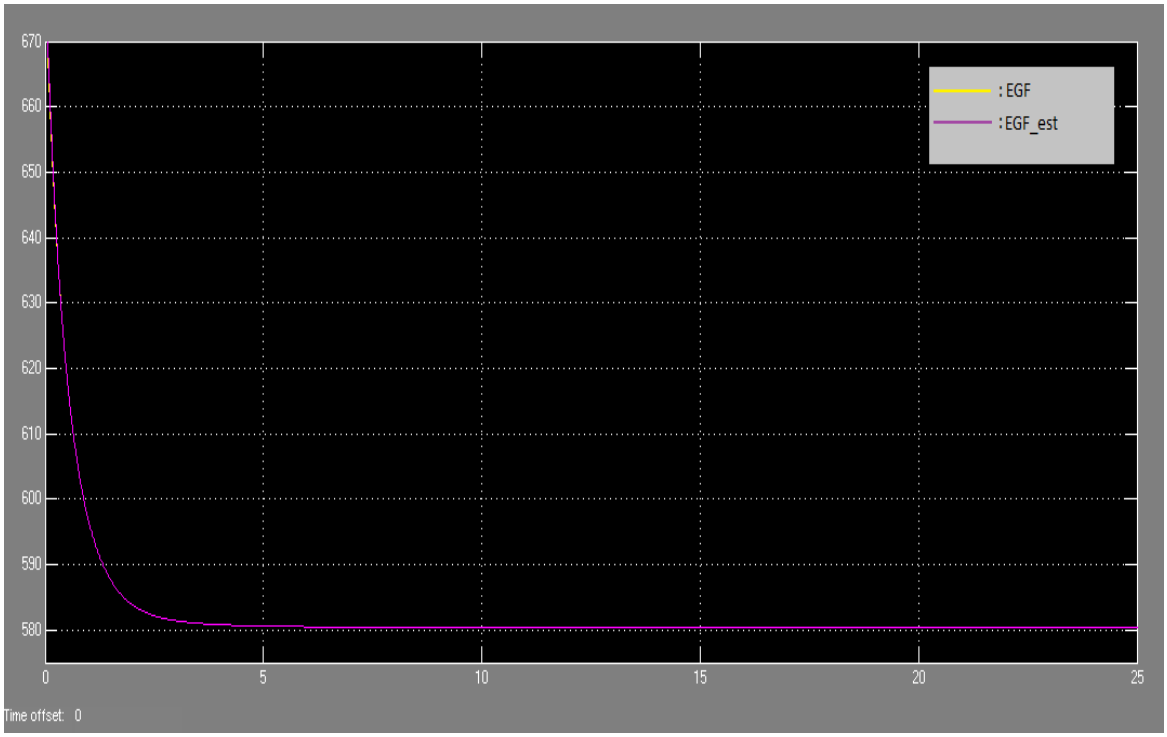




[Figure 5.8: The simulink implementation of the RISE observer.]

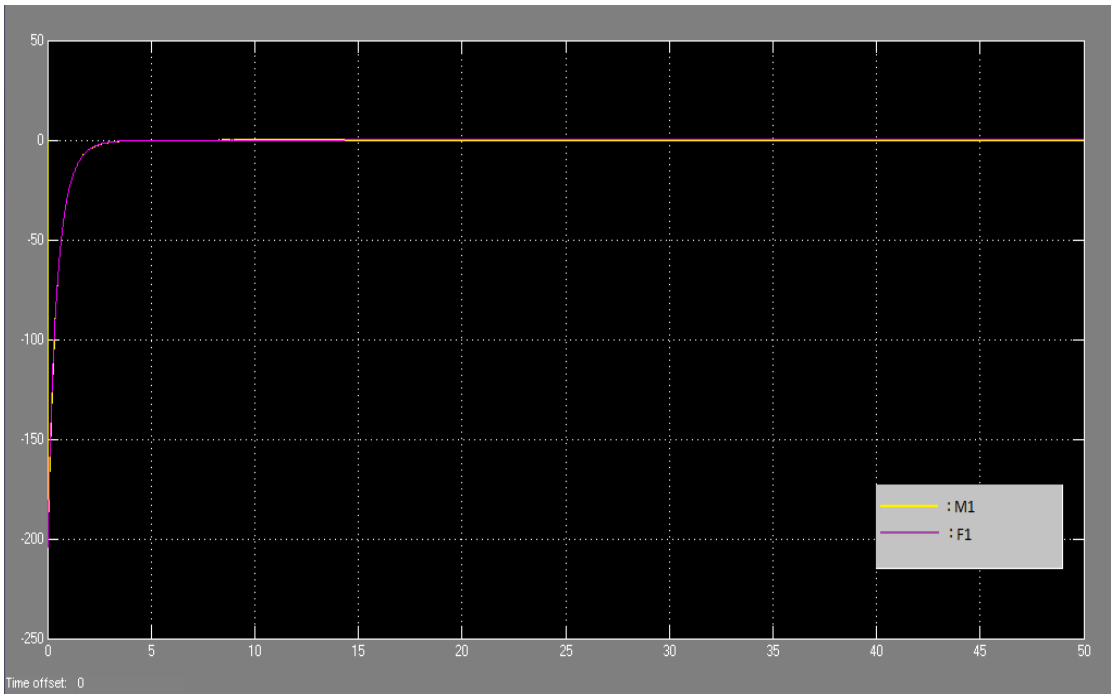
5.2 Identification and Simulink Results

With the values of parameters which we mentioned in chapter 3, we begin the identification of the EGFR pathway. After our implementation of the model in Matlab and Simulink we can watch both outputs EGF_est, which is the RISE observers output and EGF from the EGFR pathway for EGF protein. These outputs (EGF_est and EGF) are the integral results and have the major importance of this system. Our system, together EGFR pathway and RISE Observer runs until 4300 seconds.
For EGF protein we can see that the RISE observer output follows closely the real curve and at time 10 sec matches it (Fig.5.9).



[Figure 5.9:f(t,EGF,EGF_est)]

On our implementation we have added two more outputs F and M (for EGF protein its F1 and M1). These are the outputs without the integral and have a minor importance in our system. In figure 5.10 we can see these two curves where M1 is representing the observer's output for EGF protein, without integration, and F1 is the EGF output without integration, from the EGFR pathway.
F1 starts from -200 and M1 is identifying until 5 sec where it matches it.

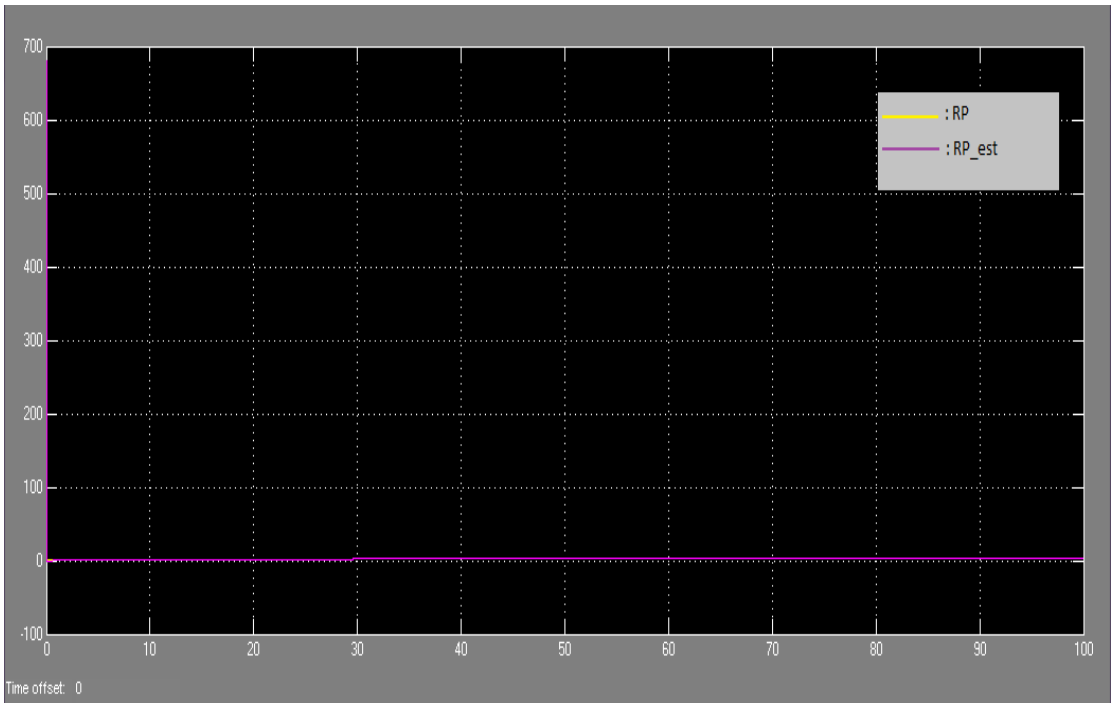


[Figure 5.10:f(t,M1,F1)]

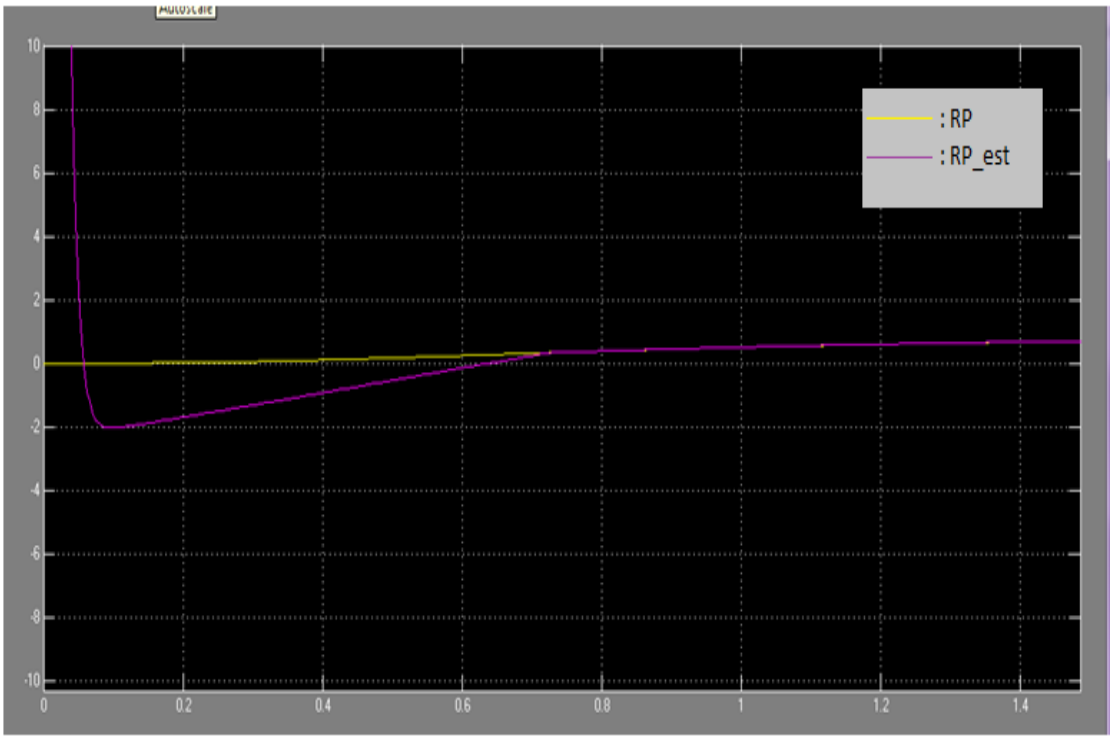
It is interesting to see also other proteins and how their concentrations change in time.
One important protein is RP. In figure 5.11 the curves are almost the same with an exception for the first 0.7 seconds where the observer is trying to identify the RP protein (this is more clear in the zoom figure). After 0.7 seconds RP has its stable value. In figure 5.12 we have the non integral outputs where in the zoom figure we can see the 0.7 seconds of identifying.
In figure 5.13 we have the R-Sh protein and its non-integral outputs in figure 5.14. PLCγ and its non integral outputs in figure 5.15 and 5.16 respectively.

SOS protein is very important also in EGFR pathway and we can see its simulink outputs in figure 5.17 and in 5.18 its non-integral outputs.
Finally another protein worth mentioned is Sh-G-S protein where we can see in figure 5.19 its integral outputs and in figure 5.20 its non-integral outputs.

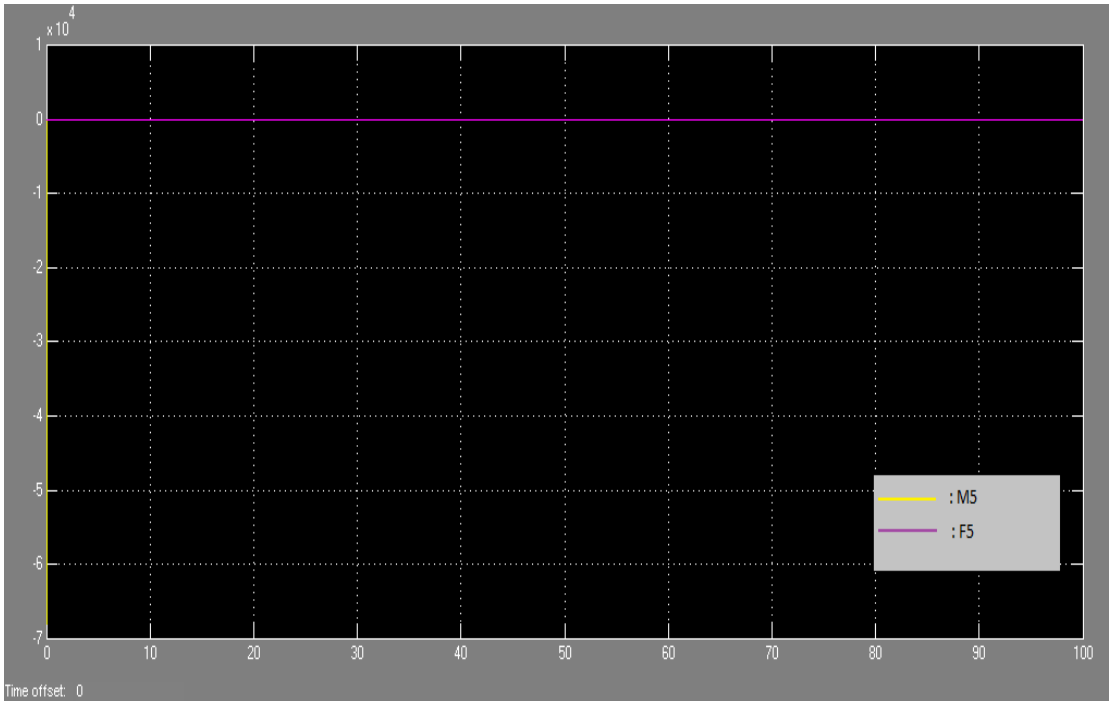
It is obvious that after a short time the concentrations reach to a stable value and that the time they need to do so, although it is not the same for each protein, does not vary sensibly.



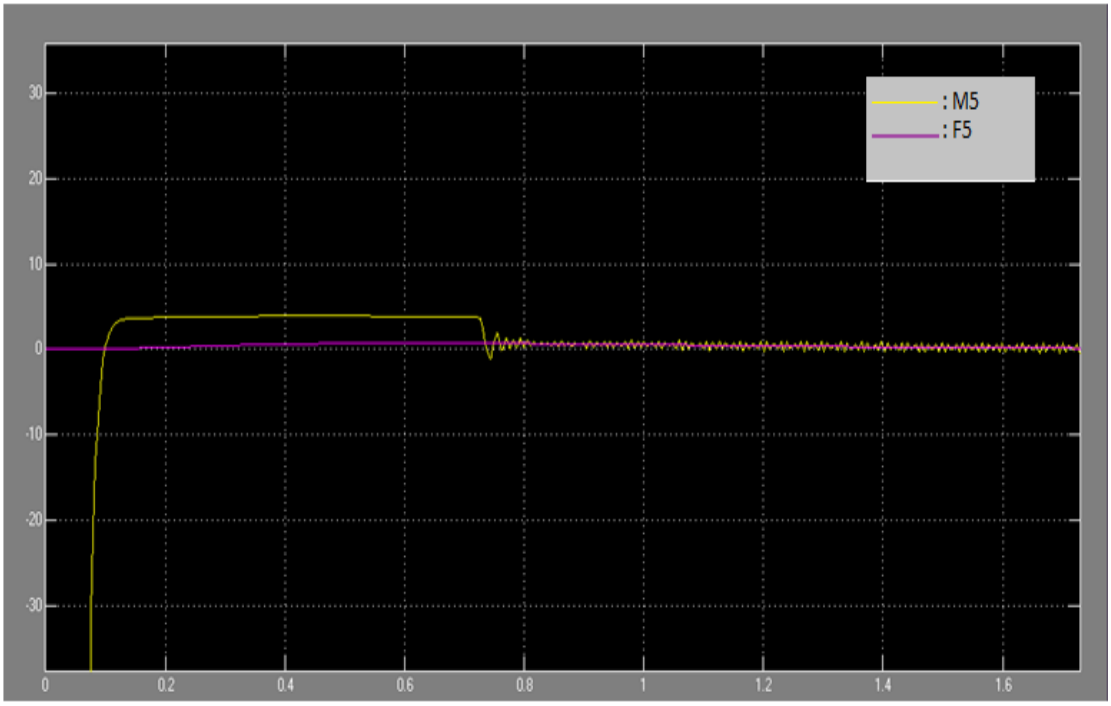
[Fiure 5.11a : f(t,RP,RP_est)]



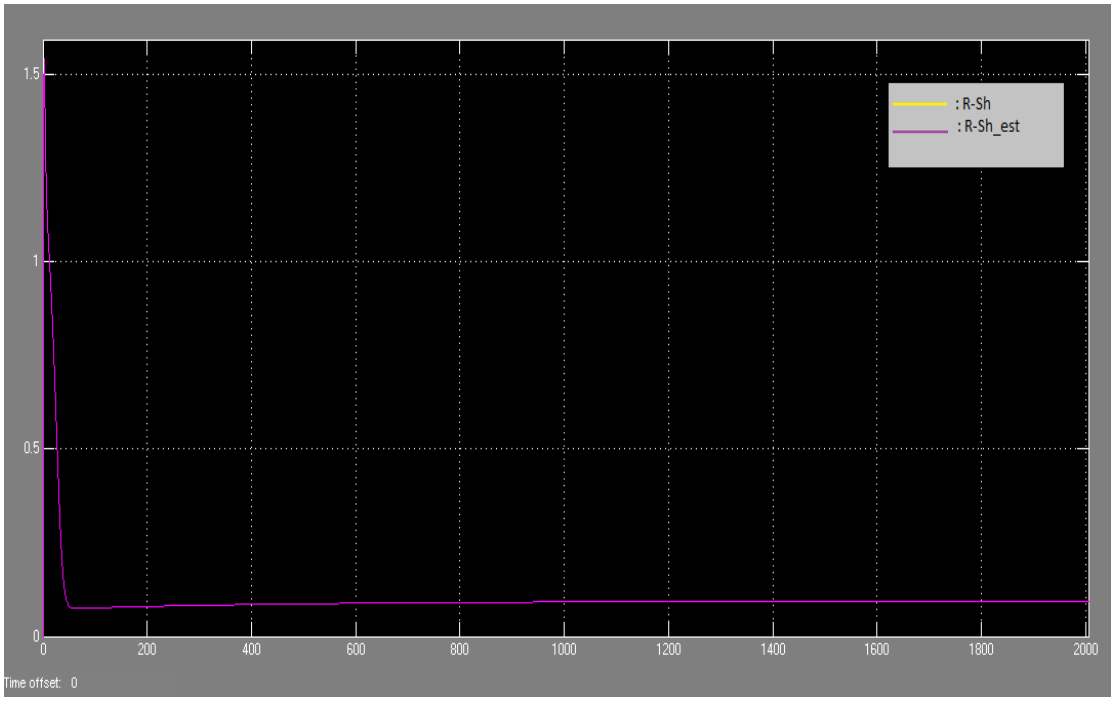
[Figure 5.11b: f(t,RP,RP_est) with zoom on both axes]



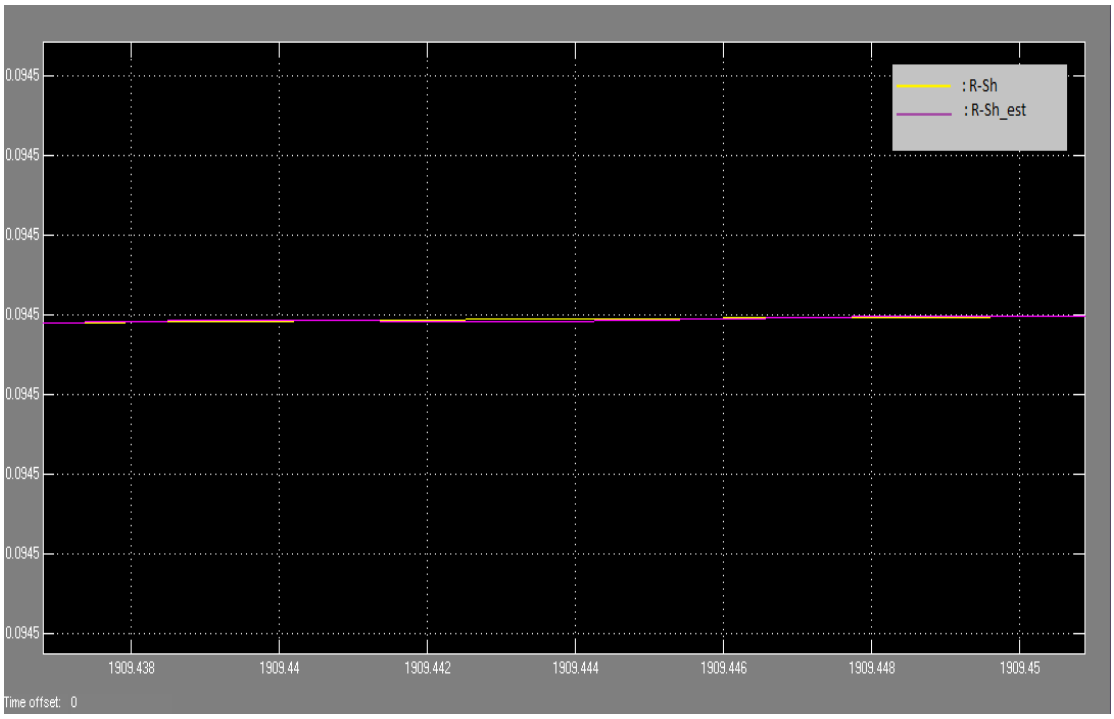
[Figure 5.12a:f(t,M5,F5)]



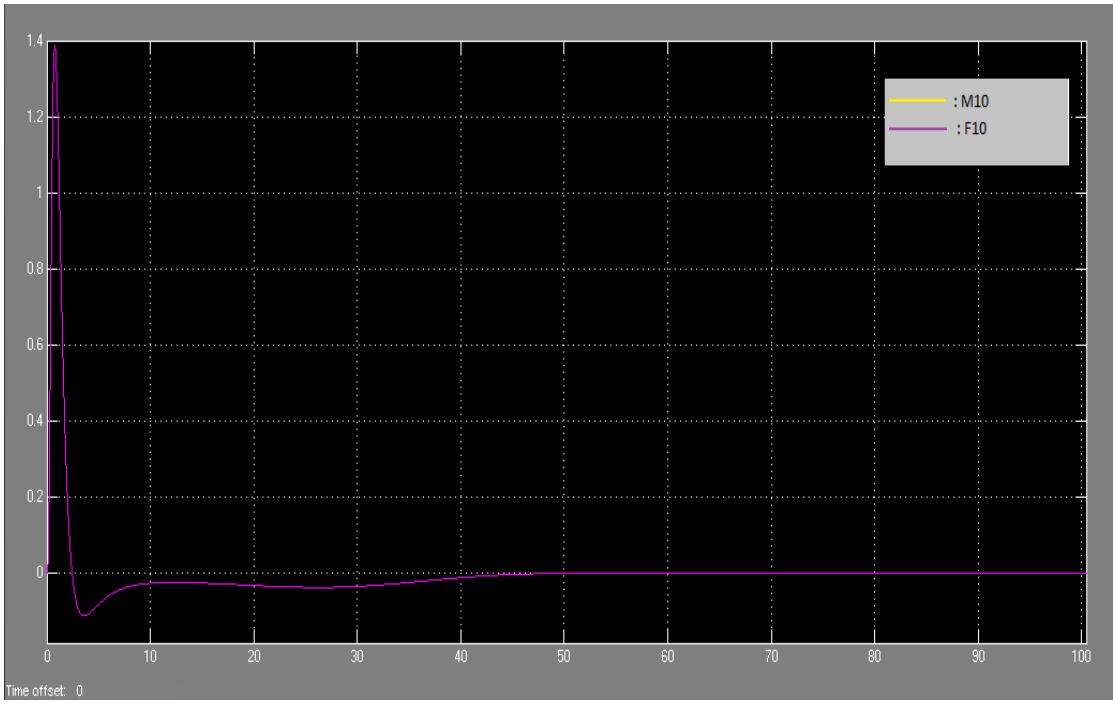
[Figure 5.12: f(t,M5,F5) with zoom on both axes]



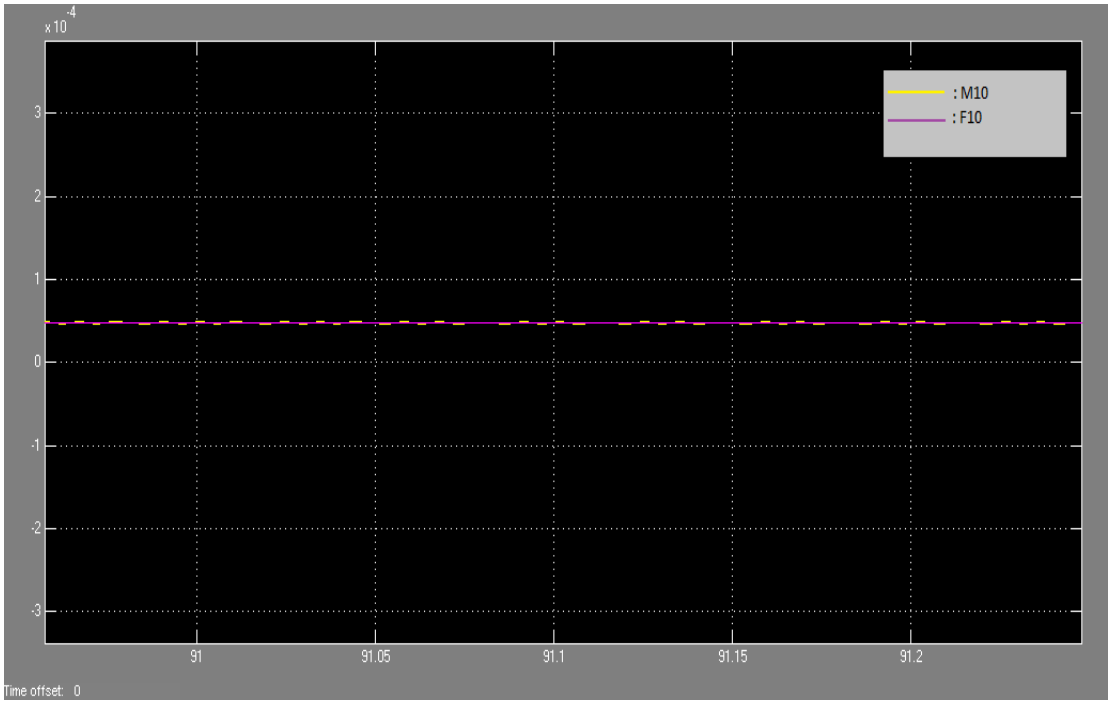
[Figure 5.13a : f(t,R-Sh,R-Sh_est)]



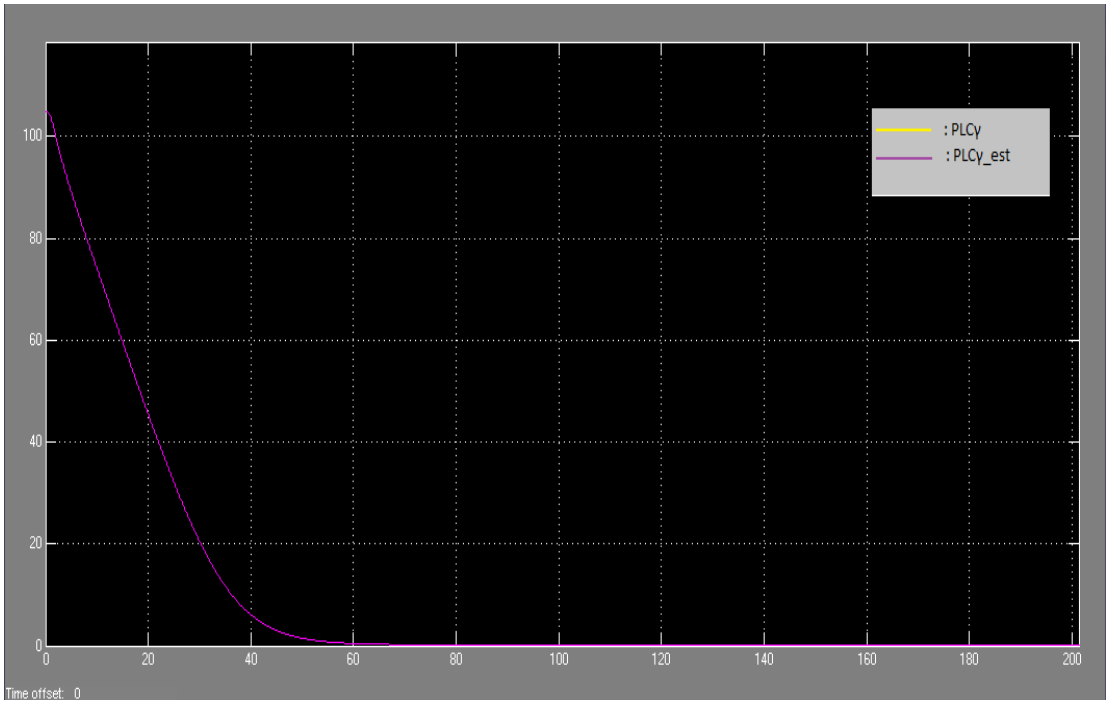
[Figure 5.13b: f(t,R-Sh,R-Sh_est) with zoom on both axes]



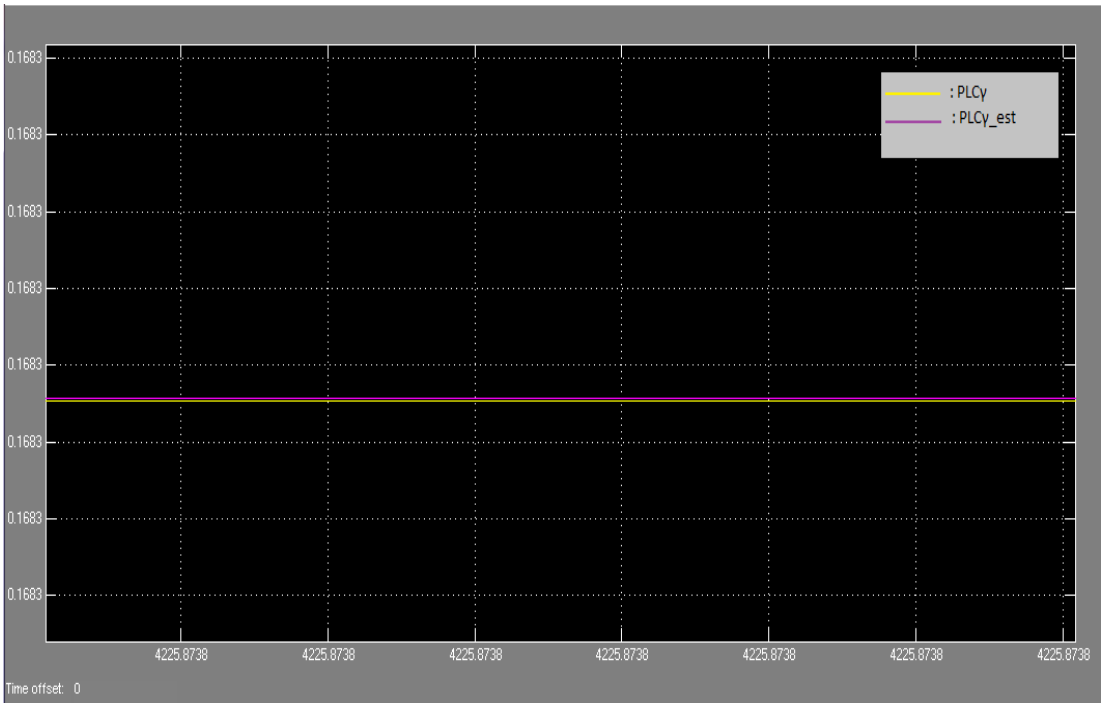
[Figure 5.14a:f(t,M10,F10)]



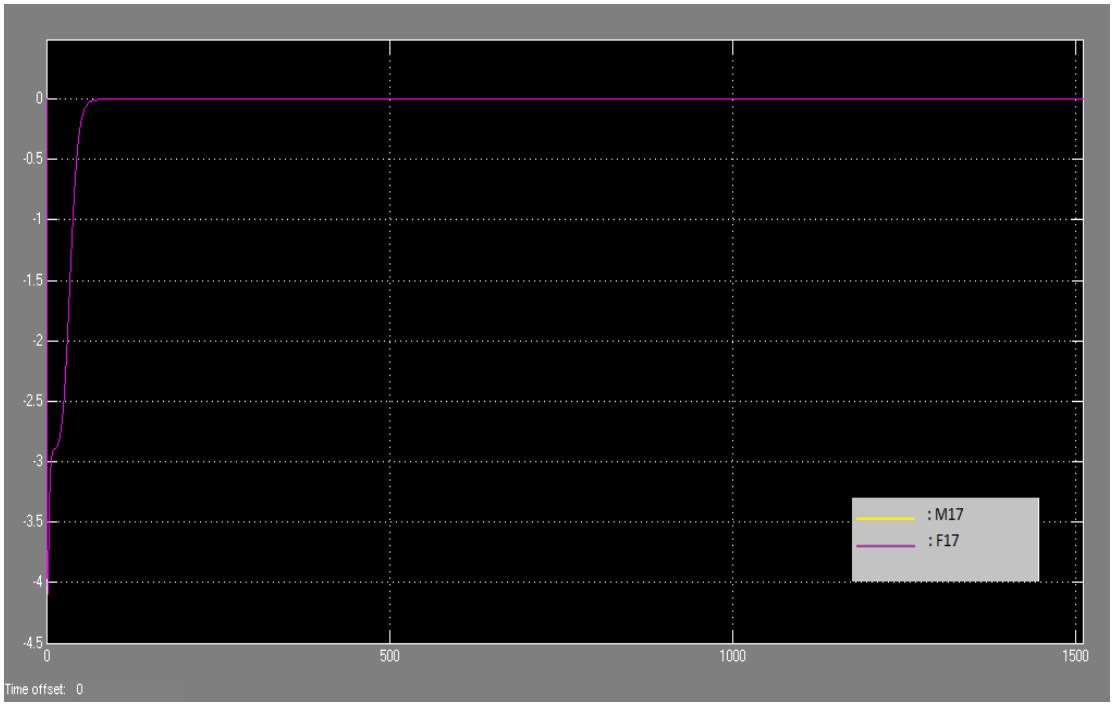
[Figure 14b: f(t,M10,F10) with zoom on both axes]



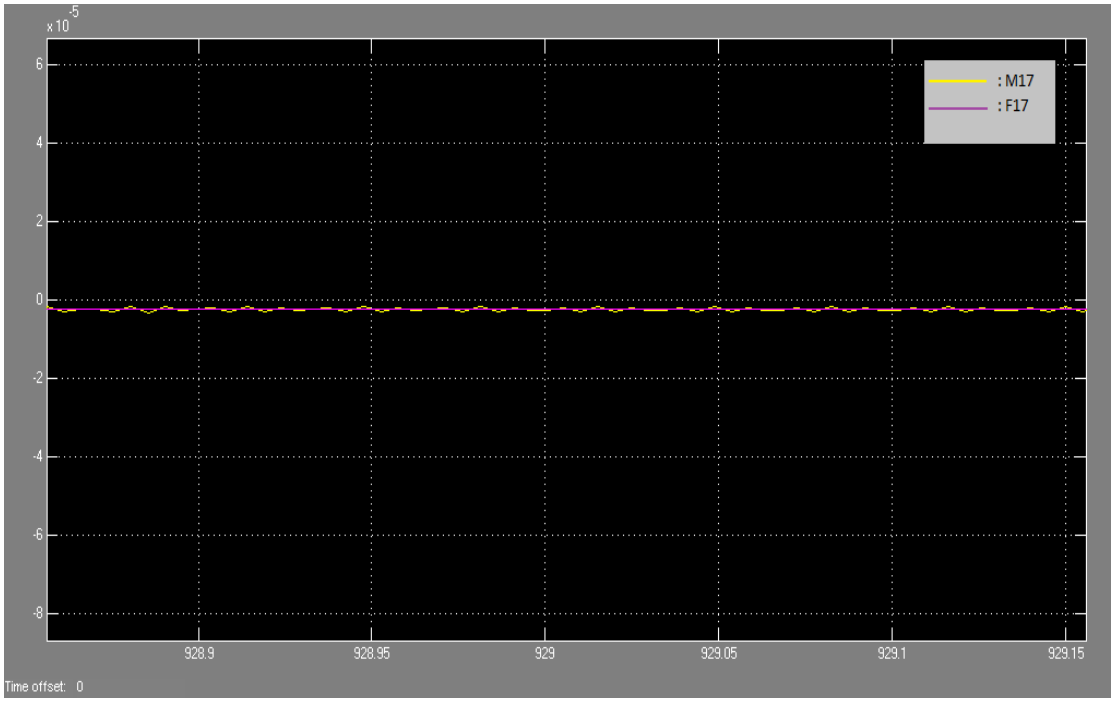
[Figure 5.15a : f(t,PLCγ,PLCγ_est)]



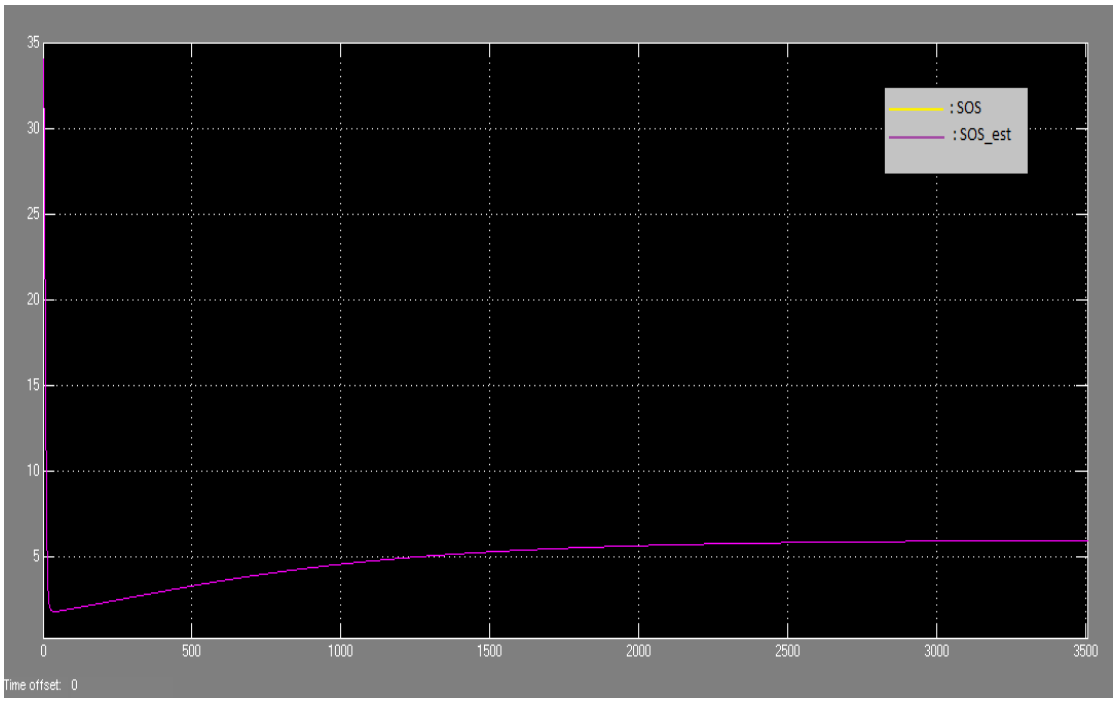
[Figure 5.15b: f(t,PLCγ,PLCγ_est) with zoom on both axes]



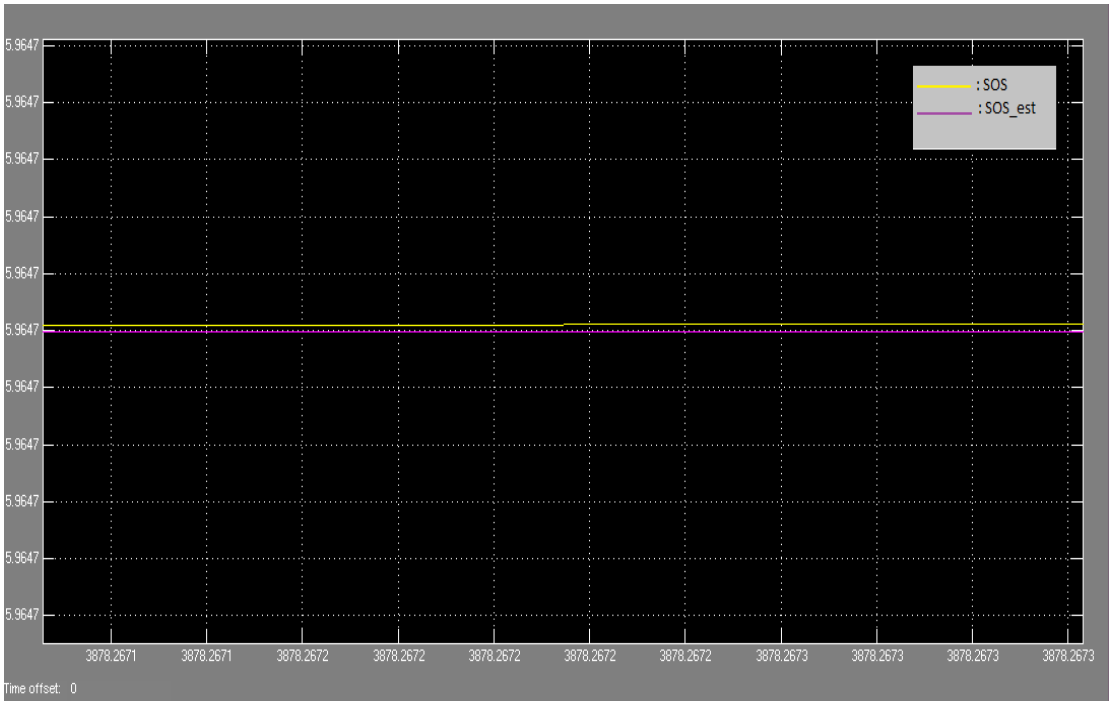
[Figure 5.16a : f(t,M17,F17)]



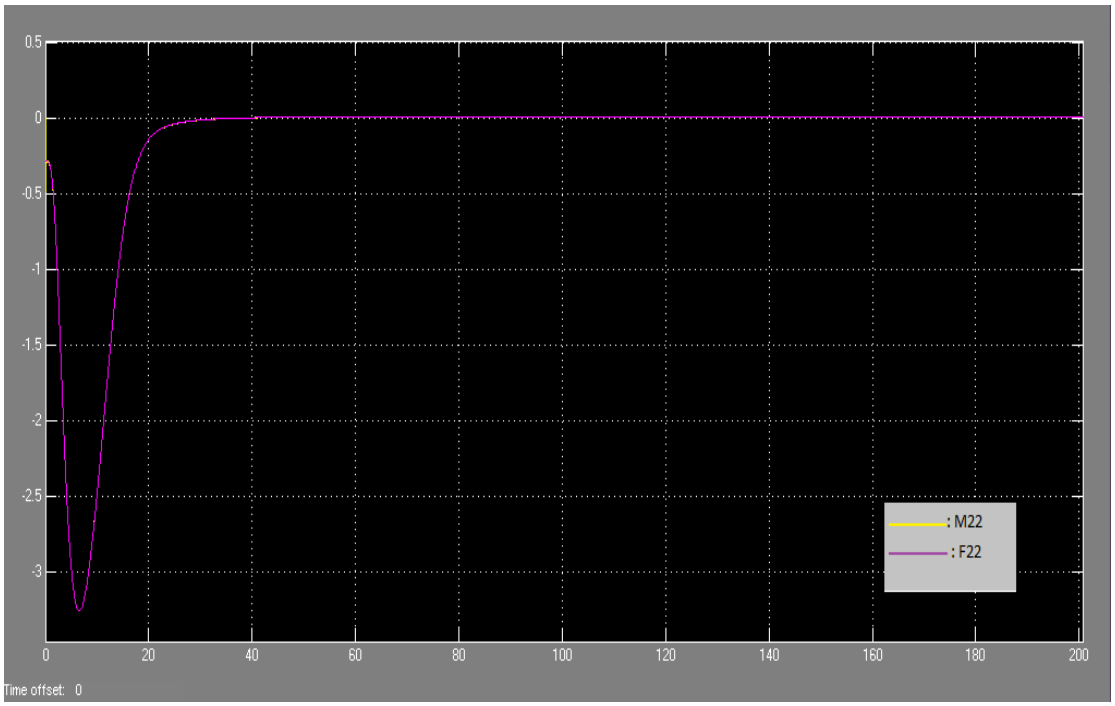
[Figure 5.16b: f(t,M17,F17) with zoom on both axes]



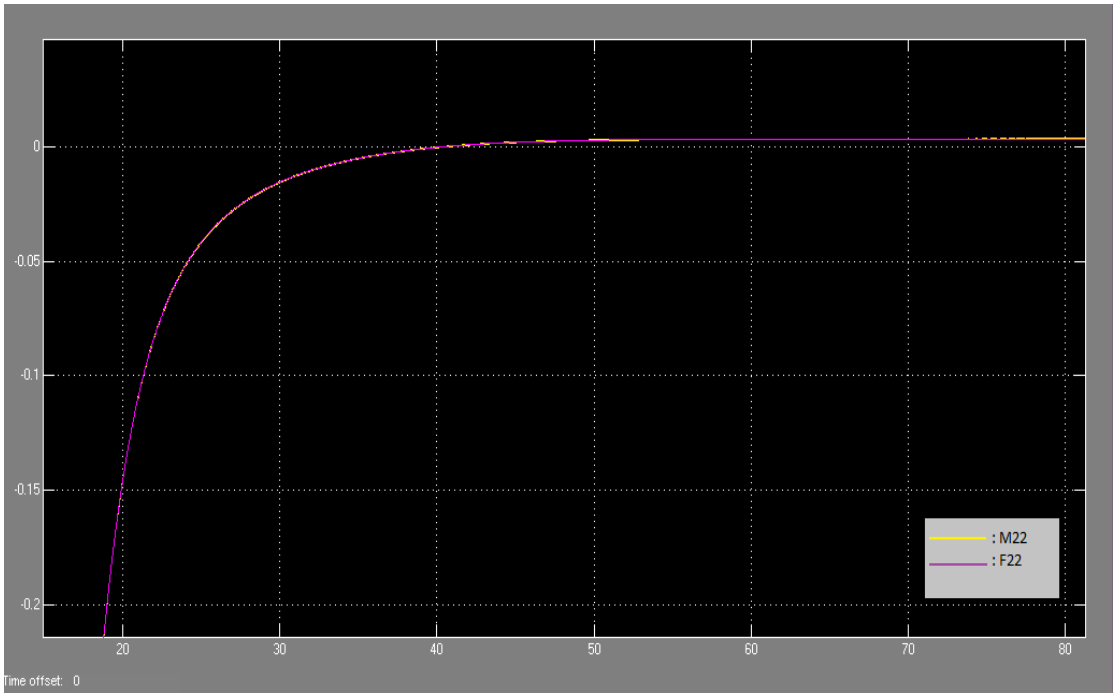
[Figure 5.17a : f(t,SOS,SOS_est)]



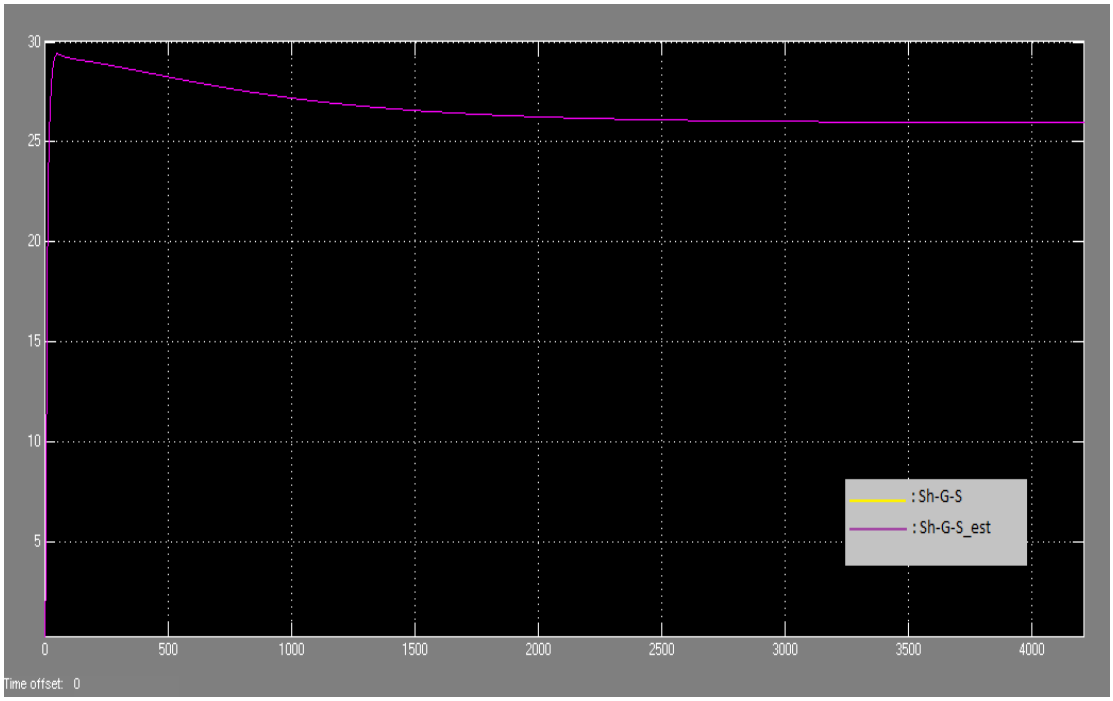
[Figure 5.17b: f(t,SOS,SOS_est) with zoom on both axes]



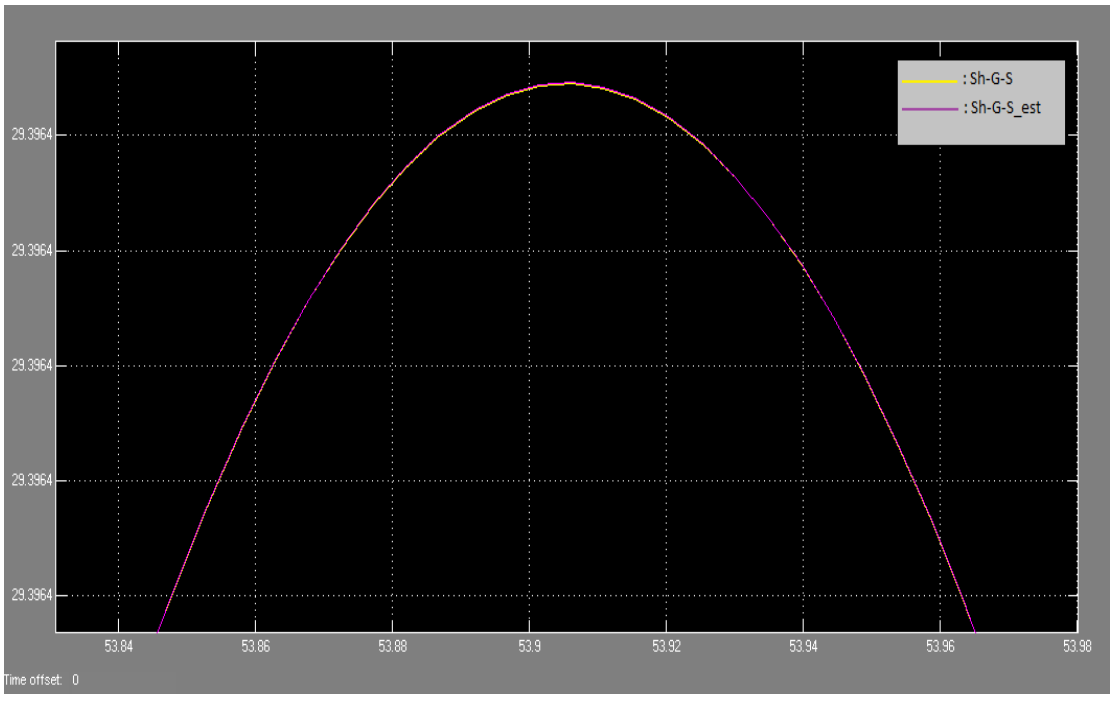
[Figure 5.18a : f(t,M22,F22)]



[Figure 5.18b : f(t,M22,F22) with zoom on both axes]

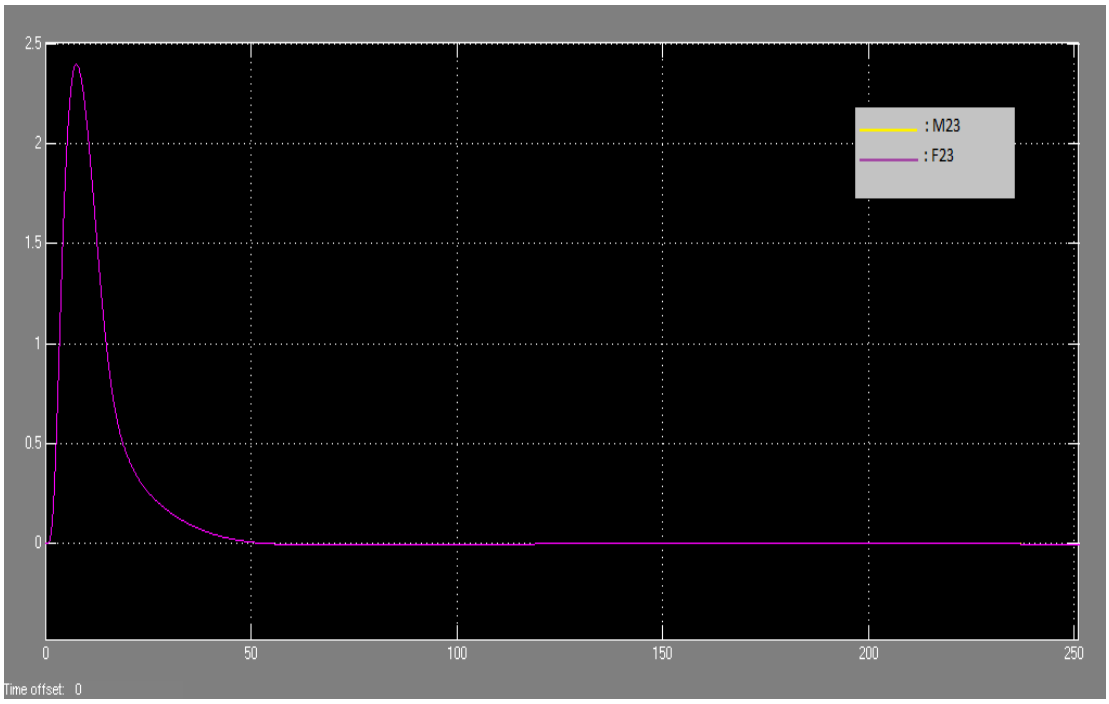


[Figure 5.19a : f(t,Sh-G-S,Sh-G-S_est)]

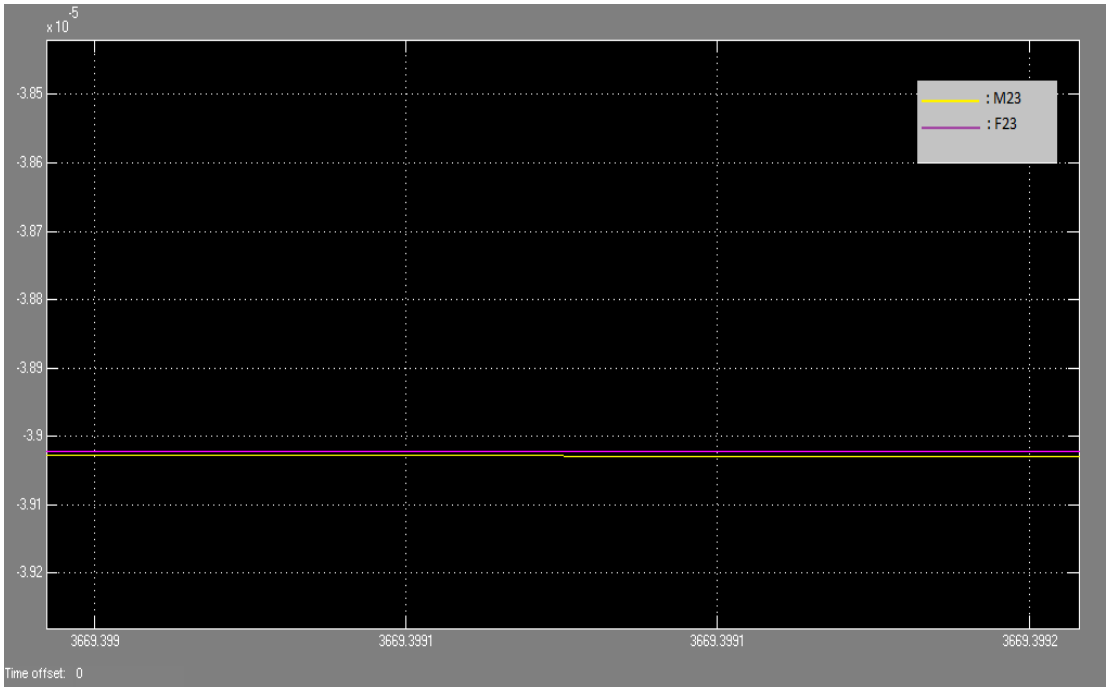


[Figure 5.19b: f(t,Sh-G-S,Sh-G-S_est) with zoom on both axes

]



[Figure 5.20a : f(t,M23,F23)]



[Figure 5.20b: f(t,M23,F23) with zoom on both axes]

Chapter 6

Conclusions and Future Developments

In this thesis we are investigating the identification of the EGFR signalling pathway using the RISE observer. The EGFR seems to be very significant for cell's function and essential in cancer research. A first model of it was constructed by Kholodenko, from the calculation of the kinetics and the concentrations of the proteins in the EGFR signalling pathway. In order to create the kinetic model, someone has to know everything about the structure of the system to identify. The more real data are available the better the identification is. The RISE observer identifies the EGFR signalling pathway perfectly and this is obvious from the stimulus response curves in simulink.

The next step relevant to our work could also be the evaluation of the right control to pathways that do not work properly in order to bring the system to the right dynamical behaviour. In medical words, this might mean "finding the treatment to a disease". We think

that it can be very helpful, in such a control procedure, the Rise Observer identification scheme we saw in this thesis to be used as the first part in a control algorithm.

The experiments in the cell are very difficult and complicated. One could easily imagine thus how hard is for the biologists to understand only from the chemical experiments the interaction between two or more cells. Control theory provides powerful mathematical tools in order to make this enormous number of data useful and easy to handle. To sum up, technical knowledge coming from control engineers should be used in cooperation with biological knowledge in order to produce significant results in medicine.

References

1. Cell and Molecular Biology by S.C. Rastogi
2. Essential cell biology: " an introduction to the molecular biology of the cell" by Bruce Alberts
3. Cell biology: a short course by Steven R. Bolsover
4. http://library.thinquest.org/C004535/prokaryotic_cells.html
5. http://www.biologymad.com/resources/AS%20Cells.pdf
6. http://ghr.nlm.nih.gov/handbook/basics/dna
7. From genes to genomes: concepts and applications of DNA technology by Jeremy Dale, Malcolm von Schantz
8. http://rarediseases.about.com
9. http://nobel.scas.bcit.ca/resource/dna/station2.htm

10. www.accessexcellence.org/RC/VL/GG/protein_synthesis.php

11. www.sysbio.de/info/backround/index.shtml

12. Systems biology: integration technology, biology and computational, Leroy Hood

13. www.systemsbiology.org

14. Olaf Wolkenhauer (2005). “Systems Biology: Dynamic Pathway Modeling In Preparation”

15. E.D. Sontag, (2004) Systems Biology, 1, 9-18

16. Pace NR (January 2001). "The universal nature of biochemistry". Proc. Natl. Acad. Sci. U.S.A. 98 (3): 805–8.

17. science.howstuffworks.com/cellular-microscopic-biology

18. Cotton, Frank Albert; Geoffrey Wilkinson, Carlos A. Murillo (1999). Advanced Inorganic Chemistry

19. Rhoades RA, Pflanzer RG (2002). Human Physiology (4th Ed.). Thomson Learning

20. http://people.ku.edu/~jbrown/antibody.html

21. Nils Bluthgen (2001). “Dynamical Models of signal Transduction and the Influence of Feedback Loops” Diploma Thesis, Institute of theoretical Biology University of Berlin

22. Online.redwoods.c.c.ca/us/instruct/damold/deproj/spo2/sauceGent/Kinetics_Prese ntation.pdf: College of the Redwoods Math SS, Differential Equations Michaelis-Menten Enzyme Kinetics.

23. *Mathematical Models in Biology*. By *Leah Edelstein-Keshet*, *McGraw-Hill*, Boston, (1988)

24. *Garret and Grisham*. Biochemistry. Saunders College Publishing, 1988

25. *More Mechanism*: A model of Enzyme Kinetics Department of Physics, University of Pennsylvania

26. Liu X, Pawson T. Biochemistry of the Src protein-tyrosine kinase: regulation by SH2 and SH3 domains. Recent Prog Horm Res. 1994;49:149–160

27. Levitzki A, Gazit A. Tyrosine kinase inhibition: an approach to drug development. Science 1995. 267, 1982

28. Cell Signaling by Receptor Tyrosine Kinases. Joseph Schlessinger Department of Pharmacology and The Skirball Institute New York University Medical Center (2000) 100,293

29. Signal transduction by receptors with tyrosine kinase activity. **Axel Ullrich and Joseph Schlessinger , Cell (1990) 61,203**

30. V.Radha, S.Nambirajan, G.Swarup (March 1996). ”Association of Lyn tyrosine kinase with the nuclear matrix and cell-cycle-dependent changes in matrix-associated tyrosine kinase activity.” 352-359

31. http://faculty.plattsburgh.edu/donald.slish/tyrosinekinase

32. http://users.rcn.com

33. Bublil EM and Yarden Y. (2007). "The EGF receptor family: spearheading a merger of signaling and therapeutics". Curr. Opin. Cell Biol. 19 (2): 124–134.

34. Robinson CJ, Stringer SE (2001). The splice variants of vascular endothelial growth factor (VEGF) and their receptors. J Cell Sci 114:853–865.
35. Baloh RH, Enomoto H, Johnson EM Jr, Milbrandt J.(2001) The GDNF family ligands and receptors - implications for neural development.103-110.
36. **R.P. Araujo, E.F. Petricoin and L.A. Liotta.** (2005).A mathematical model of combination therapy using the EGFR signaling network.
37. **Daly RJ. (1999)”Take your partners, please--signal diversification by the erbB family of receptor tyrosine kinases. Cancer Research Program, Garvan Institute of Medical Research, St Vincent's Hospital, Sydney, Australia. 255-263**
38. A Wells (1999). Molecule in Focus: EGF receptor. International Journal of Biochemistry and Cell Biology 31,637-643.
39. **Rosette, C. Karin, M. Ultraviolet light and osmotic stress: activation of the JNK cascade through multiple growth factor and cytokine receptors. Department of Pharmacology, Program in Biomedical Sciences, School of Medicine, University of California at San Diego**
40. Xiaolan Qian, William C. Vass, Alex G. Papageorge, Pieter H. Anborgh, and Douglas R. Lowy (1998) ″N Terminus of Sos1 Ras Exchange Factor: Critical Roles for the Dbl and Pleckstrin Homology Domains.” Mol Cell Biol. 1998 February; 18(2): 771–778.

41. Boris N.Kholodenko, Oleg V. Demin, Gisela Moehren and Jan B. Hoek “Quantification of Short Term Signaling by the Epidermal Growth Factor Receptor.”

42. Neural Networks. Herve Abdi,Dominique Valentin, Betty Edelman
43. http://www.learnartificialneuralnetworks.com/
44. Open Clinical: Knowledge management for medical care, “Neural Networks: an introduction”, May 2005
45. “Neural networks” Statistica April 2005
46. www.gc.ssr.upm.es/inves/neural/
47. *“An observer Approach for Deterministic Learning with Patchy Neural Networks”*
48. *J.Farrell,” Stability and approximator convergence in nonparametric nonlinear adaptive control,” IEEE Trans. Neural Netw. vol. 9, no.5, pp.1008-1020, Sep. 1998.*
49. V.P. Vapnik, “The Nature of Statistical Learning Theory”, 2nd ed. New York: Springer-Verlag, 2000.
50. C.Wang, D.J.Hill and G.Chen, “Deterministic learning of nonlinear dynamical systems,” in Proc.18th IEEE Int. Symp. Intell. Control, Houston, TX, pp8792, Oct.2003.
51. C.Wang and D.J.Hill,”Learning from neural control,” IEEE Trans. Neural Netw., vol. 17, no.1, pp.130146, Jan 2006.
52. C.Wang and D.J.Hill, “Deterministic learning and rapid dynamical pattern recognition,” IEEE Trans. Neural Netw., vol. 18, no.3, pp.617630, May 2007.

53. C.Wang and D.J.Hill, Deterministic learning theory for identification, recognition and control. ISBN 978-0-8493-7553-8, CRC Press, Boca Raton, 2010.
54. R.M.Sanner and J.E.Slotine, “Gaussian networks for direct adaptive control,” IEEE Trans. Neural Netw., vol.3, no.6, pp.837863, Nov.1992.
55. D.Jiang and J.Wang, “On-line learning of dynamical systems in the presence of model mismatch and disturbances,” IEEE Trans. Neural Netw., vol. 11, no. 6, pp.12721283, Nov.2000.
56. S.S.Ge and C.Wang,”Direct adaptive NN control of a class of nonlinear systems,” IEEE Trans. Neural Netw., vol. 13, no. 1, pp. 214221, Jan 2002.
57. J.Farrel,”Stability and approximator convergence in nonparametric nonlinear adaptive control,” IEEE Trans. Neural Netw., vol.9, no. 5, pp. 10081020, Sep.1998.
58. G.A.Rovithakis and M.A.Christodoulou, “Adaptive Control with Recurrent High-Order Neural Networks: Theory and Industrial Applications.” London UK: Springer-Verlag, 2000.
59. P.M.Patre, W.MacKunis, K.Kaiser and W.E.Dixon,”Asymptotic tracking for uncertain dynamic systems via a multilayer neural network feed forward and RISE feedback control structure,” IEEE Trans. Automat. Contr., vol.53, no. 9, pp. 2180-2185, Oct.2008.
60. B.Xian, D.M.Dawson, M.S. de Queiroz, and J.Chen, “A continuous asymptotic tracking control strategy for uncertain nonlinear systems,” IEEE Trans. Automat. Control, vol. 49, no. 7, pp. 1206-1211, 2004.
61. P.M.Patre, W.MacKunis, C.Makkar and W.E.Dixon ,”Asymptotic tracking for systems with structured and unstructured uncertainties,” IEEE Trans. Contr. Syst. Technol., vol. 16, no. 2, pp. 373-379, 2008.
62. J.E.Slotine and W.Li, Applied Nonlinear Control. Englewood Cliffs, NJ: Prentice Hall, 1991.
63. E.B.Kosmatopoulos, M.M.Polycarpou, M.A.Christodoulou and P.A.Ioannou, “High-order neural network structures for identification of dynamical systems,” IEEE Trans. Neural Netw., vol. 6, pp. 422-431, 1995.
64. P.A.Ioannou and J.Sun, Robust Adaptive Control. Englewood Cliffs, NJ: Prentice Hall, 1996.