Technical University of Crete

SIMULINK Identification Model for M.A.P. Kinase Cascade

Diploma Thesis in Systems Biology

From

Giotis Thomas

Supervisor:
Prof. E. Christodoulou

Examining Committee:
Prof. M. Zervakis
Prof. Dr. E. Georgiou

October 2005

Contents

1 Introduction

2 Biological Background
- 2.1 DNA – Proteins
  - 2.1.1 Genes And Their Relation With DNA
  - 2.1.2 Replication Of The DNA
  - 2.1.3 Usage Of Genetic Information
  - 2.1.4 Proteins And Their Role In The Cell
- 2.2 Cells As I/O Systems
- 2.3 Signaling Pathways
  - 2.3.1 The Reason Of Their Creation
  - 2.3.2 Definition – Specification Of Their Role
  - 2.3.3 Phosphorylation In Signaling Pathways
  - 2.3.4 Signaling Pathways And Cancer Diseases
  - 2.3.5 Kinds Of Signaling Pathways

3 Systems Biology
- 3.1 History
- 3.2 Characteristics
- 3.3 Modeling And Simulation
- 3.4 Modeling Techniques
  - 3.4.1 Example: Modeling of Ras/Raf-1/MEK/ERK Signal Transduction Pathway

4 MAPK Cascade
- 4.1 Functioning Of MAPK Pathways
- 4.2 Models
- 4.3 Reaction Kinetics
  - 4.3.1 Mass Action Law
  - 4.3.2 Enzyme Kinetics
- 4.4 Simulation Using SIMULINK
- 4.5 Steady-State Properties
- 4.6 Ultrasensitivity In The MAPK Cascade

5 Conclusions

Appendix
- A.1 Model According To Huang And Ferrell
- A.2 Parameters

A.3 Reactions
A.4 Differential Equations Using Explicit Kinetics

References

1 Introduction

Cells are able to receive many different chemical signals from their surrounding, and have the capability to react to signal pattern in an appropriate way. The signals are processed by the intracellular signaling network, which is mainly constructed by proteins which react with each other. In the past, many different modules of this network have been identified and qualitatively studied [HARTWELL ET AL. 99]. Recently, the "Era of Pathway Quantification" [KOSHLAND 98] began, with simulation of signaling networks and quantitative measurement of reaction rates. Beside bacterial chemotaxis, calcium oscillations, and cellcycle control, the MAPK-cascade, a three molecule module present in all eucaryotes, has become a model system for quantitative analysis of signaling pathways [LAUFFENBURGER 00]. This module is of special interest, because it is well characterized and many reaction parameters are estimated. Moreover, it plays a significant role in the so called RAS pathway, which has high influence on cell growth and cell survival.

Several methods of modeling intracellular signal transduction have appeared: reaction systems using ordinary differential equations [KHOLODENKO ET AL. 97], stochastic models [GILLESPIE 77, GIBSON & MJOLSNESS 00], Petrinets [HOFESTÄDT & THELEN 98], neural networks [BRAY 95], rule based systems [BRUTLAG ET AL. 91], and Boolean networks [THIEFFRY & ROMERO 99, THOMAS & KAUFMAN 01]. In this thesis, the published model of MAPK according to Huang and Ferrell [HUANG & FERRELL 96] is investigated by modeling the reaction dynamics with ordinary differential equations.

In chapter 2, the biological background of our subject is introduced. We discuss about DNA, proteins, cells and finally we focus on signaling pathways and the importance of their role. Chapter 3 introduces a new era in sciences, Systems Biology, showing its characteristics, the way of modeling and a common example. In chapter 4, MAPK cascade is discussed, which is actually the subject of this thesis. We explain its functioning, we present the models found in the literature, we show how someone can work on MAPK using explicit kinetics, and how we implemented this model using SIMULINK of Matlab. The presentation of the steady-state properties of the model leads as to its ultrasensitivity. After our conclusions in chapter 5, in the Appendix there can be found the parameters used in my implementation of the model, the reactions of the model and the ordinary differential equations.

2 Biological Background

The development of multicellular organisms makes it necessary to have mechanisms for communication between cells. The organized structure and

ability to maintain different tissues in these organisms is mainly due to the fact that these cells can communicate with each other [KRAUSS 97]. The basis for intercellular communication is that the cells can receive and compute stimuli reaching the cell membrane, which is called intracellular signaling. In this chapter we also discuss about DNA and proteins since they are the basis for a better understanding of biological issues.

2.1 DNA – Proteins[CLARK & RUSSELL 00]

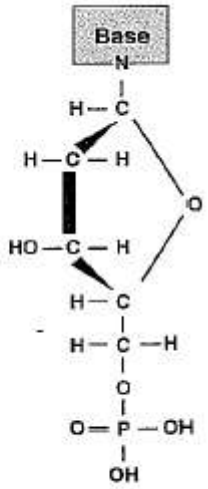
2.1.1 Genes And Their Relation With DNA

The fundamentals of modern genetics were laid when Mendel found that hereditary information is made up of discrete fundamental units which we now call genes. The discovery that atoms are made of subatomic particles ushered in the nuclear age. Similarly, the realization that genes are made up of molecules that obey the laws of chemistry has opened the way both to a deeper understanding of life and to its artificial alteration by genetic engineering.

The unit of heredity is known as a gene. Each gene is responsible for a single inherited property or characteristic of the organism. Certain properties of higher organisms, such as height or skin color, are due to the combined action of multiple genes. Consequently, in these cases there is a gradation of the property. Such multi-gene characteristics at first caused a lot of confusion and they are still difficult to analyze, especially if more than two or three genes are involved.

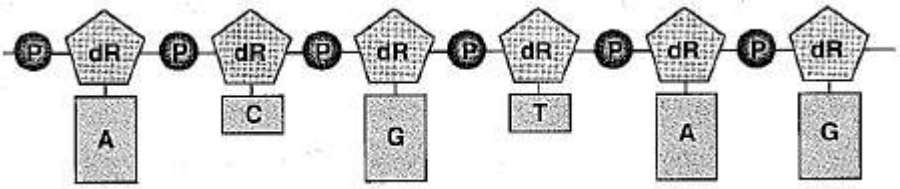
Genes are made of **DNA**, or deoxyribonucleic acid. Each gene is found in linear order and is a major component of structures known as chromosomes. Each chromosome has some accessory protein molecules which help maintain its structure and is an exceedingly long single molecule of DNA. The DNA of the chromosome is divided into segments. Many of these segments are actual genes. In front of each gene is a regulatory region of DNA involved in switching the gene ‘on’ or ‘off’. Between genes are spacer regions of DNA often referred to as intergenic regions. DNA is a polymer (molecule of similar repeating units which are linked together by a common bonding mechanism) made up of a linear arrangement of subunits known as nucleotides (the subunits from which DNA is built). Each nucleotide has three components: a phosphate group, a sugar and a nitrogen-containing base (alkaline chemical substance, in particular the cyclic nitrogen compounds found in DNA and RNA).

A Simple View

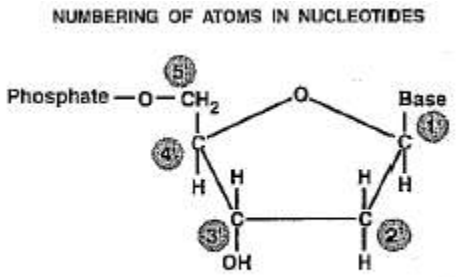


A Chemist's View

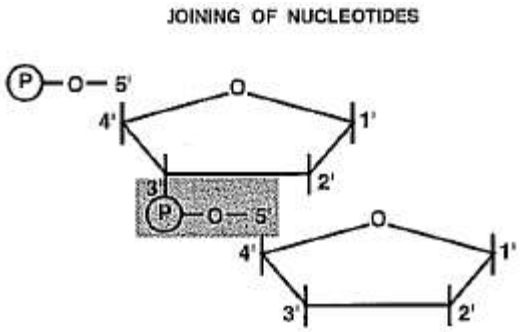
In DNA, the sugar is always deoxyribose. The different types of nucleotide differ only in the nature of the nitrogen-containing base. In DNA there are four alternative bases: adenine, thymine, guanine and cytosine. When writing out genetic information these bases are abbreviated by convention to: A, T, G and C. The phosphate groups and the deoxyribose sugars form the backbone of each strand of DNA. The bases are joined to the deoxyribose and stick out sideways. A single strand of DNA is shown in the next scheme.



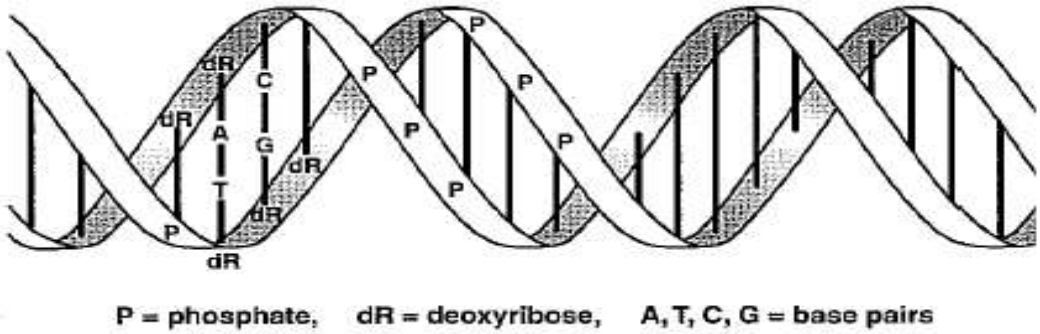
To understand how nucleotides are joined, we must clarify the situation by numbering the carbon atoms of the sugar molecule. The next scheme shows the convention for numbering nucleotides.



Nucleotides are joined by linking the phosphate on the 5' end of the deoxyribose of one to the 3' position of the next, as it's shown.



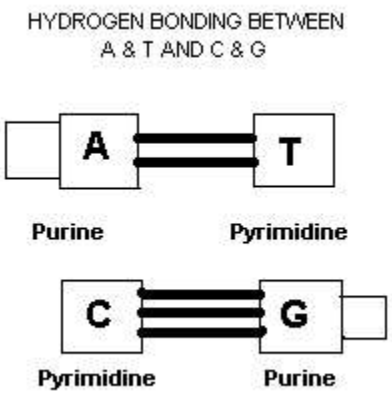
In practice, DNA is normally found as a double stranded molecule. Not only is DNA double stranded, but the two separate strands are wound around each other in a helical arrangement. This is known as the double helix.



Double helix

In double stranded DNA, the bases of one strand are paired with the bases in the other strand. Adenine (A) in one strand is always paired with thymine (T) in the other and guanine (G) is always paired with cytosine (C). The bases A and G are referred to as the purine bases as they contain a double ring structure known as a purine ring. The other two bases, C and T, are the pyrimidine bases, since they contain a single, pyrimidine ring. Each base pair consists of one double size purine base paired with a smaller pyrimidine base. So, although the bases themselves differ in size, all of the allowed base pairs are the same width. This is necessary to allow them to fit neatly into the double helix.

In double stranded DNA each base pair is held together by linkages known as hydrogen bonds. The A – T base pair has two hydrogen bonds and the G – C base pair is held together by three. Hydrogen bonds are very weak, but since a molecule of DNA usually contains millions of base pairs, the added effect of millions weak bonds is strong enough to keep the two strands together. The hydrogen bonding in DNA base pairs uses either



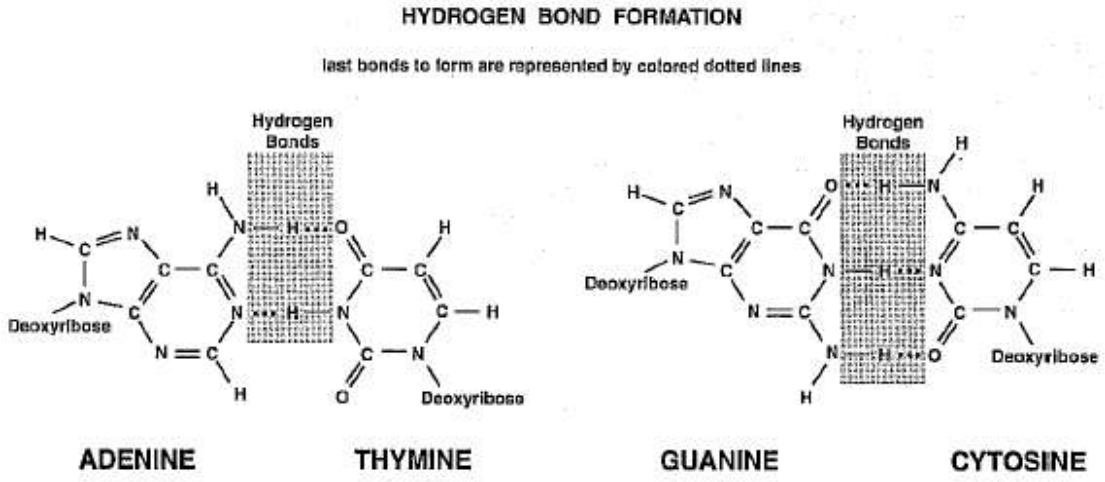
oxygen (O) or nitrogen (N), giving three alternative arrangements. In each case the hydrogen (H) is held between the other two atoms and serves to link them together. Before hydrogen bonds form and the bases pair off, the hydrogen atom is found attached to one or the other of the two bases, as it's shown with complete lines in the scheme below. During base pairing, the hydrogen binds to an atom of the second base (dashed lines).

THREE POSSIBLE TYPES
OF HYDROGEN BONDS

O - H - N

O - H - O

N - H - N



2.1.2 Replication Of The DNA

Since each cell needs a complete set of genes, it is necessary for the original cell to duplicate its genes before dividing. Because the genes are made of DNA and make up the chromosomes, this means that each chromosome must be accurately copied. Upon cell division, both daughter cells will receive identical sets of chromosomes, each with a complete set of genes.

In molecular terms this means that the DNA of the original, or mother, cell is duplicated to give two identical copies. This process is known as replication. Upon cell division each of the descendants gets one complete copy of the DNA. The original genes of the mother cell are on a double

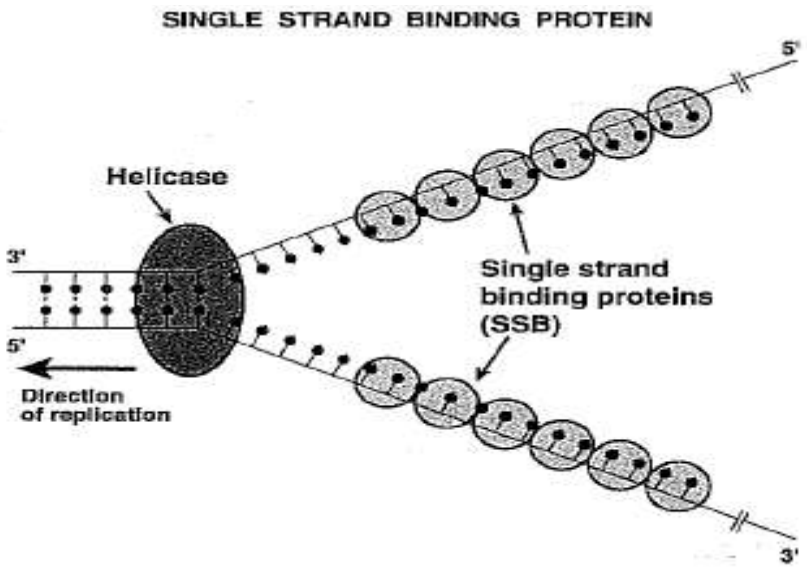
stranded DNA molecule so the first step in replication is to separate the two strands of the DNA double helix.

The next step is to build a complementary strand on each of the two original strands. Since A only pairs with T, and since G only pairs with C, the sequence of each strand dictates the sequence of its complementary strand. We now have two double stranded DNA molecules, both with sequences identical to the original one. One of these daughter molecules has the original left strand and the other daughter has the original right strand. This is known as semi-conservative replication since each of the progeny conserves half of the original DNA molecule.

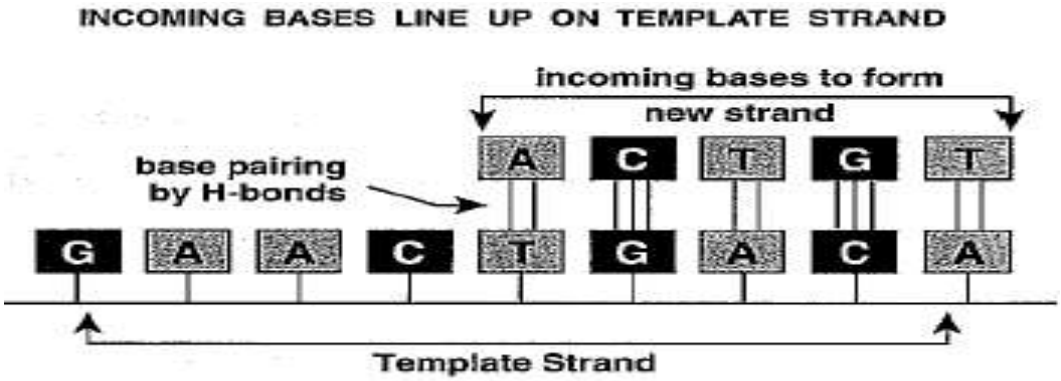
Because the two strands forming a DNA molecule are held together by hydrogen bonding and twisted around each other to form a double helix, they cannot simply be pulled apart. Worse still, the DNA inside a cell is also supercoiled to pack it into a small space. Before separating the strands, both the supercoils and the double helix must be unwound.

This is done in two stages. First the supercoils are unwound by an enzyme known as DNA gyrase. The gyrase cuts both strands of double stranded DNA to give a double stranded break. However, it keeps hold of all of the cut ends. The two halves of the gyrase then rotate relative to each other and the ends are rejoined. This untwists the supercoils. Each rotation costs the cell a small amount of energy. Once the supercoils have been untwisted, the double helix is unwound by the enzyme DNA helicase. Helicase does not break the DNA chains; it simply disrupts the hydrogen bonds holding the base pairs together.

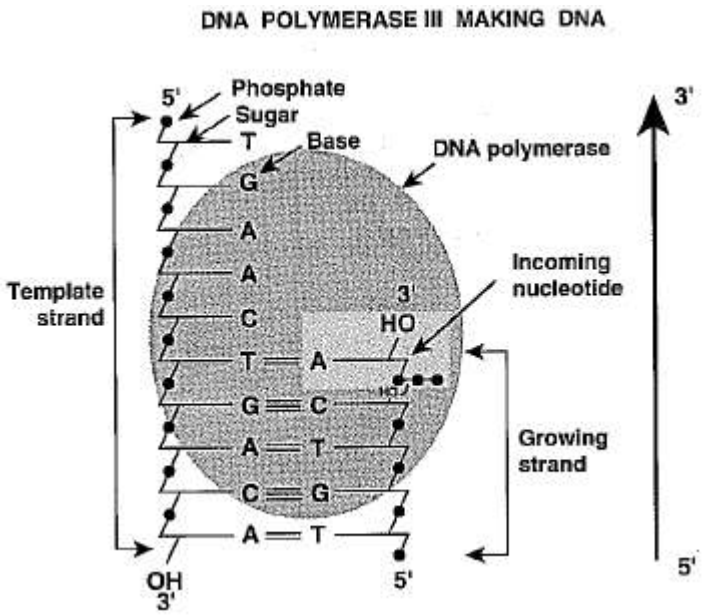
The two separated strands of the parental DNA molecule are complementary to each other. Consequently all of their respective bases are capable of pairing off and binding to each other. In order to manufacture the new strands, the two original strands, despite their desire to cling together, must somehow be kept apart. This is done by means of a special ‘divorce’ protein which binds to the unpaired single stranded DNA and prevents the two parental strands from getting back together. This is known as single strand protein or SSB.

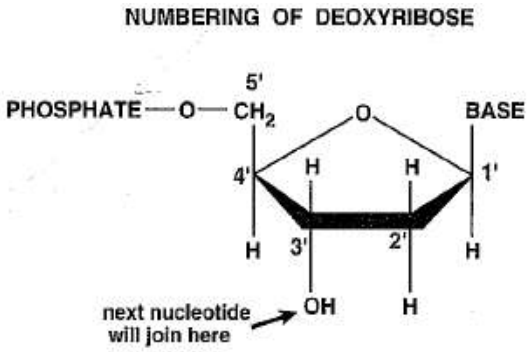


The critical issue in replication is the base pairing of A with T and of G with C. Each of the separated parental strands of DNA serves as a template strand for the synthesis of a new complementary strand. The incoming nucleotides for the new strand recognize their partners by base pairing and so are lined up on the template strand.

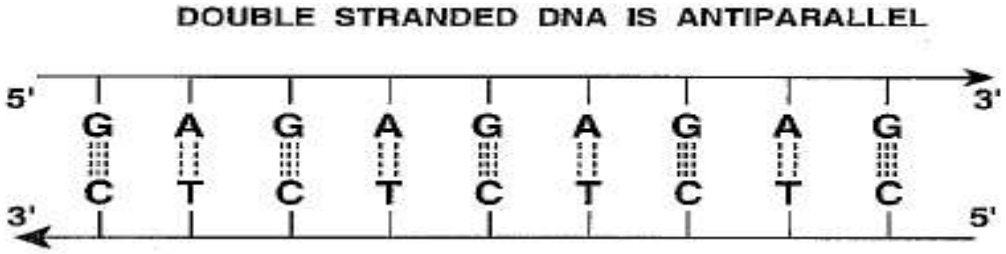


Actually, things are a bit more complicated. Although hydrogen bonding alone would match bases correctly 99 percent of the time, this is not good enough. The enzyme that links the nucleotides, known as DNA polymerase III or Pol III can also sense if bases are correctly paired. If not, the mismatched base pair is rejected. The nucleotides are then joined together by the enzyme. This DNA polymerase has two subunits. One of these is the synthetic subunit and is responsible for manufacturing new DNA. The other subunit is shaped like a doughnut and slides up and down like a curtain ring on the template strand of DNA. This 'sliding clamp' subunit binds the synthetic subunit to the DNA.



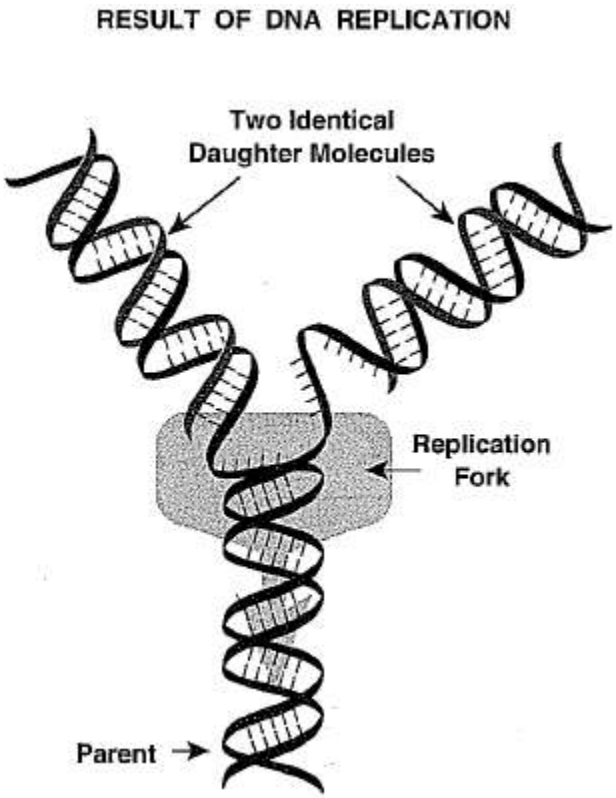


About synthesis, nucleotides, as we've already said, have three components: a phosphate group, a sugar and the base. In DNA the sugar is deoxyribose, and is joined to the base at position 1' and to the phosphate group at position 5'. The carbon atoms of the deoxyribose sugar are numbered with prime marks to distinguish them from those of the base which have plain numbers. When a new nucleotide is added it is joined, via its own phosphate group on position 5' to the 3' position as indicated by the arrow. New DNA strands always start at the 5' end and grow in the 3' direction. In fact, all nucleic acids, whether DNA or RNA, are always made in the 5' to 3' direction. However, DNA is normally double stranded, and it happens that the two strands run in opposite directions; that is, if one goes 5' to 3' then its complementary partner will run from 3' to 5'. The strands are said to be antiparallel.



Since DNA is always made in the 5' to 3' direction, and since the two strands of double helical DNA are antiparallel, this means that during DNA replication the two new strands must be synthesized in opposite directions. Because of this, one strand is made continuously and is referred to as the leading strand and the other strand can only be made in short segments and is known as the lagging strand. As the two new strands of DNA are synthesized, two double helical DNA molecules are produced, each with one

old and one new strand. Once replication fork has moved past, the double stranded DNA molecule automatically rewinds into a helix.



2.1.3 Usage Of Genetic Information

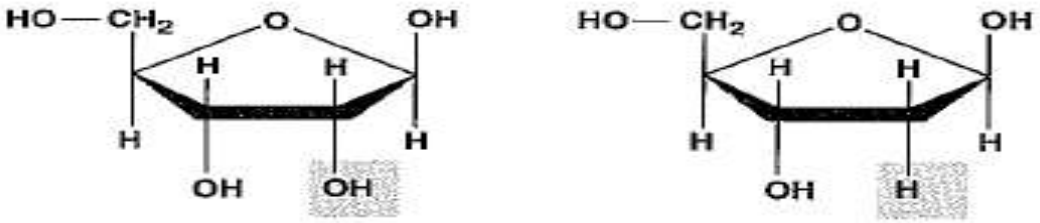
During the day-to-day life of a cell, working copies of the genes are used. The DNA molecule that carries the original copy of the genetic information is regarded as sacred and is not used as a direct source of instructions to run the cell. Genetic information can be carried by two kinds of nucleic acid molecules, DNA or RNA. The working copies of genes are made of RNA, or ribonucleic acid, which is very similar in chemical structure to DNA. The particular type of RNA molecule that carries genetic information from the genes into the rest of the cell is known as messenger RNA, usually abbreviated to mRNA. The transfer of information from DNA to messenger RNA is known as transcription.

For a gene to be transcribed, the DNA, which is double stranded, must first be pulled apart temporarily. Then a molecule of single stranded RNA is

made. This is the messenger RNA and it has a base sequence which is complementary to that of the DNA strand used as a template.

There are two related kinds of nucleic acid (DNA) and ribonucleic acid (RNA). The first difference between them is that in DNA the sugar is always deoxyribose, whereas in RNA the sugar is ribose, as it's shown in the scheme below. As its name suggests, deoxyribose has one less oxygen atom than ribose. It is this initial difference which gives the D in DNA versus the

RIBOSE VERSUS DEOXYRIBOSE



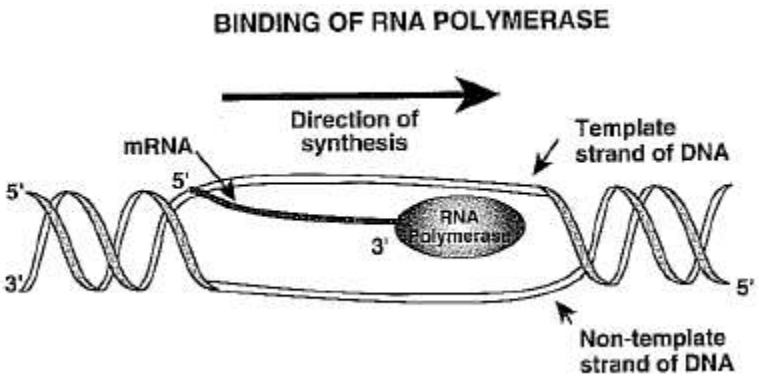
Ribose Deoxyribose

R in RNA! The second difference is that in RNA, the base thymine (T) is replaced by the closely related base uracil (U) (next scheme). Wherever you find thymine in DNA, you get uracil in RNA. Hence, uracil in RNA and thymine in DNA convey the same genetic information. So, if you include RNA with DNA, the genetic alphabet has five letters (A, C, G, T and U). The third and final difference between DNA and RNA is that DNA is double stranded (ds), whereas RNA is normally single stranded (ss). Thus, when a gene made of dsDNA is transcribed into an RNA message, only one of the strands of DNA is copied. The sequence of the RNA message is complementary to the template strand of the DNA upon which it is synthesized. Apart from the replacement of thymine in DNA with uracil in RNA, this means that the sequence of the new RNA molecule is identical to the sequence of the coding strand of DNA, the one not actually used during transcription.

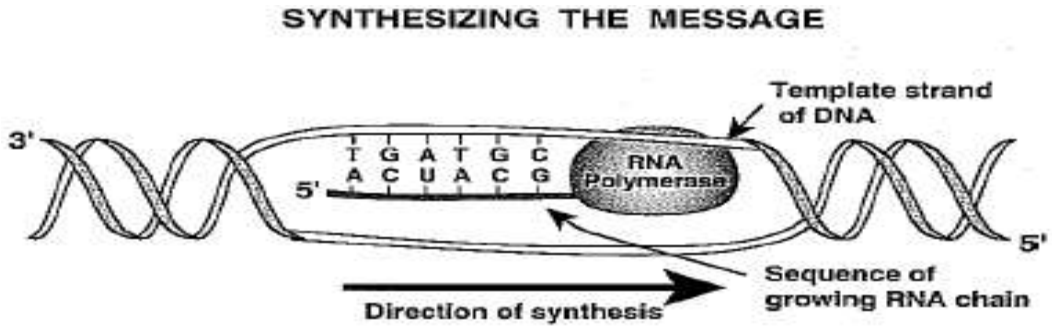
THYMINE VERSUS URACIL

Thymine Uracil

RNA is made by an enzyme called RNA polymerase. This enzyme binds to the DNA at the start of a gene and opens the double helix. It then goes on to manufacture an RNA message. The region where RNA polymerase binds is known as the promoter. Once sigma has found a promoter and the RNA



polymerase has successfully bound to it, the sigma subunit drops off. The remaining part of bacterial RNA polymerase, known as the core enzyme, then makes the mRNA, The DNA double helix is opened up and a single strand of RNA is generated using one of the DNA strands as a template for matching up the bases.



Proteins have an important role in the synthesis of messenger RNA. There are activator proteins which help turn genes on, but there are also proteins that can turn genes off. Historically, these negative regulators were actually discovered first, they are known as repressors and they work in a similar way to activators except they have the opposite effect.

2.1.4 Proteins And Their Role In The Cell

Proteins are biological polymers that carry out most of the cell's day-to-day functions. Some proteins are merely structural or take part in cell movement, others help take up nutrients, others generate energy and yet others carry out biochemical reactions, including the synthesis of nucleotides and their assembly into nucleic acids.

Molecules whose primary role is to carry information (nucleic acids like DNA and messenger RNA) are basically linear molecules with a regular repeating structure. Molecules that form cellular structures or have active

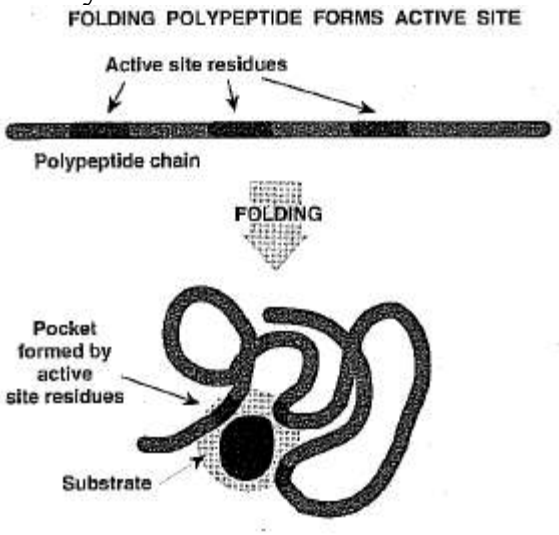
roles carrying out reactions are normally folded into three-dimensional (3-D) structures. These include both proteins and certain specialized RNA molecules (rRNA and tRNA).

Proteins are made from a linear chain of monomers, known as amino acids, and are folded into a variety of complex 3-D shapes. A chain of amino acids is called a polypeptide chain. The difference between a polypeptide chain and a protein is that some proteins consist of more than one polypeptide chain. We can subdivide proteins into four main categories:

1) *structural proteins,*
2) *enzymes,*
3) *regulatory proteins,*
4) *transport proteins*.

Structural proteins are found making up many subcellular structures. The flagella with which bacteria swim around, the microtubules used to control traffic flow inside cells of higher organisms, the fibers inside a muscle cell, and the outer coats of viruses are some few examples of structures built using proteins.

Enzymes are proteins that carry out chemical reactions. An enzyme first binds another molecule, known as its substrate, and then performs some chemical operations with it. Some enzymes bind only a single substrate molecule; others may bind two or more, and react them together to make the final product. In any case, the enzyme needs an active site, a pocket or cleft in the protein, where the substrate binds and the reaction occurs. The active site is produced by folding up the polypeptide chain correctly so that amino acid residues that were spread out at great distances in the linear chain now come together and will cooperate in the enzyme reaction.



The most famous enzyme in molecular biology is β-galactosidase, encoded by the *lacZ gene* of the bacterium *Escherichia coli*. This enzyme is so easy to assay that it is widely used in genetic analysis. The natural substrate of β-galactosidase is the sugar lactose, made by linking together the two simple sugars, glucose and galactose. There is not much else to do with lactose except to split it into these two simpler sugars, so that is exactly what β-galactosidase does.

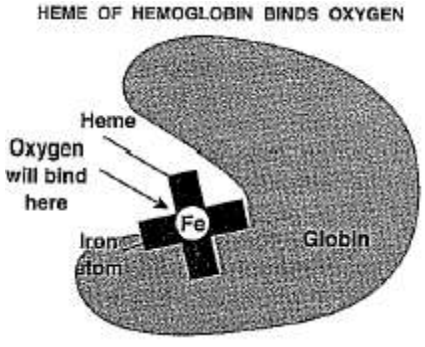
Analogs are molecules resembling natural substances well enough to fool the enzymes that use them. Some analogs bind but do not react and simply block the active site and inhibit the enzyme. Such analogs are known as competitive inhibitors since they complete with the true substrate for the attention of the enzyme. Other analogs do react. β-galactosidase splits many molecules in which galactose is linked to something else. We can take advantage of this by giving it ONPG (*ortho-nitro-phenyl-galactoside*), which consists of ortho – nitrophenol linked to galactose. When ONPG is split, we get galactose, which is colorless, and ortho – nitrophenol, which is bright yellow. Using ONPG allows us to monitor the level of β-galactosidase by measuring the appearance of the yellow color. Similarly, X-gal is split by β-galactosidase into a blue dye and galactose.

Although regulatory proteins and transport proteins are not enzymes, they also bind other molecules and so they also need 'active sites' to accommodate these.

Regulatory proteins vary enormously. Many of them can bind both small signal molecules and DNA. The presence or absence of the signal molecule determines whether or not the gene is switched on.

Transport proteins are found mostly in biological membranes where they carry material from one side to the other. Nutrients, such as sugars, must be transported into cells of all organisms, whereas waste products are deported. Multicellular organisms also have transport proteins to carry materials around the body. An example is hemoglobin, which carries oxygen in blood.

To function properly many proteins need extra components, cofactors or prosthetic groups, which are not themselves proteins. Many proteins use single metal atoms as cofactors; others need more complex molecules. Strictly speaking, prosthetic groups are fixed to a protein, whereas cofactors are free to wander around from protein to protein; however, the terms are often used loosely. A protein without its prosthetic group is referred to as an apoprotein. For example, oxygen carrier proteins such as hemoglobin have a cross-shaped cofactor with a central iron atom, called heme. The heme is bound in the active site of the apoprotein, in this case globin, and so we get hemoglobin. Oxygen binds to the iron atom at the center of the heme and the hemoglobin carries it around the body. Prosthetic groups are often shared by more than one protein; for example, heme is shared by hemoglobin and by myoglobin, which receives oxygen and distributes it inside muscle cells.



2.2 Cells As I/O Systems[SONTAG 04]

One may view cell life as a collection of "wireless networks" of interactions among proteins, RNA, DNA and smaller molecules involved in signaling 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 has provided, and will continue to produce, a wealth of data describing the elementary components of such networks, as well the mapping of intra and inter-cellular signaling networks. The genome encodes, through a particular ordering of the four possible (A,T,C,G) bases in its DNA sequence, a parts list for the proteins that are potentially present in every cell of a given organism. Genomics research has as its objective the complete decoding of this information, both the parts common for a species as a whole as well as the cataloging of differences among individual members. The shape of proteins is what largely determines their function, and thus the elucidation of their three-dimensional structure is a goal of proteomics research. Proteins, which interact with each other through lego-like fitting of parts in lock and key fashion, are the primary components of living things. Among other roles, they form receptors that endow the cell with sensing capabilities, actuators that make muscles move (myosin, actin), detectors for the immune response, enzymes that catalyze chemical reactions, and switches that turn genes on or off. They also provide structural support and help in the transport of smaller molecules, as well as in directing the breakdown and reassembly of other cellular elements such as lipids and sugars. (An intermediate link between genetic information and the proteins that DNA encodes for is RNA. Until recently, RNA was not believed to be a direct player in cell control mechanisms, but research into microRNA conducted within the past two years is forcing a complete rethinking of their role.) Massive amounts of data are being generated by genomics and proteomics projects, facilitated by sophisticated genetic engineering tools (gene knock-outs and insertions, PCR), and measurement technologies (green fluorescent protein, microarrays, FRET), and there is a widely recognized need to organize and interpret these data.

The control and systems-theory paradigm of input/output systems, built out of simpler components that are interconnected according to certain rules,

is a most natural one in this context. Cells receive external information through inputs that may be physical (UV or other radiation, mechanical, or temperature) as well as chemical (drugs, growth factors, hormones, nutrients), and their measurable outputs include chemical signals to other cells, the movement of flagella or pseudopods, the activation of transcription factors, and so forth. Each cell can be though of, in turn, as composed of a large number of subsystems, involved in processes such as cell growth and maintenance, division, and death. Indeed, an important theme in the current molecular biology literature [HARTWELL ET AL. 99, LAUFFENBURGER 00] is the attempt to understand cell behavior in terms of cascades and feedback interconnections of elementary "modules".

As a simple illustration, consider the diagram shown in Chapter 4, Figure 1, extracted from the paper on cancer research [HANAHAN &WEINBERG 00], which describes the wiring diagram of the growth signaling circuitry of the mammalian cell. Of course, such a figure leaves out a lot of information, some known but omitted for simplicity, and some unknown: much of the system has not been identified yet, and the numerical values of most parameters as well as the functional forms of interactions are only very approximately known. However, data is being collected at an amazing rate and better and better models are being constantly obtained.

Many of the natural systems-theoretic questions that one would normally pose for such a system are precisely those that leading biologists are asking, if sometimes in a different language: What is special about the information-processing capabilities, or input/output behaviors, of such networks, and how does one characterize these behaviors? How do the different signal transduction pathways interact? How does one find the forms of reactions, and values of parameters (identification, reverse engineering)? Once these forms of reactions are known, how does one estimate time-varying internal states, such as the concentrations of proteins and other chemical substances, from input/output experiments (observer problem)? What subsystems appear repeatedly? Where lie the main sensitivities affecting robustness of the system? What is the reason that there are cascades and feedback loops? More generally, what can one say, if anything, about stability, oscillations, and other dynamical properties of such complex systems? In addition to analysis questions, there are, of course, also synthesis ones, dealing with the control of cellular systems through drugs or genetic modifications. However, the fact is that the field of Systems Biology, as we will see with more details in Chapter 3, is still in its infancy, and a major and long-term research effort will continue toward the solution of the above types of problems.

2.3 Signaling Pathways

2.3.1 The Reason Of Their Creation

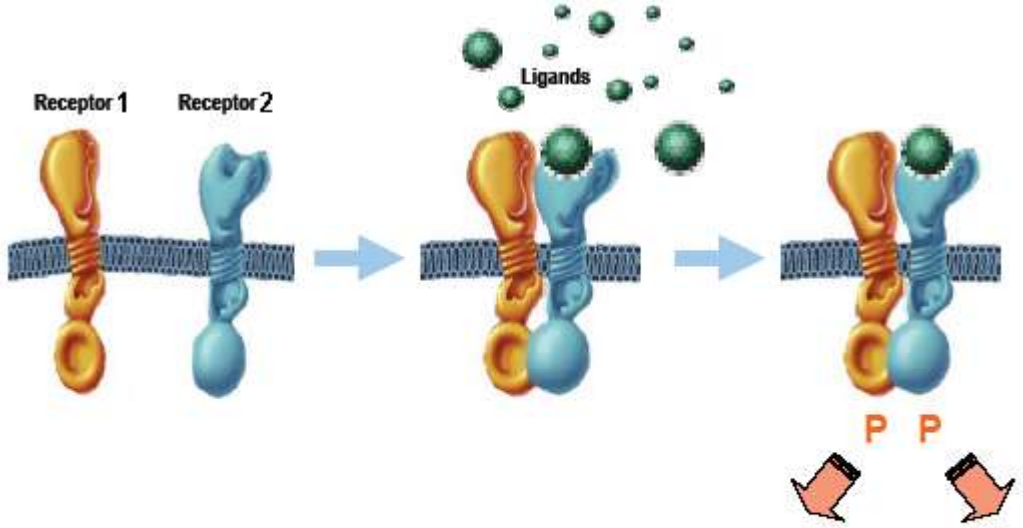
Advancements in molecular biology, genomics and proteomics have yielded a wealth of information regarding the underlying molecular basis of cancer. While the precise nature and mechanisms that underlie this disease are not known, a well-established body of scientific evidence has demonstrated that signaling pathways, and communication between them, are critical to the genesis and proliferation of cancer. Based on these advancements, the concept of "personalized medicine" has emerged, promising the ability to develop and administer specifically targeted treatments for specific types of cancer in specific individuals. While the potential this approach holds for the treatment of cancer remains strong, current prognostic and diagnostic methods have not proven effective in identifying, profiling and targeting the most appropriate patient sub-populations to enable the consistent link between treatment and patient that the personalized medicine approach promises.

With the recent deciphering of the Human Genome, research in the area of pharmacogenomics is expanding rapidly. Complementing pharmacogenomics is a growing need to both quantify and measure a broad range of protein targets and activation events in normal and diseased cells. In both cases, scientists are working to correlate the expression of a gene or protein with a particular disease in the hope of using this information to develop new and more effective targeted treatments. While researchers conduct both gene and protein analysis to gain a more comprehensive profile of tumor cell activity, only through an understanding of the protein interactions taking place on a tumor cell's surface and in its internal pathways, can targeted treatments be both developed and delivered with optimal effectiveness. While gene analysis may someday predict what can happen, proteins provide more direct insight into what is actually occurring within the cells and signaling pathways.

Protein targets may be found within a cell, secreted from a cell or on a cell surface and may be associated, or interact with, many other proteins. Analysis of several protein targets and their interactions is necessary to gain a complete understanding of a particular tumor's biology, and hence the type of targeted treatment that can be used to treat it.[ACLARA 04]

2.3.2 Definition - Specification Of Their Role

Signal transduction is a means by which one cell communicates with another and to external stimuli. The communication between the two cells involves a molecular messenger (**ligand**) from the sender and a site (**receptor**) on the cell membrane that receives the signal. When the signal is received, it is passed along and the message is communicated from the outer cell surface to the cell's nucleus. A signaling pathway transmits information through a cell via a series of steps involving assembly and disassembly of protein-protein complexes.



1. Ligands cause two receptors to dimerize.

2. Phosphorylation occurs, activating pathways in the cell

Messages can be healthy or harmful. For example, some messages might be used by the immune system to increase the amount of white blood cells needed to stop an infection. In other cases, signals may cause cells to store materials such as fatty acids, which is healthy in moderation, but when uncontrolled, can lead to obesity. In all cells, some level of growth (signaling) is normal and a part of the regular cell cycle. It is the overexpression, or overactivation of these signals – or the failure to counterbalance or block those signals – that leads to uncontrollable growth (i.e. cancer). Signaling cascades regulate cell growth, differentiation and survival as a function of complex extracellular triggers. Cancer results from a dysregulation of such processes. Hence, the rationale for carrying out basic

research in signaling cascades and cell biology is that this will eventually lead to new approaches to prevention, diagnosis and treatment of human cancer. The post-genomic era will enable the delineation of complex signaling pathways and disclose the intricate interactions between DNA repair, replication, transcription, chromatin dynamics, cell cycle progression and apoptosis.

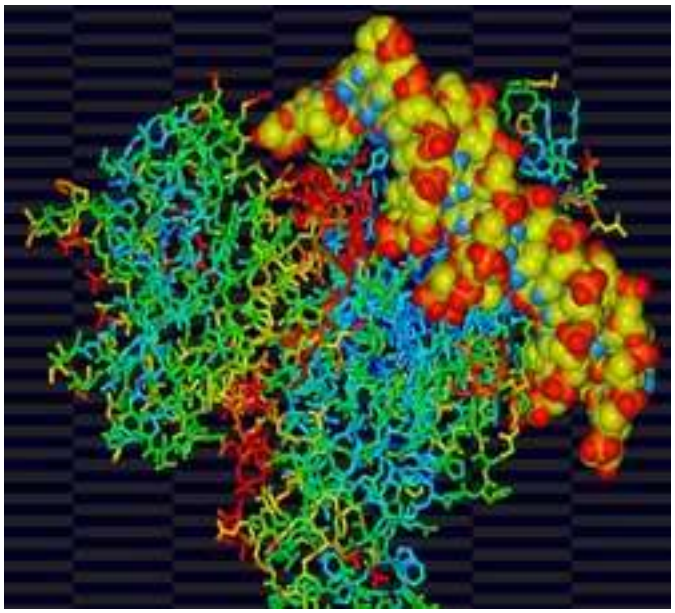
Identifying a signaling pathway is just the first, crucial step towards understanding a specific disease. Activation or inhibition of a pathway leads to a change in the amount or location of specific protein complexes. While genetic knowledge is important, understanding the protein interactions occurring within these pathways is critical to determining both the specific disease and how the pathway is activated. A key reaction of cellular signal processing is provided by reversible protein phosphorylation, catalyzed by a variety of protein kinases. Using protein kinase A (*Kinzel*) as paradigm, the mechanism of protein phosphorylation is studied at atomic resolution to develop ATP-competitors which block specific protein kinase sites. The existence of ecto-protein kinases and their substrates shows that phosphorylation serves as a powerful tool to inhibit growth of transformed cells and the underlying mechanism is characterized. Protein kinase CK2 has important roles in regulation of the cell cycle. The regulation of mammalian CK2 is examined by promoter analysis as well as by identifying proteins binding to CK2 tetramers. Protein kinase C (PKC) isoenzymes regulate cell-specifically cell growth and differentiation and are potential mediators of tumor-promoters.[NIEHRS & BIRCHMEIER 05]

2.3.3 Phosphorylation In Signaling Pathways

As we've already said, phosphorylation is the main key for the pathway's activation. It is defined as the process of adding a phosphate group to a molecule. This often activates the molecule by altering its shape or charge and allows it to participate in a chemical reaction. The activity of many proteins is regulated by phosphorylation of hydroxyl containing residues (serine, threonine, tyrosine) by various protein kinases. In eukaryotes, protein phosphorylation is probably the most important regulatory event. Many enzymes and receptors are switched "on" or "off" by phosphorylation and dephosphorylation. Phosphorylation is catalyzed by various specific protein kinases, whereas phosphatases *de*phosphorylate. An example of the important role that phosphorylation plays is the p53 tumor suppressor gene,

which—when active—stimulates transcription of gene that suppress the cell cycle, even to the extent that it undergoes apoptosis. However, this activity should be limited to situations where the cell is damaged or physiology is disturbed. To this end, the p53 protein is extensively regulated. In fact, p53 contains more than 18 different *phosphorylation* sites.[IRUSTA ET AL. 02]

P53, also known as TP53 or tumor protein is a gene that codes for a protein that regulates the cell cycle and hence functions as a tumor suppressor. It is very important for cells in multicellular organisms to suppress cancer. P53 has been described as "the guardian of the genome", referring to its role in conserving stability by preventing genome mutation. The name is due to its molecular mass: it is in the 53 kilodalton fraction of cell proteins.



Human p53 protein bound to a short DNA fragment. Protein atoms are represented as sticks, the DNA helix is in spacefill mode.

Upon the deactivating signal, the protein becomes dephosphorylated again and stops working. This is the mechanism in many forms of signal transduction, for example the way in which incoming light is processed in the light-sensitive cells of the retina.

The network underlying phosphorylation can be very complex. In some cellular signalling pathways, a protein A phosphorylates B, and B phosphorylates C, but A also phosphorylates C directly, and B can phosphorylate D, which may in turn phosphorylate A. More particular, in signaling pathways, receptor tyrosine kinases (RTKs) are transmembrane proteins that regulate numerous aspects of cell physiology including

proliferation and survival. Binding of a soluble ligand to the extracellular domain of these receptors typically induces receptor dimerization and transphosphorylation of the cytoplasmic catalytic domain. This tyrosine phosphorylation stimulates the intrinsic tyrosine kinase activity of the receptor and generates binding sites for signaling proteins containing SH2 domains. Although ligand-induced dimerization is an important trigger of receptor activation, receptor activity is also subject to additional levels of regulation. For example, the cytoplasmic juxtamembrane region of receptor tyrosine kinases, which is located between the transmembrane domain and the kinase domain, has been implicated in regulation of receptor enzymatic activity.[WIKIPEDIA]

2.3.4 Signaling Pathways And Cancer Diseases

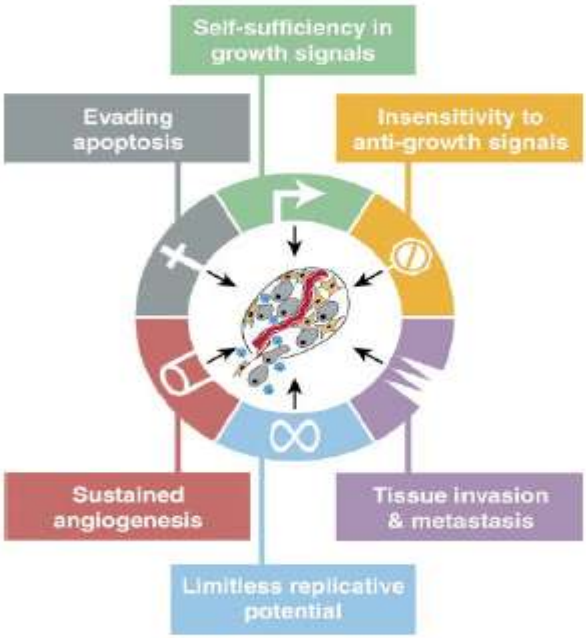
After a quarter century of rapid advances, cancer research has generated a rich and complex body of knowledge, revealing cancer to be a disease involving dynamic changes in the genome. The foundation has been set in the discovery of mutations that produce oncogenes with dominant gain of function and tumor suppressor genes with recessive loss of function; both classes of cancer genes have been identified through their alteration in human and animal cancer cells and by their elicitation of cancer phenotypes in experimental models.

Several lines of evidence indicate that tumorigenesis in humans is a multistep process and that these steps reflect genetic alterations that drive the progressive transformation of normal human cells into highly malignant derivatives. Many types of cancers are diagnosed in the human population with an age-dependent incidence implicating four to seven rate-limiting, stochastic events. Pathological analyses of a number of organ sites reveal lesions that appear to represent the intermediate steps in a process through which cells evolve progressively from normalcy via a series of premalignant states into invasive cancers.

Most cancer, if not all, have acquired the same set of functional capabilities during their development, albeit through various mechanistic strategies, as it's shown in the next scheme. Normal cells require mitogenic growth signals (GS) before they can move from a quiescent state into an active proliferative state. These signals are transmitted into the cell by transmembrane receptors that bind distinctive classes of signaling molecules: diffusible growth factors, extracellular matrix components, and cell-to-

celladhesion/interaction molecules. To our knowledge, no type of normal cell can proliferate in the absence of such stimulatory signals. Many of the oncogenes in the cancer catalog act by mimicking normal growth signal in one way or another.

Dependence on growth signaling is apparent when propagating normal cells in culture, which typically proliferate only when supplied with appropriate diffusible mitogenic factors and a proper substratum for their integrins. Such behavior contrasts strongly with that of tumor cells, which invariably show a greatly reduced dependence on exogenous growth stimulation. The conclusion is that tumor cells generate many of their own growth signals, thereby reducing their dependence on stimulation from their normal tissue microenvironment. This liberation from dependence on exogenously derived signals disrupts a critically important homeostatic mechanism that normally operates to ensure a proper behaviour of the various cell types within a tissue.



Acquired GS autonomy was the first of the six capabilities to be clearly defined by cancer researchers, in large part because of the prevalence of dominant oncogenes that have been found to modulate it. Three common molecular strategies for achieving autonomy are evident, involving alteration of extracellular growth signals, of transcellular transducers of those signals, or of intracellular circuits that translate those signals into action. While most soluble mitogenic growth factors (GFs) are made by one cell type in order to

stimulate proliferation of another—the process of heterotypic signaling—many cancer cells acquire the ability to synthesize GFs to which they are responsive, creating a positive feedback signaling loop often termed autocrine stimulation. Clearly, the manufacture of aGF by a cancer cell obviates dependence on GFs from other cells within the tissue. The production of PDGF (platelet-derived growth factor) and TGFa (tumor growth factor a) by glioblastomas and sarcomas, respectively, are two illustrative examples.

The cell surface receptors that transduce growth stimulatory signals into the cell interior are themselves targets of deregulation during tumor pathogenesis. GF receptors, often carrying tyrosine kinase activities in their cytoplasmic domains, are overexpressed in many cancers. Receptor overexpression may enable the cancer cell to become hyperresponsive to ambient levels of GF that normally would not trigger proliferation. For example, the epidermal GF receptor (EGF-R/*erbB*) is upregulated in stomach, brain, and breast tumors, while the HER2/*neu* receptor is overexpressed in stomach and mammary carcinomas.[HANAHAN &WEINBERG 00]

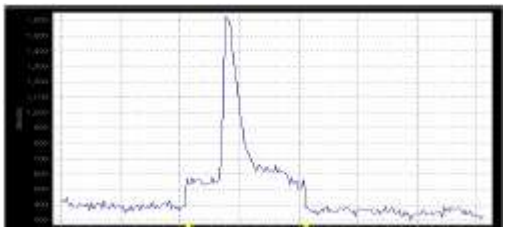
Lung cancer is the most frequent cause of cancer-related death in men and women and accounts for approximately more than a million deaths yearly worldwide. NSCLCs (non-small cell lung cancer) constitute 75% of primary lung cancers and are comprised of large-cell undifferentiated carcinomas, epidermoid carcinomas, and adenocarcinomas including bronchoalveolar lung cancers. Nearly 65% of NSCLCs exhibit significant heterogeneity, with 45% containing both adeno and squamous features. Phosphatidylinositol 3-Kinase/PTEN/Akt Kinase Pathway is strongly related with this kind of disease, having an important role in Tumor Necrosis Factor-related Apoptosis-inducing Ligand-induced Apoptosis in Non-Small Cell Lung Cancer Cells. Apoptosis is the programmed cell death and is involved in the maintenance of tissue homeostasis in normal physiology. The process of apoptosis has to be strictly controlled, as excessive or diminished apoptosis may contribute to various pathological conditions. [DJERBI 03] A reduced propensity to undergo cell death may for instance promote the development and/or progression of cancer or autoimmune syndromes. Tumor necrosis factor (TNF)-related apoptosis-inducing ligand (TRAIL)/APO-2L is a member of the TNF superfamily of signaling pathways and has been shown to have selective antitumor activity. TRAIL does not induce apoptosis in some non-small cell lung cancer (NSCLC) cells. These cells are resistant to TRAIL because of the phosphatidylinositol 3_-kinase (PI3-K)-dependent activation of Akt/protein kinase B. The expression of phospho-Akt varies at the functional level but not at the

mRNA level in NSCLC cells. Akt induces cell survival in NSCLC cells by blocking the Bid cleavage, upstream of cytochrome *c* release in the mitochondrial-dependent apoptotic pathway. The use of PI3-K inhibitors, Wortmannin or LY-294002, down-regulates the active Akt and reverses cellular resistance to TRAIL. In addition, genetically altering Akt expression by transfecting dominant negative Akt, sensitizes NSCLC cells to TRAIL. Conversely, transfection of constitutively active Akt into cells that express low, constitutively active Akt, increases TRAIL resistance. Alternate to this approach, transfection with PTEN, a lipid phosphatase, promotes sensitivity to TRAIL, whereas a PTEN mutant (PTEN-G129E) at the catalytic site is inactive in dephosphorylating active Akt. Furthermore, the loss of PTEN activity or overexpression of PI3-K-dependent Akt/protein kinase B activity promotes the survival of NSCLC cells. Modulation of Akt activity by combining pharmacological drugs or genetic alterations of the Akt expression induces cellular responsiveness to TRAIL. Thus, TRAIL can be used to treat NSCLC-resistant cells when combined with agents that down-regulate Akt activity.[KARTHIKEYAN & RAKESH 02]

Moreover, studies of inherited cancer syndromes have implicated numerous signaling pathways in colorectal carcinogenesis, but the relationship between these pathways remains poorly understood. A new mouse model of juvenile polyposis syndrome identifies a molecular mechanism for clonal epithelial cell expansion and links several pathways with an established role in polyposis. Intestinal polyps arise from single epithelial precursor cells by clonal expansion of their progeny. An increase in the population of affected daughter cells occurs through multiplication of affected intestinal crypts and spreading of cells over the unaffected neighbouring crypts. Classically, intestinal polyps are categorized into two classes based on histology. Adenomatous polyps are neoplasms showing clonal expansion of dedifferentiated epithelial cells and are considered precursors to adenocarcinomas. Hamartomatous polyps, in contrast, are lesions in which expansion of epithelial cells coincides with expansion of other mucosal cell types. Unlike adenomatous polyps, epithelial cells in hamartomatous polyps seem to differentiate normally and are not necessarily neoplastic in nature, although secondary adenomatous transformation does occur in these polyps. Juvenile polyposis syndrome (JPS) is a rare autosomal dominant syndrome characterized by development of multiple hamartomatous polyps throughout the gastrointestinal tract. Some families with JPS carry mutations in the genes encoding the BMP receptor BMPR1A and its signaling intermediate SMAD4. As BMP signaling has been shown to regulate apoptosis of intestinal epithelial cells, mutations in BMP pathway components could interfere with programmed cell death, resulting in the

epithelial overgrowth observed in JPS. But the mechanism underlying the increase in the rate of crypt fission in JPS remains unknown. The mice had hamartomatous polyps similar to those seen in individuals with JPS, with amplification of epithelial precursor cell members and an increased rate of crypt fission. The authors were focused on the molecular mechanism underlying this precursor cell expansion and observed that both phosphatidyl inositol 3 (PI3) kinase-Akt and β-catenin-T-cell factor (TCF) signaling seem to occur exclusively in intestinal stem cell (ISC). They further showed that, in BMPR1A-negative polyps, the number of ISCs is greatly increased. That was a kind of proof about the linking between these three pathways and the regulating ISC expansion.[GIJS 04]

Furthermore, techniques have been described that allow mechanical forces to be applied to specific cell surface receptors, such as integrins, via the use of micron-scale magnetic particles coated in specific ligands or antibodies. There were investigations about the potential of using magnetic particle based techniques to mechanically condition cells via the activation of specific mechanotransduction pathways. The result was that the magnet field alone had no effect on either cell type. The presence of magnetic particles in the absence of an applied magnetic field also had little effect on calcium signalling, although there was a slight increase in the percentage of cells exhibiting Ca2+ transients compared to normal cells. Both osteoblasts and bone marrow stromal cells exhibited increased levels of calcium activity in response to magnetic loading. There were however significant differences in both the background levels of spontaneous calcium activity and the characteristics of magnetically stimulated calcium responses between the different cell types. Bone marrow stromal cells typically responded with a single Ca2+ transient whereas responsive osteoblast cells demonstrated both single and oscillating Ca2+ transients. Time to onset of response was also found to vary between cell types. [HUGHES ET AL. 03]



Ca2+ trace from hBMSc exposed to magnetic loading. Magnet applied between yellow arrows. Image every 2 sec.

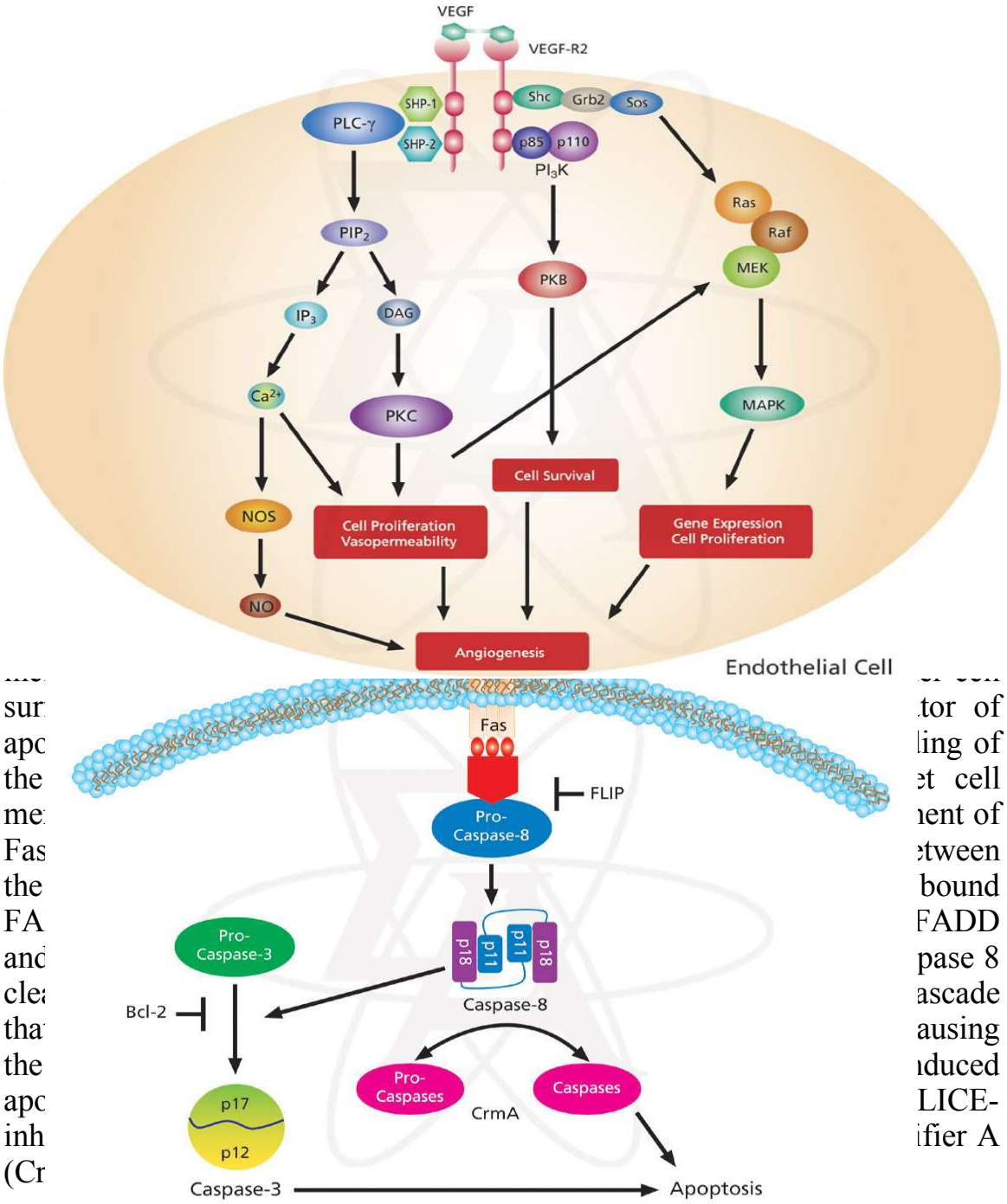
Finally, intense investigation into the molecular basis of angiogenesis is rapidly revealing novel signaling pathways involved in the generation of new vasculature. These range from elucidation of the mechanism by which hypoxia initiates expression of a proangiogenic gene repertoire via the hypoxia-inducible transcription factors (HIFs) to molecular pathways involved in extra- and intracellular signaling during new vessel formation. Extracellular pathways include those of the Notch/delta, ephrin/Eph receptor

and roundabout/slit families, and intracellular pathway members of the hedgehog and sprouty families. Angiogenesis has been a topic of vigorous research for more than a decade, a situation stimulated by the discovery of key angiogenic growth factors vascular endothelial growth factor (VEGF) and basic and acidic fibroblast growth factors. There have been many recent reviews on the processes involved in angiogenesis and its role in cancer and as a therapeutic target, focused on more recently described pathways of the past two to three years, particularly those initially identified in embryonic vascular development and differentiation. These pathways are important in normal and pathological angiogenesis (development of blood vessels from a preexisting vasculature) and also in vasculogenesis (development of blood vessels from progenitor cells).[BICKNELL & HARRIS 04]

2.3.5 Kinds Of Signaling Pathways

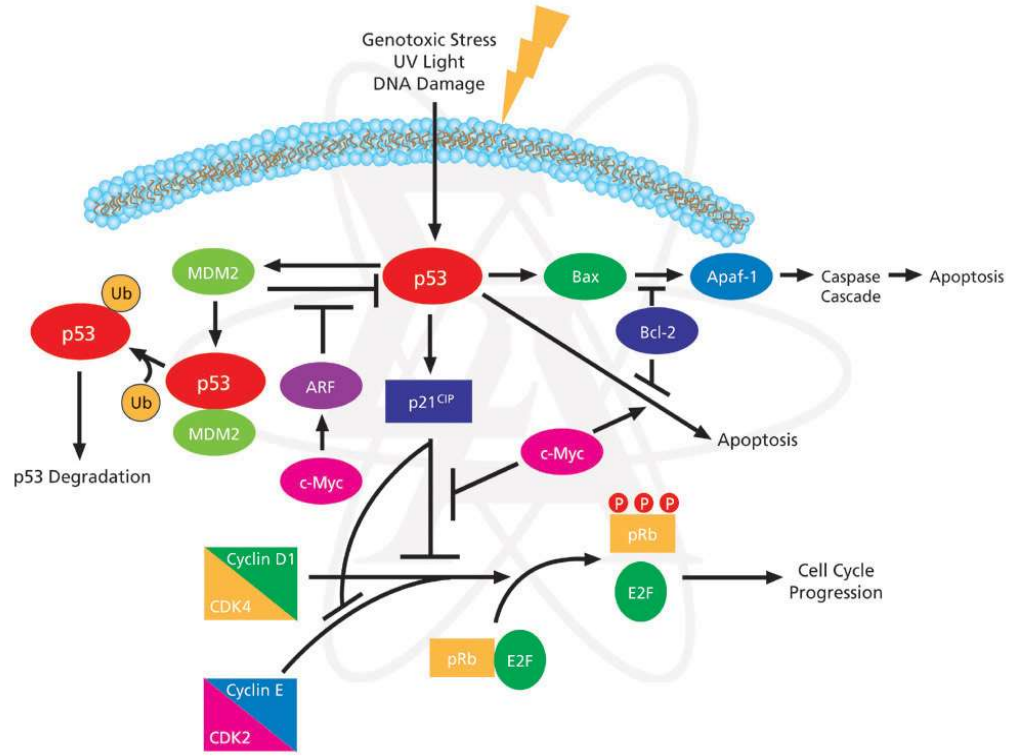
There are several kinds of signaling pathways, depending on the way of their activation and the type that they interact to the cell. Some of them are shown below:

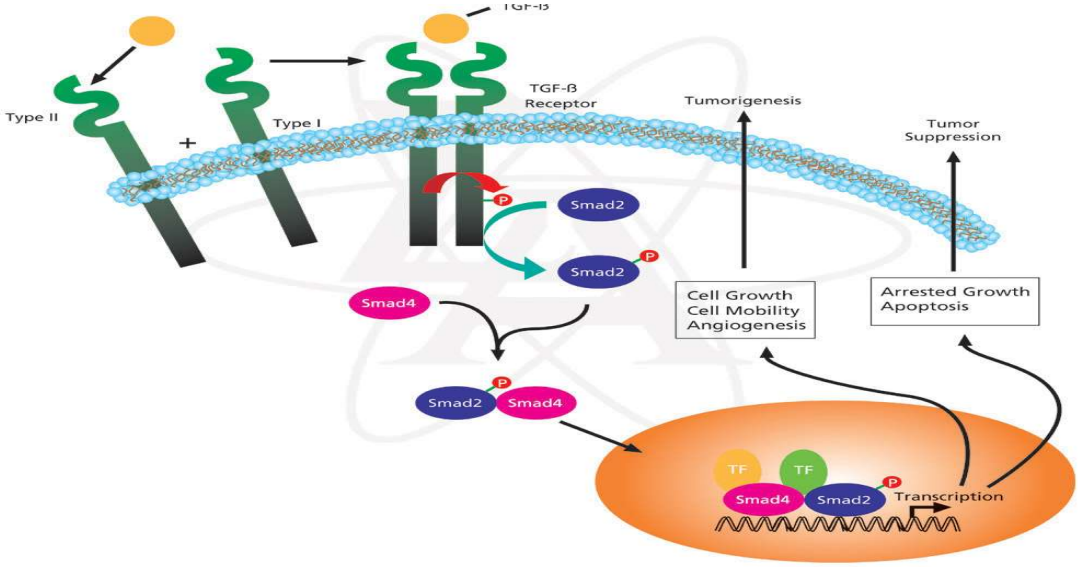
- **Signaling Pathways Activated by VEGF**:



- **The p53 Signaling Pathway:**

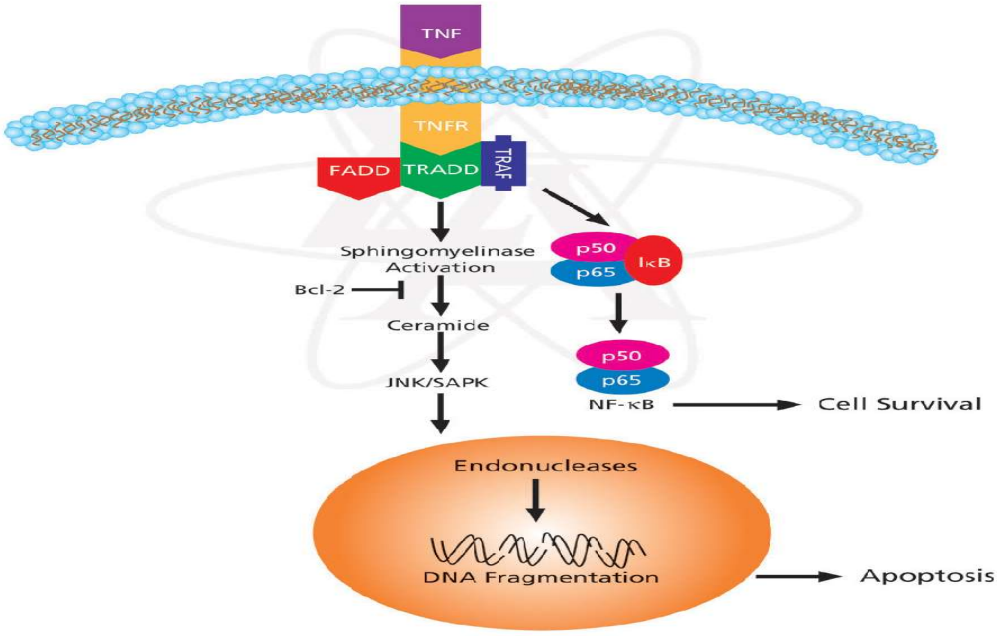
bin)NA-
ind and
gen gnals
dar DNA
res that
to help
arr r by
act ation
and ds to
reti n of
G1- g the
der y the
thr olled
p53 12 to
and irrest





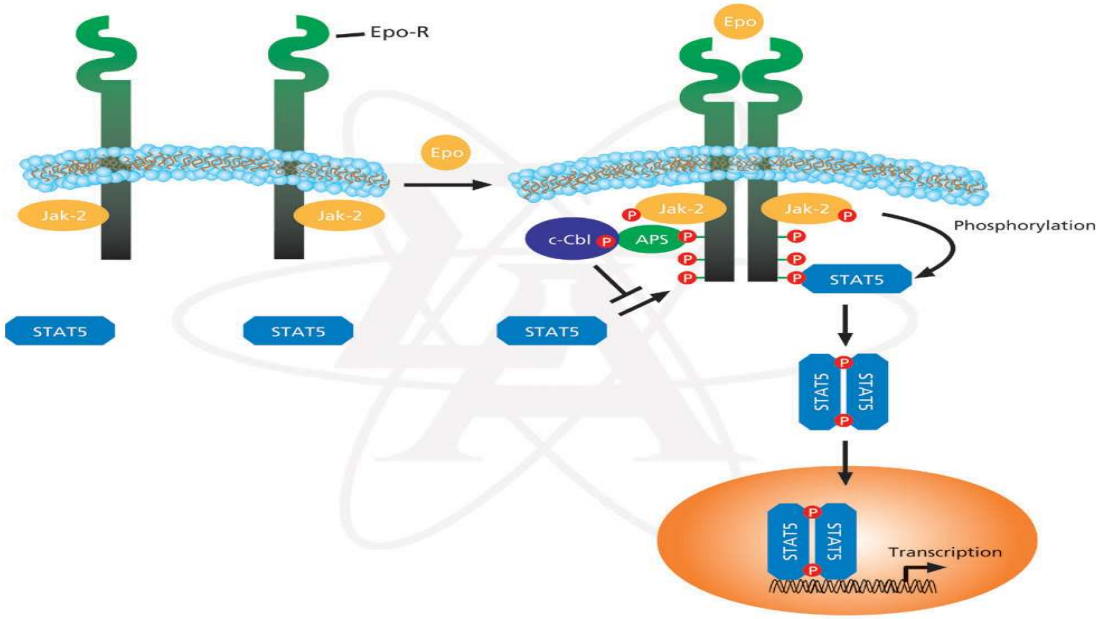
ma
sub
fam
cell
rec
act
tra
tra
tha
gen
[K

- **TNF Signaling Pathway:**



JFR) F is leath and ed a)), a totic otion B at clear ls in ergo may and ET AL. 98]

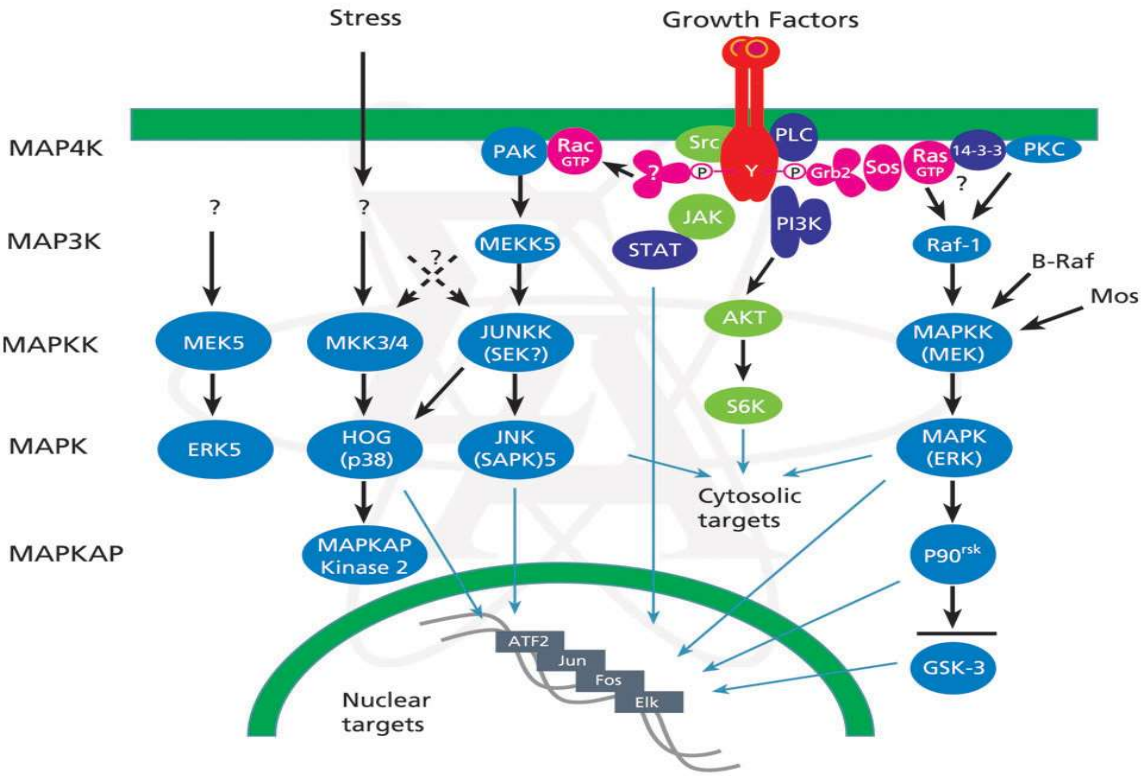
- **The JAK/STAT Signaling Pathway:**



tran Ma sup and rec JAI tail the the AL SH the ele AP can inhibit the JAK- STAT pathway by binding to the cytoplasmic domain of the receptor where it is phosphorylated (activated) by JAK. Activated APS binds to c-Cbl and blocks STAT activation.[SCHINDLER 99]

- **The Mitogen-activated Protein Kinase (MAPK) Cascades:**

inc
and
[LC
cell
gro
the
MA
pho
also
cas
diff
sig
foc



Systems theory or systems science has never managed to achieve widespread and independent status in curricula, departments, and journals but instead acts as an umbrella for a number of research activities across the physical and engineering sciences. Now, with revolutionary developments in the life sciences, there is renewed interest in systems thinking. In this chapter, we survey opportunities and challenges for the application of systems theory to biology in the postgenomic era—a new area of research also referred to as *Systems Biology*.

3.1 History

Although generally considered to be a new area of research, systems biology is not without history, and as early as the 1960s the term was used to describe the application of systems and control theory to biology [WOLKENHAUER 01]. At the time, Mesarovic wrote: "In spite of the considerable interest and efforts, the application of systems theory in biology has not quite lived up to expectations. ... one of the main reasons for the existing lag is that systems theory has not been directly concerned with some of the problems of vital importance in biology." Today, scientists in this field are motivated by the availability of experimental data, including, for example, DNA microarray time series, and interdisciplinary collaborations are widely supported. In fact, the importance of interdisciplinary research and close collaborations between biologists and physical scientists is evident in the many multidisciplinary research centers that are being built around the world, gently forcing researchers to interact by confining them into purpose-built housing. Mesarovic further suggested that progress could be made by more direct and stronger interactions of biologists with system scientists: "The real advance in the application of systems theory to biology will come about only when the biologists start asking questions which are based on the system-theoretic concepts rather than using these concepts to represent in still another way the phenomena which are already explained in terms of biophysical or biochemical principles. ...then we will not have the 'application of engineering principles to biological problems' but rather a field of *Systems Biology* with its own identity and in its own right."
Molecular characterization has led to very accurate spatial representations of cellular components, and biochemical modeling has been the main approach to studying cellular processes. However, the future lies in extending this knowledge to observations at higher organizational levels. There are few examples of a concerted effort to "translate" biological representations of gene expression and regulation into the language of the system scientist [KREMLING ET AL. 00], [KITANO], and all indications are that the field is going to provide the vital interface between basic cell biology, physiology, and biotechnological applications such as in metabolic engineering.

Systems biology has new technologies available to generate data from the genome, transcriptome, proteome, and metabolome, in addition to the physiome. However, while bioinformatics is usually associated with vast amounts of data available in databases, the systems-biological description of cellular processes often suffers from a lack of data.

3.2 Characteristics

Systems Biology is an emerging field expected to have major impact on the future of biological and medical research. It aims at system-level understanding of biological processes employing mathematical analysis and computational tools to integrate the information content obtained in experimental biology. In order to generate accurate descriptions of the interaction and operation of different components and to understand at a quantitative level the relationship among genotype and phenotype it was necessary to develop a research environment in which experimental and theoretical scientists work together. Hence, there have been an integration of experimental research with efforts to generate and optimize computer models of cellular networks and processes. This integration represents a real challenge for researchers in the future. In particular, in order to describe the *dynamic operation* of biological systems, the generation of quantitative data, time courses and spatial information were needed. In fact, generating the data to answer questions such as "Which proportion of protein X is located in compartment Y at time Z" or "How many molecules of metabolite A are present in compartment B" constituted a true experimental challenge. Therefore, Systems Biology will encompass the development of tools and approaches to generate such data. This type of information will help to better understand diseases and hence Systems Biology will become an integral part of drug target identification and drug design.

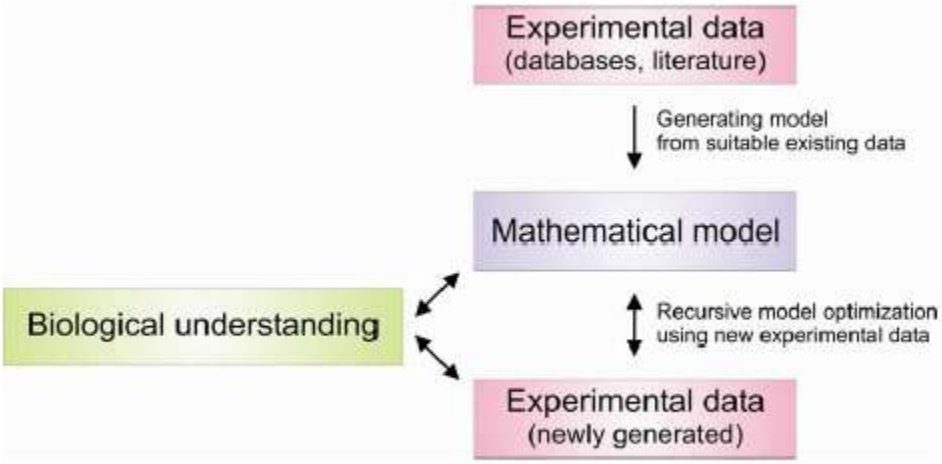
Generally, we can define Systems Biology as having the three basic characteristics:

- Multidisciplinary research with a close interaction between experimental research (biology, chemistry, physics…) and mathematical modelling and analysis (mathematics, bioinformatics, engineering…).
- Generation and iterative improvement of mathematical models, which describe realistically cellular networks or dynamic biological processes. Models have the ability to elucidate systems properties and to predict the outcome of perturbations.
- Where possible use of existing and generation of novel data, in particular quantitative data, time courses and spatial information of high definition.

Analyzing these characteristics, as we are going to do below, makes clear the importance of Systems Biology creation. Functional genomics, i.e. global approaches on gene and protein expression, metabolomics as well as interaction of cellular components, is generating huge amounts of data at increasing pace. Bioinformatics is required to evaluate and organize the data and make them available to researchers. The yeast community is realizing

that it studies an incredibly complex network of interconnecting pathways and processes. In order to interpret the maltitude of data in terms of biological functionality, i.e. networks and processes (systems), mathematical models are required. Such models should be realistic descriptions or "replicas" of the biological system. The models should be able to elucidate system properties and assist phrasing of hypotheses for experimental studies and predict the outcome of (even subtle) perturbations, such as mutations or drug effects. Therefore, mathematical models are tools for data interpretation and experimental planning as well as being a vehicle for exploring insights on fundamental principles behind biological systems.

In order for the models to describe *dynamics* of cellular processes realistically, such as flow of metabolites or information through pathways, quantitative data, time courses and spatial information is required. Assessing the information available on even the best studied cellular system, *S. cerevisiae*, most data are of the "yes/no" character, i.e. qualitative. Those have been useful to define pathways and describe functions but they are largely inadequate to describe processes. To obtain quantitative data and to capture subtle changes is trivial in some instances (such as relative levels of certain metabolites, proteins, mRNA) but in many cases a technological challenge (quantification of absolute levels of biomolecules, for instance). However, *quantitative* differences, i.e. if a protein kinase is 30% or 60% "on", may make a major difference in cell behavior and drug effects. Ultimately, we would like to be able to measure (or predict) AIMS-4D (Amount, Interactions, Modifications, Spatial movements at each XYZ+T coordinate) for every component. For these reasons we need to advance technology to assess quantitative data.



The use of mathematical models is part of an important transition that is already occurring in biological research: moving from description to understanding. Much of genetic and molecular data *describe* a system but do not *explain* why, for instance, a specific feedback loop is implemented or why the system is robust against a certain mutation. Modelling helps providing a logical *explanation* and in many instances is the only way to discover such regulatory features in the first place.

Understanding biological systems is crucial for understanding diseases. Hence, one of the major promises of Systems Biology is to help understanding how diseases alter biological processes, identify the appropriate drug targets, and circumvent feedback control and robustness in drug development and to design the appropriate treatment for the individual patient.

Systems Biology operates as collaboration between experimentalists from different disciplines such as molecular biology, chemistry and physics on the one hand and bioinformaticians, computer scientists, mathematicians and engineers on the other to generate and improve models iteratively, and to explore fundamental principles. To achieve such collaboration, interdisciplinary projects, networks and even institutes are being formed all over the world. The yeast community has the expertise, excellence and the capacity to form an interdisciplinary network of Systems Biology. The Network is needed to instrumentalize such an effort and provide the platform to integrate data acquisition, data generation, modelling and recursive model optimization. [HOHMANN ET AL. 04]

3.3 Modeling And Simulation

In the physical sciences, besides theoretical and experimental studies, modelling and simulation has emerged as the third indispensable approach because not all hypotheses are amenable for confirmation or rejection by experimental observations. In biology, researchers are facing the same or maybe even worse situation. On one hand experimental study is unable to produce enough data for theoretical interpretation; on the other hand, due to data insufficiency, theoretical research cannot provide substantial guidance and insights for experimentation. Therefore, computational modelling takes a more important role in biological research by integrating experimental data, facilitating theoretical hypotheses, and addressing what if questions.

Biological modelling is both old and new. Having its genesis in the physical sciences and engineering, it has a history of several decades. However, due to the distinctive differences between biological systems and physical systems, biological modelling comes with additional challenges and calls for new strategies and tools. To model biological systems at various levels i.e., molecular, cell, tissue and organ, different strategies and techniques are needed. In this paper, we briefly review the evolution of challenges, strategies, and methods of biological modelling, with a focus on molecular networks because, with the increasingly large sequence repositories, physiomics, rather than genomics or proteomics, will emerge as the new focus of biomedical research. Revealing how individual genes and proteins are organized into networks and how molecular networks temporally and spatially evolve in cells to perform different functions will raise new challenges in systems biology.

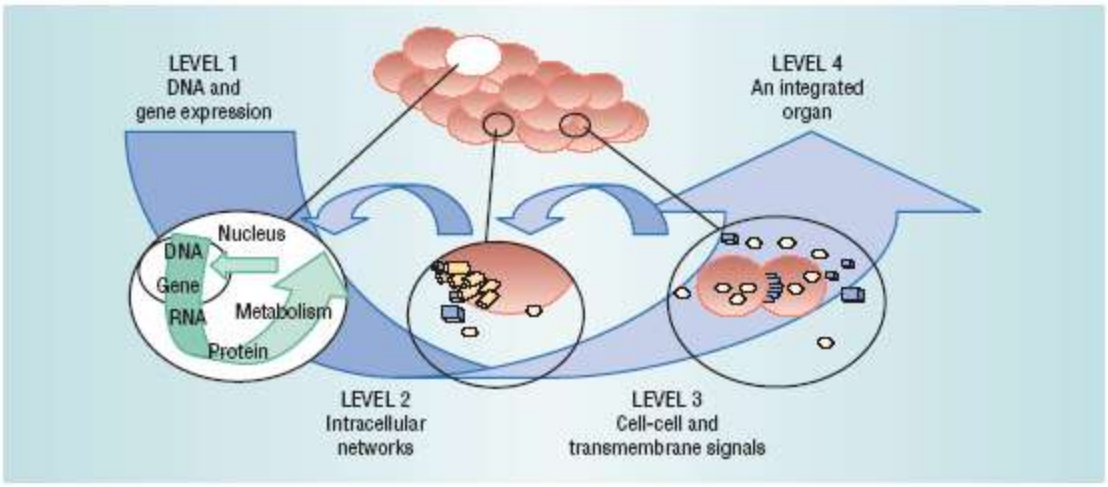
Modelling and simulation appeared on the scientific horizon much before the emergence of molecular and cellular biology. Early on the objective of modelling was to explore the features of complex biological systems that we treates as black boxes e.g., heart, brain, and circulation system, a concept borrowed from physical sciences and engineering. In such a scenario, the main challenge was to understand and predict the behavior of a system without knowing the microscopic details. Typical questions included how pathological ECG waveforms as signatures of the WPW syndrome (Wolf-Parkinson-White syndrome) were formed. The strategy was to reproduce observed phenomena at high level with simplified description of internal structures. Though inferring the microscopic details was necessarily the major goal, one needed to model the behavior of the system as a whole and utilize this understanding in clinical practice. A shining example include the inverse modelling of cardioelectrical and cerebroelectrical activity, whose simulation results were used to improve diagnosis of heart and brain diseases. [PAWAN ET AL. 04]

To understand the role of modelling in biological systems, it is useful to give an example on cell-cell interactions. Suppose we had a catalog of all the gene sequences, how they translate to make proteins, and which proteins interact with each other. Further, assume we know the way in which the protein backbones fold—whether into sheets, helices, or other shapes with differing properties. For several reasons, we would not be able to put them into a functionally meaningful framework simply from the data.

First, all proteins undergo post-translational modification that adds side chains like sugars to make, for example, *glycoproteins*—important constituents of cell membranes. These additions influence the shape and properties of proteins and hence their function and behavior. Further, just

because two proteins can interact in principle does not mean that they do so in real cells. Also, metabolic processes synthesize many small, functionally important molecules. For example, many neurotransmitters are made by cells, not translated from RNAs. Biological systems are so enormously complicated that, however much we learn about them, it will be impossible to create a full simulation based on complete understanding.

Thus, a bottom-up, data-driven strategy will not work. We cannot build an understanding of biological systems from an understanding of the components alone. We must seek other approaches. Modeling lies at the heart of systems biology. We can use experimental information to build models at different biological scales, integrating them to create an orchestrated assemblage ranging from gross models of physiological function through detailed models that build directly on molecular data. As the scheme below shows, in principle these models should span from DNA and gene expression to intracellular networks, to cell-to-cell and transmembrane signals, and through to the organ level. Tenuously, we might eventually construct such models at the organism level.



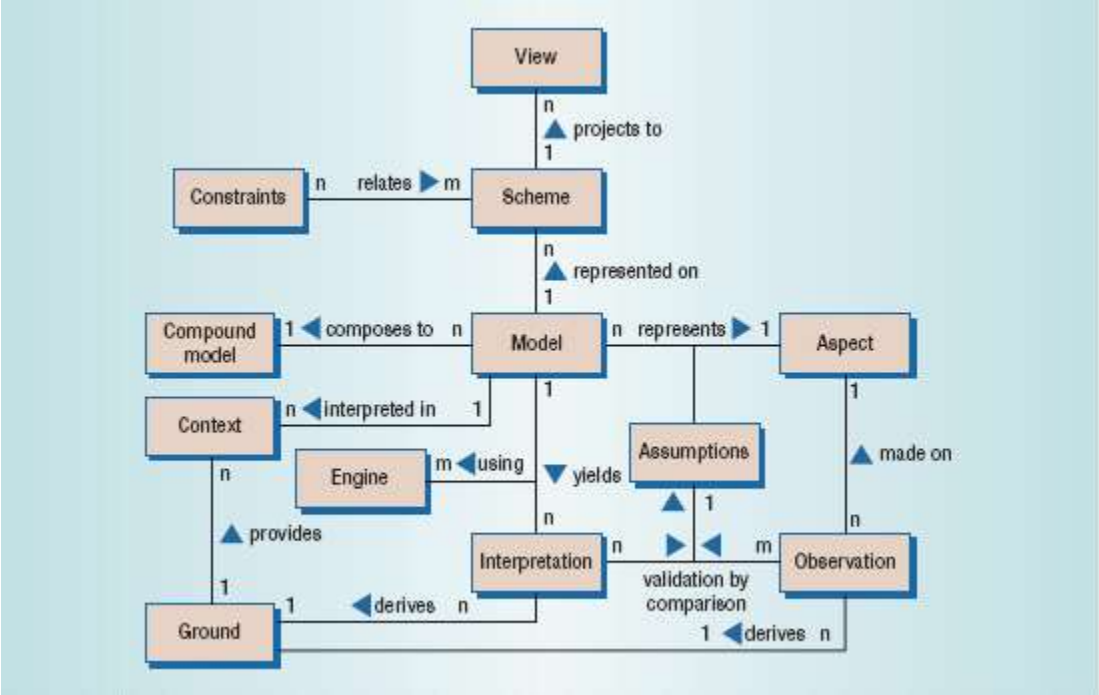
Building models in systems biology. The models should span from DNA and gene expression to Intracellular networks to cell to cell and transmembrane signals and through to the organ level.

Scientists thus introduced two key concepts for systems biology, methodologies forced by the peculiar complexity of biological systems. First, they acknowledge the importance of simplification because biological complexity requires them to model, not simulate. Second, they acknowledge the importance of both modularity and the integration of modules. Biological complexity requires them to break their systems into manageable components, but it also requires them to reassemble them because

behaviours can emerge that they cannot understand from the components alone.

The resulting models can provide coarse-grained prediction, be used as a scaffold for our emerging understanding of the data, identify gaps in our biological knowledge, and, if the models are good, predict new behaviours that we can explore experimentally. Iteration between model and experiment provides the key to ensuring that models are realistic. Given that researchers may need a different technique to study each component, it is difficult if not impossible to undertake physiological studies of whole systems in which the individual components are monitored simultaneously.

To map out the systems biology space more systematically, and to identify the computational challenges more precisely, we used the high-level information model shown below, as a general representable model. The metamodel is presented using a stripped-down entity relationship modelling convention.



Systems biology metamodel, presented using entity-relationship modelling. This model identifies key concepts in systems biology and their relationships.

Our information model has three overlapping regions, each representing a key concern in systems biology and consisting of several components:

- *construction*—the model, compound model, scheme, constraints, and view components;

- *analysis*—the model, context, engine, interpretation, and ground components; and
- *validation*—the model, aspect, observation, assumptions, and interpretation components.

Models represent ***aspects,*** a term that denotes a coherent set of properties or phenomena of biological interest. The aspect anchors the model in the real world. We establish a correspondence through an *ontology*, an explicit formal specification of how to represent the objects, concepts, and other entities assumed to exist in the biological domain being studied and the relationships that hold among them. The model and appropriate elements must then be linked to elements in the ontology.

Assumptions condition or determine the relationship between models and the aspects they represent. Assumptions underpin model construction, constitute the rationale for the model, and must be precisely documented and connected to the model for it to have meaning beyond the immediate use to which it has been put.

Experimental biologists make ***observations*** about phenomena of biological interest. Classically, these observations are used to validate interpretations derived from models. Commonly, however, models yield interpretations that prompt further observations or, when compared with observations, question the validity of the assumptions. Researchers document the observations in the scientific literature and in data resources associated with the experiments.

Models, once instantiated, yield ***interpretations*** through analysis. This can be a dynamic simulation process or a static mathematical reasoning process. The engine that both encompasses and executes a model determines the analytic process. Researchers can analyze the same model in many different ways using different procedures. The engine thus conditions an interpretation. We must precisely specify the engine to anchor the interpretation. In short, defining the model is insufficient—we must define how we use the model. Analysis can require significant computational resources.

Context is the data required to produce a model instance—it is the input to the model. Researchers could derive a context from observation, as in the straightforward case where experimental results provide a ***ground*** for data supplied to a model. In an alternative and somewhat more complex case, one model yields interpretations that constitute the context for another model. From an informational standpoint, we need to track the contexts supplied to the model and associate them with the interpretations to which they correspond. To maintain validation integrity, we must also track the context elements through their grounds.

Models are constructed in different languages, or ***representation schemes***, each appropriate to the expression of and reasoning about different sets of properties. No universal language for systems biology can capture the many different phenomena we seek to explore.

We present these schemes through ***views*** defined as projections on the underlying scheme. Modelling schemes relate to each other through ***constraints*** that define what it means for models in these schemes to be consistent with each other. Most schemes for modelling in the large provide a compositional mechanism that researchers can use to compose models and construct larger-scale ***compound models***.

Generally, three basic points should always be determined:

- defining and managing the views, languages, and constraints;
- providing the means for checking the constraints and devising modelling schemes with sound compositional mechanisms; and
- managing models that may not be consistent with each other, either across schemes or across scales.

This complex picture excludes two key dimensions, however. Models may be produced in different versions over time and by different teams. Disagreements can arise and observations can be contested. Different researchers may generate models in different versions and configurations.
These unpredictable factors mean that systems biology is unlikely to produce a set of canonical models. Rather, a complex ecology of models embedded within a framework that enables debate and collaboration among contributors will arise. Ultimately, our objective might include individualized models that account for variations in physiology, rather than generic models of biological phenomena. [FINKELSTEIN ET AL. 04]

3.4 Modeling Techniques

Biologists and mathematicians together formulated realistic mathematical models of metabolic and regulatory networks including intrinsic spatial non-homogeneity. Depending on the cellular phenomenon considered, models and methods of appropriate temporal and spatial scales were developed and then applied as models in the form of ordinary differential equations and methods for system reduction; multi-adaptive computational methods for partial differential equations (PDEs) for moderate spatial and temporal variability within a cell or an organelle; particle models describing the interaction of individual molecules and computational methods for the

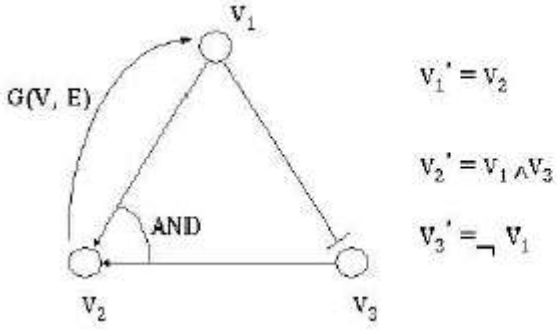
evaluation of the dynamic behaviour; and methods for integration of these different approaches into a single simulation.

The first issue involved in molecular level modelling is to effectively and faithfully describe molecular activity or to formalize a model. There have been many different methods and tools; each one developed to meet different requirements. Here follows a presentation of the strengths and the limitations of some popular techniques:

- ***Boolean Network****:* As a highly abstracted method for molecular network description, Boolean network focuses on revealing the overall, global property of large networks, especially gene regulatory networks. A Boolean network has the following features:

1) State of a system at time t+1 is determined by Boolean rules based on its current state and input.
2) Systems undergo deterministic state transition path and produce predictable behaviour.
3) Small, local perturbations produce small, local effect only.

Apparently, simplicity comes with both the strength and weakness. Deterministic dynamic properties of large systems, such as attractors and stable attractor basin, can be easily expressed with Boolean network. Even in a small network as shown below, there are a point attractor (attractor consisting of only one fix state) 1 → 2 → 3 → 3 and a dynamic attractor (attractor oscillating among several states) 1 → 2 → 3.



INPUT			OUTPUT		
v_1	v_2	v_3	v_1'	v_2'	v_3'
0	0	0	0	0	1
0	0	1	0	0	1
0	1	0	1	0	1
0	1	1	1	0	1
1	0	0	0	0	0
1	0	1	0	1	0
1	1	0	1	0	0
1	1	1	1	1	0

Example - Small Boolean Network

Despite the feasibility of using attractors to describe cell fate determination in embryonic development, in reality, both simplicity and determinism are absent. Due to the combinatorial control of transcription and the existence of enhancers and silencers the gene expression is complex, timely and precise. Thus, instead of ON/OFF states, genes often differentially express themselves during development, forming protein

gradients in tissues that guide cell differentiation. In many cases, especially single cell environment, intrinsic and extrinsic stochasticity exists. When proteins are included, Boolean network is apparently more unsuitable. So far we have not seen any practical model built with this method, reflecting another shortcoming. To overcome the first drawback, a generalized form, multi-value logical network, maybe a better choice.

- ***Rule-based Systems****:* Rule-based systems have been well studied and widely applied in computer science. Some important issues, like consistency and completeness of the rule system have been studied in depth. As early as 1990s, it was used in biological modelling. A rule-based system consists of a set of objects, a set of facts on the objects declaring their properties, and a set of rules, conditions and actions on them, e.g.:

IF (AND (condition1) (condition2))
    THEN (action1)
ELSE (action2)

An obvious advantage of rule-based system is its flexibility: nearly everything can be described with rules. In artificial intelligence applications, rule-based systems are often equipped with an inference engine to deduce conclusions automatically based on different triggering conditions. To avoid erroneous and conflicting conclusions, a precondition for such applying rules is that they should be consistent and complete. However, in very large systems, to prove and to maintain consistency and completeness is a very difficult task, thus leaving many open questions. On the other hand, when a system is complex, especially in terms of hierarchy, heterogeneity and evolution, rule description may be very inefficient. In our opinion in many situations rule-based description can be used as a strategy embedded in or combined with other methods, instead of as an independent and exclusive method.

- ***Ordinary Differential Equations (ODE)****:* In contrast to the Boolean and rule-based methods, ODE description has been practically used in many quantitative models. The general form of an ODE model can be written as

$$\frac{dx}{dt} = f_i(x)$$

where i = 1,2,…,N
xi,1 < i < n are states of molecular species.

In a molecular network model, an ODE equation is built for each molecule x quantitatively describing its relationship with all relevant

molecules and solving all equations simultaneously. There have been several platforms for ODE based modelling. Among them are Gepasi E-CELL, Virtual Cell, which share a number of features in common e.g., for chemical reactions simulation. Tools of mathematical analysis like metabolic control analysis and linear stability analysis of steady state, and parameter fitness have also been implemented. However, though metabolic reactions can be simulated by these tools, signalling activities may not be well supported.

Furthermore, signaling networks are non static and undergo evolution. Thus, modelling of the context dependent cellular processes merits a different approach. A typical example is Presenilin, a protein responsible for cleaving Notch/Delta complex, can selectively cleave a large group of membrane proteins in different contexts. Thus, to describe its behaviour with ODEs is infeasible, because:

- the biochemical equation would be very complex, and
- with the addition of a new gene or protein into the model many equations must be rewritten, an arduous work that greatly slows down the modeling process itself.

Another example of genes with complex function is the Notch gene, which takes part in intercellular communication processes. The semantics or function of its interaction with other proteins depends on its partners and the timing of interaction. In addition, in any practical model, to get complete quantitative data on gene and protein activity, such as the rate of transcription, translation, and degradation of proteins, is extremely difficult. Thus, only small or medium sized models have been reported. [PAWAN ET AL. 04]

3.4.1 Example: Modeling of Ras/Raf-1/MEK/ERK Signal Transduction Pathway[WOLKENHAUER ET AL. 03]

The Ras/Raf-1/MEK/ERK module in Figure 4 is a ubiquitously expressed signaling pathway that conveys mitogenic and differentiation signals from the cell membrane to the nucleus [YEUNG ET AL. 00],[CHO ET AL. 03]. This kinase cascade appears to be spatially organized in a signaling complex nucleated by Ras proteins. The small G protein Ras is activated by many growth factor receptors and binds to the Raf-1 kinase with high affinity

when activated. This induces the recruitment of Raf-1 from the cytosol to the cell membrane. Activated Raf-1 then phosphorylates and activates MAPK/ERK kinase (MEK), a kinase that in turn phosphorylates and activates extracellular signal regulated kinase (ERK), the prototypic mitogen-activated protein kinase (MAPK).

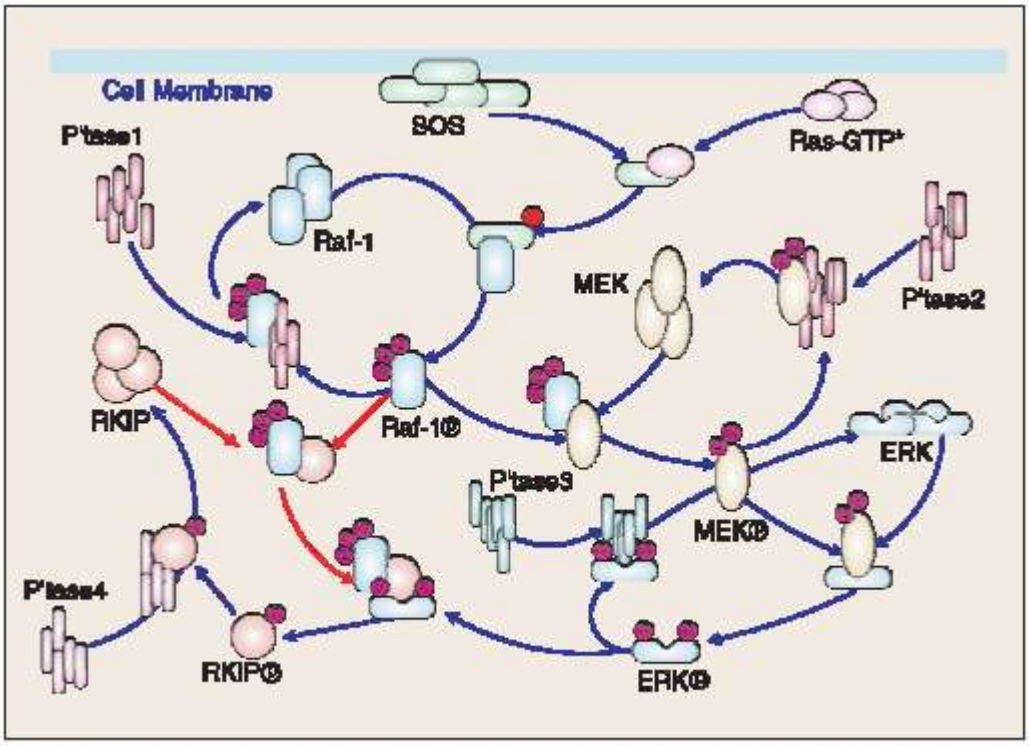


Figure 4. *Biologist's drawing for the Ras/Raf-1/MEK/ERK signal transduction pathway.*

Activated ERKs can translocate to the nucleus and regulate gene expression by the phosphorylation of transcription factors. This kinase cascade controls the proliferation and differentiation of different cell types. The specific biological effects are crucially dependent on the amplitude *and* history of ERK activity. The adjustment of these parameters involves the regulation of protein interactions within this pathway and motivates a systems biological study. Figures 5 and 6 describe the "circuit diagrams" of the biokinetic reactions for which a mathematical model is used to simulate the influence of ligand variations on the pathway

$$S + E \underset{k_2}{\overset{k_1}{\rightleftarrows}} SE \xrightarrow{k_3} E + P.$$

Signal transduction pathways can be represented as sequences of enzyme kinetics reactions which turn a substrate S into a product P via an

intermediate complex *SE* and regulated by an enzyme *E*. The rate by which the enzyme-substrate complex *SE* is formed is denoted by $k1$. The complex *SE* holds two possible outcomes in the next step. It can be dissociated into *E* and *S* with a rate constant $k2$ or it can further proceed to form a product *P* with a rate constant $k3$. It is required to express the relations between the rate of catalysis and the change of concentration for the substrate, the enzyme, the complex, and the product. Based on this reaction kinetics [ROBERT & TOM 01], we first consider a basic modeling block of signal transduction pathways.

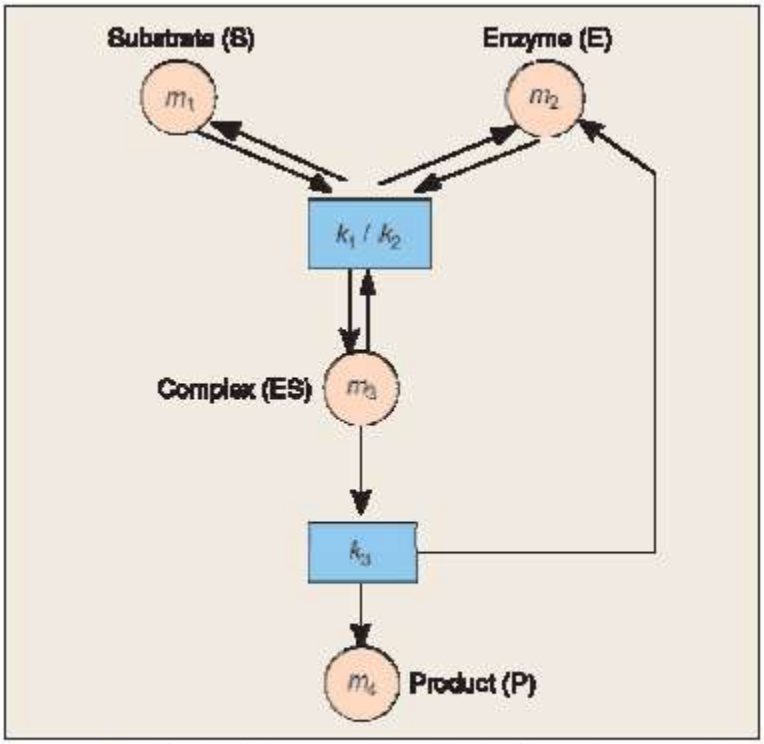


Figure 5. *Basic pathway modeling block. The pathway model is constructed from basic reaction modules like this enzyme kinetic reaction for which a set of four differential equations is required.*

This basic modeling block is illustrated in Figure 5 and can be described by the following set of nonlinear ordinary differential equations:

$$\frac{dm_1(t)}{dt} = -k_1 m_1(t) m_2(t) + k_2 m_3(t)$$
$$\frac{dm_2(t)}{dt} = -k_1 m_1(t) m_2(t) + k_2 m_3(t) + k_3 m_3(t)$$
$$\frac{dm_3(t)}{dt} = k_1 m_1(t) m_2(t) - k_2 m_3(t) - k_3 m_3(t)$$
$$\frac{dm_4(t)}{dt} = k_3 m_3(t).$$

From these we have

$$m_2(t) + m_3(t) = C_1,$$
$$m_1(t) + m_3(t) + m_4(t) = C_2.$$

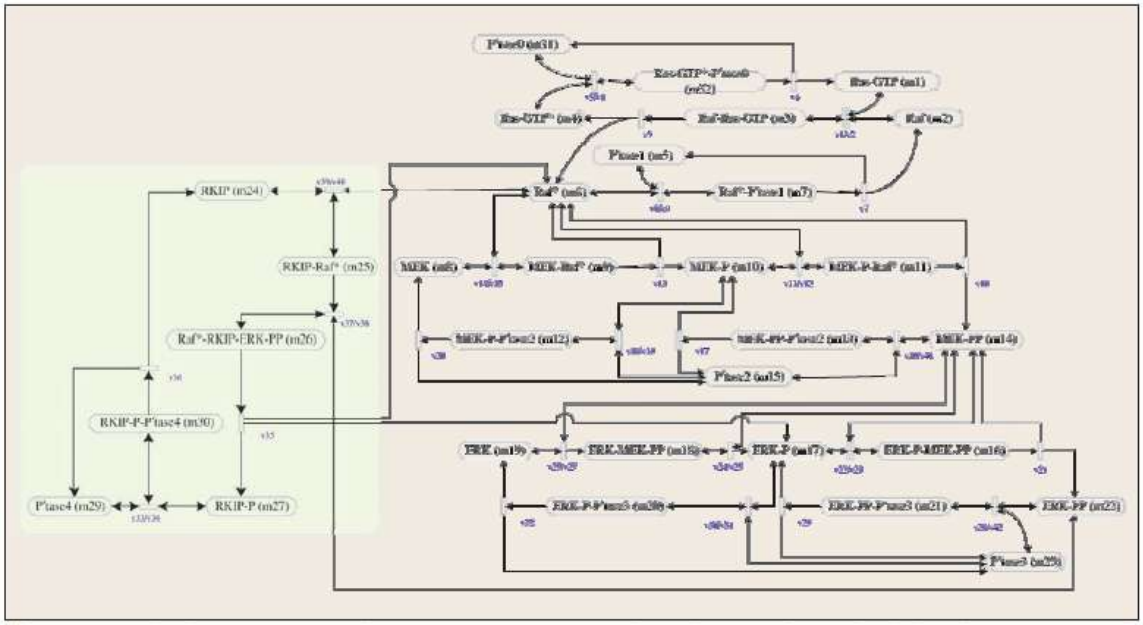


Figure 6. *Graphical representation of the Ras/Raf-1/MEK/ERK signal transduction pathway (the shadowed part represents the suppression by RKIP): a circle ○ represents a state for the concentration of a protein and a bar ▯ a kinetic parameter of the reaction to be estimated. The directed arc (arrows) connecting a circle and a bar represents a direction of a signal flow. The bidirectional thick arrows represent an association and a dissociation rate at the same time. The thin unidirectional arrows represent a production rate of products.*

Hence we can describe the basic reaction module by two nonlinear equations subject to two algebraic conditions. In general, for a given signal transduction system, the whole pathway can be modeled by a set of

nonlinear differential equations and a set of algebraic conditions in the following form:

$$\frac{d\mathbf{m}(t)}{dt} = \mathbf{f}(\mathbf{m}(t), \mathbf{k}(t)),$$

$$\sum_{i \in [1,p]} m_i(t) = C_j,$$

where $\mathbf{m}(t) = [\mathbf{m}_1(t), \mathbf{m}_2(t), \ldots, \mathbf{m}_p(t)]$, $\mathbf{k}(t) = [\mathbf{k}_1(t), \mathbf{k}_2(t), \ldots, \mathbf{k}_q(t)]$, p is the number of proteins involved in the pathway, q is the required number of parameters, and $j \in \{1, \ldots, J\}$ with the number of algebraic conditions $J<p$.

Parameter estimation is widely regarded as a major problem in dynamic pathway modeling [BOCK 81], [HEGGER ET AL. 98]. A simple method first discretizes the nonlinear differential equations into algebraic

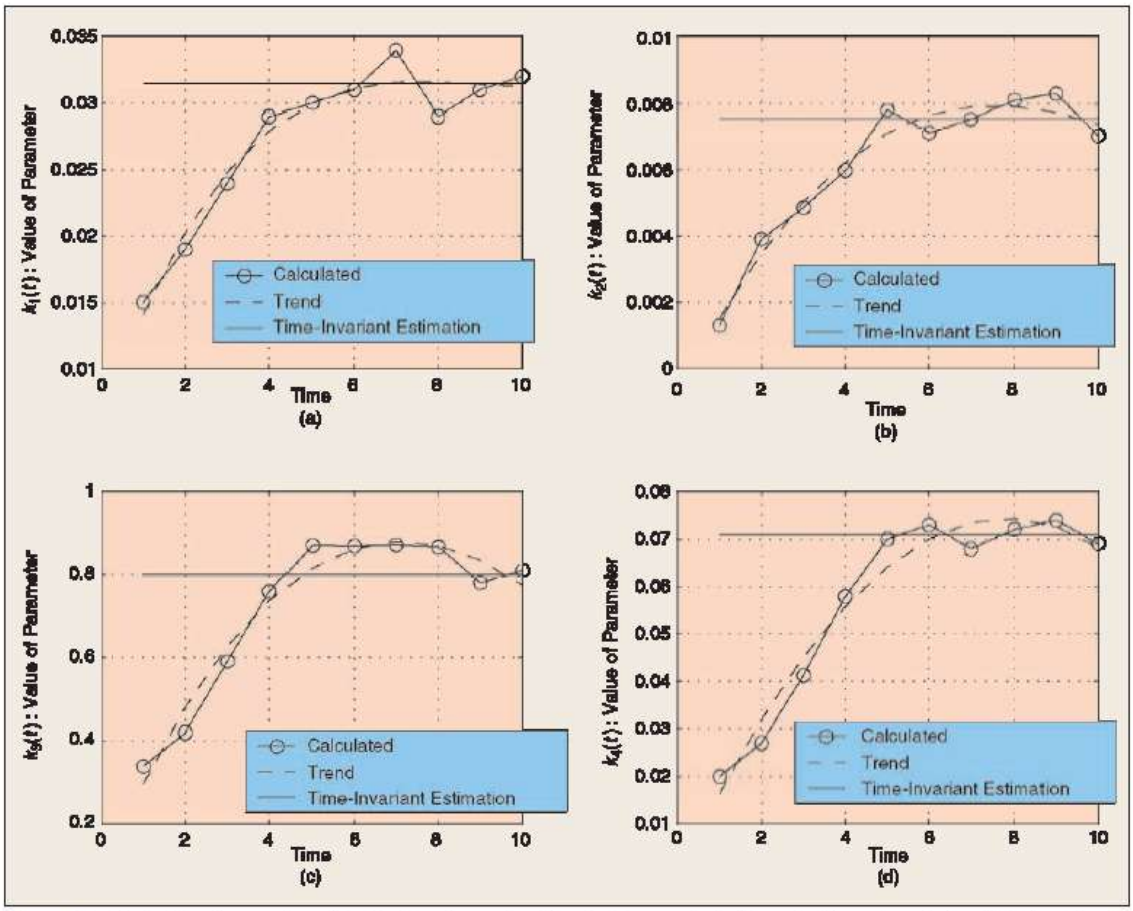


Figure 7. *Illustration of parameter estimation from time series data: Each parameter is determined from the value to which the estimates converge (shown by the horizontal line). (Note that any experimental noise can be further eliminated by regression techniques if multiple experimental replicates at each time point are available.)*

difference equations that are linear with respect to the parameters and then solve the transformed linear algebraic difference equations to obtain the parameter values at each sampling time point. We can then estimate the required parameter values by employing curve fitting, calculation of steady state values, and regression techniques. For this purpose of parameter estimation, the previous equations are transformed into

$$\mathbf{k}(t) = \mathbf{g}\left(\mathbf{m}(t), \frac{d\mathbf{m}(t)}{dt}\right),$$

and this can be further transformed into a set of algebraic difference equations by approximating the differential operator vector **g** via a difference operator vector **h** as

$$\mathbf{k}(t) \equiv \mathbf{h}(\mathbf{m}(t), \mathbf{m}(t-1), \ldots, \mathbf{m}(t-r))$$

where r depends on the order of approximation. Without loss of generality, $\mathbf{k}(t)$ can be approximated by $\mathbf{k}$ since most of the signal transduction systems can be regarded as slowly time varying systems compared with the measurement windows in time scale. Hence we have

$$\mathbf{k} \equiv \mathbf{h}(\mathbf{m}(t), \mathbf{m}(t-1), \ldots, \mathbf{m}(t-r)),$$

which implies the parameter estimates based on time course measurements.

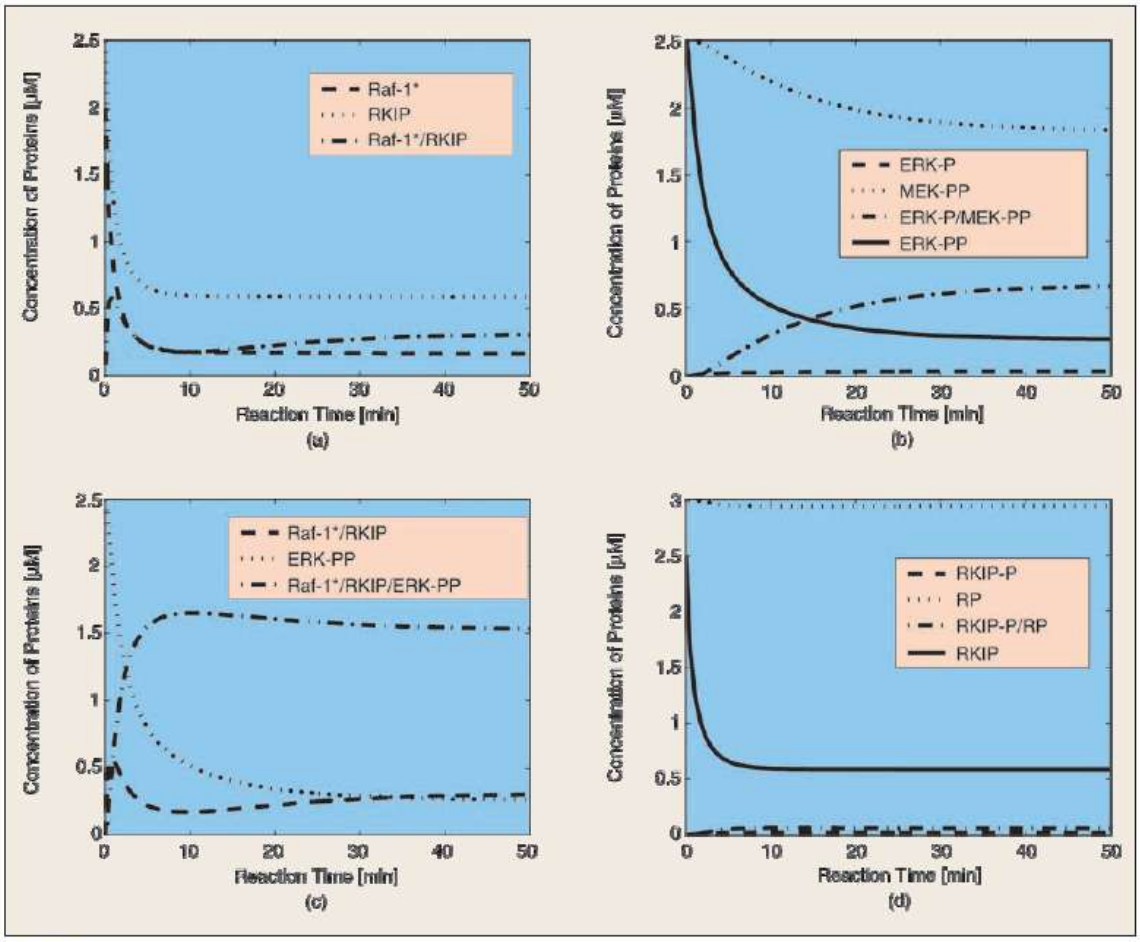


Figure 8. *The simulation results for fixed initial conditions: (a) shows the binding of RKIP to Raf-1*, (b) shows the binding of MEK-PP to ERK-P, (c) shows the binding of ERK-PP to Raf-1*/RKIP, and (d) shows the binding of RP to RKIP-P.*

The entire model, as shown in Figure 6, is constructed in this way, leading to what usually becomes a relatively large set of differential equations for which parameter values have to be identified. As illustrated in Figure 7, in the estimation of parameters from western blot data, the parameter estimates usually appear as a time dependent profile since the time course data include various uncertainties. However, since the signal transduction system itself can be considered as time invariant, the estimated parameter profile should converge to a constant value at steady state. Figure 7 illustrates this estimation procedure.

If a reasonable model is constructed, this can then be used in a variety of ways to validate and generate hypotheses, or to help experimental design [CHO ET AL. 03], [CHO ET AL. 02]. Based on the mathematical model illustrated in Figure 6 and the estimated parameter values as, for example, obtained using a discretization of the nonlinear ordinary differential equations (as illustrated in Figure 7), we can perform simulation studies to validate the signal transduction mechanism as illustrated in Figure 5 and also

to analyze the signal transduction system with respect to the sensitivity for the ligand (via simulation of variable initial conditions) as illustrated in Figure 9.

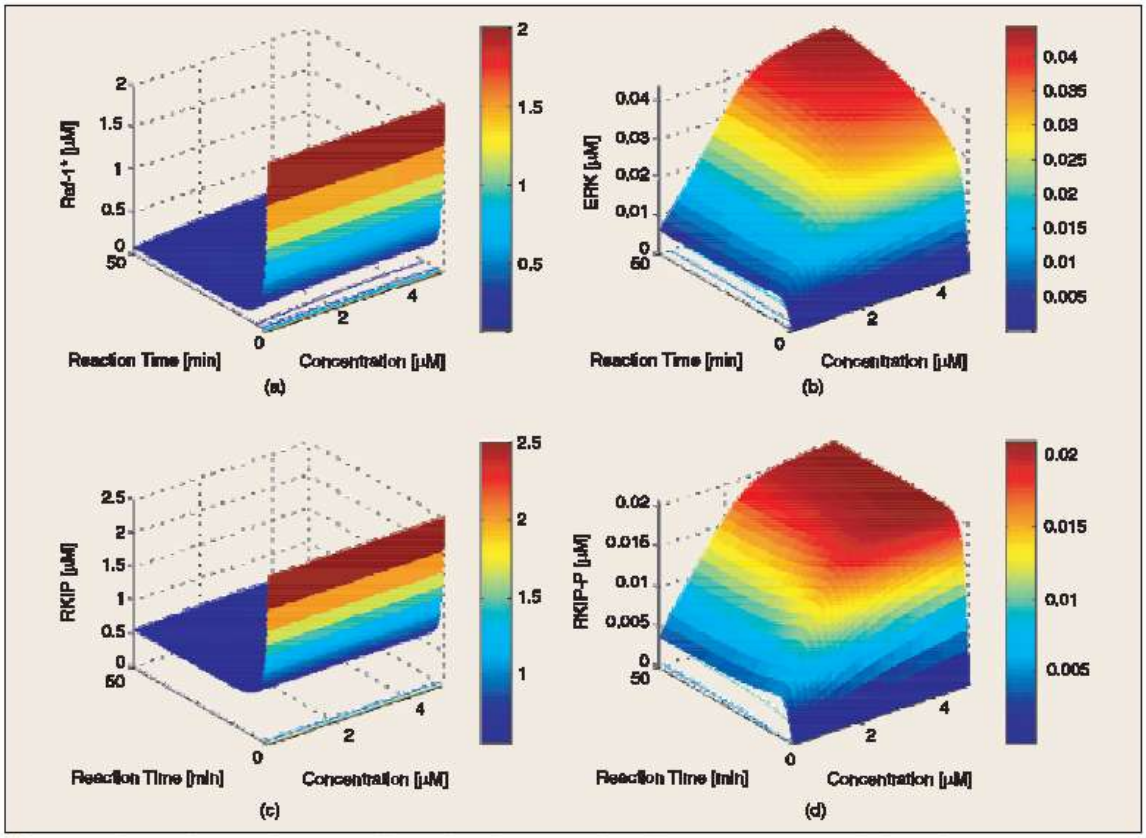


Figure 9. *The simulation results for variable initial conditions: (a) shows the variation of Raf-1*, (b) shows the variation of ERK, (c) shows the variation of RKIP, and (d) shows the variation of RKIP-P.*

4 MAPK Cascade

Mitogen-activated protein kinase (MAPK) cascades represent a biological module, or subcircuit, which is ubiquitous in eukaryotic cell signal transduction processes [HUANG & FERRELL 96, LAUFFENBURGER 01, WIDMANN ET AL. 99] and is a critical component of pathways involved in cell proliferation, differentiation, movement, and death. They are activated by diverse stimuli (cytokines,

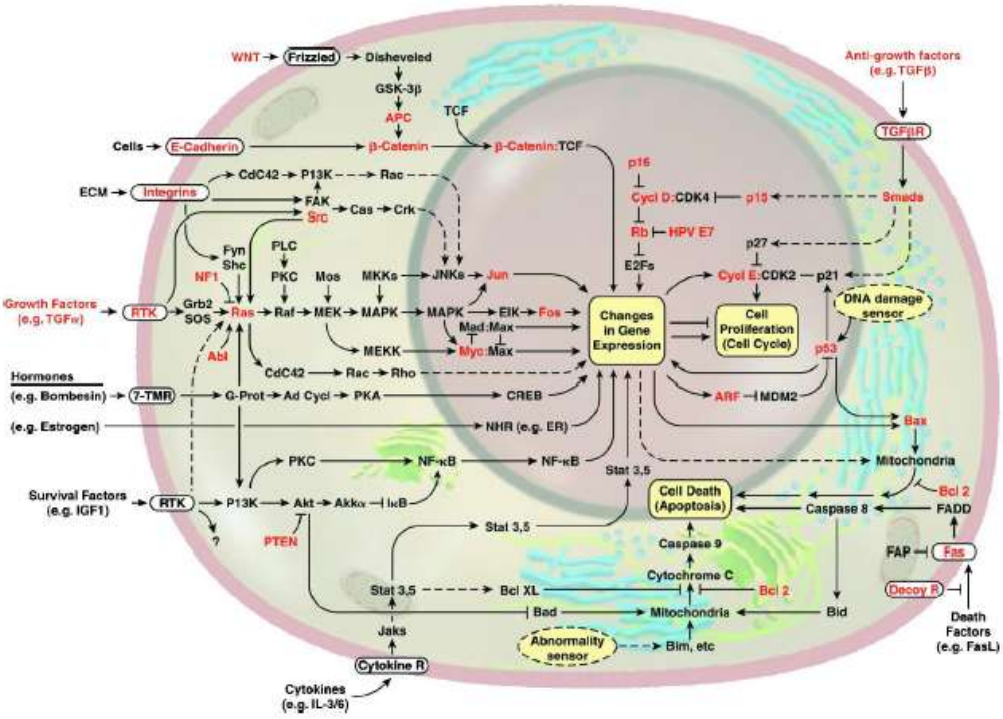


Figure 1: ***Part of the signal transduction network in human cells. The pathways shown are known to play a role in cancer development. This thesis focusses on one module involved in this network, the mitogen-activated protein kinase (MAPK) cascade. It is part of the growth factor signaling path, shown on the left side of this picture. The core of this module containers three protein kinases, RAS/RAF (MAPKKK), MEK (MAPKK) and MAPK. This module is mainly activated by the receptor tyrosine kinase (RTK) via son of sevenless (SOS) protein. Figure taken from [HANAHAN&WEINBERG 00].***

growth factors, neurotransmitters, hormones, cellular stress, cell adherence), and their outputs drive different cellular responses, including DNA transcription.

4.1 Functioning Of MAPK Pathways

The MAPK cascade motif appears in different forms in distinct organisms from yeast to humans, and even in any given cell, and different MAPK cascades involve different chemicals, but their basic architecture is conserved. A MAPK pathway is a three-component module, consisting of three kinases that establish a sequential activation pathway comprising a MAPK kinase kinase (MKKK), a MAPK kinase (MKK), and the final MAPK itself [WIDMANN ET AL. 99]. (A kinase is an enzyme that catalyzes the transfer of phosphate groups from a highenergy phosphate-containing molecule, such as ATP or ADP, to a substrate.) The diagram in Figure 2 illustrates one standard biological model for this process. The output of the system is represented by the "activated" form of MAPK, denoted as MAPK** in the

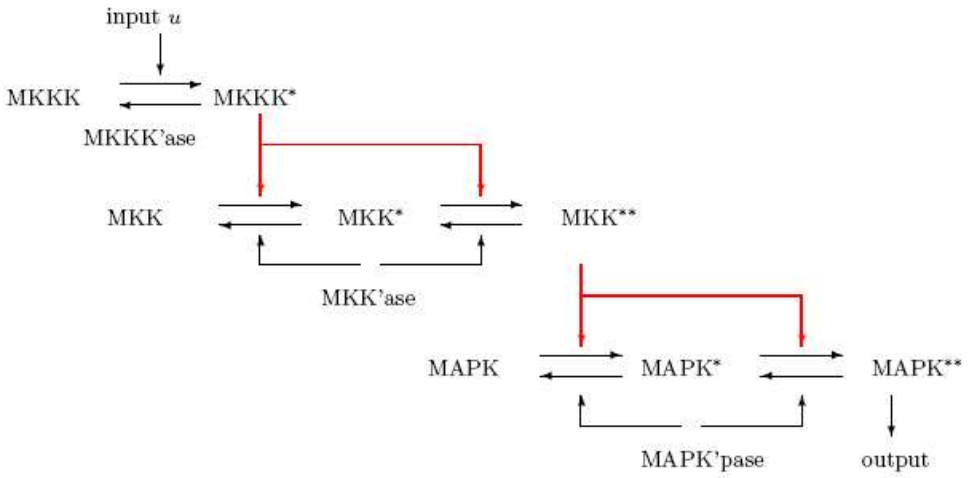


Figure 2: *Schematic of a MAPK signaling module. The output of the system is represented by the concentration of the "activated" form of MAPK, denoted as MAPK**. Red arrows indicate that this concentration is in turn driven by the time-varying concentration of MKK**, which in turn depends on the previous level of the cascade.*

diagram. The precise meaning of this "activation" need not concern us here, but we remark that MAPK must be phosphorylated on both a threonine and

tyrosine residue for its activation, a dual phosphorylation catalyzed by activated-MKK. The concentration MAPK**(t) of this activated form is controlled by the amount MKK**(t) of activated MKK present at any given time. (The intermediate MAPK*(t) corresponds to a partially-activated form of MAPK, where only one phosphorylation has taken place.) A reverse reaction of dephosphorylation takes place as well, not controlled by activated MKK, but controlled by a phosphatase (an enzyme that removes phosphate groups), whose concentration is assumed to be constant in the time scale being studied. Similarly, the concentrations of partially and completely activated MKK's are controlled by the concentration of activated MKKK. The input u to the complete subsystem represents the concentration of a different kinase (e.g. PKC) or of a small GTP-binding protein (like in the RAS pathway, see Figure 2.2), that phosphorylates, and hence activates,

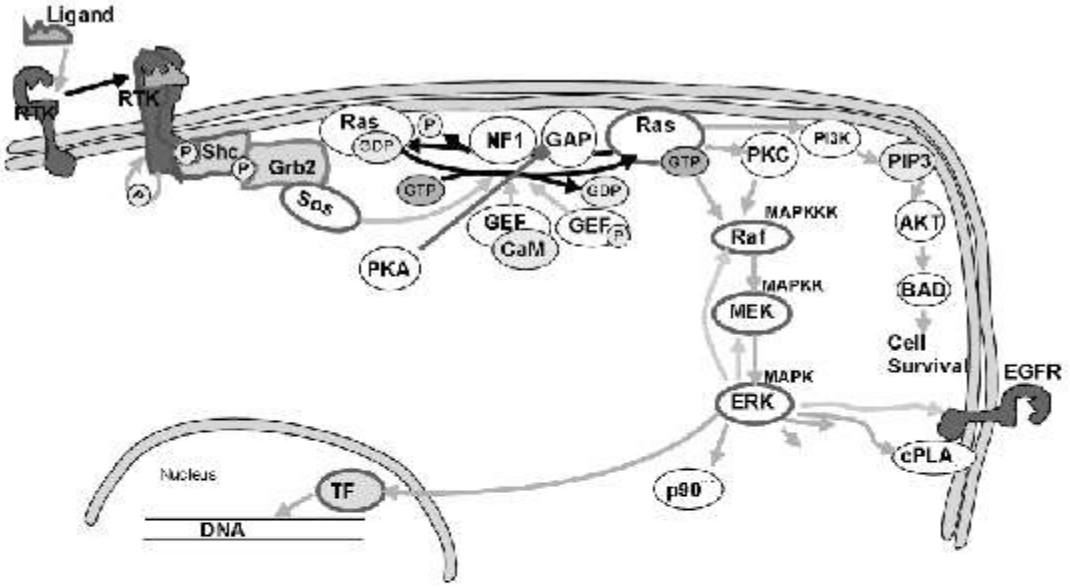


Figure 2.2: The RAS pathway: If the ligand binds to receptor tyrosine kinase (RTK), the receptor gets phosphotorylated, and binds to several other proteins (SHC, GRB2 and SOS). This complex activates RAS, RAS activates PKC [DIAZ-MECO ET AL. 94] and binds to phosphorylated RAF. PKC activates RAF by phosphotorylation [KOLCH ET AL. 93] which serves in complex with GTP-bound RAS as a MAPKKK [MARSHALL 95b] and activates MEK (MAPKK) and ERK (MAPK). ERK activates transcription factors (TF). ERK has also an influence on the activity of the receptor, and can also influence the activity of RAF and MEK. Additionally, RAS also plays an important role in the cell survival pathway. Figure taken from TRANSPATH database [WINGENDER ET AL. 00] and modified.

MKKK. This three-step cascade is present in all eukaryotes and has a wide range of functions in signal transduction, such as stressresponse, cellcycle-control, cellwallconstruction, osmosensing, growth and differentiation.

But why does the module have three steps? It is obvious that one possible function is to generate high amplification. Another role of the three steps might be to generate a switchlike response [HUANG& FERRELL 96]. This seems reasonable, since the MAPK module is used for many important decisions in cells. Moreover, Asthagiri and Laufenburger argue that the integral of activated MAPK** over time may serve as a reasonable metric for the output rather than the steady state of the system [ASTHAGIRI &LAUFFENBURGER 01]. In this case, the MAPK cascade can work as a feedback controller capable to adapt to different stimuli. Such a robust adaptation has been demonstrated in bacterial chemotaxis [YI ET AL. 00]. Furthermore, the three steps might be used to integrate many inputs, as known in osmosensors [WIDMANN ET AL. 99].

4.2 Models[BLUTHGEN 02]

Huang and Ferrell [HUANG & FERRELL 96] developed a model to describe MAPK activation in Xenopus oocytes, and focused on the role of the MAPK cascade in allornone decisions [FERRELL &XIONG 01]. Their model is shown in Figure 2.3 on the right hand side. The stimulus in this model is a hypothetical enzyme E1. This model shows a pronounced amplification from the stimulus (in the order of tenth of nM) to the response (in the order of μM). Parameters and ODEs of this model can be found in the appendix. Moreover, in a recent article Ferrell proposed, that there might be a positive feedback loop to produce ultrasensitivity, which is not included in his model. Within a large model of second messenger cascades in neurons, Bhalla and Iyengar [BHALLA & IYENGAR 99] also consider the MAPK module, which is shown on the left side of Figure 2.3. It has a negative feed back loop via doublephosphorylation of the MAPKKK (RAF). Moreover, RAF is only active in a complex with GTP bound RAS. Here PKC serves as an input. However, they focus on properties of a large network rather than on features of small modules like the MAPK cascade. A rather different model is described by Asthagiri and coworker [ASTHAGIRI ET AL. 00, ASTHAGIRI & LAUFFENBURGER 01], where they included a negative feedback loop, and showed that the MAPK cascade can show adaptative behavior. These papers do not consider the double phosphorylation, but

focus on the desensitization of the receptor. Kholodenko [KHOLODENKO 00] developed a model with a negative feedback loop, which can cause oscillations in the concentration of activated MAPK**. He used a model which was very similar to the Huang/Ferrell model [HUANG & FERRELL 96]. He included a negative feedback loop acting upstream of MAPKKK with an adiabatic approach by reducing the stimulus activity by the concentration of activated MAPK. In this model it was shown, that a combination of ultrasensitivity and negative feedback can cause oscillations.

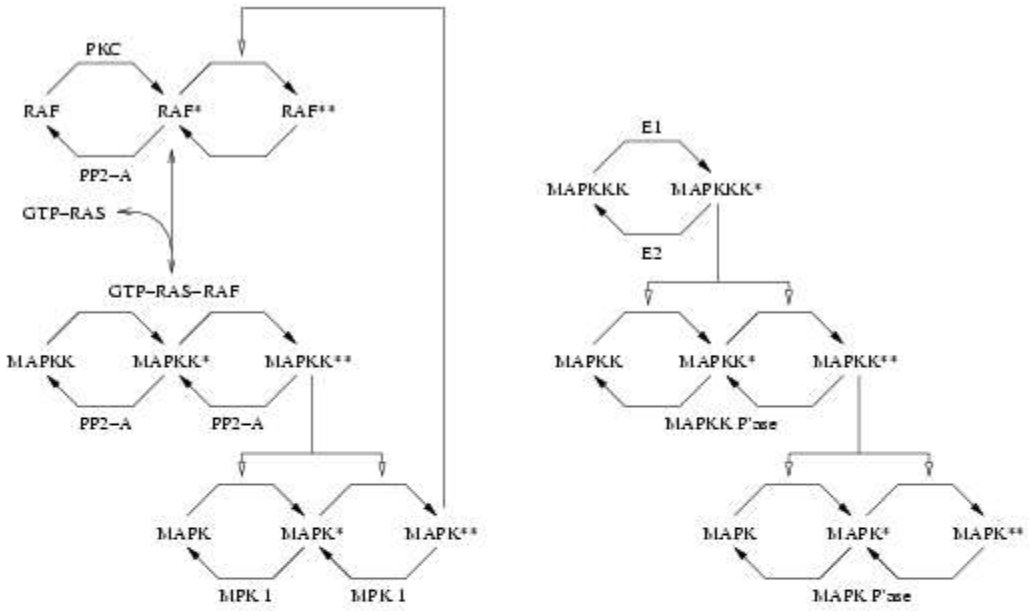


Figure 2.3: A schematic view of the MAPK-cascade models analysed in this thesis. Stars indicate phosphorylated enzymes. The left graph shows a model according to Bhalla and Iyengar [BHALLA & IYENGAR 99]. RAF, the MAPK kinase kinase, has three states. It is only active in the monophosphorilated state, and when it is bound to GTP-RAS. RAF's first phosphorylation is catalyzed by PKC, the second by MAPK**. In the model according to Huang and Ferrell [HUANG & FERRELL 96] (right figure), MAPKKK is activated by E_1, and it is active without binding to another protein. In both models, MAPKK and MAPK have three states, and only their double-phosphorylated states (MAPK** and MAPKK**) are active. The dephosphorylation in catalyzed by phosphatases.

4.3 Reaction Kinetics[BLUTHGEN 02]

In order to construct a model of the MAPKcascade, the reaction kinetics has to be discussed. I am going to introduce the mass action law as a fundamental concept of kinetic modeling and the Michaelis Menten mechanism, where the enzymatic modification of a substrate can be broken down into simple reactions, which can be described by the mass action law. Furthermore, the MichaelisMenten equation will be introduced.

4.3.1 Mass Action Law

Two main types of reactions can be distinguished:
unimolecular reactions

$$A \rightarrow P \tag{4.3.1}$$

and bimolecular reactions

$$A + B \rightarrow P. \tag{4.3.2}$$

Higher order reactions such as (A+B+C → P) are extremely rare, if they occur at all [CORNISHBOWDEN 99].

If diffusion is fast compared to time scales of the reaction, the concentration of reactants can be assumed to be homogenous. In this case, the reaction rate becomes obviously proportional to the molar concentration of the reactant in unimolecular reactions. It is proportional to the product of the concentration of reactants in bimolecular reactions, because the collision probability of both reactants should be proportional to either concentration. Moreover, if one of the reactants is fixed (for example to the cell wall), this changes the constant of proportionality (also called rate constant), but not the bilinear dependency between the rate of the reaction and the concentration [KHOLODENKO ET AL. 00].

The linear or bilinear dependency of the reaction rate on the concentration is called the mass action law. The rate constant for a unimolecular reaction has the dimension of (1/time), and for bimolecular

reaction (1/concentration) x (1/time). Concentrations are measured in molar (M):

$$\mathrm{M} = \frac{\mathrm{mol}}{\mathrm{l}}. \qquad (4.3.3)$$

Unimolecular reactions are also called firstorder reaction, while bimolecular reactions are called secondorder reaction. In some cases, the reaction rates are independent of the concentration of reactants (for example when enzymes are saturated, see GoldbeterKoshland switch [GOLDBETER & KOSHLAND 81]). These reactions are called zeroorder reactions.

4.3.2 Enzyme Kinetics

Enzymatic reactions can be deduced to basic bi- and unimolecular reactions:

$$\mathrm{E} + \mathrm{A} \leftrightarrow \mathrm{EA} \rightarrow \mathrm{E} + \mathrm{P}. \qquad (4.3.4)$$

In the first step, enzyme E binds reversibly with substrate A, and in a second step, the enzyme releases the modified substrate P. The second step is assumed to be irreversible. This is a reasonable assumption for phosphorylation, because phosphorylation consumes energy in the form of ATP and it cannot be reversed, once the ADP is released. And also dephosphorylation can be modeled by this, because the phosphorylated state of a protein is on a energetically higher level, therefore the probability of returning to the phosphorylated state is very small.

The reaction scheme (4.3.4) is called MichaelisMenten mechanism [CORNISHBOWDEN 99]. The rates for the three reactions (v1, v2, v3) are:

$$\mathrm{E} + \mathrm{A} \rightarrow \mathrm{EA} \ : \ v1 = \alpha[\mathrm{E}][\mathrm{A}] \qquad (4.3.5)$$

$$\mathrm{E} + \mathrm{A} \leftarrow \mathrm{EA} \ : \ v2 = d[\mathrm{EA}] \qquad (4.3.6)$$

$$\mathrm{EA} \rightarrow \mathrm{E} + \mathrm{P} \ : \ v3 = V[\mathrm{EA}] \qquad (4.3.7)$$

where [E], [A] and [EA] are the concentrations of enzyme E, substrate A, and enzymesubstrate complex EA. a, d and V are rate constants. Thus, there

are two unimolecular and one bimolecular reactions. The ordinary differential equations for this system are given by:

$$d[A]/dt = v2 - v1 \quad (4.3.8)$$

$$d[EA]/dt = v1 - v2 - v3 \quad (4.3.9)$$

$$d[E]/dt = v2 + v3 - v1 \quad (4.3.10)$$

$$d[P]/dt = v3 \quad (4.3.11)$$

with the initial conditions:

$$[E](t=0) = [E_{tot}] \quad (4.3.12)$$

$$[A](t=0) = [A_{tot}] \quad (4.3.13)$$

$$[P](t=0) = 0 \quad (4.3.14)$$

$$[EA](t=0) = 0 \quad (4.3.15)$$

Given the two conservation laws for the molar concentrations,

$$[E_{tot}] = [E] + [EA] \quad (4.3.16)$$

$$[A_{tot}] = [A] + [EA] + [P] \quad (4.3.17)$$

the system can be reduced to two equations:

$$d[EA]/dt = \alpha[A]([E_{tot}] - [EA]) - d[EA] - V[EA] \quad (4.3.18)$$

$$d[A]/dt = -\alpha([E_{tot}] - [EA])[A] + d[EA] \quad (4.3.19)$$

The MichaelisMenten equation can be derived in case that the amount of enzymesubstrate complex is in pseudo steadystate (d[EA]/dt = 0). Collecting terms in [EA] and rearranging leads to:

$$[\mathrm{EA}] = \frac{a[\mathrm{E_{tot}}][\mathrm{A}]}{d+V+a[\mathrm{A}]} = \frac{[\mathrm{E_{tot}}][\mathrm{A}]}{\frac{d+V}{a}+[\mathrm{A}]}. \qquad (4.3.20)$$

Conventionally, a parameter called MichaelisMenten constant k_m is introduced,

$$k_m = (d + V)/\alpha, \qquad (4.3.21)$$

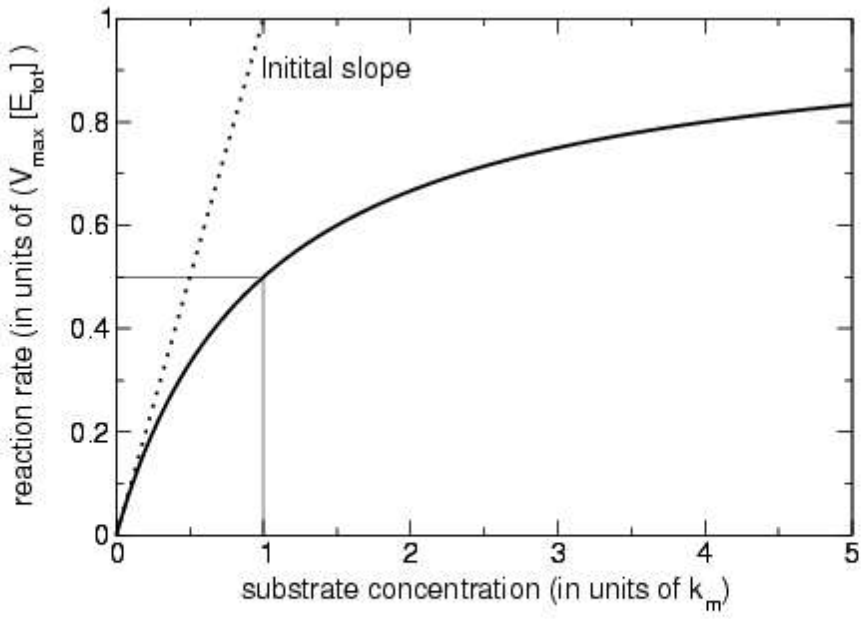


Figure 3.1: Solid line: plot of the Michaelis-Menten equation. the reaction rate is normalized by the maximum reaction rate ($V_{max} \times [\mathrm{E_{tot}}]$) and the concentration of the substrate is normalized to k_m. Dotted line: initial slope for small concentrations ($V_{max}[\mathrm{E_{tot}}][\mathrm{A}]k_m^{-1}$)

with the dimension of a concentration, and the parameter V is called V_{max}. Introducing these parameters, the steadystate of [EA] can be written as:

$$[\mathrm{EA}] = \frac{[\mathrm{E_{tot}}][\mathrm{A}]}{k_m+[\mathrm{A}]}. \qquad (4.3.22)$$

The rate of production of P is given by:

$$v_3 = V[\mathrm{EA}] = \frac{V_{max}[\mathrm{E_{tot}}][\mathrm{A}]}{k_m + [\mathrm{A}]}. \tag{4.3.23}$$

Equation (3.23) is known as MichaelisMenten equation.

The MichaelisMenten equation is linearly dependent on the concentration of the enzyme, while the rate of the reaction saturates as a function of substrate. Maximum velocity is given by $V_{max}[E_{tot}]$, and half of this is reached at $[A] = k_m$. The velocity for small concentration of A ($[A] << k_m$) is given by $v = (V_{max}/k_m)[E_{tot}][A]$ (see figure 3.1).

A timescale for approaching the pseudo steadystate can be easily derived if [A] is large compared to the enzyme concentration. In this case, the concentration of A may be treated as a constant, and equation (4.3.18) can be integrated [CORNISHBOWDEN 99]:

$$[\mathrm{EA}] = \frac{[\mathrm{E_{tot}}][\mathrm{A}]\left(1 - e^{-(a[\mathrm{A}]+d+V)t}\right)}{k_m + [\mathrm{A}]}. \tag{4.3.24}$$

Thus, the pseudo steadystate is reached in the timescale of

$$\tau_{EA} = (a[\mathrm{A}] + d + V)^{-1}. \tag{4.3.25}$$

Another more detailed analysis of the steadystate is made by Schnell and Mendoza [SCHNELL &MENDOZA 97]: Assuming that $[A] >> [E_{tot}]$, the timescale is estimated as

$$\tau_{EA} = \frac{1}{a[\mathrm{A_{tot}}] + d + V} \tag{4.3.26}$$

$$= \frac{1}{a(k_m + [\mathrm{A_{tot}}])}. \tag{4.3.27}$$

Furthermore, they also estimate the time for the change of the substrate concentration as

$$\tau_A = \frac{[A_{tot}]}{|d[A]/dt|_{max}} \tag{4.3.28}$$

$$= \frac{k_m + [A_{tot}]}{V_{max}[E_{tot}]}. \tag{4.3.29}$$

Therefore, the ratio of the two timescales is given by:

$$\frac{\tau_{EA}}{\tau_A} = \frac{V_{max}[E_{tot}]}{a(k_m + [A_{tot}])^2}. \tag{4.3.30}$$

A more general conditon for the pseudo steadystate assumption is given by Segel (see [SCHNELL & MENDOZA 97]):

$$\frac{[E_{tot}]}{k_m + [A_{tot}]} \ll 1. \tag{4.3.31}$$

Therefore, the steadystate assumption is reasonable, if the enzyme concentration is small compared to the concentration of the substrate. A problem might occur, if an enzyme acts in different reactions. In this case, the two conservation equations (4.3.16) and (4.3.17) are no longer valid. Nevertheless, if the saturation of the enzyme is small, they might be a good approximation, because the concentration of bound enzyme is small.

In this thesis, I simulated the model by Huang and Ferrell [HUANG & FERRELL 96]. The ordinary differential equations for the concentrations of the molecules in the MAPK cascade can be found in the appendix, as well as the reaction parameters taken from [HUANG & FERRELL 96], which were not changed in order my work to be comparable to published results.

4.4 Simulation Using SIMULINK

After modeling the MAPK Cascade with differential equations, we created the SIMULINK model based on these equations. Each subcircuit in

the model corresponds to one differential equation; totally we have 22 equations and subcircuits. In Figures 4.4.1 and 4.4.2 we suggestively present the implementation for the equations referred to the concentrations of the input activating enzyme E1 and the double phosphorylated MAPK**.

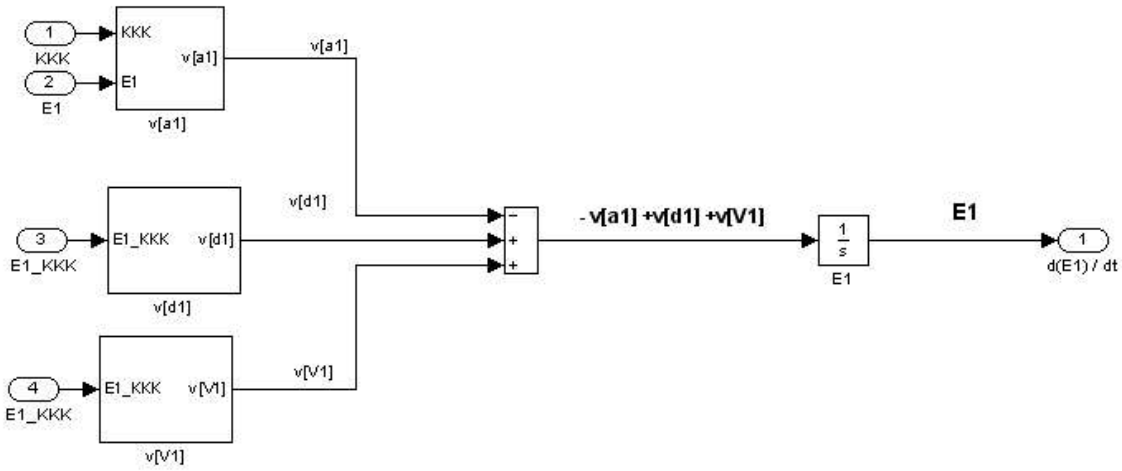


Figure 4.4.1: ***Implementation in SIMULINK for the differential equation*** **d[E1]/dt = -α1[KKK][E1] + d1[E1_KKK] + V1[E1_KKK],** ***referred to the concentration of the input activated enzyme E1.***

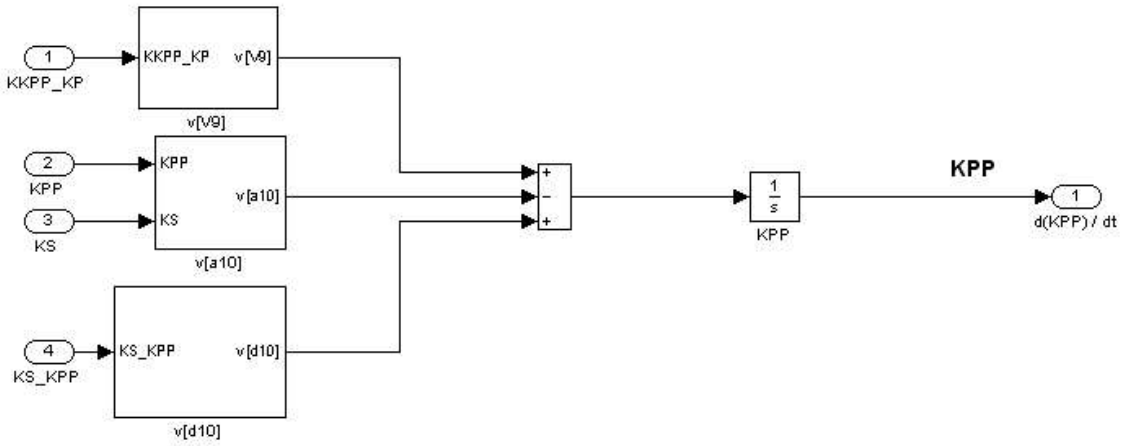


Figure 4.4.2: ***Implementation in SIMULINK for the differential equation*** **d[K**]/dt = V9[KK**_K*] – α10[K**][KS] + d10[KS_K**],** ***referred to the concentration of the double phosphorylated MAPK**.***

By simulating this SIMULINK implementation, we can find the steady-state of the three kinases (MAPK**, MAPKK** and MAPKKK*) involved in the MAPK cascade module for the model according to Huang/Ferrell [HUANG & FERRELL 96], as a function of the concentration of the activating enzyme E1. The results are shown in Figures 4.4.3, 4.4.4 and 4.4.5 respectively.

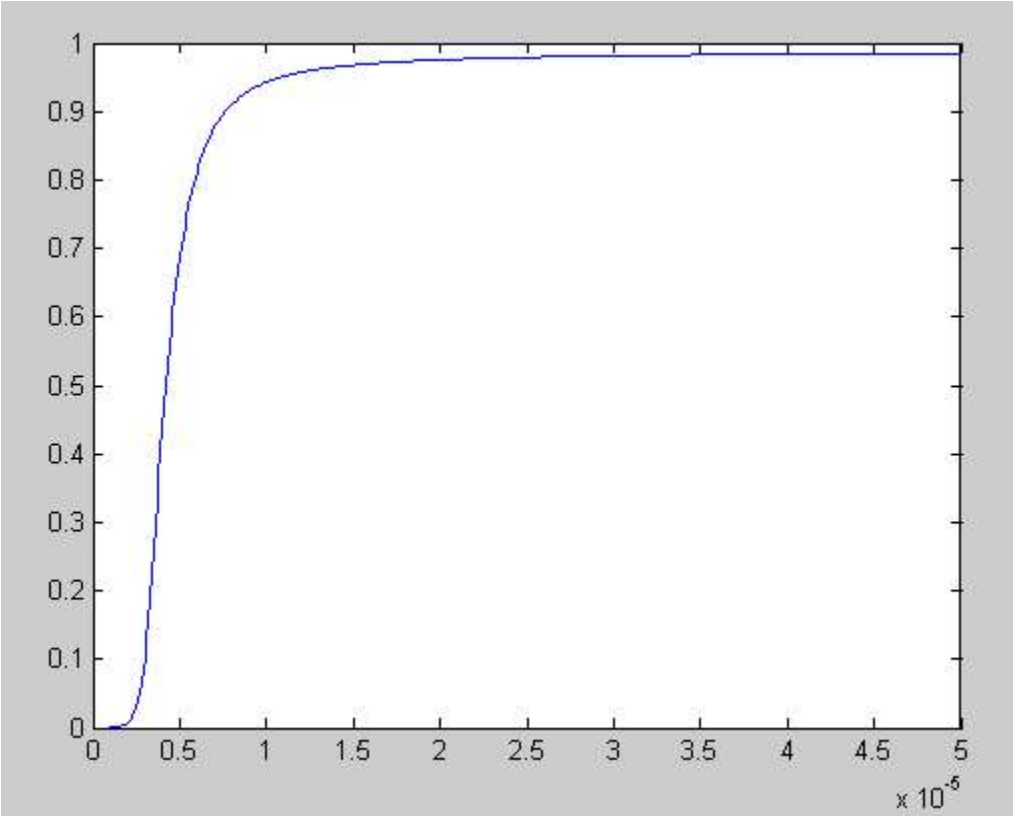


Figure 4.4.3: ***Steady-state activation of MAPK**. Concentrations in μM***

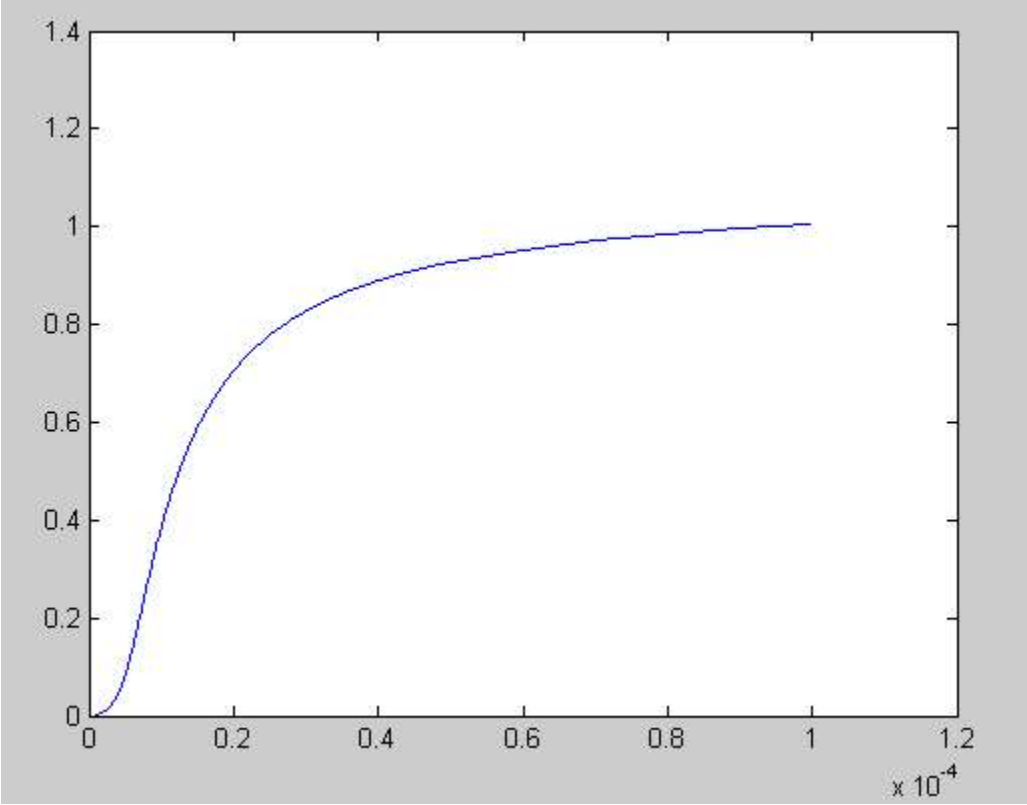


Figure 4.4.4: ***Steady-state activation of MAPKK**. Concentrations in μM***

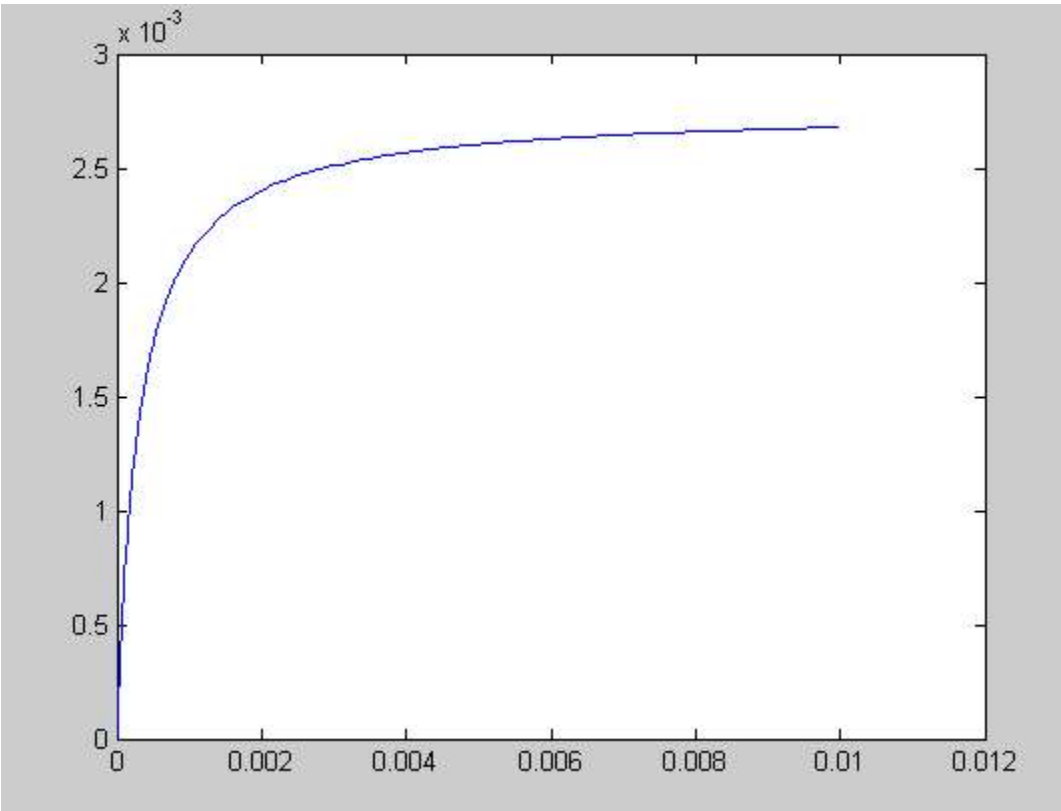


Figure 4.4.5: ***Steady-state activation of MAPKKK*. Concentrations in µM***

4.5 Steady-State Properties

The steadystate stimulusresponse curve in intracellular signaling is called ultrasensitive, if it is sigmoidal. This means that relative changes in the stimulus cause small relative changes in the response at a low stimulus, but high changes at higher stimuli. It is believed that ultrasensitivity plays an important role in intracellular signaling, because it can reduce noise and convert gradual stimuli into allornon decisions [GOLDBETER & KOSHLAND 81, FERRELL 96].

4.6 Ultrasensitivity In The MAPK Cascade

The steadystate of the double phosphorylated MAPK** as a function of the concentration of the enzyme which phosphorylates MAPKKK serves as a stimulusresponse function of the system. The plot of this function for the model according to Huang/Ferrell [HUANG& FERRELL 96] shows that this inputresponsecurve is sigmoidal (see Figure 4.4.3).

Furthermore, when plotting the steadystate of the active forms MAPKKK* and MAPKK** as a function of the MAPKKK activating enzyme, the curve for MAPKKK* is not sigmoidal and the curve of MAPKK** is strongly sigmoidal.

If the concentration of enzymesubstratecomplex is negligible, the three steps of the cascade are well separated, and the stimulusresponsecurve of the entire cascade can be treated as a threestep chain of functions:

$$\text{MAPK**(Input)} = \text{MAPK**(MAPKK**(MAPKKK*(Input)))} \qquad (4.6.1)$$

Therefore, one can analyze each layer of the cascade separately and study the steadystates of the active form for the three kinases as a function of the activating kinase.

We saw before the steady-state activation of the three kinases involved in the MAPK cascade module for the model according to Huang/Ferrell [HUANG & FERRELL 96]. It is interesting to see also how do these concentrations change in time, with a stable input (for our example I chose [E1] = 0.0003μM as proposed in [HUANG & FERRELL 96]), after our implementation of the model in Matlab (Figure 4.6.1 for MAPK**, Figure 4.6.2 for MAPKK** and Figure 4.6.3 for MAPKKK*):

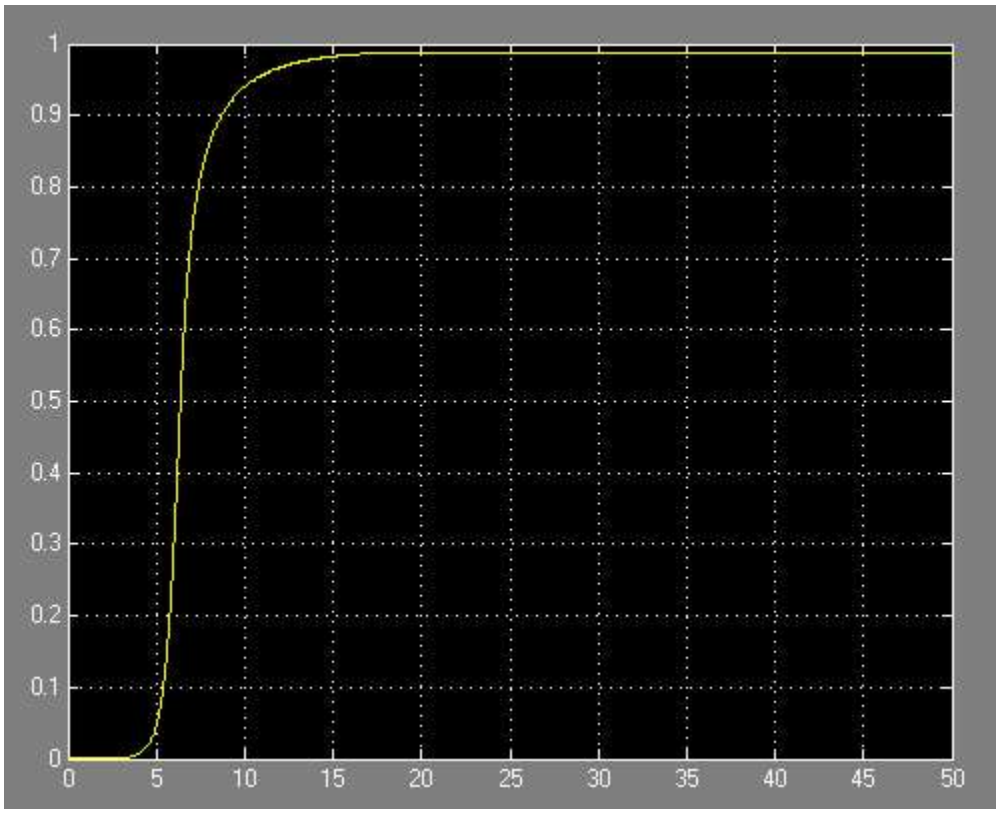


Figure 4.6.1: ***f(t,MAPK**)***

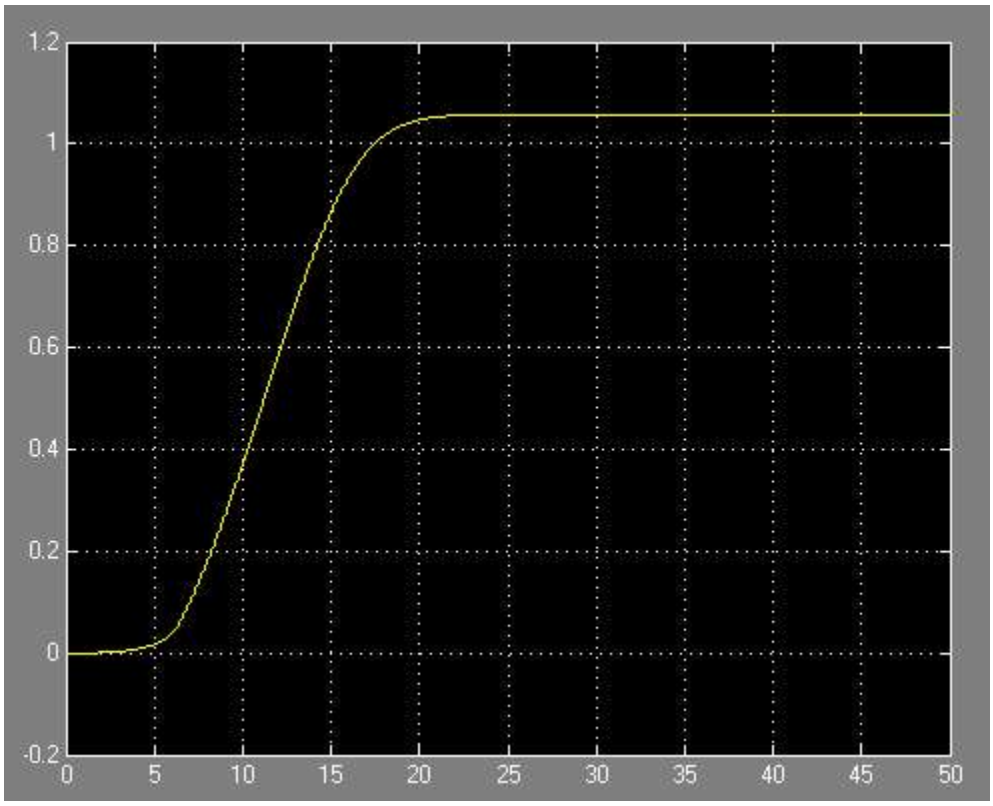


Figure 4.6.2: ***f(t,MAPKK**)***

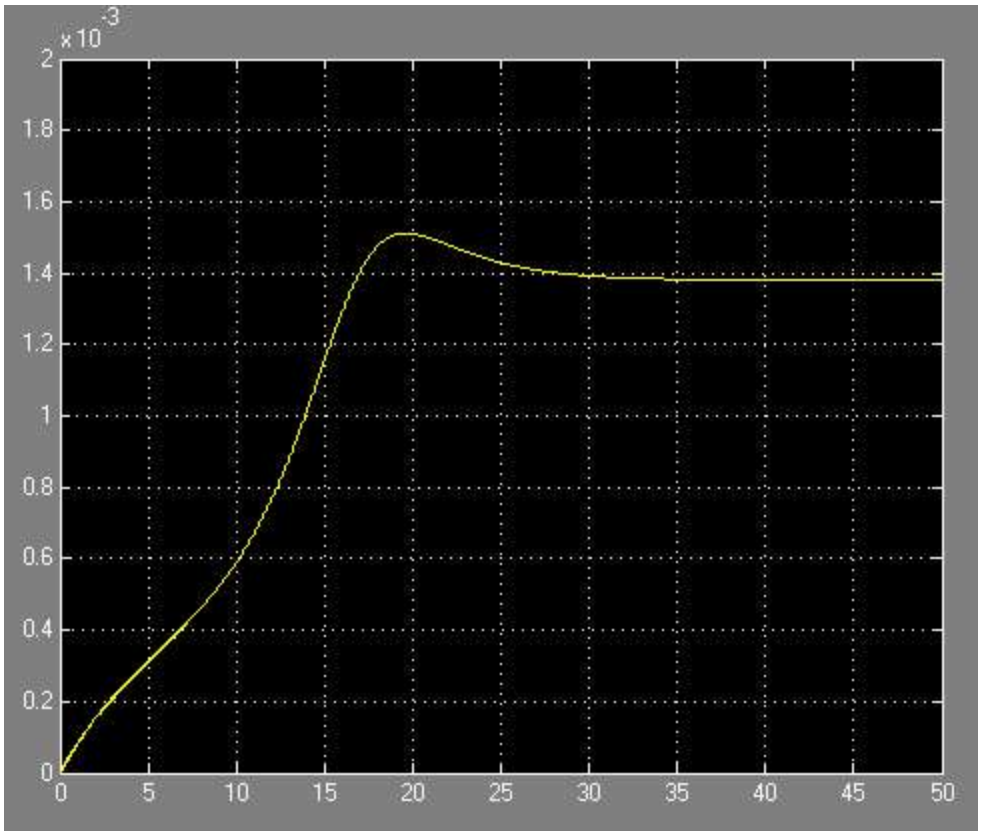


Figure 4.6.3: ***f(t,MAPKKK*)***

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 kinase, does not vary sensibly.

5 Conclusions

In this thesis stimulusresponse curves of the mitogenactivated protein kinase (MAPK) cascade (published model by Huang/Ferrell [HUANG & FERRELL 96]) are investigated. It is proven that the sigmoidality of the stimulusresponsecurves is a result of the doublephosphorylation and saturation of the enzymes acting in the second and third layer of the cascade. We know that if we add a negative feedback loop (see model by Bhalla/Iyengar [BHALLA & IYENGAR 99]), the cascade shows oscillations. If the stimulusresponsecurves of the layers in the cascade are not too sigmoidal, the oscillations are damped and have been interpreted as adaptation, which seems to have physiological relevance and which has been experimentally observed. It is known from the literature that a strong sigmoidal stimulusresponsecurve may lead to sustained oscillations, due to the delay in the signal transduction [KHOLODENKO 00]. The physiological relevance of sustained oscillations in the MAPKcascade is debated, and there is no experimental evidence yet [KHOLODENKO 00].

For future work, the effect of fluctuations of the concentrations might be studied. It will get important when more about the total concentrations and the phosphatase activities is known. Additionally, the effect of scaffold molecules might be of interest [PAWSON & SCOTT 97] because they are predicted to reduce the ultrasensitivity of the stimulusresponsecurve [LEVCHENKO ET AL. 00].

To resume, it can be argued that the highly conserved MAPKcascade is a “multipotent” module in eukaryotic signal transduction. It may act as a (irreversible) switch, a signal amplifier and an integral feedback controller. The function of the MAPKcascade depends on the context, therefore the role of MAPKcascade might be different in different organisms, developmental stages and pathways. If the MAPKcascade will be considered in a model of a whole pathway including receptor and transcription, it might be useful to investigate experimentally the role of the cascade and afterwards develop a simplified model of the cascade in order to keep the numbers of equations and parameters small.

Appendix

In order to keep variable names short, MAPK is abbreviated as K. An S at the end indicates phosphotases, wherever it is neccessary, a hyphen indicates a complex.

Stars show the phosphorylation states.

A.1 Model According To Huang And Ferrell

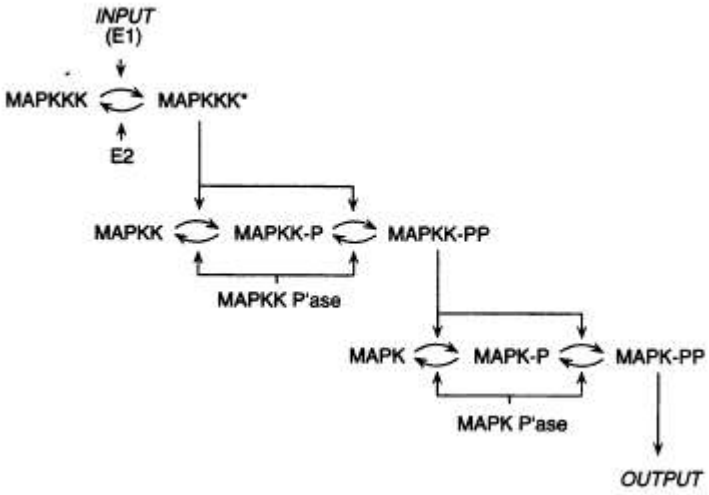


Figure A.1

A.2 Parameters

Total Concentration [μM]

Protein	H/F
Input-enzyme	0.0003
MAPKKK	0.003
MAPKK	1.2
MAPK	1.2
E2	0.0003
MAPKK-Phosphatase	0.0003
MAPK-Phosphatase	0.12

Table A.2.1: ***Concentrations taken from [HUANG &FERRELL 96]***

V_{max} [1/min]	
Reaction	**H/F**
MAPKKK → MAPKKK*	150
MAPKKK ← MAPKKK*	150
MAPKK → MAPKK*	150
MAPKK ← MAPKK*	150
MAPKK* → MAPKK**	150
MAPKK* ← MAPKK**	150
MAPK → MAPK*	150
MAPK ← MAPK*	150
MAPK* → MAPK**	150
MAPK* ← MAPK**	150

Table A.2.2: ***V_{max} values taken from [HUANG &FERRELL 96]***

k_m [μM]	
Reaction	**H/F**
MAPKKK → MAPKKK*	0.3
MAPKKK ← MAPKKK*	0.3
MAPKK → MAPKK*	0.3
MAPKK ← MAPKK*	0.3

MAPKK* → MAPKK**	0.3
MAPKK* ← MAPKK**	0.3
MAPK → MAPK*	0.3
MAPK ← MAPK*	0.3
MAPK* → MAPK**	0.3
MAPK* ← MAPK**	0.3

Table A.2.3: ***k_m values taken from [HUANG & FERRELL 96]***

A.3 Reactions

From the MAPK model according to Huang and Ferrell shown in Figure A.1, we have the following reactions:

$$KKK + E1 \Leftrightarrow KKK.E1 \rightarrow KKK^* + E1 \quad \text{(A.3.1)}$$

$$KKK^* + E2 \Leftrightarrow KKK^*.E2 \rightarrow KKK + E2 \quad \text{(A.3.2)}$$

$$KK + KKK^* \Leftrightarrow KK.KKK^* \rightarrow KK^* + KKK^* \quad \text{(A.3.3)}$$

$$KK^* + KKS \Leftrightarrow KK^*.KKS \rightarrow KK + KKS \quad \text{(A.3.4)}$$

$$KK^* + KKK^* \Leftrightarrow KK^*.KKK^* \rightarrow KK^{**} + KKK^* \quad \text{(A.3.5)}$$

$$KK^{**} + KKS \Leftrightarrow KK^{**}.KKS \rightarrow KK^* + KKS \quad \text{(A.3.6)}$$

$$K + KK^{**} \Leftrightarrow K.KK^{**} \rightarrow K^* + KK^{**} \quad \text{(A.3.7)}$$

$$K^* + KS \Leftrightarrow K^*.KS \rightarrow K + KS \quad \text{(A.3.8)}$$

$$K^* + KK^{**} \Leftrightarrow K^*.KK^{**} \rightarrow K^{**} + KK^{**} \quad \text{(A.3.9)}$$

$$K^{**} + KS \Leftrightarrow K^{**}.KS \rightarrow K^* + KS \quad \text{(A.3.10)}$$

A.4 Differential Equations Using Explicit Kinetics

The parameters α_n, d_n and V_n were chosen in order to fulfill the k_m and V_{max} values in Tables A.2.3 and A.2.2:

$$\alpha_n = V_n(1+r)/k_{m,n} \quad \text{(A.4.1)}$$

$$d_n = rV_n \quad \text{(A.4.2)}$$

$$V_n = V_{max,n} \quad \text{(A.4.3)}$$

The ratio of d and V_{max}(r) was chosen as 4 as in [HUANG & FERRELL 96].

We can notice from Tables A.2.2 and A.2.3 that the values of k_m and V_{max} are the same for each one of the ten reactions. So, it is obvious from the equations A.4.1, A.4.2 and A.4.3 that we have:
$\alpha_1 = \alpha_2 = \ldots = \alpha_{10} = \alpha$,
$d_1 = d_2 = \ldots = d_{10} = d$ and
$V_1 = V_2 = \ldots = V_{10} = V$.

Finally, the ODEs for our model are the following:

$$d[KKK]/dt = -\alpha1[KKK][E1] + d1[E1\_KKK] + V2[E2\_KKK^*] \quad \text{(A.4.4)}$$

$$d[E1]/dt = -\alpha1[KKK][E1] + (d1+V1)[E1\_KKK] \quad \text{(A.4.5)}$$

$$d[E1\_KKK]/dt = \alpha1[KKK][E1] - (d1+V1)[E1\_KKK] \quad \text{(A.4.6)}$$

$$\begin{aligned} d[KKK^*]/dt = {} & V1[E1\_KKK] - \alpha2[KKK^*][E2] + d2[E2\_KKK^*] \\ & - \alpha3[KK][KKK^*] + (d3+V3)[KKK^*\_KK] \\ & - \alpha5[KK^*][KKK^*] + (d5+V5)[KKK^*\_KK^*] \end{aligned} \quad \text{(A.4.7)}$$

$$d[E2]/dt = -\alpha2[KKK^*][E2] + (d2+V2)[E2\_KKK^*] \quad \text{(A.4.8)}$$

$$d[E2\_KKK^*]/dt = \alpha2[KKK^*][E2] - (d2+V2)[E2\_KKK^*] \quad \text{(A.4.9)}$$

$$d[KK]/dt = -\alpha3[KK][KKK^*] + d3[KKK^*\_KK] + V4[KKS\_KK^*] \quad (A.4.10)$$

$$d[KKK^*\_KK]/dt = \alpha3[KK][KKK^*] - (d3+V3)[KKK^*\_KK] \quad (A.4.11)$$

$$d[KK^*]/dt = V3[KKK^*\_KK] - \alpha4[KK^*][KKS] + d4[KKS\_KK^*] - \alpha5[KK^*][KKK^*] + d5[KKK^*\_KK^*] + V6[KKS\_KK^{**}] \quad (A.4.12)$$

$$d[KKS]/dt = -\alpha4[KK^*][KKS] + (d4+V4)[KKS\_KK^*] - \alpha6[KK^{**}][KKS] + (d6+V6)[KKS\_KK^{**}] \quad (A.4.13)$$

$$d[KKS\_KK^*]/dt = \alpha4[KK^*][KKS] - (d4+V4)[KKS\_KK^*] \quad (A.4.14)$$

$$d[KKK^*\_KK^*]/dt = \alpha5[KK^*][KKK^*] - (d5+V5)[KKK^*\_KK^*] \quad (A.4.15)$$

$$d[KK^{**}]/dt = V5[KKK^*\_KK^*] - \alpha6[KK^{**}][KKS] + d6[KKS\_KK^{**}] - \alpha7[K][KK^{**}] + (d7+V7)[KK^{**}\_K] - \alpha9[K^*][KK^{**}] + (d9+V9)[KK^{**}\_K^*] \quad (A.4.16)$$

$$d[KKS\_KK^{**}]/dt = \alpha6[KK^{**}][KKS] - (d6+V6)[KKS\_KK^{**}] \quad (A.4.17)$$

$$d[K]/dt = -\alpha7[K][KK^{**}] + d7[KK^{**}\_K] + V8[KS\_K^*] \quad (A.4.18)$$

$$d[KK^{**}\_K]/dt = \alpha7[K][KK^{**}] - (d7+V7)[KK^{**}\_K] \quad (A.4.19)$$

$$d[K^*]/dt = V7[KK^{**}\_K] - \alpha8[K^*][KS] + d8[KS\_K^*] - \alpha9[K^*][KK^{**}] + d9[KK^{**}\_K^*] + V10[KS\_K^{**}] \quad (A.4.20)$$

$$d[KS]/dt = -\alpha8[K^*][KS] + (d8+V8)[KS\_K^*] - \alpha10[K^{**}][KS] + (d10+V10)[KS\_K^{**}] \quad (A.4.21)$$

$$d[KS\_K^*]/dt = \alpha8[K^*][KS] - (d8+V8)[KS\_K^*] \quad (A.4.22)$$

$$d[KK^{**}\_K^*]/dt = \alpha9[K^*][KK^{**}] - (d9+V9)[KK^{**}\_K^*] \quad (A.4.23)$$

$$d[K^{**}]/dt = V9[KK^{**}\_K^*] - \alpha10[K^{**}][KS] + d10[KS\_K^{**}] \quad (A.4.24)$$

$$d[KS\_K^{**}]/dt = \alpha_{10}[K^{**}][KS] - (d_{10}+V_{10})[KS\_K^{**}] \quad (A.4.25)$$

References

[ACLARA 04] ACLARA BioSciences Inc. REALIZING THE PROMISE OF PERSONALIZED MEDICINE,Industry Overview,(2004).

[ASTHAGIRI &LAUFFENBURGER 01] A.R. Asthagiri and D.A. Lauffenburger, A computational study of feedback effects on signal dynamics in a mitogenactivated protein kinase (MAPK) pathway model., Biotechnol. Prog. 17 (2001), 227--39.

[ASTHAGIRI ET AL.00] A.R. Asthagiri, C.A. Reinhart, A.F. Horwitz, and D.A. Lauffenburger, The role of transient ERK2 signals in fibronectin and insulin mediated DNA synthesis., J. Cell Sci. 113 (2000), 4499--510.

[BICKNELL & HARRIS 04] Roy Bicknell and Adrian L. Harris, NOVEL ANGIOGENIC SIGNALING PATHWAYS AND VASCULAR TARGETS, Annual Review of Pharmacology and Toxicology Vol. **44**, page 219-239 (2004).

[BHALLA & IYENGAR 99] U.S. Bhalla and R. Iyengar, Emergent properties of networks of biological signaling pathways., Science 283 (1999), 381--7.

[BLUTHGEN 02] Nils Bluthgen, Diplomarbeit im Studiengang Physik, "Dynamical Models of Signal Transduction and the Influence of Feedback Loops" (2002)

[BOCK 81] H.G. Bock, "Numerical treatment of inverse problems in chemical reaction
kinetics," in *Modeling Chemical Reaction Systems,* K.H. Ebert, Ed.
Springer-Verlag: New York, vol. 18, pp. 102-125, 1981.

[BRAY 95] D. Bray, Protein molecules as computational elements in living cells., Nature 376 (1995), 307--12.

[BRUTLAG ET AL.91] D.L. Brutlag, A.R. Galper, and D.H. Millis, Knowledgebased simulation of DNA metabolism: prediction of enzyme action., Comput. Appl. Biosci. 7 (1991), 9--19.

[CHO ET AL. 02] K.-H. Cho, S.-Y. Shin, H.-W. Lee, and O. Wolkenhauer, "Investigations into the analysis and modeling of the TNF α mediated NF-κ B signaling pathway," in *Proc. 3rd Int. Conf. Systems Biology*, Stockholm, Sweden, p. 87, 2002 and *Genome Research* (Special Issue on Systems Biology), to be published.

[CHO ET AL. 03] K.-H. Cho, S.-Y. Shin, H.-W. Kim, O. Wolkenhauer, B. McFerran, and W.Kolch, *Mathematical Modeling of the Influence of RKIP on the ERK Signaling Pathway* (Lecture Notes in Computer Science). Milan, Italy: Springer-Verlag,2003.

[CLARK & RUSSELL 00] David P. Clark, Lonnie D. Russell, MOLECULAR BIOLOGY made simple and fun, Vol. **2,** page 25-196, (2000).

[CORNISHBOWDEN 99] A. CornishBowden, Fundamentals of Enzyme Kinetics, Portland Press, London, 1999.

[DIAZ-MECO ET AL.94] M.T. DiazMeco, J. Lozano, M.M. Municio, E. Berra, S. Frutos, L. Sanz, and J. Moscat, Evidence for the in vitro and in vivo interaction of Ras with protein kinase C zeta., J. Biol. Chem. 269 (1994), 31706--10.

[DJERBI 03] Mounira Djerbi, APOPTOSIS IN CANCER AND AUTOIMMUNITY **1**, June 5, (2003).

[FERRELL &XIONG 01] J.E. Ferrell and W. Xiong, Bistability in cell signaling: How to make continuous processes discontinuous, and reversible processes irreversible, Chaos 11 (2001), 227--35.

[FERRELL 96] J.E. Ferrell, Tripping the switch fantastic: how a protein kinase cascade can convert graded inputs into switchlike outputs., Trends Biochem. Sci. 21 (1996), 460--6.

[FINKELSTEIN ET AL. 04] Anthony Finkelstein, James Hetherington, Linzhong Li, Ofer Margoninski, Peter Saffrey, Rob Seymour, Anne Warne, Computational Challenges of Systems Biology, IEEE Computer Society, pages 26-33, May (2004).

[GIBSON &MJOLSNESS 00] M.E. Gibson and E. Mjolsness, Computational Methods in Molecular & Cellular Biology: from Genotype to Phenotype, ch. Modeling the Activity of Single Genes, MIT Press, 2000.

[GIJS 04] Gijs R van den Brink, Linking pathways in colorectal cancer. NATURE GENETICS. Vol. 36,1038-1039, October (2004).

[GILLESPIE 77] D.T. Gillespie, Exact Stochastic Simulation of Coupled Chemical Reactions, J. Phys. Chem. 81 (1977), 2340--61.

[GOLDBETER &KOSHLAND 81] A. Goldbeter and D.E. Koshland, An amplified sensitivity arising from covalent modification in biological systems., Proc. Natl. Acad. Sci. USA 78 (1981), 6840--4.

[GU ET AL. 00] Gu, J., et al., Identification of a sequence element from p53 that signals for Mdm2-targeted degradation. Mol. Cell Biol., **20,** 1243-1253 (2000).

[HANAHAN &WEINBERG 00] Douglas Hanahan and Robert A. Weinberg, The Hallmarks of Cancer, Cell, Vol. **100**, 57–70, January 7, (2000).

[HARTWELL ET AL. 99] L.H. Hartwell, Hopfield, J.J., Leibler, S. and Murray, A.W.: 'From molecular to modular cell biology', *Nature*, 1999, **402**, pp. C47-C52

[HEGGER ET AL. 98] R. Hegger, H. Kantz, F. Schmuser, M. Diestelhorst, R.P. Kapsch, and H. Beige, "Dynamical properties of a ferroelectric capacitor observed through nonlinear time series analysis," *Chaos*, vol. 8, pp. 727-736, 1998.

[HOFESTÄDT & THELEN 98] R. Hofestädt and S. Thelen, Quantitative modeling of biochemical networks., In. Silico. Biol. 1 (1998), 39--53.

[HOHMANN ET AL. 04] Stefan Hohmann, Jens Nielsen, Hiroaki Kitano, White Paper - Yeast Systems Biology Network, page 1-8, May (2004).

[HUANG & FERRELL 96] Huang, C.-Y.F. and Ferrell, J.E.: `Ultrasensitivity in the mitogen-activated protein kinase cascade',
Proc. Natl. Acad. Sci. USA, 1996, 93, pp. 10078-10083.

[HUANG ET AL. 99] Huang, D.C., et al., Activation of Fas by FasL induces apoptosis by a mechanism that cannot be blocked by Bcl-2 or Bcl-XL . Proc. Natl. Acad. Sci. USA., **96,** 14871-14876 (1999).

[HUGHES ET AL. 03] Steven Hughes, Jon Dobson and Alicia El Haj., Mechanical Stimulation of Calcium Signaling Pathways in Human Bone Cells Using Ferromagnetic Micro-particles: Implications for Tissue Engineering, EUROPEAN CELLS AND MATERIALS Vol. **6.** page 43 September 2 (2003).

[IRUSTA ET AL. 02] Pablo M. Irusta1,2, Yue Luo1, Omar Bakht3, Char-Chang Lai4, Steven O. Smith3,and Daniel DiMaio, Definition of an inhibitory juxtamembrane WW-like domain in the platelet-derived growth factor β receptor, JBC PAPERS IN PRESS, August 13 (2002).

[JIMENEZ ET AL. 99] Jimenez, G.S., et al., p53 regulation by post-translational modification and nuclear retention in response to diverse stresses. Oncogene, **18,** 7656-7665 (1999).

[JUO ET AL. 99] Juo, P., et al., FADD is required for multiple signaling events downstream of the receptor Fas. Cell Growth Differ., **10,** 797-804 (1999).

[KARTHIKEYAN & RAKESH 02] Karthikeyan Kandasamy and Rakesh K. Srivastava1, Role of the Phosphatidylinositol 3_-Kinase/PTEN/Akt Kinase Pathway in Tumor Necrosis Factor-related Apoptosis-inducing Ligand-induced Apoptosis in Non-Small Cell Lung Cancer Cells, CANCER RESEARCH 62, 4929–4937, September 1, (2002).

[KAWABATA & MIYAZONO 99] Kawabata, M., and Miyazono, K., Signal transduction of the TGF-β superfamily by Smad proteins. J. Biochem. (Tokyo), **125**, 9-16 (1999).

[KHOLODENKO 00] B.N. Kholodenko, Negative feedback and ultrasensitivity can bring about oscillations in the mitogenactivated protein kinase cascades., Eur. J. Biochem. 267 (2000), 1583--8.

[KHOLODENKO ET AL.97] B.N. Kholodenko, J.B. Hoek, H.V. Westerhoff, and G.C. Brown, Quantification of information transfer via cellular signal transduction pathways., FEBS Letters 414 (1997), 430--4.

[KHOLODENKO ET AL.00] B.N. Kholodenko, J.B. Hoek, and H.V. Westerhoff, Why cytoplasmic signalling proteins should be recruited to cell membranes., Trends. Cell. Biol. 10 (2000), 173--8.

[KING & CIDLOWSKI 98] King, K.L., and Cidlowski, J.A., Cell cycle regulation and apoptosis. Annu. Rev. Physiol., **60,** 601-617 (1998).

[KITANO] Caltech ERATO Kitano, “Systems Biology Workbench,” [Online].
http://www.cds.caltech.edu/erato/

[KOLCH ET AL.93] W. Kolch, G. Heidecker, G. Kochs, R. Hummel, H. Vahidi, H. Mischak, G. Finkenzeller, D. Marme, and U.R. Rapp, Protein kinase C alpha activates RAF1 by direct phosphorylation., Nature 364 (1993), 249--52.

[KOSHLAND 98] D.E. Koshland, The era of pathway quantification., Science 280 (1998), 852--3.

[KRAUSS 97] G. Krauss, Biochemie der Regulation und Signaltransduktion, WileyVCH, Weinheim, 1997.

[KREMLING ET AL. 00] A. Kremling,K. Jahreis, J.W. Lengeler, and E.D. Gilles, "The organization of metabolic reaction networks: A signal-oriented approach to cellular models,"*Metabolic Eng.*, vol. 2, no. 3, pp. 190-200, July 2000.

[LAUFFENBURGER 00] Lauffenburger, D.A.: 'Cell signaling pathways as control modules: Complexity for simplicity', *Proc. Nat. Acad. Sci. USA*, 2000, **97**, pp. 5031-5033

[LAUFFENBURGER 01] Lauffenburger, D.A.: `A computational study of feedback e_ects on signal dynamics in a mitogenactivated protein kinase (MAPK) pathway model' Biotechnol. Prog., 2001, 17, pp. 227-239.

[LEVCHENKO ET AL.00] A. Levchenko, J. Bruck, and P.W. Sternberg, Scaffold proteins may biphasically affect the levels of mitogenactivated protein kinase signaling and reduce its threshold properties., Proc. Natl. Acad. Sci. USA 97 (2000), 5818--23.

[LOWES ET AL. 02] Lowes, V.L., et al., Integration of signals from receptor tyrosine kinases and G protein-coupled receptors. Neurosignals, **11**, 5-19 (2002).

[MARSHALL 95] C.J. Marshall, "Specificity of receptor tyrosine kinase signaling: Transient versus sustained extracellular signal-regulated kinase activation," *Cell*,vol. 80, pp. 179-185, Jan. 1995.

[MARSHALL 95b] M.S. Marshall, Ras target proteins in eukaryotic cells., FASEB J. 9 (1995), 1311--8.

[NIEHRS & BIRCHMEIER 05] Christof Niehrs, Walter Birchmeier, Signaling Pathways, Cell Biology and Cancer in the HGF, Division of Molecular Embryology, HGF Association,(2005).

[PAWAN ET AL. 04] Pawan K. Dhar, Hao Zhu, Chee Meng Tan, Sandeep Somani, Ye Li, Santosh K. Mishra, Modeling and Simulation in Systems Biology: The Genesis and Evolution of Challenges, Strategies and Methods, Bioinformatics Institute of Singapore, 6 June, (2004).

[PAWSON & SCOTT 97] T. Pawson and J.D. Scott, Signaling through scaffold, anchoring, and adaptor proteins., Science 278 (1997), 2075--80.

[PIMENTEL-MUINOS & SEED 99] Pimentel-Muinos, F.X., and Seed, B., Regulated commitment of TNF receptor signaling: a molecular switch for death or activation. Immunity, **11**, 783-793 (1999).

[PLUMPE ET AL. 00] Plumpe, J., et al., NF- κB determines between apoptosis and proliferation in hepatocytes during liver regeneration. Am. J. Physiol. Gastrointest. Liver Physiol., **278**, G173-G183 (2000).

[ROBERT & TOM 01] D.P. Robert and M. Tom, "Kinetic modeling approaches to in vivo imaging,"*Nature Rev.: Molecular Cell Biol.*, vol. 2, pp. 898-907, Dec. 2001.

[SCHINDLER 99] Schindler, C., Cytokines and JAK-STAT signaling. Exp. Cell Res., **253**, 7-14 (1999).

[SCHNELL &MENDOZA 97] S. Schnell and C. Mendoza, Closed Form Solution for Timedependent Enzyme Kinetics, J. theor. Biol. 187 (1997), 207--212.

[SCHWANDER ET AL. 98] Schwandner, R., et al., TNF receptor death domain-associated proteins TRADD and FADD signal activation of acid sphingomyelinase. J. Biol. Chem., **273**, 5916-5922 (1998).

[SEGER & KREBS 95] Seger, R., and Krebs, E.G., The MAPK signaling cascade. FASEB J., **9**, 726-735 (1995).

[SONTAG 04] Eduardo D. Sontag, Some new directions in control theory inspired by Systems Biology. (2004)

[TAMURA ET AL. 02] Tamura, S., et al., Regulation of stress-activated protein kinase signaling pathways by protein phosphatases. Eur. J. Biochem., **269**, 1060-1066 (2002).

[THIEFFRY &ROMERO 99] D. Thieffry and D. Romero, The modularity of biological regulatory networks., Biosystems 50 (1999), 49--59.

[THOMAS &KAUFMAN 01] R. Thomas and M. Kaufman, Multistationarity, the basis of cell differentiation and memory. II. Logical analysis of regulatory networks in terms of feedback circuits., CHAOS 11 (2001), 180--95.

[WAKIOKA ET AL. 99] Wakioka, T., et al., APS, an adaptor protein containing Pleckstrin homology (PH) and Src homology-2 (SH2) domains inhibits the JAK-STAT pathway in collaboration with c-Cbl. Leukemia, **13**, 760-767 (1999).

[WANG ET AL. 00] Wang, J., et al., Inhibition of Fas-mediated apoptosis by the B cell antigen receptor through c-FLIP. Eur. J. Immunol., **30,** 155-163 (2000).

[WIDMANN ET AL. 99] Widmann, C., Spencer, G., Jarpe, M.B. and Johnson, G.L.: 'Mitogen-activated protein kinase: Conservation of a three-kinase module from yeast to human', Physiol. Rev., 1999, 79, pp. 143-180.

[WIKIPEDIA] Wikipedia, the free encyclopedia.

[WINGENDER ET AL.00] E. Wingender, X. Chen, R. Hehl, H. Karas, I. Liebich, V. Matys, T. Meinhardt, M. Pruss, I. Reuter, and F. Schacherer, TRANSFAC: an integrated system for gene expression regulation., Nucleic Acids Res. 28 (2000), 316--9.

[WOLKENHAUER 01] O.Wolkenhauer, "Systems biology: The reincarnation of systems theory applied in biology?," *Briefings Bioinform.*, vol. 2, no. 3, pp. 258-270, 2001.

[WOLKENHAUER ET AL. 03] Olaf Wolkenhauer, Hiroaki Kitano, and Kwang-Hyun Cho, 'Systems Biology: Looking at opportunities and challenges in applying systems theory to molecular and cell biology'. (August 2003)

[WRANA 98] Wrana, J.L., TGF-β receptors and signalling mechanisms. Miner. Electrolyte Metab., **24**, 120-130 (1998).

[YEUNG ET AL. 00] K. Yeung, P. Janosch, D.W. Rose, H. Mischak, J.M. Sedivy, and W. Kolch,"Mechanism of suppression of the Raf/MEK/Extracellular Signal-Regulated Kinase pathway by the Raf Kinase Inhibitor Protein," *Mol. Cell. Biol.*, vol. 20,pp. 3079-3085, May 2000.

[YEUNG ET AL. 99] K. Yeung, T. Seitz, S. Li, P. Janosch, B. McFerran, C. Kaiser, F. Fee, K.D.Katsanakis, D.W. Rose, H. Mischak, J.M. Sedivy, and W. Kolch, "Suppression of Raf-1 kinase activity and MAP kinase signaling by RKIP," *Nature*, vol. 401, pp.173-177, Sept. 1999.

[YI ET AL.00] T.M. Yi, Y. Huang, M.I. Simon, and J. Doyle, Robust perfect adaptation in bacterial chemotaxis through integral feedback control., Proc. Natl. Acad. Sci. USA 97 (2000), 4649--53.