STRUCTURAL ANALYSIS OF EPIDERMAL GROWTH FACTOR RECEPTOR AND YEAST SIGNALING NETWORKS

by

Saliha Durmuş Tekir

B.S., Chemical Engineering, Boğaziçi University, 2004

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science

Graduate Program in Chemical Engineering

Boğaziçi University

2007

ACKNOWLEDGEMENTS

The research in this thesis was performed at the Chemical Engineering Department of the Boğaziçi University between September 2004 and January 2007 under the supervision of Prof. Dr. Kutlu Ülgen.

I offer my deepest thanks to Prof. Dr. Kutlu Ülgen for her support and counsel throughout this study and also for her patience in evaluating the numerous drafts of this thesis.

I would like to thank all the girls in the laboratory; Esra, Betül, Senem, Ayça, Aylin, Nil and Hilal for enjoyable coffee breaks. My special thanks to our boys, Yalçın and Tunahan for answering all my questions.

ABSTRACT

STRUCTURAL ANALYSIS OF EPIDERMAL GROWTH FACTOR RECEPTOR AND YEAST SIGNALING NETWORKS

The recent progress in high-throughput technologies is allowing the reconstruction of large-scale models of signaling networks and development of corresponding analysis methods. Network properties of the most comprehensive published maps of the epidermal growth factor receptor (EGFR) signaling with related pathways and signal transduction mechanisms in yeast (*Saccharomyces cerevisiae*) were analyzed structurally using graph theoretic and pathway analysis techniques. The EGFR signaling pathway controls growth, survival, proliferation and differentiation of mammalian cells. It is one of the most investigated and best-understood signal transduction systems and now it is being used to understand the general mechanisms of signal transduction processes. The graph theoretic analyses of the EGFR and yeast signaling networks indicated that they have small-world topologies. The connectivity distributions of the signaling molecules in the networks have scale-free natures following power law model. The pathway analysis of EGFR signaling network indicated that as a result of signal transduction apoptosis may be the most frequent phenotype of the EGFR signaling having the smallest path lengths for nearly same number of linear paths for all phenotypes. The tumorigenesis and mitogenesis are found to be the least robust phenotypes having limited routes from the ligands that pass over the same six signaling molecules each time. Analysis via linear paths revealed that G1/S cell wall remodeling and mating are the most and least frequent phenotypes of yeast signaling network, respectively. The critical signaling molecules in the networks that participate in most of the linear paths from the ligands to the phenotypes were identified using pathway analysis. The cofactor GTP and the complex G□/G□are involved in more than 70 % of the linear paths through all phenotypes in the network indicating that the pathways going to all phenotypes of EGFR signaling network are crosstalking through these molecules. Similarly the proteins Rvs167 and Sla1 have roles in the linear paths going to all phenotypes in signal transduction network of yeast.

ÖZET

EPIDERMAL BÜYÜME FAKTÖRÜ ALMACI VE MAYA SİNYAL İLETİ AĞYAPILARININ YAPISAL ANALİZİ

Hızlı-tarama teknolojilerindeki yeni gelişmeler büyük ölçekli sinyal ileti ağyapı modellerinin oluşturulmasına ve bu modelleri analiz etmekte kullanılan metodların geliştirilmesine olanak sağlamaktadır. Bu tez çalışması kapsamında epidermal büyüme faktörü almacı sinyal ileti ağyapısı ve maya hücresindeki sinyal mekanizmaları grafik teorisi ve yolizi analizi yöntemleri ile yapısal olarak incelenmiştir. Epidermal büyüme faktörü almacı sinyal ileti ağyapısı memeli canlı hücrelerinde büyüme, yaşama, çoğalma ve başkalaşma mekanizmalarını düzenlemektedir. Bu ağyapı, en fazla incelenmiş ve en iyi anlaşılmış sinyal ileti sistemlerindendir ve günümüzde sinyal iletim sürecindeki genel mekanizmaları anlamak amacıyla kullanılmaktadır. Grafik teorisi teknikleri kullanılarak yapılan analiz sonucunda EGFR ve maya sinyal ileti ağyapılarının küçük-dünya yapısına sahip olduğu bulunmuştur. Bu ağyapıların içerisindeki sinyal moleküllerinin birbirleriyle olan bağlarının dağılımının ölçek-bağımsız olduğu ve güç-yasası modeline uyum sağladığı anlaşılmıştır. EGFR sinyal ileti ağyapısına yolizi analizi uygulandığında sinyal iletimi sonucunda en sık meydana gelen fenotipin programlanmış hücre ölümü olduğu gözlemlenmiştir. Bu fenotip diğer fenotiplerle yaklaşık olarak aynı sayıdaki doğrusal yola en kısa yol uzunluğu ile ulaşmaktadır. Diğer fenotipler arasında ise tümör oluşumu ve mitojenez fenotiplerinin çevresel ve genetik değişikliklere karşı en dayanıksız olduğu bulunmuştur. Hücre dışı sinyal proteinlerinden bu fenotiplere giden bütün doğrusal yollar aynı altı sinyal molekülünden geçmektedir. Maya sinyal ileti ağyapısında ise sahip oldukları doğrusal yolların sayısına paralel şekilde, hücrenin G1 ve S fazları arasındaki geçişten sonra hücre duvarının düzenlenmesi en sık görülen, çiftleşme ise en az görülen fenotip olarak bulunmuştur. Yolizi analizi ile doğrusal yollar içerisinde en fazla görev alan moleküller belirlenerek ağyapıdaki en önemli sinyal molekülleri bulunmuştur. EGFR sinyal ileti ağyapısında kofaktör GTP ve kompleks Gβ/Gγ'nın ağyapının bütün fenotiplerine giden doğrusal yolların % 70'inden fazlasında görev aldığı gözlemlenmiştir. Mayadaki sinyal ileti ağyapısındaki bütün fenotiplere giden doğrusal yollarda görev alan proteinler ise Rvs167 ve Sla1 olarak bulunmuştur.

TABLE OF CONTENTS

ACKNOWLEDGEMENT iii
ABSTRACT iv
ÖZET v
LIST OF FIGURES viii
LIST OF TABLES x
1. INTRODUCTION 1
2. BACKGROUND ASPECTS 4
2.1. Cell Signaling 4
2.1.1. Ligand 5
2.1.2. Receptor 6
2.1.3. Features of Signaling 7
2.1.3.1. Specificity 7
2.1.3.2. Amplification 8
2.1.3.3. Desensitization/Adaptation 8
2.1.3.4. Integration 8
2.1.4. Types of Cell Signaling 9
2.1.5. Signaling Networks 10
2.1.5.1. Signaling Network Reconstruction 11
2.1.5.2. Mathematical Modeling of Signaling Networks 12
2.2. Epidermal Growth Factor Receptor Signaling 16
2.3. Yeast Signaling 18
3. MATERIALS AND METHODS 24
3.1. Maps of Signaling Networks 24
3.1.1. Comprehensive Pathway Map of EGFR Signaling 24
3.1.2. Molecular Interaction Map of Yeast Signaling 30
3.2. Structural Analysis Techniques 32
3.2.1. Graph Theory 32
3.2.2. Metabolic Pathway Analysis 35
4. STRUCTURAL ANALYSIS OF EGFR SIGNALING NETWORK 37

4.1. Graph Theoretical Analysis 37
4.2. Pathway Analysis 40
4.2.1. Network Diameter 40
4.2.2. Phenotype Frequency 41
4.2.3. Network Crosstalk 42
4.2.4. Specific Molecules in Each Phenotypes 45
4.2.5. Participation of Molecules..... 47
4.2.6. Mitogen-Activated Protein Kinase Signaling 48
5. STRUCTURAL ANALYSIS OF YEAST SIGNALING NETWORK 51
5.1. Graph Theoretical Analysis 51
5.2. Pathway Analysis 53
5.2.1. Network Diameter 54
5.2.2. Phenotype Frequency 55
5.2.3. Network Crosstalk 56
5.2.4. Specific Proteins in Each Phenotypes 59
5.2.5. Participation of Proteins 60
6. CONCLUSIONS AND RECOMMENDATIONS 62
6.1. Conclusions 62
6.2. Recommendations 64
REFERENCES 65

LIST OF FIGURES

Figure 2.1.	Simplified view of an intracellular signaling pathway……..……	5
Figure 2.2.	A more detailed intracellular signaling pathway ……..……	6
Figure 2.3.	Specificity of receptor ……..…….	7
Figure 2.4.	Signal amplification ……..…….....	8
Figure 2.5.	Feedback of signaling ……..…….	8
Figure 2.6.	Signal integration ……..……	9
Figure 2.7.	Classification of signal transduction input-output relationships ……..	10
Figure 2.8.	Different approaches in reconstruction of signaling networks	12
Figure 2.9.	Three levels of resolution in reconstructions	13
Figure 2.10.	Structural analyses of signaling networks …….	14
Figure 2.11.	Integrative and iterative process of cellular signaling network reconstruction …….....	16
Figure 3.1.	Comprehensive map of EGFR signaling network	25
Figure 3.2.	The bow-tie architecture of EGFR signaling network …….	27

Figure 3.3. Main symbols adopted by CellDesigner ver. 2.1.1. …….. 29

Figure 3.4. Expression of the inner structures and states in the EGFR map 30

Figure 3.5. Signal transduction and carbohydrate metabolic pathways in yeast 31

Figure 3.6. Attributes of generic network structures 33

Figure 4.1. Degree (connectivity) distribution of the nodes in EGFR signaling network 38

Figure 4.2. Log of connectivities versus log of P(k) (probability of having connectivity, k) in EGFR signaling network 39

Figure 4.3. log k versus log P(k) in EGFR signaling network when the outlier is excluded 39

Figure 4.4. Dynamic model of EGFR signal transduction pathway ……..... 49

Figure 5.1. Degree (connectivity) distribution of the nodes in yeast signaling network 52

Figure 5.2. Log of connectivities versus log of P(k) (probability of having connectivity, k) in yeast signaling network 52

LIST OF TABLES

Table 4.1. Shortest path lengths of the phenotypes of EGFR signaling network 41

Table 4.2. Approximately same number of linear paths with different maximum path lengths for the phenotypes of EGFR signaling network 41

Table 4.3. Around 200 linear paths of the phenotypes of EGFR signaling network with maximum path lengths ……. 42

Table 4.4. Shared molecules of crosstalking pathways of EGFR signaling network 44

Table 4.5. Specific molecules of the phenotypes of EGFR signaling network ……..... 46

Table 4.6. The top five molecules that participate in most of the linear paths of each phenotype in EGFR signaling network …… 48

Table 5.1. Signal transduction through the phenotypes of yeast signaling network …. 53

Table 5.2. Shortest path lengths of the phenotypes of yeast signaling network …….. 55

Table 5.3. The number of linear paths with same maximum path lengths for the phenotypes of yeast signaling network ……..... 56

Table 5.4. The number of linear paths with the order of magitude 10^3 in yeast signaling network…….. 57

Table 5.5. Shared proteins of crosstalking pathways of yeast signaling network …..... 58

Table 5.6. Specific proteins in each phenotype of yeast signaling network …….. 59

Table 5.7. The top five proteins that participate in the linear paths of each phenotype in yeast signaling network ……... 61

1. INTRODUCTION

Mammalian cells integrate information from complex intracellular signal transduction pathways to make decisions in response to changes in environment. Signaling molecules and their networks also play important roles in molecular diagnosis and therapy of diseases as almost all known diseases exhibit some kind of dysfunction at gene level and hence there has been a great deal of enthusiasm to identify novel drug targets based on knowledge of key signal transduction components which have links to diseases. Understanding cellular signaling therefore is central for gaining insight into the molecular mechanisms behind diseases as well as adaptation of living cells to changes in the environment. A very important signaling network in mammalian cells is the epidermal growth factor receptor (EGFR) signaling pathway, which regulates processes such as growth, survival, proliferation and differentiation. The malfunctions of the signal transduction in this pathway may lead to uncontrolled cell proliferation resulting in tumor formations. It may also cause diseases with immunological, inflammatory and degenerative syndromes (Orton *et al.*, 2005). Therefore the understanding of EGFR signaling mechanism is vital for the identification of potential drug targets for these diseases and also for the understanding of design principles of signaling mechanisms in development and tissue homeostasis (Wells, 1999; Moghal and Sternberg, 1999).

In recent years the high-throughput techniques such as transcriptomics, proteomics and metabolomics enabled the collection of genome scale experimental data and lead to the reconstruction of large-scale signaling networks (Papin and Palsson, 2004b; Oda *et al.*, 2005a; Oda *et al.*, 2005b; Oda and Kitano, 2006). In the last 40 years, the mechanism of the EGFR signaling has been analyzed thoroughly both experimentally and computationally to uncover the unknowns about general principles of signal transduction processes including receptor-mediated endocytosis, oncogenesis, mitogen activated protein kinase (MAPK) signaling pathways, multi-receptor family interactions, autocrine loops and receptor transactivation (Carpenter, 2000). Lots of computational models have been developed for the EGFR signaling system (Gex-Fabry and DeLisi, 1984; Wiley and Cunnigham, 1981; Wiley and Cunningham, 1984; Chang *et al.*, 1991; Chang *et al.*, 1993; Opresko *et al.*, 1995; Kholodenko *et al.*, 1999; Brightman and Fell, 2000; Schoeberl *et al.*, 2002; Shvartsman *et al.*, 2002; Kholodenko, 2003; Oda *et al.*, 2005a). Nevertheless, only a

limited part of the EGFR-mediated signaling pathway could be modeled until recent years (Kholodenko, 2003) and the most comprehensive pathway map for EGFR-induced and related pathways was published in 2005 (Oda *et al.*, 2005a). Small-scale dynamic models were used as the basis for a dynamic analysis of EGFR signaling with determined kinetic parameters (Starbuck and Lauffenburger, 1992; Brightman and Fell, 2000). Large-scale stoichiometric models, on the other hand, give information on the architectural feature of the signaling network (Oda *et al.*, 2005a) and these structure-oriented analyses usually provide valuable insight into the mechanisms of the networks by defining function-structure relationships.

In the last decade, methods derived from graph theory have been developed to understand the structure of the systems of molecular interactions (Jeong *et al.*, 2000; Barabasi and Oltvai, 2004). The graph theoretic analysis can detect the components that are well or poorly connected in the interaction network and thus it is a useful guide for experimental studies on protein interactions. The recent comprehensive EGFR signaling network map (Oda *et al.*, 2005a) has the overall bow-tie structure that is a characteristic of the robust evolvable systems (Kitano, 2004). A variety of ligands bind to corresponding receptors leading to diverse phenotypes via activations of the intermediate signaling molecules in the network (Oda *et al.*, 2005a). In the present study, the topological structures of the signaling networks were captured as a system of linear paths connecting each ligand to a phenotype. The investigation of the signaling network topology in terms of phenotype frequency, network crosstalk and participation of reactions in signaling pathways in EGFR and yeast signaling networks were performed by analyzing their linear paths of which any network state can be defined as a unique, nonnegative, linear combination.

The second chapter answers the question; “What is cell signaling?” introducing general signal transduction pathways and signaling molecules. Cell signaling studies are given in two parts as reconstruction of signaling networks and their analysis both dynamically and structurally. Second and third sections of this chapter include information about analyzed networks of epidermal growth factor receptor (EGFR) and baker’s yeast (*Saccharomyces cerevisiae*) signaling, respectively. The third chapter gives the models of signaling networks and the computational methods used in this research. The analyzed

maps of the EGFR and yeast signaling networks as well as the structural analysis techniques (graph theory and pathway analysis) are explained in detail. The fourth and fifth chapters focus on the results of topological analyses of the signaling networks. The graph theoretical properties of the networks and the information from linear paths (obtained by pathway analysis) are discussed. Finally, in the sixth chapter of "Conclusions and Recommendations", the summary of the main results and main contributions to the research area are given with the recommendations for future work.

2. BACKGROUND ASPECTS

2.1. Cell Signaling

Signaling is the study of how cells communicate, and it impinges on all aspects of biology, from development to disease. Like all living things, all cells must continually sense their surrounding environment and make decisions on the basis of that information. Single cell organisms must be able to sense which toxic compounds and nutrients are nearby and regulate their metabolic processes accordingly. Cells in multicellular organisms such as human must sense the presence of neighboring cells and hormones when making decisions such as to whether to proliferate, move or die. The mechanism from sensing to responding requires the transfer of biological information. Signal transduction is the study of the mechanisms by which this transfer of biological information comes about. Signaling can be studied at the level of the individual cell or the whole organism. For individual cells, signaling is crucial to decisions about division, specialization, death and metabolic control. In more specialized cells it is central to immunity and the transmission of nerve impulses. At the level of whole multicellular organisms, signaling controls growth and development as well as aspects of metabolism and behavior. Not surprisingly, then, signaling malfunctions underlie many human diseases (i.e. cancer, AIDS, obesity, diabetes, asthma and infectious diseases). We can hope that in depth characterization of signaling pathways will lead eventually to an ability to intervene in diseases in which those pathways are defective and define related signaling molecules as drug targets (Downward *et al.*, 2001).

In general, a trans-membrane protein complex works as a receptor in an intracellular signaling pathway. Extracellular signaling molecule (ligand) binds to the receptor and activates the signal transmission. A cascade of intracellular signaling proteins transmits the signal to the target proteins and finally initiates a response (Fig. 2.1). The response may be alteration of metabolism if the target protein is a metabolic enzyme. On the other hand, if the target protein functions as a transcriptional regulator, the response may be changes in gene expression. Also, the signal may be transmitted to a cytoskeletal protein, which alters the shape or location of the cell.

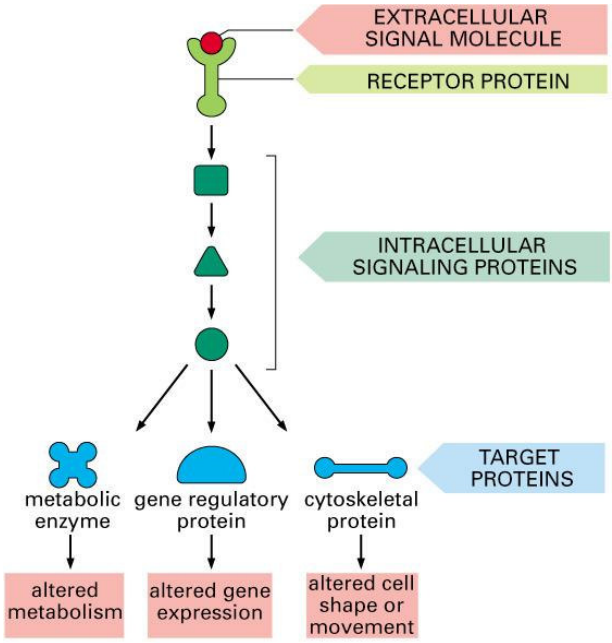


Figure 2.1. Simplified view of an intracellular signaling pathway (Alberts *et al.*, 2002)

The general organization is very similar for all intracellular signaling pathways. In Fig. 2.2, the complexity of intracellular signaling pathways is viewed in general terms. As seen in the figure, the complexity comes from the different levels of signal transmission, large number of elements present in the mechanism and the differentiation of the individual proteins with the same function.

2.1.1 Ligand

An extracellular signaling molecule is called 'ligand'. Yeast cells communicate with one another for mating by secreting several kinds of small peptides. However, in higher animals, cells communicate by means of hundreds of signaling molecules, including proteins, small peptides, amino acids, nucleotides, steroids, retinoids, fatty acid derivatives, and even dissolved gases such as nitric oxide and carbon monoxide. Most of these signaling molecules are secreted from the signaling cell by exocytosis or released by diffusion through the plasma membrane. On the other hand, some signaling molecules remain tightly bound to the cell surface and influence only cells that contact the signaling cell (Alberts *et al.*, 2002).

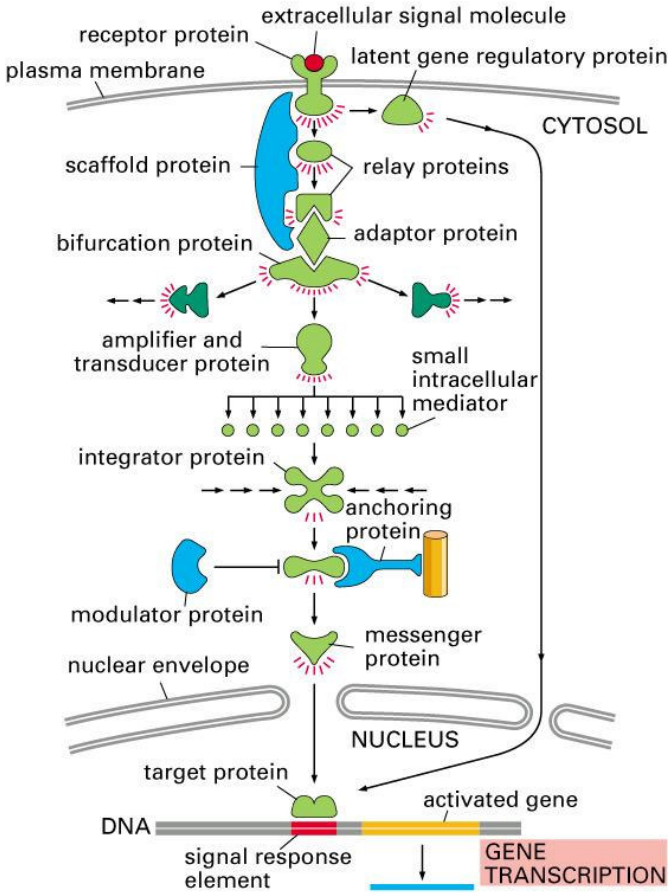


Figure 2.2. A more detailed intracellular signaling pathway (Alberts *et al.*, 2002)

2.1.2 Receptor

Regardless of the nature of the signal, the target cell responds by means of a specific protein called a receptor. It specifically binds the signaling molecule and then initiates a response in the target cell. Many of the extracellular signaling molecules act at very low concentrations (typically $< 10^{-8}$ M), and the receptors that recognize them usually bind with high affinity (affinity constant $> 10^{8}$ liters/mole). In most cases the receptors are trans-membrane proteins on the target-cell surface (called cell-surface receptors); when they bind an extracellular signaling molecule, they become activated so as to generate a cascade of intracellular signals that alter the behavior of the cell. All water-soluble signaling molecules, as well as some lipid-soluble ones, bind to specific receptor proteins on the surface of the target cells they influence. These cell surface receptor proteins act as

signal transducers: they bind the ligand with high affinity and convert this extracellular event into one or more intracellular signals that alter the behavior of the target cell.

There are three known classes of cell-surface-receptor proteins: ion-channel linked, G-protein linked, and enzyme linked. In some cases, the receptors can be inside the target cell which are called intracellular receptors. Ion-channel linked receptors are involved in rapid synaptic signaling between electrically excitable cells in which ion permeability of the plasma membrane changes. G-protein linked receptors act indirectly to regulate the activity of target protein, called a G-protein. Enzyme linked receptors either functions directly as enzymes or they are associated with enzymes. When receptors are inside the target cell, the ligand has to enter the cell to activate them. These signaling molecules therefore must be sufficiently small and hydrophobic to diffuse across the plasma membrane (Alberts *et al.*, 2002).

2.1.3 Features of Signaling

The main features of signaling are specificity, amplification, desensitization/ adaptation and integration, which are explained shortly in the following subsections.

2.1.3.1 Specificity: A signaling molecule fits specifically to the binding site on its complementary receptor (Figure 2.3); other signaling molecules do not fit. Each cell is programmed to respond to specific combinations of signaling molecules. Different cells can respond differently to the same chemical signal since a single signaling molecule often has different effects on different target cells. (Nelson and Cox, 2004)

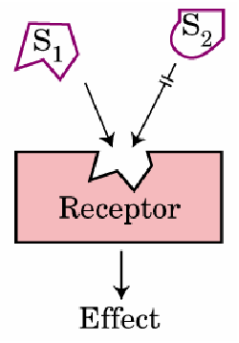


Figure 2.3. Specificity of receptor (Nelson and Cox, 2004)

2.1.3.2 Amplification: When enzymes activate enzymes, the number of affected molecules increases geometrically in an enzyme cascade (Figure 2.4) (Nelson and Cox, 2004).

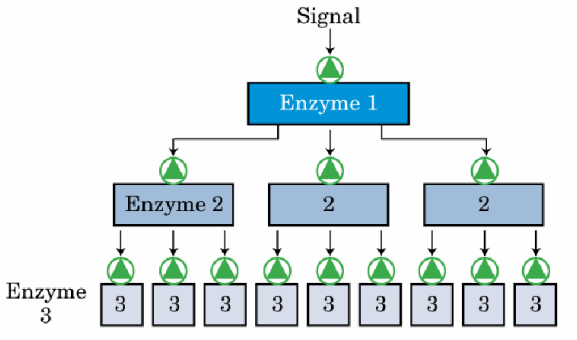


Figure 2.4. Signal amplification (Nelson and Cox, 2002)

2.1.3.3 Desensitization/Adaptation: Receptor activation triggers a feedback circuit that shuts off the receptor or removes it from the cell surface (Figure 2.5) (Nelson and Cox, 2004)

.

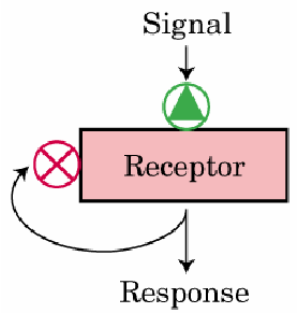


Figure2.5. Feedback of signaling (Nelson and Cox, 2004)

2.1.3.4 Integration: When two signals have opposite effects on a metabolic characteristic such as the concentration of a second messenger, or the membrane potential, the regulatory outcome results from the integrated input from both receptors (Figure 2.6) (Nelson and Cox, 2004).

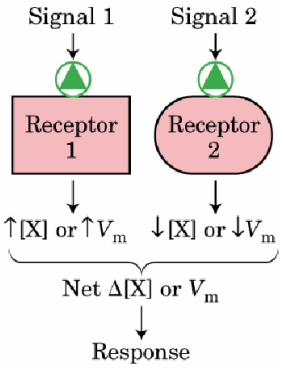


Figure 2.6. Signal integration (Nelson and Cox, 2004)

2.1.4 Types of Cell Signaling

Synaptic signaling occurs between electrically excitable cells as in the case of nerve cells. In endocrine signaling, hormones transmit the signals between cells through the blood. In paracrine signaling, a regulatory cell secretes a signaling molecule, which diffuses into another cell. The opposite case is the autocrine signaling in which a cell produces the signal by which the same cell is activated. The last form of the cell to cell signaling is by direct contacts via receptors or gap junctions on the cell membranes (Alberts *et al.*, 2002). In biological systems, signal transmission occurs mostly through two mechanisms: (i) protein-protein interactions and enzymatic reactions such as protein methylation, phosphorylation and dephosphorylation (post-translational modifications) or (ii) protein degradation or production of intracellular messengers (Bhalla and Iyengar, 1999).

There are four basic categories of signal transduction events (Figure 2.7). The classical case of a transduced signal relates a single input to a single output. Some outputs require the concatenation of multiple inputs. Other signaling interactions occur in which the transduction of a single input generates multiple outputs, a type of signal pleiotropy. Complex signaling events arise as multiple inputs trigger interacting signaling cascades that result in multiple outputs (Papin and Palsson, 2004a).

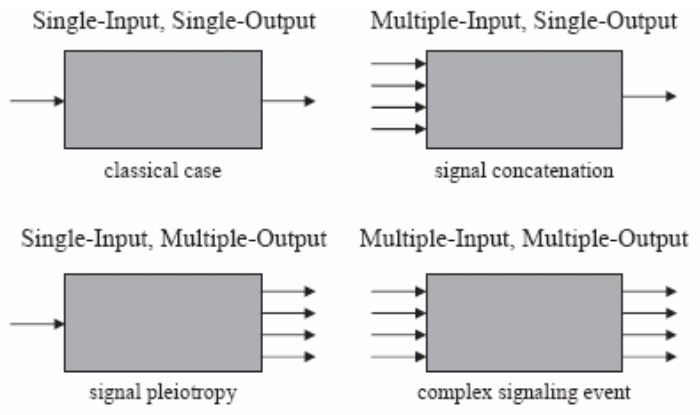


Figure 2.7. Classification of signal transduction input–output relationships (Papin and Palsson, 2004a)

2.1.5 Signaling Networks

The cell signaling mechanisms have been studied extensively over the past 20 years. The advent of high-throughput technologies is enabling the reconstruction of large scale signaling networks. After careful reconstruction of signaling networks, their properties are analyzed with mathematical modeling. The network reconstruction with high-throughput technologies and mathematical modeling to analyze network properties are two main parts of this section. The study of cellular signaling networks in transcriptional regulation, mechanotransduction, cytoskeletal organization, organelle assembly and metabolism is an active area of research in recent years (Papin *et al.*, 2005).

Recent genomic technologies have resulted in increasingly more detailed descriptions of signaling mechanisms, which have generated reconstructions of ever-larger signaling networks (Papin and Palsson, 2004a; Oda *et al.*, 2005a; Oda *et al.*, 2005b; Oda and Kitano, 2006). Such reconstructions will enable a systemic understanding of signaling network function, which is crucial for studying diseases as diverse as asthma and cancer (Finkel and Gutkind, 2003). The characterization of signaling properties that arise from whole-cell function requires integrated, mathematical descriptions of the relationships between different cellular components (Levchenko, 2003; Weng *et al.*, 1999).

2.1.5.1 Signaling Network Reconstruction: A network reconstruction includes a chemically accurate representation of all of the biochemical events that are occurring within a defined signaling network, and incorporates the interconnectivity and functional relationships that are inferred from experimental data. Network reconstructions provide the framework for the application of mathematical methods that can quantitatively describe the properties of signaling networks. Network reconstruction involves the integration of several sources of data to describe the biochemical transformations that occur in a given network. Contextual specificity is a crucial consideration in answering five questions for signaling network reconstruction: What proteins and other network components participate? What are the ligand-receptor interactions? What are the receptor-intracellular component interactions? What are the intracellular component interactions? What are the intracellular component-DNA interactions? Genome annotation, biochemical experimentation, cell-physiology characterizations, expression arrays, and other such data sources each provide different types of datum that answer these questions and contribute to the reconstruction of a given cellular signaling network (Papin *et al.*, 2005)

Signaling network reconstruction can be performed in three different ways (Fig. 2.8). The first approach consists of reconstructions of highly connected nodes in networks. Such reconstructions involve comprehensively listing the compounds and reactions that are associated with a given protein, ion or metabolite. The second approach to network reconstruction involves forming linear pathways that connect signaling inputs to signaling outputs. For example, such a pathway might be delineation of all of the steps from the binding of a growth factor to its receptor through to the subsequent activation of a transcription factor that induces the expression of target genes. The third approach consists of identifying signaling modules. Such modules historically consist of groups of compounds and proteins that function together under certain conditions. These modules have led to detailed kinetic analyses that traced the concentrations of various effector proteins and helped to understand processes such as feedback mechanisms (Papin *et al.*, 2005). For example, the EGF-receptor system has been extensively analyzed, and the effects of receptor internalization and autocrine signaling loops have been described in detail (Wiley *et al.*, 2003).

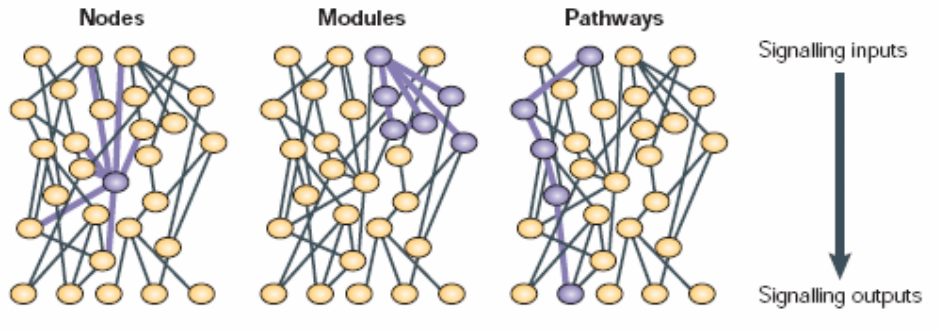


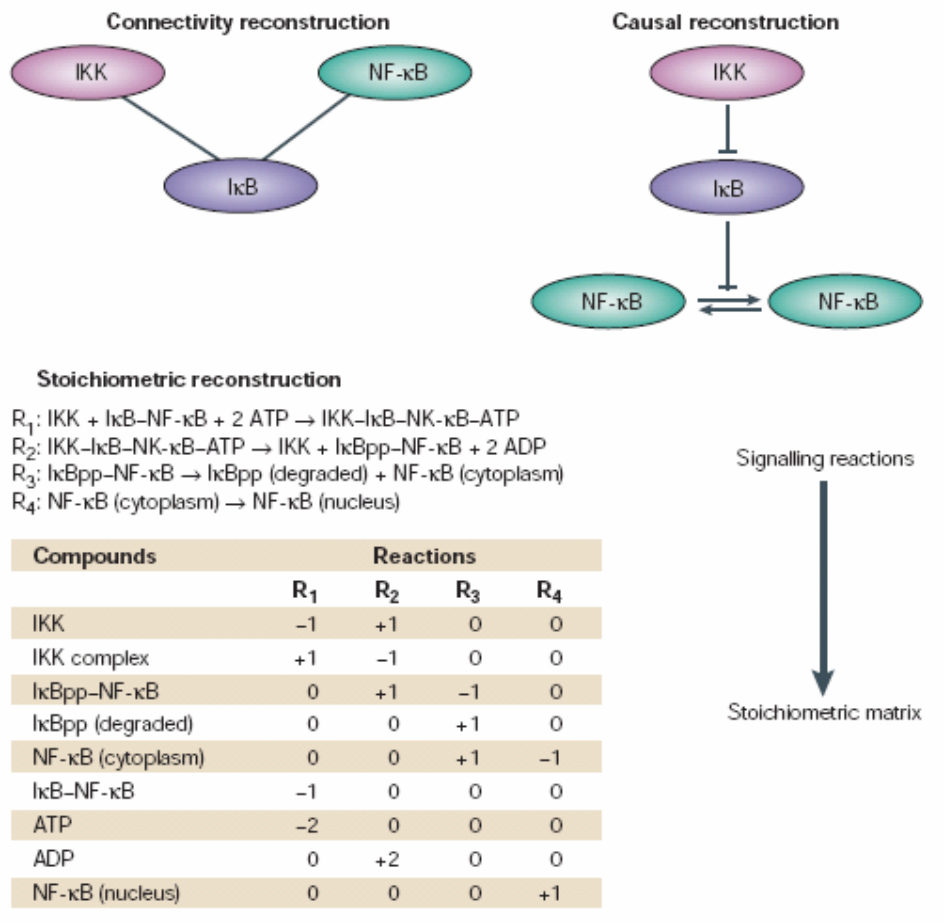
Figure 2.8. Approaches in reconstruction of signaling networks (Papin *et al.*, 2005)

Reactions amongst components in signaling networks are chemical transformations. There are three levels of resolution in reconstructions (Fig. 2.9). A connectivity reconstruction lists the associations between network components (for example, nuclear factor (NF)-κB is functionally connected to IκB kinase (IKK) through the inhibitor of NF-κB (IκB)). A more detailed causal reconstruction describes cause-and-effect relationships and is often analyzed with differential equations (for example, IKK interacts with the IκB–NF-κB complex such that NF-κB is activated). As signaling reactions are chemical transformations, they can also be represented by a more mechanistic description – for example, a stoichiometric matrix. This representation accounts for all chemical events that occur in a given network. For example, one IKK complex binds to and phosphorylates one IκB-NF-κB complex with two ATP molecules, which leads to the degradation of IκB and the nuclear localization of NF- κB (R_1-R_4). This relationship can be written out as a series of stoichiometric equations and its accompanying matrix (Papin *et al.*, 2005).

2.1.5.2 Mathematical Modeling of Signaling Networks: Large-scale signaling networks are complex. Their complexity necessitates the use of methods from systems sciences, which are quite mathematical to understand the network properties of cellular signaling. Structural and dynamic analyses measure the time-invariant/topological and the time-variant properties of a network, respectively. They can provide different results that can be integrated to characterize the properties of reconstructed signaling networks (Papin *et al.*, 2005).

Large-scale networks can undergo structural analysis in their entirety, as this does not require an extensive knowledge of the parameters that have been determined from

detailed experimentation. Structural analyses of connectivity reconstructions can generate hypotheses regarding the structure of the global network as well as the function of individual proteins. Recently published examples illustrate the analyses that have led to hypotheses concerning global, modular and individual protein function (Jeong *et al.*, 2001; Rives *et al.*, 2003; Bu *et al.*, 2003). Initial structural analyses of causal reconstructions of signaling networks have also highlighted the value of these analyses in describing network properties (Schuster *et al.*, 2000). So far, stoichiometric analyses of signaling networks are limited, owing to a lack of corresponding reconstructions. This analysis of a stoichiometric reconstruction has led to descriptions of protein synthesis requirements and energy demands of signaling networks, as well as mathematical definitions of network properties such as crosstalk and pathway redundancy (Papin *et al.*, 2005).



Compounds	Reactions			
	R_1	R_2	R_3	R_4
IKK	−1	+1	0	0
IKK complex	+1	−1	0	0
IκBpp–NF-κB	0	+1	−1	0
IκBpp (degraded)	0	0	+1	0
NF-κB (cytoplasm)	0	0	+1	−1
IκB–NF-κB	−1	0	0	0
ATP	−2	0	0	0
ADP	0	+2	0	0
NF-κB (nucleus)	0	0	0	+1

Figure 2.9. Three levels of resolution in reconstructions (Papin *et al.*, 2005)

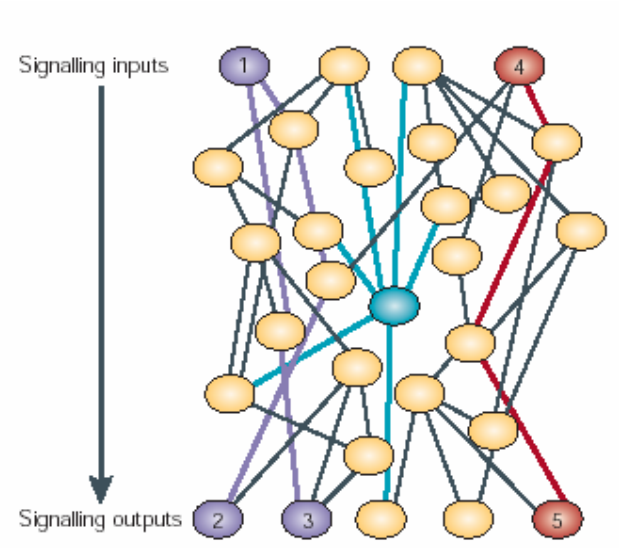


Figure 2.10. Structural analyses of signaling networks (Papin *et al.*, 2005)

Structural analyses can identify components that are well or poorly connected (and therefore of potential interest for drug targeting). For example, the green component is the most highly connected in the schematic of a signaling network (Fig. 2.10), and so drugs that inhibit the activity of this hypothetical component could have the broadest effect on the functions of the network. Structural analyses can also characterize which signaling inputs generate which signaling outputs. For example, in the schematic of signaling network (Fig. 2.10) the signaling inputs 1 and 4 can generate signaling outputs 2, 3 and 5.

A dynamic analysis of a reconstructed signaling network can be carried out once the associated kinetic parameters are known. The timescales that are associated with signaling processes can be estimated, and crudely divided into two groups: signaling activities and signaling responses. Signaling activities typically occur rapidly. For example, most protein conformational changes, kinase/phosphatase reactions (Goodman *et al.*, 1998; Vuong *et al.*, 1991), and the physical movement of signaling compounds by diffusion or cytoskeleton-dependent mechanisms (Teruel *et al.*, 2000; Theurkauf *et al.*, 1994) occur over a timeframe that ranges from fractions of a second to seconds. However, signaling responses can occur over a wider range of timescales. Signaling responses that are coupled with metabolic processes or intermediate phenotypes can occur over a timeframe of fractions of a second (Stryer, 1995; Neves *et al.*, 2002) as can elements of chemotactic and

mechanotransduction behaviour (Stryer, 1995). However, other signaling responses occur over a timescale that is an order of magnitude slower. For example, transcriptional events (Zubay, 1973; McAdams and Arkin, 1998), cellular growth (Alberts *et al.*, 2002) and receptor internalization (Bomsztyk *et al.*, 1989; Chang *et al.*, 1996; Jullien *et al.*, 2002) require several minutes, or longer, in response to a signal. This timescale separation is a crucial consideration for dynamic network analyses and can lead to simplifications that enable more thorough analyses, which would otherwise be difficult (Papin *et al.*, 2005).

As numerical values for kinetic parameters are typically difficult to obtain (Bailey *et al.*, 2001), dynamic analyses are usually only carried out for causal and stoichiometric reconstructions (Fig. 2.9) of smaller cellular signaling network reconstructions. These studies have analyzed complex network properties. The coupling of experimental data with these mathematical analyses can enable the identification of previously unknown signaling mechanisms. An elegant study that shows the benefit of integrative experimental and mathematical analyses deciphered the importance of particular IκB isoforms in feedback loops that involved the NF-κB signaling module. Predictions were experimentally verified in knockout mouse models (Hoffman *et al.*, 2002). The WNT signaling module, which is important for development as well as oncogenesis, was recently represented using an extensive set of kinetic reactions (Lee *et al.*, 2003). Predictions were made for dynamic profiles of concentrations of β-catenin and other signaling mediators, and these matched experimental results. These dynamic analyses of signaling modules show the complex properties that can be studied once the reconstruction of only a limited number of reactions have been completed and experimental data are integrated with the model predictions (Papin *et al.*, 2005).

The existing biological knowledge for a given system is composed of several types of datum (Figure 2.11). Each datum type provides unique information that can be incorporated into a chemically accurate reconstruction (for example, a stoichiometric matrix). The first type of datum includes the identification of components and endpoints of a network (for example, genome sequencing or genome-wide location analysis). The second type of datum characterizes the interactions between network components (for example, yeast two-hybrid and immunoprecipitation data identify protein–protein interactions and protein complexes). The third datum type describes the network behaviour

of the integrated components (for example, perturbation analysis and cDNA arrays delineate how entire networks function under various conditions). Each of these results provides unique types of datum that can be used to generate a cellular network reconstruction (for example, genome sequencing enables the annotation of the genes that are present in a given organism). With a network reconstruction, dynamic and structural analysis techniques can be used to describe emergent properties of the network, and generate new hypotheses. These characterizations then expand and revise the foundation of biological knowledge for the given system. This process can be iterated to offer increasingly more accurate descriptions of a given biochemical network (Papin *et al.*, 2005).

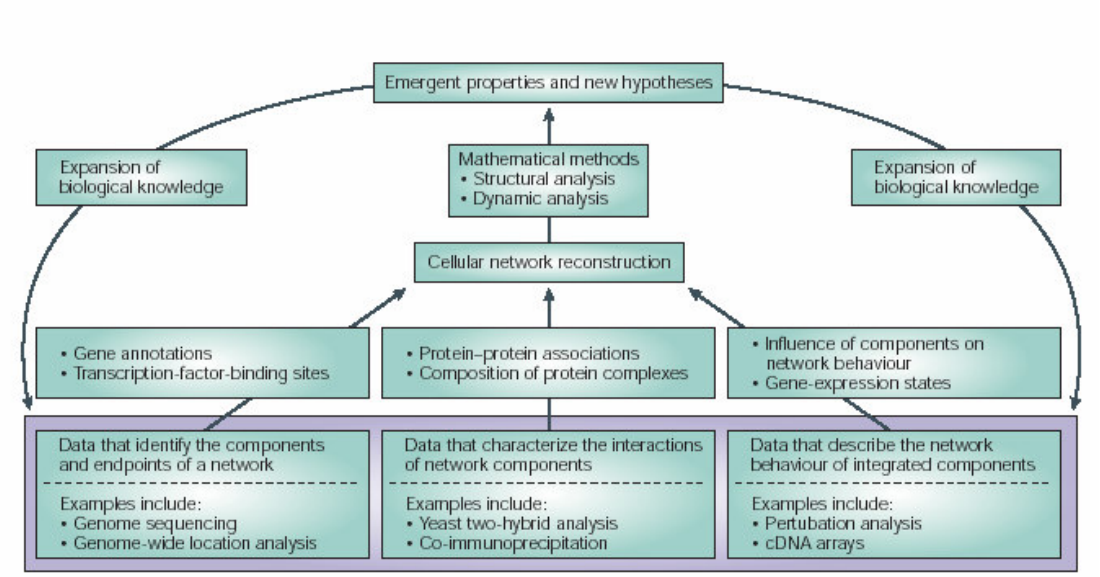


Figure 2.11. Integrative and iterative process of cellular signaling network reconstruction (Papin *et al.*, 2005)

2.2 Epidermal Growth Factor Receptor Signaling

The epidermal growth factor (EGF) receptor (EGFR) signaling pathway is one of the most important pathways that regulate growth, survival, proliferation, migration and differentiation. Reflecting this importance, it is one of the best-investigated signaling systems, both experimentally and computationally (Oda *et al.*, 2005a). Mutations in the EGFR system have been associated with cancer in humans causing uncontrolled cell proliferation as well as diseases with immunological, inflammatory, and degenerative

syndromes (Orton *et al.*, 2005). The studies of EGFR signaling system aim to identify drug targets for these diseases. Inhibitors of the EGFR are being pursued as potential cancer therapies.

The EGFR has a tyrosine kinase cytoplasmic domain, a single transmembrane domain and an extracellular domain involved in EGF binding and receptor dimerization. The proliferative effects of EGF are signaled through several pathways. Binding of EGF results in EGFR dimerization, autophosphorylation of the receptor, and tyrosine phosphorylation of other proteins. The EGFR activates Ras and the MAPK pathway, ultimately causing phosphorylation of transcription factors such as c-Fos to create AP-1 and ELK-1 that contribute to proliferation. Activation of STAT-1 and STAT-3 transcription factors by JAK kinases in response to EGF contributes to proliferative signaling. Phosphatidylinositol signaling and calcium release induced by EGF activate protein kinase C, another component of EGF signaling. Crosstalk of EGF signaling with other pathways makes the EGF receptor a junction point between signaling systems (www.biocarta.com/pathfiles/h_egfPathway.asp).

The EGFR system has been used in pioneering studies of fundamental processes such as receptor-mediated endocytosis, oncogenesis, MAPK signaling pathways, multi-receptor family interactions, autocrine loops and receptor transactivation (Carpenter, 2000). The EGFR and other members of its family have been the targets of succesful therapies against cancer and continue to be sources of discoveries about the cell signaling mechanisms involved in development, tissue homeostasis and disease (Wells, 1999; Moghal and Sternberg, 1999).

The original models of EGFR binding and internalization were derived from classic enzyme kinetics (Gex-Fabry and DeLisi, 1984; Wiley and Cunningham, 1981; Wiley and Cunningham, 1982). Receptor mediated endocytosis was modeled as an enzymatic process in which surface-associated EGF was considered to be a substrate, internalized EGF was the product and coated pits were the enzyme. These early studies demonstrated that the EGFR system behaves as a robust, proportional control system. Even though many complex processes occur in response to EGFR activation, the final output (cell responses)

is remarkably simple. Simple behavior that arise from complex processes is characteristic of robust systems (Wiley *et al.*, 2003).

It is essential, of course, that signaling models be coupled with binding and trafficking models to capture correctly the effects of receptor distribution on signaling pathway activities. The model of Schoeberl *et al.*, 2002 is the most comprehensive in that it includes the fullest range of dynamic processes and makes several interesting and significant predictions that can be validated by experimental tests. Chief among these is the prediction that the activation amplitude of the MAPK extracellular-regulated kinase (ERK) is a non-linear function of EGFR-ligand complex levels. One implication is that ERK-mediated cell responses might be surprisingly difficult to suppress by blocking EGFR ligand binding (e.g. using anti-EGFR antibodies) or inhibiting EGFR kinase activity (e.g. with small-molecule kinase inhibitors).

In any event, pursuing this kind of combined modeling and quantitative experimental approach with inclusion of binding, trafficking and signaling processes could significantly enhance efforts to identify useful drug targets (Wiley *et al.*, 2003).

2.3 Yeast Signaling

The signaling network of *Saccharomyces cerevisiae* is composed of the pheromone response pathway, the filamentous/invasive growth pathway, the high osmolarity glycerol (HOG) pathway, the cell wall integrity pathway, the TOR signaling, the autophagy/Cvt pathway, and the glucose sensing pathway.

The MAPK cascade includes critical pathways (the pheromone response pathway, the filamentous/invasive growth pathway, the HOG pathway, the cell wall integrity pathway, spore wall assembly). MAPK (Mitogen-activated protein kinases) are serine-threonine protein kinases that are activated by diverse stimuli ranging from cytokines, growth factors, neurotransmitters, hormones, cellular stress, and cell adherence. MAPKs are expressed in all eukaryotic cells. The basic assembly of MAPK pathways is a three-component module conserved from yeast to humans. The MAPK module includes three kinases that establish a sequential activation pathway comprising a MKKK (MAPK Kinase

Kinase), MKK (MAPK Kinase), and MAPK. Presently, five MAPK pathways have been well characterized in the budding yeast, *Saccharomyces cerevisiae*. The four MAPKs present in vegetative cells, Fus3, Kss1, Hog1, and Slt2/Mpk1, are involved in the mating-pheromone response, filamentation-invasion pathway, high osmolarity glycerol, and cell integrity pathway, respectively. The fifth one, Smk1, is believed to play a role in spore wall assembly (Widmann *et al.*, 1999).

The best-defined yeast MAPK pathway in *S. cerevisae* is involved in the mating of haploid cells. The haploid cells have two sexual phenotypes characterized by the expression of a set of genes involved in mating that are not expressed in diploids. The mating response to generate diploids is controlled by the α-pheromones and a-pheromones that bind to their respective receptors that are coupled to a heterotrimeric G-protein. Pheromone binding to its receptor leads to G-protein activation and the dissociation of the Beta-Gamma subunit complex from α-GTP. The *S. cerevisiae* genes whose disruption inhibited mating and caused sterility were designated as sterile (Ste) genes. The transmembrane receptors for the α- and a-factors are designated as Ste2 and Ste3, respectively, and are coupled to a heterotrimeric G-protein. Pheromone activation of the G-protein induces the dissociation of the heterotrimeric G-protein subunits designated Gpa1 (α subunit), Ste4 (β subunit), and Ste18 (γ subunit). The released β–γ subunit complex (Ste4/Ste18) activates Ste20 and interacts with the scaffolding protein Ste5, resulting in the stimulation of the MAPK module MKKK Ste11/MKK Ste7/MAPK Fus3. The activated form of Fus3 is thought to translocate to the nucleus, where it mediates pheromone induction of transcription of PRE (Pheromone Response Element) containing genes through phosphorylation and activation of at least three nuclear proteins: Dig1 (also called Rst1)/Dig2 (also called Rst2), Ste12 and Far1. These targets effect changes in gene expression and block cell cycle progression. Another physiological effect of pheromone, reoriented cellular polarity, requires a different biochemical module, Cdc42 (Cell Division Cycle-42), a p21 GTPase of the Ras superfamily. The membrane tethered, activated G-protein is thought to lead to localized activation of the GEF (Guanine nucleotide Exchange Factor) for Cdc42 and, thereby, to localized activation of Cdc42. Exchange of GDP for GTP on Cdc42 activated by Cdc24, and the hydrolysis of the Cdc42-bound GTP to GDP is predicted to be regulated by the GAPs (GTPase-Activating Proteins), Bem3 and Rga1. Once activated, Cdc42 can organize the actin cytoskeleton as it does in vegetative cells.

Cdc42 may also activate Ste20 and influence signaling through the MAP kinase cascade. Bem1, like Cdc42, interacts with several proteins important for the function of the actin cytoskeleton in polarized growth. Bem1, like Cdc42, interacts with several proteins important for the function of the actin cytoskeleton in polarized growth (Van Drogen *et al.*, 2001).

Filamentous invasive growth of *S. cerevisiae* also requires multiple elements of the MAPK signaling cascade that are also components of the mating pheromone response pathway. The MAPK cascade mediates signal transduction in filamentation-invasion pathway from two small GTP binding proteins, Ras2 and Cdc42. Signaling from Ras2 requires the 14-3-3 proteins Bmh1 (Brain modulosignalin homolog-1) and Bmh2 (Brain modulosignalin homolog-2) and possibly Sho1 receptor. Cdc42 acts downstream of Ras2 and is required for the function of the Ste20 in the filamentation-invasion pathway. Cdc42-Ste20 then transmits signal to the MAPK cascade. This cascade contains the MEKK Ste11 and the MEK Ste7. The MAPK for the filamentation-invasion pathway is Kss1. The MAPK Kss1 has a dual role in regulating filamentous invasive growth of the yeast *Saccharomyces cerevisiae*. The stimulatory function of Kss1 requires both its catalytic activity and its activation by the MEK (MAPK/ERK kinase) Ste7; in contrast, the inhibitory function of Kss1 requires neither. Unphosphorylated Kss1 binds directly to the transcription factor Ste12 and forms a protein complex that also contains Tec1, and the inhibitory proteins Dig1 or Dig2. Upon phosphorylation through a MAPK cascade, Kss1 dissociates from the complex, thereby destabilizing the Ste12-Dig association. Activated Kss1 phosphorylates and activates Ste12, leading to binding of Ste12 in combination with Tec1 to genes containing a Ste12/Tec1 composite binding site, referred to as a FRE (Filamentous and invasive growth Response Element). Fre, a combination of Tcs (TEA/ATTS consensus sequence) and Pre (Pheromone response element), mediates the binding of the heterodimer formed by the association of the transcriptional activators Tec1 and Ste12. Thus, the MAPK Kss1 plays a key role in the transcriptional control of genes regulated by Fre both by derepression and activation (Palecek *et al.*, 2002).

In addition to mating and filamentous invasive growth, yeasts respond to their environment with metabolic changes that involve MAPK pathways. For example, under conditions of high osmolarity Ste11 can lead to activation of Hog1 but does not induce

mating-specific genes. The Hog pathway is activated predominantly by two independent mechanisms that lead to the activation of either the Ssk2 and Ssk22 or the Ste11 MAPKKKs, respectively. The first mechanism involves a 'two-component' osmosensor, composed of the Sln1-Ypd1-Ssk1 proteins. The Sln1 transmembrane protein has intrinsic histidine kinase activity and is a homologue of bacterial two-component signal transducers. Using a phospho-relay mechanism involving the Ypd1 and Ssk1 proteins, Sln1 is able to control the activity of Ssk1, which in turn interacts with and regulates the Ssk2 and Ssk22 MAPKKKs and subsequent Pbs2 activation. Pbs2 activation can also be achieved by a second, independent mechanism that involves the transmembrane protein Sho1, the MAPKKK Ste11, the Ste11-binding protein Ste50, the Ste20 PAK (p21-Activated Kinase) and the small GTPase Cdc42. Activation of Pbs2 by Ste11 requires the interaction of Pbs2 with Sho1 and, although this interaction is thought to be regulated, the basic activation mechanism for this remains unclear. Once activated, Pbs2 phosphorylates and activates the Hog1 MAPK. In budding yeast, Hog1 MAPK plays a key role in global gene regulation. These osmostress-regulated genes are implicated mainly in carbohydrate metabolism, general stress protection, protein production and signal transduction. This global change in transcription could account, at least in part, for the metabolic adjustments required for osmostress adaptation. In yeast, five transcription factors are known to be controlled by the Hog1 MAPK. Hot1, Smp1, Msn2 and Msn4 activate, whereas Sko1 represses or activates, different subsets of osmotic-inducible and Hog1 regulated genes. The Hog pathway also plays a role in mediating the hyperosmotic stress-induced expression of stress response genes, recovery of cell morphogenesis, and repression of the pheromone response pathway (Tatebayashi *et al.*, 2003).

Another MAPK cascade is found in budding yeast as part of the cell integrity pathway. This pathway mediates cell cycle-regulated cell wall synthesis and responds to different signals including cell cycle regulation, growth temperature, changes in external osmolarity, and mating pheromone. The pathway is under the control of PKC. Signals activating the pathway are detected by sensors located at the cell surface, such as Slg1/Wsc1 (also called Hsc77), Wsc2, Wsc3 and Mid2. Information is then transduced via the GDP/GTP exchange factor Rom2 to the small GTPase Rho1. The latter, like all small GTPases, is considered active in its GTP-bound and inactive in its GDP-bound state. Sac7 and Lrg1 act as GAPs (GTPase-Activating Protein) for Rho1 and thus function as negative

regulators. Further GAP functions have been assigned to Bag7 and Bem2 but seem to be less crucial for Rho1 function. Similar to many other small GTPases, Rho1 has a set of different target proteins. In its GTP bound state it binds to and thereby activates the Beta-1, 3-glucan-synthase complex. It is also involved in regulation of the actin cytoskeleton by interacting with Bni1. In addition, interaction with Skn7, a regulator of oxidative stress response, has been reported. With respect to signals ensuring cellular integrity, the main effector of Rho1 is PKC1. This kinase then activates a MAP-kinase cascade consisting of the MAPKKK Bck1, the MAPKKs MKK1 and/or MKK2 and the MAPK Slt2, also referred to as Mpk1. Rlm1 and the Sbf complex (consisting of Swi4 and Swi6) have been reported as targets of the MAP kinase Slt2. Rlm1 regulates transcription of a specific set of genes. Swi4 is the DNA binding subunit and transcriptional activator of Sbf and is required for normal expression of the G1 cyclin genes Cln1, Cln2, Pcl1, and Pcl2 at the G1/S transition. Swi6 is more of a regulatory subunit, because loss of Swi6 leads to constitutive intermediate levels of Cln1 and Cln2 expression. Cln1 and Cln2 are G1 cyclins that complex with the cyclin-dependent kinase Cdc28 and thereby activate the G1/S transition (Verna *et al.*, 1997).

Another yeast MAPK, Smk1, is recently discovered, which is believed to be involved in sporulation. Upon starvation for carbon and nitrogen sources, diploid yeast cells enter meiosis to generate four haploid spores. Upon completion of meiosis, the four-haploid nuclei, which still remain within a single nuclear membrane, are enveloped by the double membranous prospore wall. The spore wall is then deposited from the space between the layers of the prospore wall. The final differentiated spore wall consists of four layers. The two inner layers appear indistinguishable from the vegetative cell wall, whereas the third layer is a spore-specific structure composed primarily of chitosan and chitin. A putative MAPK module, which employs MAPK Smk1 regulate the sporulation response. Smk1 is one of the middle sporulation genes and is therefore expressed during the latter stages of meiosis and during the time when spore wall formation occurs. The Smk1-containing spore wall assembly pathway is thus required for completion of a developmental pathway that has previously been induced. Other signaling proteins that may be part of the Smk1 pathway are two protein kinases, Sps1 and Cak1. Sps1, a MKKKK homolog, is similar in sequence to members of the PAK subfamily of protein kinases. Like Smk1, Sps1 is a middle sporulation gene and is required for proper spore

wall assembly. Whether the Smk1 pathway contains a MEKK or MEK has not yet been determined. Cak1 is an essential protein kinase that is required during vegetative growth for progression through the cell cycle. The cell cycle function of Cak1 is probably related to the ability of Cak1 to activate the cyclin dependent protein kinase Cdc28 by phosphorylation. Cak1 also appears to regulate the Smk1 pathway. Cak1 is expressed to a higher level during the same time that Smk1 is expressed. Whether Cak1 plays a supporting or instructive role on the Smk1 pathway has not yet been determined. The mechanism by which Smk1, Sps1, and Cak1 coordinate the assembly of the spore wall is unknown. Although each MAPK cascade contains a conserved set of three protein kinases, the upstream activation mechanisms for these cascades are diverse, including a trimeric G protein, monomeric small G proteins, and a prokaryotic-like two-component system. Recently, it became apparent that there is extensive sharing of signaling elements among the MAPK pathways; however, little undesirable cross-talk occurs between various cascades. The formation of multi-protein signaling complexes is probably centrally important for this insulation of individual MAPK cascades (Pierce *et al.*, 1998).

The recent finding that cell growth control, regardless of the eukaryotic organism or the physiological context, seems always to involve the same protein - the target of rapamycin TOR protein – and its namesake signaling network. TOR is a highly conserved protein kinase and the target of the immunosuppressive and anti-cancer drug rapamycin. The TOR signaling network controls cell growth by activating an array of anabolic processes including protein synthesis, transcription, and ribosome biogenesis, and by inhibiting catabolic processes such as bulk protein turnover (autophagy) and mRNA degradation, all in response to nutrients. Dysfunction of signaling pathways controlling cell growth results in cells of altered size and, in turn, causes developmental errors and a wide variety of pathological conditions. An understanding of the TOR signaling network may lead to novel drugs for the treatment of, for example, cancer, diabetes, inflammation, muscle atrophy, learning disabilities, depression, obesity, and aging (Crespo and Hall, 2002).

3. MATERIALS AND METHODS

3.1. Maps of Signaling Networks

3.1.1 Comprehensive Pathway Map of EGFR Signaling

Although there have been several computational models developed for the analysis of the EGFR signaling (Kholodenko *et al.*, 1999; Schoeberl *et al.*, 2002; Shvartsman *et al.*, 2002) a limited part of the EGFR signaling system could only be modeled (Kholodenko *et al.*, 2003). The model has been improved by incorporating various intracellular dynamics and a comprehensive pathway map for EGFR signaling (Fig. 3.1) has been constructed based on published scientific papers (Oda *et al.*, 2005a). The map includes EGFR endocytosis followed by its degradation or recycling, small guanosine triphosphatase (GTPase)-mediated signal transduction such as mitogen-activated protein kinase (MAPK) cascade, phosphatidylinositol polyphosphate (PIP) signaling, cell cycle, and G protein-coupled receptor (GPCR)-mediated EGFR transactivation via intracellular Ca^{2+} signaling. The map was created using CellDesigner (http://celldesigner.org/), a software package that enables users to describe molecular interactions using a well-defined and consistent graphical notation (Funahashi *et al.*, 2003; Kitano, 2003). The data of molecular interactions are stored in Systems Biology Markup Language (SBML; http://sbml.org/) (Hucka *et al.*, 2003). Since SBML is a standard machine-readable model representation format, all the information can be used for a range of computational analysis, including computer simulation.

The map comprises 222 reactions and 329 species. A ‘species’ is a term defined by SBML as ‘an entity that takes part in reactions’ and it is used to distinguish the different states that are caused by enzymatic modification, association, dissociation, and translocation. In the map there are clusters of reactions that are involved in specific functions, such as endocytosis, degradation, recycling of EGFR, small GTPase signaling, MAPK cascade, PIP signaling, cell cycle, Ca^{2+}signaling, and GPCR-mediated EGFR transactivation (Oda *et al.*, 2005a).

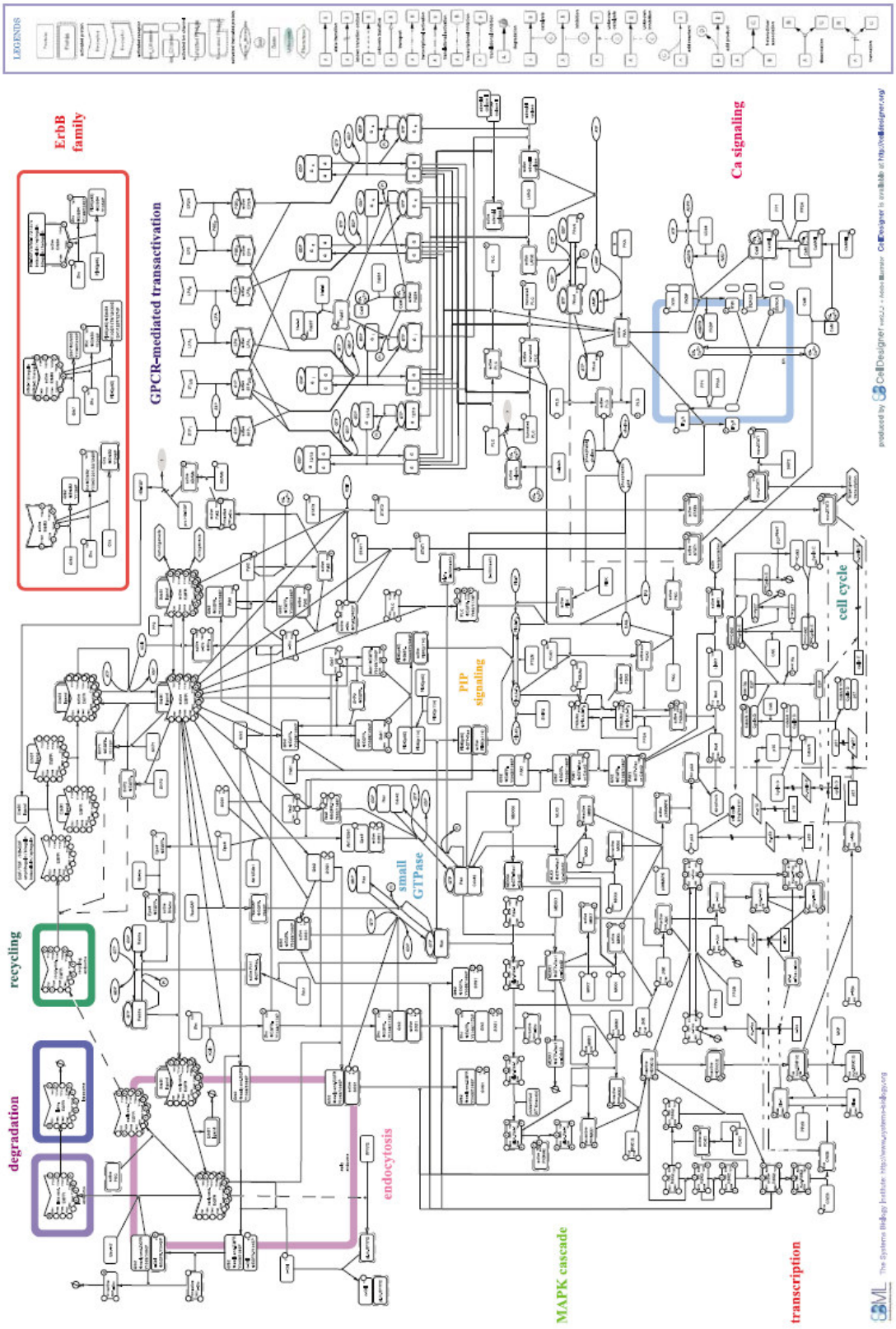


Figure 3.1. Comprehensive map of EGFR signaling network (Oda *et al.*, 2005a)

Although the EGFR map cannot yet be the basis for a dynamic simulation until a series of kinetic parameters have been identified, it can help us to understand the architectural feature of the signaling network. It can be seen from the map (Fig. 3.1) that a variety of ligands bind to corresponding subtypes of erythroblastic leukemia viral (v-erb-b) oncogene homolog (ErbB) receptors that activate molecules in an extensive network of receptor complexes, and then converge into a handful of molecules, such as nonreceptor tyrosine kinase (non-RTK), small GTPase, and PIPs, which activate a variety of cascades leading to diverse responses including transcriptional regulation (Oda *et al.*, 2005a). This architecture, also called a bow-tie (or hourglass) structure, is a characteristic feature for robust evolvable systems (Kitano, 2004). Typically, it has diverse molecules for input and output that are connected to the conserved core with highly redundant and extensively crosstalking pathways and feedback control loops in various places in the pathway (Oda *et al.*, 2005a).

Figure 3.2 illustrates the overall bow-tie structure of molecular interactions included in the EGFR map in Figure 3.1. The arrows in Figure 3.2 represent an informal notation of flow of reaction. Various ligands bind to diverse receptor heterodimers, which then converge into a handful of molecules building a conserved core. Activities of these molecules play important roles in controlling diverse responses. Notable interactions are color-coded: red, positive feedback loop; blue, negative feedback loop; purple, inhibitory feed-forward path; green, crosstalk from GPCR cascade to EGFR cascade via calcium release.

As input signals, 15 members of the endogenous EGF ligand family have been identified, that is, amphiregulin, betacellulin, biregulin, EGF, epiregulin, HB-EGF, heregulin α/β, neuregulin (NRG) 1α/1β/2α/2β/3/4, and transforming growth factor alpha (TGFα) (Jones *et al.*, 1999; Olayioye *et al.*, 2000; Yarden and Sliwkowski, 2001). While the ligands overlap with respect to binding to ErbB receptors, they have their own specificities and affinities for the respective receptors. The redundant and overlapping nature of ligand receptor binding enhances robustness in sensing the molecules in the environment, as dysfunction in one of the receptors may be compensated for by other receptors that have an affinity for the overlapping ligand molecule (Oda *et al.*, 2005a).

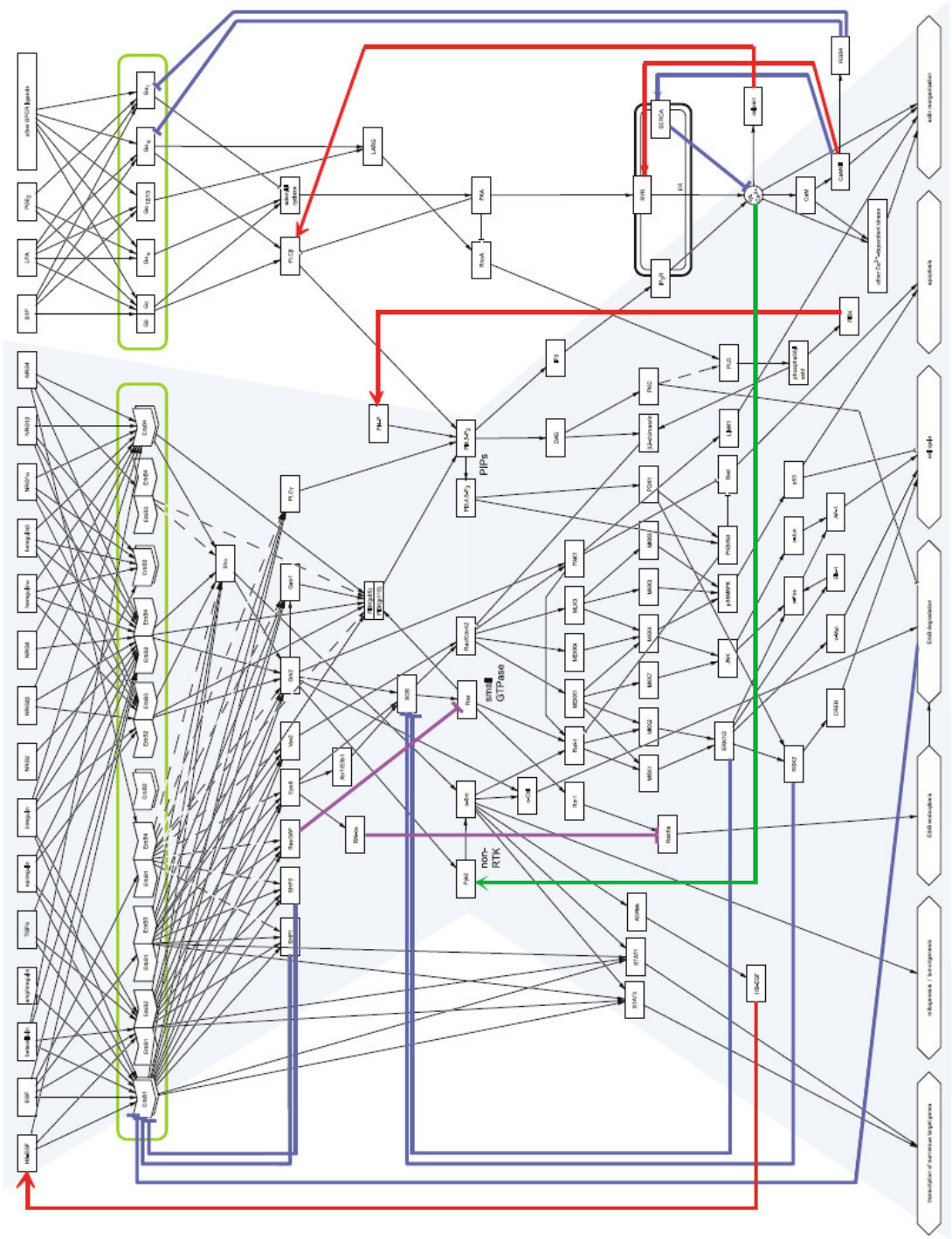


Figure 3.2. The bow-tie architecture of EGFR signaling network (Oda *et al.*, 2005a).

The binding of ligands induces homo- and heterodimerization of four ErbB family receptors: EGFR (ErbB1), ErbB2, ErbB3, and ErbB4 (Yarden and Schlessinger, 1987; Yarden and Sliwkowski, 2001). Although 10 combinations of ErbB receptor dimers are mathematically possible, only a subset of these is biologically meaningful which result in state transitions of the molecules in the network. Activities of these molecules control diverse phenotypes including apoptosis (programmed cell death), cell cycle progression, actin reorganization, target genes transcription, tumorigenesis (production of a tumor) and mitogenesis (induction of mitosis) which are cell responses (phenotypes) of EGFR signaling network (Oda *et al.*, 2005a).

There are two positive feedback loops in the ErbB bow-tie structure. Firstly, Pyk2/c-Src activates ADAMs (a disintegrin and metalloproteases), which shed pro-HB-EGF (Dikic *et al.*, 1996; Li *et al.*, 1996; Poghosyan *et al.*, 2002), so that the amount of HB-EGF will be increased and enhance the signaling. This Pyk2/c-Src mediated feedback loop is further enhanced by the Ca^{2+} mediated crosstalk from the GPCR signaling cascade (shown by a green line in Figure 3.2) (Prenzel *et al.*, 1999; Carpenter, 2000; Shi *et al.*, 2000; Schafer *et al.*, 2004). Second, active PLCβ/γ produces diacylglycerol (DAG) from PI4,5-P2, which results in the cascading activation of protein kinase C (PKC) (Mellor and Parker, 1998), phospholipase D (PLD) (Exton, 2002), and phosphatidylinositol-5-kinase (PI5K) (Moritz *et al.*, 1992). PI5K phosphorylates PI4-P resulting in an increase of PI4,5-P2.

There are six negative feedback loops. In two of these, protein tyrosine phosphatases (SHP-1 and SHP-2) inhibit EGFR at the input wing of the bow tie. In three others, a son of sevenless (SOS) homolog (Rozakis-Adcock *et al.*, 1995; Douville and Downward, 1997) is inhibited (by extracellular signal-regulated kinase (ERK) 1, ERK2, or ribosomal protein S6 kinase (RSK 2)), starting from the output wing to SOS, which localizes near the core of the bow tie. In the sixth, ErbB is degraded via the activity of Casitas B-lineage lymphoma proto-oncogene (c-Cbl), which is recruited by growth factor receptor-bound protein (Grb) 2 (Levkowitz *et al.*, 1999; Yokouchi *et al.*, 1999; Ravid *et al.*, 2004); here, feedback starts from the very end of the output wing, moving toward the initial input wing of the bow tie. In addition, a number of local inhibitory control exists that use phosphatases to control kinase activities.

There are cases where both activation and inhibition are directed to the same protein. For example, EGFR provides both positive signaling to Ras activation, and negative regulation through the recruitment of Ras GTPase-activating protein (Ras-GAP) (Agazie and Hayman, 2003). RAS-associated protein RAB5a (Rab5a) is influenced by both activation and inhibition signals from Ras interaction 1 (Rin1) (Tall *et al.,* 2001) and related to the N-terminus of tre (RN-tre) (Lanzetti *et al.*, 2000). EGFR essentially regulates both paths as it binds EGF receptor pathway substrate (Eps) 8 that activates RN-tre, and binds Grb2, which in turn stimulates Ras via SOS leading to Rin1 activation (Han *et al.*, 1997). It is interesting to note that in both cases, the length of the path for inhibition is shorter than that of activation. It is important to understand how such positive and negative controls are regulated.

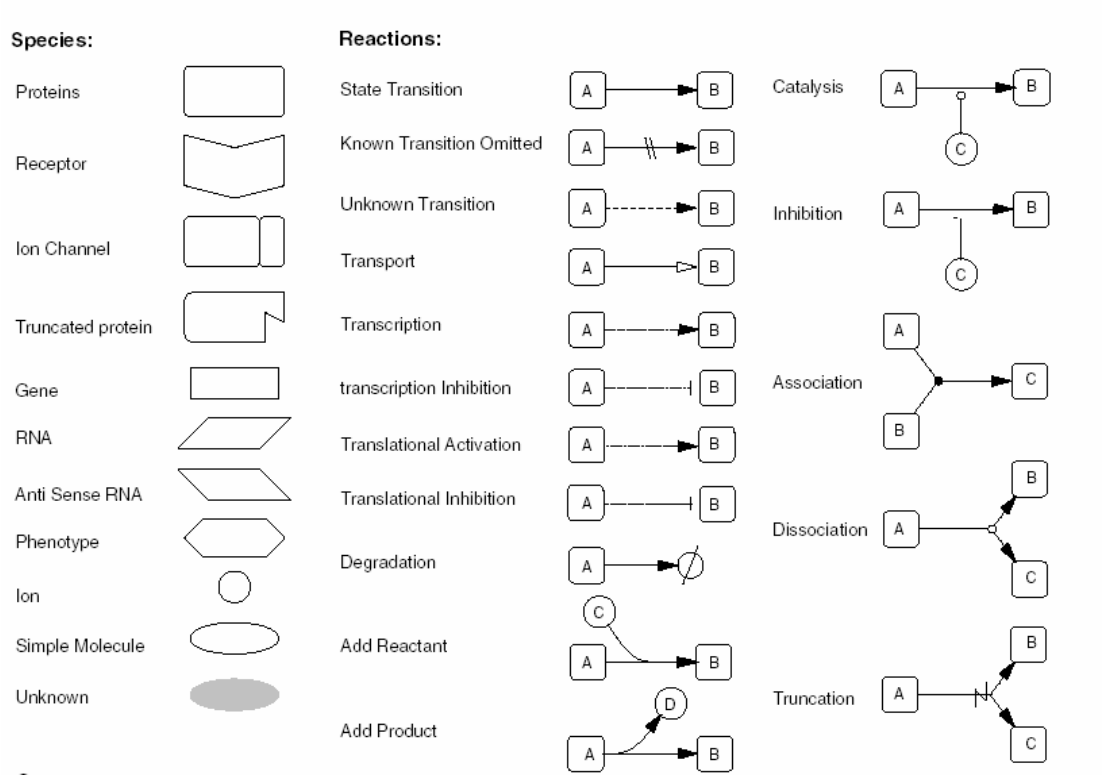


Figure 3.3. Main symbols adopted by CellDesigner ver. 2.1.1. (Oda *et al.*, 2005a)

The main symbols used to represent molecules and interactions in the map (Figure 3.1) are displayed in Figure 3.3. Kitano, (2003) proposed a graphical notation system for biological networks designed to express sufficient information in a clearly visible and

unambigious way. The EGFR map is a state transition diagram, in which one state of the system is represented in one node, and an arc from one node to another node represents a transition of the state of the system.

Figure 3.4 illustrates how the modification status of a protein is presented. Essentially, each state of a protein (i.e. phosphorylation, acetylation, etc.) can be represented such that it reflects its modification and oligomerization. The active state of the molecule is indicated by a dashed line surrounding the molecule. State changes of a component such as phosphorylation, acetylation, ubiquitination, and allosteric changes can be represented with specific information such as target residue and position. In the map, a naming convention in which the localization of protein is indicated by a prefix to the protein name, such as 'cyt.XX' and 'pl.XX' for protein XX in the cytosol and protein XX at the plasma membrane, respectively (Oda *et al.*, 2005a).

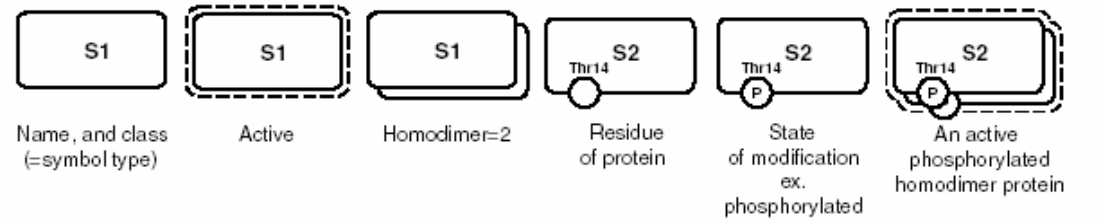


Figure 3.4. Expression of the inner structures and states in the EGFR map. (Oda *et al.*, 2005a)

3.1.2 Molecular Interaction Map of Yeast Signaling

The molecular interaction map of yeast (*Saccharomyces cerevisiae*) includes five major signal transduction and other related pathways: the pheromone response pathway, the filamentous/invasive growth pathway, the high osmolarity glycerol (HOG) pathway, the cell wall integrity pathway, the TOR signaling, the autophagy/Cvt pathway, the carbohydrate metabolism, and the glucose sensing pathway. The modules are highlighted within color boxes in Figure 3.5. (Oda *et al.*, 2005b). A recent study on the reconstruction of signal transduction network of yeast was performed by Arga *et al.* (in press). It uses a computational method (Selective Permissibility Algorithm) coupled to identify candidate proteins in signaling with protein-protein interaction data using Gene Ontology annotations (www.geneontology.org).

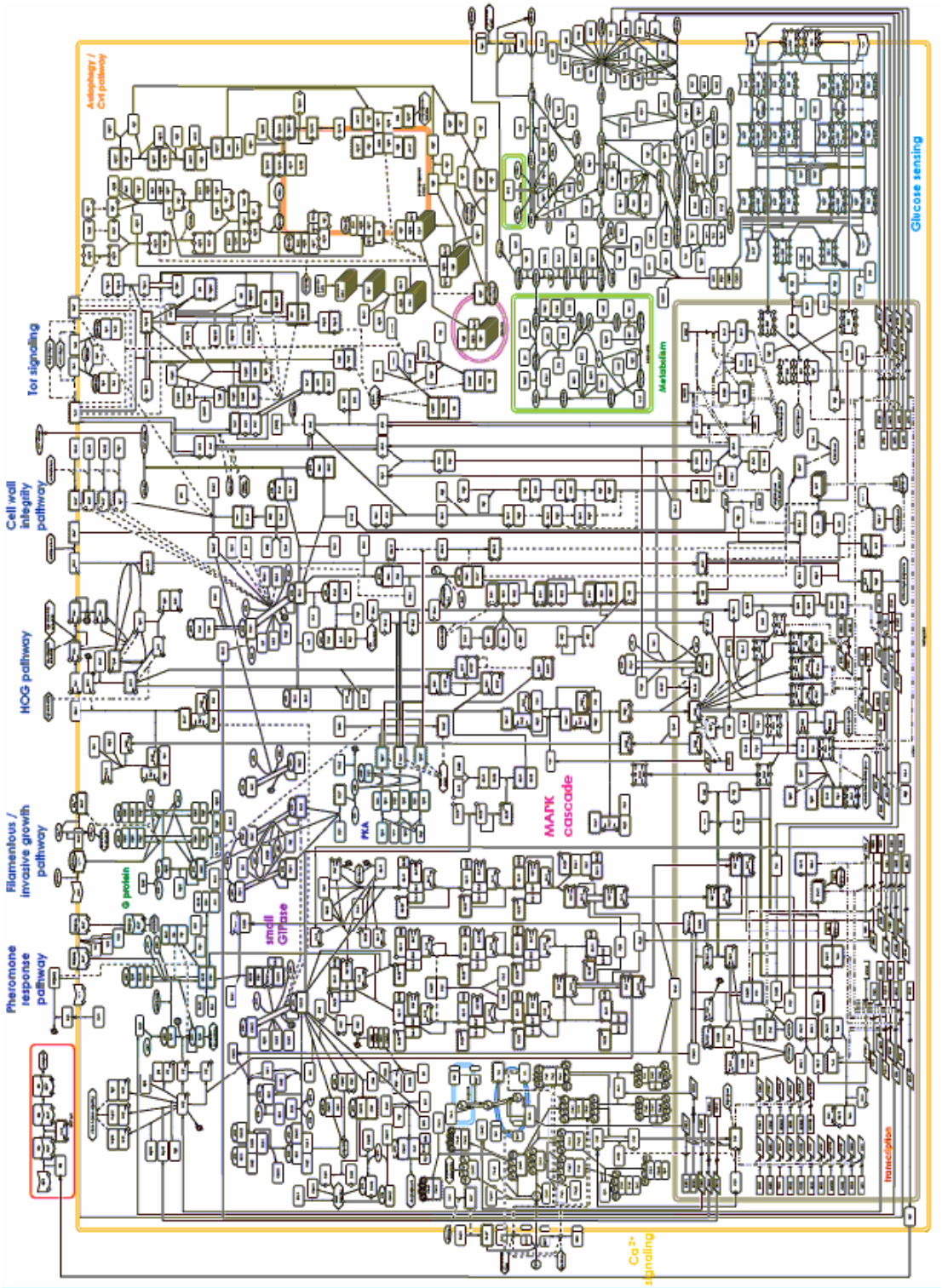


Figure 3.5. Signal transduction and carbohydrate metabolic pathways in yeast (Oda *et al.*, 2005b)

In this thesis the most comprehensive map of signal transduction mechanisms in budding yeast (Arga *et al.*, in press) was analyzed structurally by graph theoretic and pathway analysis techniques.

The analyzed reconstructed interaction map of yeast signaling (Arga *et al.*, in press) includes 1,388 proteins and 4,640 interactions among these signaling proteins. As most of the signaling networks, yeast signaling also has a bow-tie structure. Various input signals (starvation, α–pheromone, a-pheromone, high osmolority, hypotonic shock and carbon - nitrogen deficiency) activate the receptors on the plasma membrane (Sho1, Ste2, Ste3, Sln1, Wsc1,2,3, Mid2) initiating signal transduction to produce various phenotypes (filamentous invasive growth, cell cycle arrest, mating, osmoregulation, cell wall re-modeling and sporulation). The transcription factors of these phenotypes are Ste12, Tec1 (filamentous invasive growth), Fus1 (mating), Msn2, Msn4, Hot1, Sko1 (osmoregulation), Rlm1 (cell wall remodeling), Swi4, Swi6 (G1/S specific cell wall remodeling). Far1 is a final control element for cell cycle arrest and Smk1 is the known last protein in the spor wall assembly pathway.

3.2 Structural Analysis Techniques

3.2.1 Graph Theory

Studying of individual protein-protein interactions which are building blocks of biological networks enlighten the working mechanisms of cells. The first step in a graph-theory based analysis is the mathematical representation of a protein interaction network (PIN) where nodes are proteins and edges connecting them are interactions. The second step is to determine graph properties of the network using graph theoretic techniques. These properties are the degree (connectivity) of nodes, the number of hubs (highly connected nodes), the shortest path lengths between indirectly connected nodes and the key nodes (lethal proteins) in the network (Albert *et al.*, 1999; Albert and Barabasi, 2002; Wagner, 2003; Wagner and Fell, 2001; Barabasi and Albert, 1999; Barabasi and Oltvai, 2004; Strogatz, 2001; Wang, 2002; Jeong *et al.*, 2000).

Until recently, complex networks have been modeled using the classical random network theory (Erdös and Renyi, 1965; Bollobas, 1985). This model assumes that each pair of nodes (constituents) in the network is connected randomly with probability p, leading to a statistically homogeneous network in which, despite the fundamental randomness of the model, most nodes have the same number of links <k> (Figure 3.6a). In particular, the connectivity follows a Poisson distribution that peaks strongly at <k> (Figure 3.6b), implying that the probability of finding a highly connected nodes decays exponentially ($P(k) \approx e^{-k}$ for k>> <k>). On the other hand, empirical studies on the structure of the World Wide Web (Albert *et al.*, 1999; Faloutsos *et al.*, 1999), social networks (Barabasi and Albert, 1999) and scientific collaboration network (Barabasi *et al.*, 2002) have reported serious deviations from this random structure, showing that these systems are described by scale-free networks (Barabasi and Albert, 1999) (Figure 3.6c), for which P(k) follows a power-law ($P(k) \approx k^{-\gamma}$) (Figure 3.6d). Unlike exponential networks, scale-free networks are extremely heterogeneous; their topology being dominated by a few highly connected nodes (hubs), which link the rest of the less connected nodes to the system (Figure 3.6c).

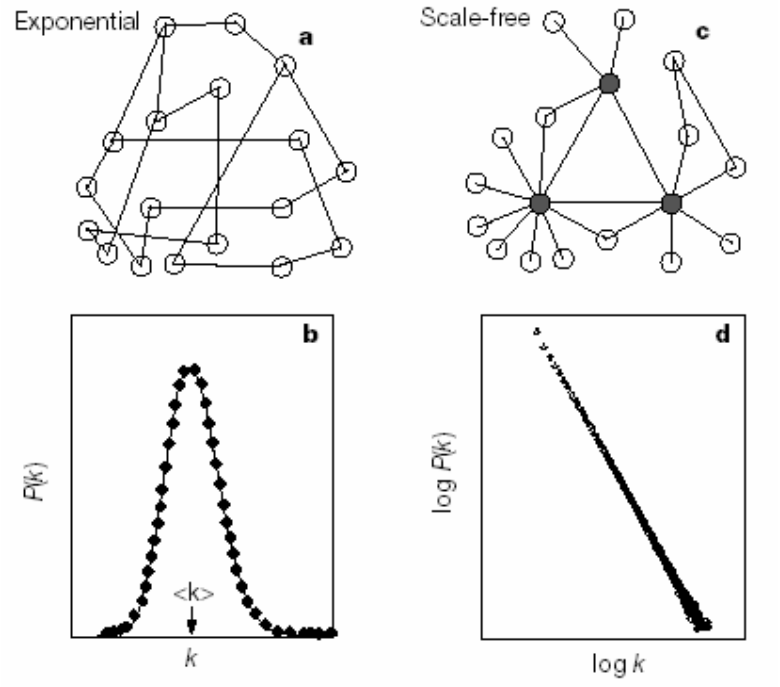


Figure 3.6. Attributes of generic network structures (Jeong *et al.*, 2000)

As the distinction between scale-free and exponential networks emerges as a result of simple dynamic principals (Amaral *et al.*, 2000; Dorogovtsev and Mendes, 2000), understanding the large-scale structure of cellular networks cannot only provide valuable and perhaps universal structural information, but could also lead to a better understanding of the dynamic processes that generated these networks. In this respect the emergence of power-law distribution is intimately linked to the growth of the network in which new nodes are preferentially attached to already established nodes (Barabasi and Albert, 1999), a property that is also thought to characterize the evolution of biological systems (Hartwell *et al.*, 1999).

In this work, the comprehensive map of EGFR signaling network (Oda *et al.*, 2005a) was converted into an undirected interaction graph by defining the reactions (state transitions of the signaling molecules) as connections between the molecules. In other words, in the graph, nodes are signaling molecules and edges are reactions in the signaling network. When two molecules are in the same reaction as reactant, product or modifier they were represented as interacting pairs (two nodes linked by an edge). The interactions are undirected meaning that when there is a connection between molecules A and B; the signal can flow both from A to B and from B to A in the interaction graph. The graph properties were determined using an algorithm (Arga *et al.*, in press) implemented in MATLAB 7.0. The input to the algorithm is the adjacency matrix (S), which is a binary, square matrix representing the edges between nodes (interactions). S(i,j)=1, if (i,j) is an interacting pair, otherwise S(i,j)=0. The constructed adjacency matrix has the dimensions of 329*329 as there are 329 signaling molecules in EGFR signaling network. If two molecules are included in the same reaction it was considered that they are interacting pairs otherwise they are not. On the other hand, undirected protein interaction graph of signaling network in yeast reconstructed by Arga *et al.* (in press) was directly used in the graph theoretic calculations. The adjacency matrix has the dimensions of 1,388*1,388.

In the algorithm used, the successive powers of the adjacency matrix are found by multiplying the matrix by itself and the element of the n^{th} power of the adjacency matrix, $S^n(i,j)$, gives the number of paths with path length of n from molecule i to molecule j. Shortest path lengths are minimum number of interactions between indirectly connected two nodes and were calculated as the minimum value of n when $S^n(i,j)$ is equal to a

nonzero value. Mean path length and network diameter were calculated as the average and maximum of the shortest path lengths in the network, respectively.

3.2.2 Metabolic Pathway Analysis

Metabolic pathway analysis (MPA) enables the screening for a number of different flux distributions or even the computation of all theoretical possible pathways within the defined metabolic network (Förster, 2003). In pathway analysis, the topology of biological networks is investigated via fundamental routes (pathways) in the networks (Schuster *et al.*, 1999). It is used to define the structure of the network and the overall metabolic capabilities of the cell (Çakır, 2006). Thus, MPA is one of the main approaches for the flux analyses of metabolic networks (Papin *et al.*, 2003; Schilling *et al.*, 1999).

There are three different approaches to find these routes. The first method constructs all possible linear paths from inputs to outputs (Seressiotis and Bailey, 1988; Mavrovouniotis *et al.*, 1990; Mavrovouniotis, 1992; Steffen *et al.*, 2002). The second approach finds a set of linearly independent basis vectors in flux space (Fell, 1990; Schilling and Palsson, 1998) and finally the concept of 'elementary flux modes' (conically independent basis vectors) was proposed (Schuster and Hilgetag, 1994; Schuster *et al.*, 1996). Any steady state flux distributions of the networks can be defined in terms of non-negative linear combination of these fundamental routes (Schuster *et al.*, 1999). The number of fundamental routes for small-world (having relatively short network diameter) large-scale networks such as EGFR and yeast signaling networks is very high and computationally unmanageable with the three above-mentioned methods. As opposed to linearly and conically independent basis vectors, the number of linear paths between inputs and outputs can be limited by specifying the maximum path length in the calculation procedure.

In order to investigate the topology of EGFR signaling network, linear paths connecting ligands to phenotypes were determined using NetSearch algorithm of Steffen *et al.* (2002). The EGFR signaling map (Oda *et al.*, 2005a) was converted into an interaction graph by representing directed interactions between reactants-products and modifiers-products of the same reaction (the direction is always towards the products of the

reactions) and that are used as the input to the NetSearch algorithm. In this representation, it was considered that there are only directed signal flows in the EGFR signal transduction network since the state transitions were defined as irreversible reactions in the map of Oda *et al.* (2005a). The signaling network in yeast was reconstructed as protein-protein interaction graph (Arga *et al.*, in press) and it was used without any modification in pathway analysis. The interactions between proteins are not directed in case of yeast signaling. In addition to this interaction (connectivity) information, inputs (ligands) and outputs (phenotypes as cellular responses) were set prior to linear paths calculation in NetSearch algorithm (Steffen *et al.*, 2002). Unlike the EGFR map, ligands and phenotypes were not defined as species in the signaling map of yeast (Arga *et al.*, in press). Therefore ligands and phenotypes cannot be specified as inputs and outputs of the system in linear path calculations for the yeast signaling network, instead receptors and transcription factors of the phenotypes are specified as the inputs and outputs, respectively.

The linear path analyses were performed to have information about phenotype frequency, network crosstalk, participation of signaling molecules in the pathways in EGFR and yeast signaling networks.

4. STRUCTURAL ANALYSIS OF EGFR SIGNALING NETWORK

Whole-network analyses are necessary to elucidate the global properties enlighten complexity of the signaling systems. Network reconstruction allows the application of mathematical methods that can quantitatively describe the properties of signaling networks. Large-scale networks can undergo only structural analysis in their entirety, as this does not require an extensive knowledge of the kinetic parameters that are determined from detailed experimentation (Papin *et al.*, 2005). In this thesis, two examples of the reconstructed large-scale signaling networks, EGFR and yeast signaling networks were structurally analyzed at whole-network level. Since the kinetic parameters of the EGFR and yeast signaling networks are incomplete yet, their dynamic analyses are not possible. Their structural analyses were here performed to generate hypotheses regarding the structure of the global networks as well as the functions of individual signaling molecules involved in these networks. Using the techniques of graph theory, the connectivity distributions of signaling molecules in the networks with values of network diameter and mean path length were obtained. By pathway analysis the information about phenotype frequency, crosstalking pathways, specific molecules and participation of molecules in the networks were obtained via linear paths connecting ligands to the phenotypes through signaling.

4.1 Graph Theoretical Analysis

The EGFR signaling network (Oda *et al.*, 2005a) was represented by an undirected interaction graph with 329 nodes and 1795 edges connecting them. Using the algorithm (Arga *et al.*, in press) mentioned in section 3.2.1 the network diameter was calculated as 11 and the mean path length as 4.7. Despite the large size of the EGFR signaling network, any two nodes in the system can be connected by relatively short paths along existing links. This means that the EGFR-mediated signal transduction network has small-world architecture that is a general feature of many complex biological networks (Watts and Strogatz, 1998). The small-world topology of the signaling network may lead to shorter transition times between different metabolic states (Wagner and Fell, 2001).

It was found that the connectivity (k, the number of links per node) distribution of the nodes in the graph has a scale-free nature (Figure 4.1 and 4.2) following nearly a power law ($P(k) \approx k^{-\gamma}$ γ=1.34 with R^2=0.63) model which is the general characteristics of biological networks (Jeong *et al.*, 2000; Jeong *et al.*, 2001) as well as World-Wide Web (Albert *et al.*, 1999; Faloutsos *et al.*, 1999), social networks (Barabasi and Albert, 1999) and scientific collaboration network (Barabasi *et al.*, 2002). It is observed in Figure 4.1 and 4.2, that the small number of nodes having only one link to the network disturbs the power law of the degree distribution causing a low correlation (R^2=0.63) of the linearity of log k vs. log P(k).

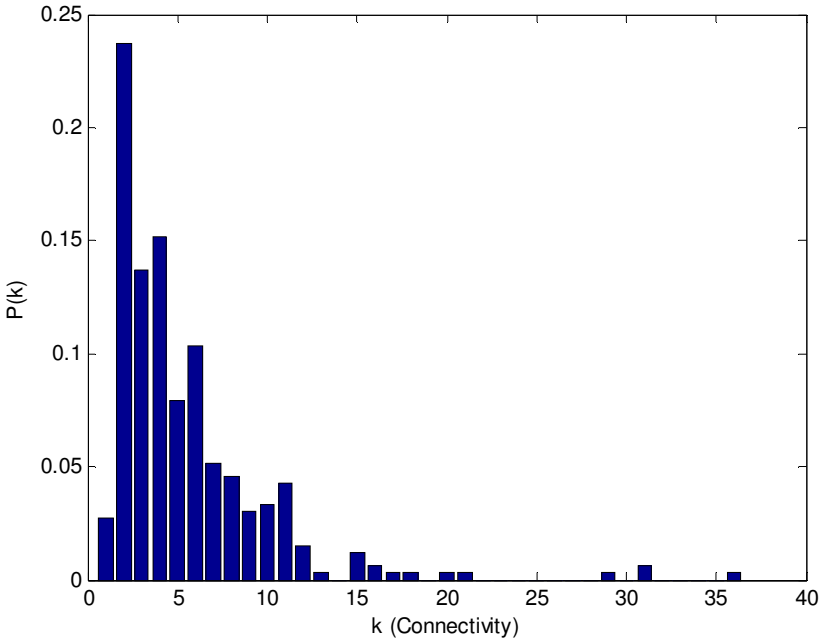


Figure 4.1. Degree (connectivity) distribution of the nodes in EGFR signaling network

The point on the y axis of the Figure 4.2 was determined as an outlier by statistical analysis via the software, essential regression ver.2.2. In the regression analysis, the residuals versus expected normal values plot gives this point. When the outlier point in Figure 4.2 is excluded, a high correlation of fitting to the power law ($P(k) \approx k^{-\gamma}$ γ=1.86 with R^2=0.84) was obtained (Figure 4.3).

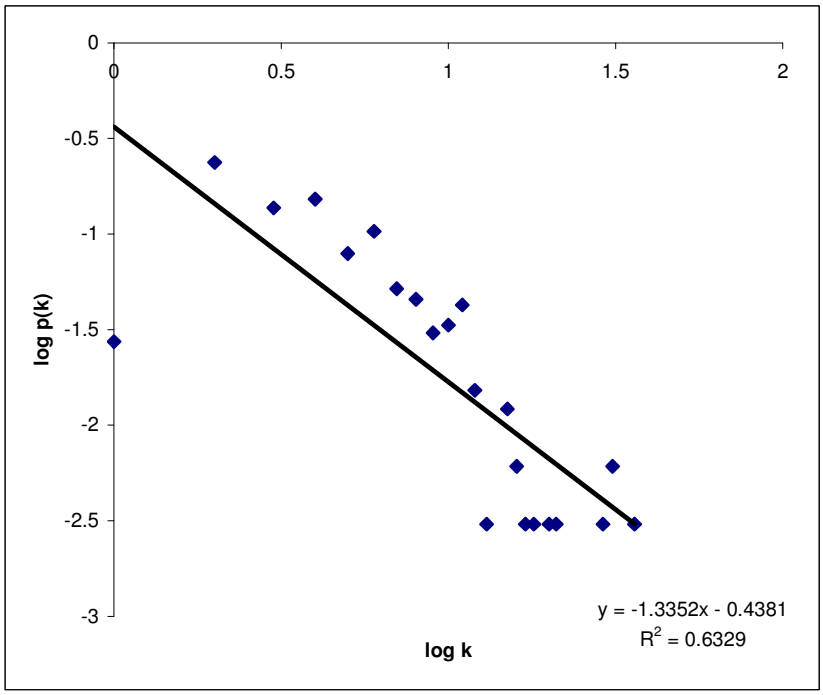


Figure 4.2. Log of connectivities versus log of P(k) (probability of having connectivity, k) in EGFR signaling network

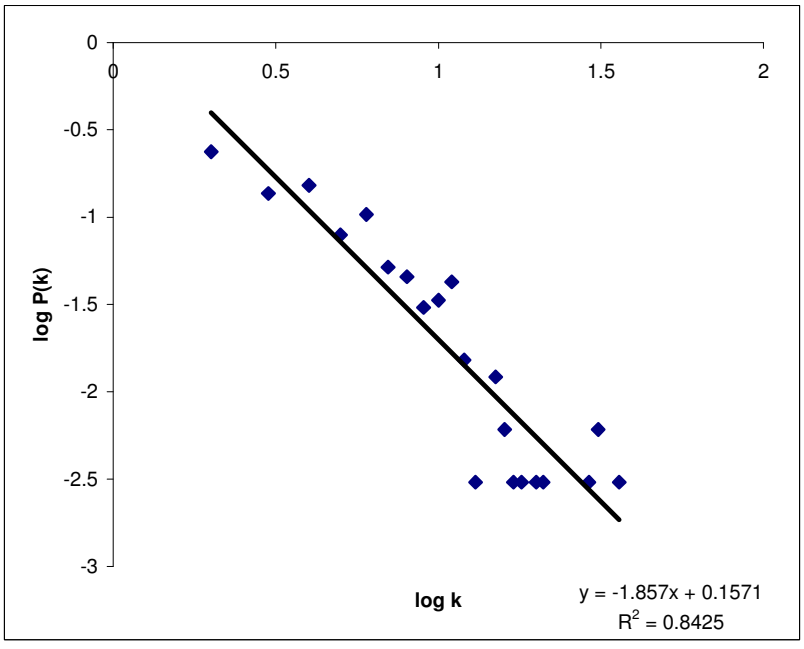


Figure 4.3. log k versus log P(k) in EGFR signaling network when the outlier is excluded

In the scale-free distribution of the degree of the nodes there are a few hubs linking less connected nodes to the signaling network. The hubs in EGFR-mediated signaling network are receptor-ligand complex (ErbB1-active EGFR), the co-factors, GTP (guanosine triphosphate) and GDP (guanosine diphosphate), and β/γ subunits of the heterotrimeric G-protein (Gβ/Gγ) having 36, 31, 31 and 29 connectivities, respectively. Any mutations in these highly connected signaling molecules may interrupt the links resulting in breaks in signal transduction as hubs are among the critical molecules in the EGFR network.

4.2 Pathway Analysis

In pathway analysis, as opposed to graph theoretical analysis, the EGFR-mediated signaling network was converted into a directed interaction graph. The molecular interactions of the network are used for a systemic analysis of input/output relationships via linear paths starting from the ligands ending at phenotypes. The NetSearch algorithm (Steffen *et al.*, 2002) gives the number of linear paths with their length (the number of interactions in each pathway).

For tumorigenesis and mitogenesis, all steps in the network are same except the last one going towards the corresponding end-phenotypes. Therefore in the pathway analysis of EGFR signaling network, tumorigenesis and mitogenesis were considered as one phenotype (tumorigenesis - mitogenesis) throughout this study.

4.2.1 Network Diameter

The shortest path lengths (Table 4.1) between the ligands and the phenotypes were calculated as described in Section 3.2.2. The maximum of the shortest path lengths gives the network diameter (14) for the phenotypes in EGFR signaling network. This means that in 14 molecular interactions (state transitions in EGFR signaling) the ligands can be linked to all phenotypes through signaling molecules. This value is higher than the network diameter of 11 obtained from graph theoretic analysis (Section 4.1). The difference in network diameters is due to the fact that the adjacency matrix (input of the graph theoretic calculations) is undirected whereas the interaction graph (input of the NetSearch

algorithm) is directed. Therefore we can conclude that the diameter of EGFR signaling network is 14 since the state transitions in the signal transduction are directed.

Table 4.1. Shortest path lengths of the phenotypes of EGFR signaling network

Phenotype	**Shortest path length**
Apoptosis	9
Cell cycle progression	14
Actin reorganization	7
Target genes transcription	7
Tumorigenesis - Mitogenesis	5

4.2.2 Phenotype Frequency

In order to obtain information about the systemic properties of the network, such as phenotype frequency, by a comparative analysis among the phenotypes, the number of linear paths were kept approximately constant while maximum path lengths were allowed to differ (Table 4.2). The NetSearch algorithm (Steffen *et al.*, 2002) allows the user to define the length of the longest linear paths in the calculation procedure.

Table 4.2. Approximately same number of linear paths with different maximum path lengths for the phenotypes of EGFR signaling network

Phenotype	**# of linear paths (~100) – max path length**	**# of linear paths (~500) -- max path length**	**# of linear paths (~1000) – max path length**
Apoptosis	89 – 12	487 – 16	993 – 18
Cell cycle progression	74 – 18	598 – 22	961 – 23
Actin reorganization	91 – 11	479 – 19	980 – 21
Target genes transcription	84 – 14	466 – 19	1,027 – 22
Tumorigenesis - Mitogenesis	118 – 18	498 – 23	1,026 – 28

The maximum path lengths are not identical for approximately same number of linear paths resulting in different phenotypes (Table 4.2). The maximum path length can be used as an indicator of frequency of the phenotype. Smaller maximum path length indicates higher frequency of the observed phenotype. One can state that apoptosis as a cell response of EGFR signaling is observed more frequently than other phenotypes because it has shorter maximum path lengths for nearly same number of linear paths. On the other hand the phenotypes of cell cycle progression and tumorigenesis - mitogenesis have longest maximum path lengths among the others meaning that they are the least frequent cellular responses of EGFR signal transduction network. For the remaining phenotypes, actin reorganization is probably more frequent than target genes transcription as it has smaller maximum path lengths.

4.2.3 Network Crosstalk

The time-invariant topological structure of the network was captured as a system of around 200 linear paths connecting ligands to each phenotype (Table 4.3). In order to comprehensively analyze the reactions and molecules included in the linear paths, the number of linear paths is constrained to 200 since this amount of paths include most of the molecules (at least once) in the EGFR signaling network and hence is expected to give statistically meaningful results for network crosstalk, molecules specific to the phenotypes and participation of molecules discussions.

Table 4.3. Around 200 linear paths of the phenotypes of EGFR signaling network with maximum path lengths

Phenotype	**# of linear paths**	**max path length**
Apoptosis	238	14
Cell cycle progression	248	20
Actin reorganization	212	15
Target genes transcription	242	17
Tumorigenesis - Mitogenesis	225	20

238 linear paths with a maximum path length of 14 are found to result in apoptosis. 248 linear paths can bind the ligands to cell cycle progression in maximum 20 interactions.

212 linear paths carry signals from extracellular signaling molecules to actin reorganization within 15 interactions. 242 linear paths with maximum path length of 17 result in target genes transcription. For tumorigenesis – mitogenesis, 225 linear paths are found within maximum 20 molecular interactions.

The pathway analysis via linear paths allows a quantitative evaluation of network crosstalk i.e. observing same species involved in linear paths leading to different phenotypes (Table 4.4). In this study, classical definition of network crosstalk (Schwartz and Baron, 1999) was accepted and identical signaling molecules used in different signaling pathways were investigated. The molecules involved in more than 50% of the linear paths were considered in each phenotype and the molecules that are shared between different pathways (i.e. cross-talking pathways) were analyzed for network crosstalk.

It was observed that the cofactor GTP and the complex Gβ/Gγ are involved in more than 70 % of the linear paths through all phenotypes in the network. All pathways in the EGFR signaling network are crosstalking through these two signaling molecules. It can therefore be stated that these molecules are crucial for the continuity of the network as their mutations may prevent most of the linear paths for all phenotypes to be completed. The linear paths resulting in apoptosis and cell cycle progression have eight common molecules that are involved in more than 50% of the paths through these phenotypes. These eight molecules were not observed in most of the linear paths of the remaining phenotypes (they are included in less than 50% or in none of the linear paths of the phenotypes other than apoptosis and cell cycle progression in EGFR signaling network). These signaling molecules are complexes Abi1 (Abscisic acid insensitivity, Eps8 binding protein)/Eps8 (EGFR pathway subsstrate)/activeSOS1 (Son of sevenless homolog), Cdc42 (Cell division cycle)/Rac (small GTP binding protein)/GTP, Cdc42/Rac/GDP, phosphatidic acid and proteins β2 chimaerin, β2 chimaerin at plasma membrane, active MKK4 (Mitogen activated kinase kinase) and gene p53 (tumor supressor). The number of shared molecules is two (simple molecule IP3 (inositol 1,4,5- trisphosphate; InsP(3) and protein IP_3R (IP3 receptor)) for actin reorganization and target genes transcription. There are five such common molecules (two different complexes of ErbB1 ligand/active EGFR, protein c-Src (a membrane-associated tyrosine kinase) and simple molecules ATP (Adenosine triphosphate) and KDI1 (selectively interfere with the EGF-induced phosphorylation of the

tyrosine residues (Buerger *et al.*,2003)) in the linear paths resulting in target genes transcription and tumorigenesis - mitogenesis.

Table 4.4. Shared molecules of crosstalking pathways of EGFR signaling network

Phenotypes	**Shared molecules**
Apoptosis Cell cycle progression	Abi1/Eps8/activeSOS1 Cdc42/Rac/GTP Cdc42/Rac/GDP phosphatidic acid β2 chimaerin β2 chimaerin at plasma membrane active MKK4 gene p53
Actin reorganization Target genes transcription	IP3 IP_3R
Target genes transcription Tumorigenesis - Mitogenesis	ErbB1 ligand/active EGFR c-Src ATP KDI1
Actin reorganization Target genes transcription Tumorigenesis - Mitogenesis	phoshorylated IP_3R Ca^{+2} in the cytosol Ca^{+2} in ER PYR
Apoptosis Cell cycle progression Tumorigenesis - Mitogenesis	DAG PLCγ at EGFR
Cell cycle progression Actin reorganization Tumorigenesis - Mitogenesis	PI4, 5-P_2

Some ternary combinations of the phenotypes have also shared molecules in their linear paths. For instance four molecules (phoshorylated IP_3R, Ca^{+2} in the cytosol, Ca^{+2} in

endoplasmic reticulum and protein PYR (Pyrimidine biosynthesis)) have roles in more than 50% of the linear paths resulting in actin reorganization, target genes transcription and tumorigenesis - mitogenesis phenotypes. Similarly, the linear paths ending with tumorigenesis - mitogenesis, apoptosis and cell cycle progression have two common molecules (lipid DAG (diacylglycerol) and protein PLCγ (Phosphatidylinositol-specific phospholipase) at EGFR). Simple molecule PI4, 5-P_2 (Phosphatidylinositol-4,5-biphosphate) is involved in more than 50% of the linear paths from ligands to cell cycle progression, actin reorganization, tumorigenesis - mitogenesis phenotypes.

4.2.4 Specific Molecules in Each Phenotype

Beyond the shared molecules in linear paths leading to different phenotypes (Table 4.4), specific non-shared molecules can also be observed in linear paths of the EGFR signaling network (Table 4.5). These specific molecules are involved in more than 50% of the linear paths of one phenotype, but do not have any role in most of the linear paths of other phenotypes.

The species that have roles in most of the linear paths resulting in apoptosis are active MKK3 (Mitogen activated kinase kinase) protein, active MKK6 protein, MAPK protein p38 (The p38 kinases were first defined in a screen for drugs inhibiting tumor necrosis factor α–mediated inflammatory responses (Lee *et al.*, 1994)), and active p38. Similarly, protein c-Jun (a typical member of the bZIP (basic zipper) family of dimeric transcriptional activators (Deng and Karin, 1992)), c-Jun at nucleus, dimer of c-Jun at nucleus, gene p53, RNA p53, protein Vav (guanosine nucleotide exchange factor that is expressed ubiquitously, as a substrate of the EGFR (Pandey *et al.*, 2000)) at EGFR, MEKK1 at GTP-Rac1 or Cdc42 protein, JNK (Jun N-terminal Kinase) protein, active MKK7 protein, active JNK, complex PGE_2 (Receptor for prostaglandin E2)/active EP_2 (PGE2 receptor type 2)/active EP_4 (PGE2 receptor type 4), complex MLK3 at Rac1 or Cdc42 and protein MKK4 are the specific molecules in the linear paths of cell cycle progression. There is no such specific molecule in the linear paths from the ligands leading to actin reorganization. There are eight such molecules for target genes transcription that

Table 4.5. Specific molecules of the phenotypes of EGFR signaling network

Apoptosis	**Cell Cycle Progression**	**Actin Reorganization**	**Target Genes Transcription**	**Tumorigenesis - Mitogenesis**
active MKK3	c-Jun	-	Pyk2	EGFR
active MKK6	c-Jun at nucleus		Grb2 at EGFR	ErbB1 ligand
p38	dimer of c-Jun at nucleus		active Pyk2	ErbB1 ligand/EGFR
active p38	gene p53		active c-Src at Pyk2	dimer of ErbB1/EGFR
	RNA p53		STAT3	active c-Src at Shc
	Vav at EGFR		active STAT3	IP3R
	MEKK1 at GTP-Rac1 or Cdc42		STAT3 at nucleus	LPA/active LPA2
	JNK		phosphorylated STAT3	early end EGFR
	active MKK7			another form of early end EGFR
	active JNK			PKC
	PGE_2/active EP_2/active EP_4			recycled end EGFR
	MLK3 at Rac1 or Cdc42			active PKC
	MKK4			active PDK1

are protein Pyk2 (Protein tyrosine kinase), complex Grb2 (Growth factor receptor bound protein) at EGFR, active Pyk2, active c-Src at Pyk2, STAT3 (signal transducer and activator of transcription) protein, active STAT3, STAT3 at nucleus and phosphorylated

STAT3. Thirteen molecules have roles in most of the linear paths resulting in tumorigenesis - mitogenesis; these are EGFR receptor protein, ErbB1 ligand protein, complex ErbB1ligand/EGFR, dimer of complex ErbB1 ligand/EGFR, active c-Src at Shc protein, IP_3R protein, complex LPA/active LPA_2, early end EGFR, another form of early end EGFR, active PKC (Protein Kinase C) protein, recycled end EGFR, PKC protein and active PDK1 (Pyruvate Dehydrogenase Kinase) protein.

4.2.5 Participation of Molecules

The linear paths from ligands to phenotypes give information about the participation of molecules in the signaling network. The role of species in linear paths can be considered as an indication of their importance in the signal transduction since any state of the signaling network is a nonnegative combination of the linear paths. One can therefore comment about the importance of each molecule for each phenotype (signaling responses of EGFR network) by its participation percentage in linear paths (Table 4.6). For instance, the simple molecule GTP is included in 84.5% of the linear paths ending with apoptosis as a signaling response of the cell. The complex Abi1/Eps8/active SOS1 has roles in 83.6% of the linear paths of apoptosis. The remaining top five molecules of the apoptosis are the complexes Cdc42/Rac/GTP, Cdc42/Rac/GDP and active MKK3 protein that are involved in 83.6%, 83.6% and 70.6% of the linear paths of the phenotype apoptosis, respectively. On the other hand, protein c-Jun (97.6%), c-Jun at the nucleus (97.6%), dimer of protein c-Jun (97.6%), gene p53 (97.6%) and RNA p53 (97.6%) are very critical since they are involved in nearly all linear paths resulting in cell cycle progression. If a deficiency occurs in any of these five molecules, 97.6% of the linear paths will be destroyed; the ligands cannot be linked to cell cycle progression and there will be very limited routes leading to this phenotype. The top five molecules for actin reorganization are GTP (92.0%), Gβ/Gγ (74.5%), active truncated PLCβ protein (70.3%), active PLCβ protein (69.3%) and protein IP_3R (67.9). For target genes transcription the key molecules that have significant roles in the linear paths of this phenotype of EGFR signaling network are the ligand-receptor complex (ErbB1/EGFR) (88.0%), protein Pyk2 (79.3%), the complex Grb2 at EGFR/active Pyk2 (79.3%), active Pyk2 (79.3%) and protein c-Src (79.3%). Finally for the phenotype tumorigenesis - mitogenesis there are six molecules (two complexes of ErbB1/EGFR, dimer of the complex ErbB1/EGFR, protein STAT1, simple molecule

KDI1, active c-Src protein at Shc and the complex ErbB1/EGFR) that are involved in all their linear paths meaning that these phenotypes are strictly dependent on these molecules.

Table 4.6. The top five molecules that participate in most of the linear paths of each phenotype in EGFR signaling network

Apoptosis	**Cell Cycle Progression**	**Actin Reorganization**	**Target Genes Transcription**	**Tumorigenesis - Mitogenesis**
GTP (84.5 %)	c-Jun (97.6%)	GTP (92.0 %)	ErbB1/EGFR (88.0 %)	two forms of ErbB1/EGFR (100 %)
Abi1/Eps8/ active SOS1 (83.6 %)	c-Jun at nucleus (97.6 %)	Gβ/Gγ (74.5 %)	Pyk (79.3 %)	dimer of ErbB1/EGFR (100 %)
Cdc42/Rac/GTP (83.6 %)	dimer of c-Jun at nucleus (97.6 %)	truncated PLCβ (70.3 %)	Grb2 at EGFR/active Pyk2 (79.3 %)	STAT1 (100 %)
Cdc42/Rac/GDP (83.6 %)	gene p53 (97.6 %)	active PLCβ (70.3 %)	active Pyk2 (79.3)	KDI1 (100 %)
active MKK3 (70.6 %)	RNA p53 (97.6 %)	IP3R (67.9 %)	c-Src (79.3 %)	active c-Src at Shc (100 %)

4.2.6 Mitogen-Activated Protein Kinase Signaling

Both EGF and NGF (Nerve Growth Factor) activate a cytosolic signal transduction pathway including the small guanine nucleotide binding protein, Ras, and a mitogen-activated protein kinase (MAPK) cascade formed by Raf, MAPK or ERK kinase (MEK) and extracellular signal-regulated kinase (ERK) protein kinases (Brightman and Fell, 2000). In this MAPK cascade, the proteins Ras, MEK and ERK are activated (Figure 4.4). In PC12 cells (rat phaeochromocytoma cell line) EGF causes transient activation whereas NGF induces sustained activation of these signaling proteins (Kaplan, 1998). The basis for

these quantitative differences was proposed as differential feedback regulation of the MAPK cascade (Brightman and Fell, 2000).

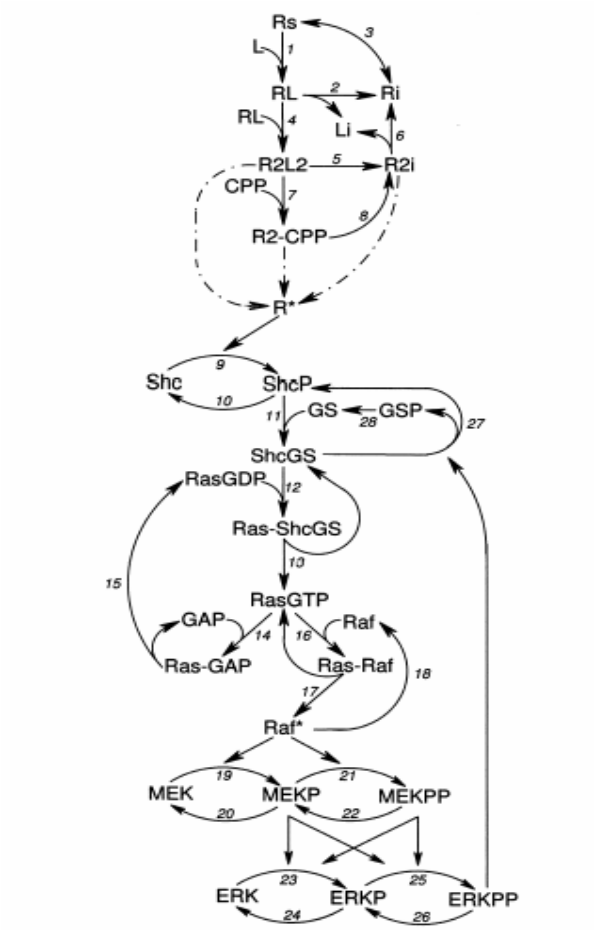


Figure 4.4. Dynamic model of EGFR signal transduction pathway (Brightman and Fell, 2000)

The differential feedback regulation of the active protein ERKPP affects (inhibits) the 27th reaction in Figure 4.4 that is dissociation of ShcGS (Shc, Src homology and collagen domain protein - GS, complex formed by Grb2, growth factor receptor binding protein and son of sevenless protein SOS) into GSP (phosphorylated GS) and ShcP (phosphorylated Shc). It means that the protein ShcGS has positive effect on the activation of the pathway. The protein Shc-Grb2-SOS in the Kitano's map was observed to participitate in linear paths going to all of the phenotypes reflecting its importance for the signal transduction in the network. Its participation percentage in the linear paths going to apoptosis is 19.3 %,

cell cycle progression is 14.5 %, actin reorganization is 5.7 %, target genes is 0.8 % and tumorigenesis-mitogenesis is 11.6%. Even though the participation in the linear paths are not so high, the protein complex Shc-Grb2-SOS is one of the few signaling molecules that are involved in linear paths from the ligands to all phenotypes of the EGFR network.

5. STRUCTURAL ANALYSIS OF YEAST SIGNALING NETWORK

Due to the huge number of necessary kinetic parameters of the large scale signaling network of yeast *Saccharomyces cerevisiae* (1,388 proteins), the dynamic investigation of the signaling mechanisms in the network is not possible for today. However the architectural characteristics of a network can give lots of information about the working mechanisms by function-structure relationships. The connectivity distribution of the proteins in the undirected interaction graph was achieved by graph theoretical calculations. The information about phenotype frequency, crosstalking pathways, proteins specific to phenotypes and protein participation in the network was obtained observing the proteins involved in the linear paths.

5.1 Graph Theoretical Analysis

The undirected interaction graph of the yeast signaling network was analyzed using the graph theoretical method described in section 3.2.1. The reconstructed signaling network includes 1,388 proteins and 4,640 interactions among these proteins. Therefore there are 1,388 nodes and 4,640 edges between the nodes.

The structural analysis of the yeast network revealed that the network has small-world properties since the distance measures (mean path length of 6.81 and network diameter of 9) are orders of magnitude smaller than the number of proteins (1,388) in the network. It was characterized by scale-free connectivity distribution since the complex structure was characterized by the presence of few highly connected proteins while the majority of the proteins have very few connections. Previous studies have also shown that the connectivity distribution of the yeast protein interaction network, as well as many other biological networks, has small-world properties (Jeong *et al.*, 2001; Maslov and Sneppen, 2002) with scale-free topology (Wagner, 2001; Jeong *et al.*, 2001). It was found that the connectivity (k, the number of links per node) distribution of the nodes in the graph has a scale-free nature (Figure 5.1 and 5.2) following nearly a power law ($P(k) \approx k^{-\gamma}$ γ=1.66 with R^2=0.93) model which is the general characteristics of biological networks (Jeong *et al.*, 2000; Jeong *et al.*, 2001).

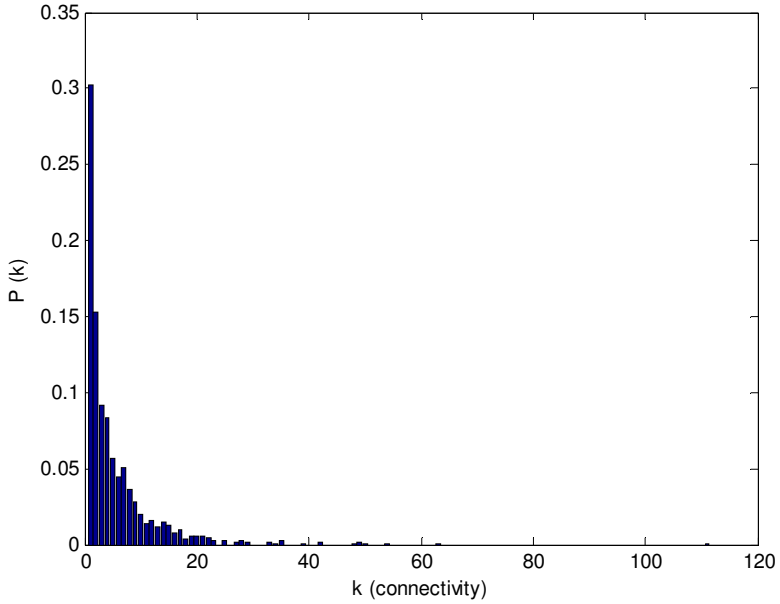


Figure 5.1. Degree (connectivity) distribution of the nodes in the yeast signaling network

The hubs in scale-free yeast signaling network are Cdc28 (Cell division cycle; catalytic subunit of the main cell cycle cyclin-dependent kinase (CDK)), Ypt6 (Yeast protein two) and Ric1 (Ribosome control) having 111, 63 and 54 interactions, respectively.

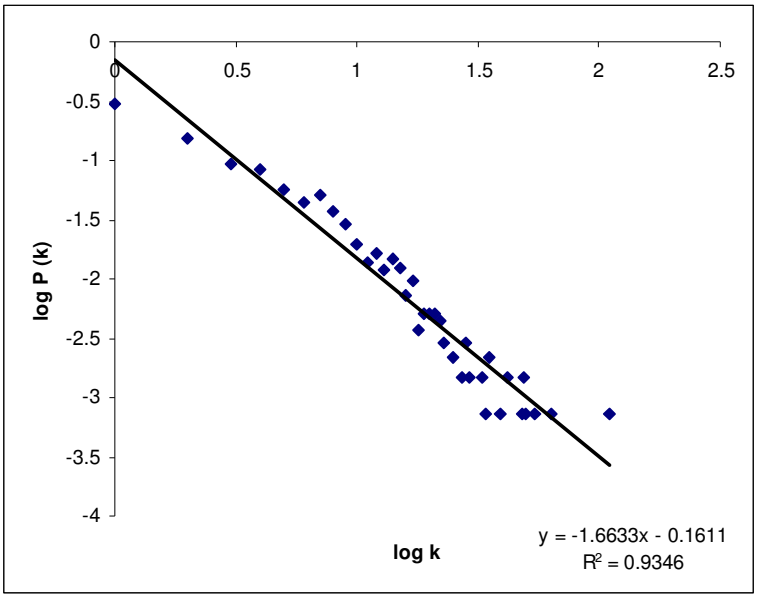


Figure 5.2. Log of connectivities versus log of P(k) (probability of having connectivity, k) in yeast signaling network

5. 2 Pathway Analysis

The bow-tie structured protein-protein interaction map of yeast signaling network (7 ligands, 7 phenotypes) was analyzed by pathway analysis to find its linear paths from ligands to phenotypes. As opposed to EGFR signaling network, the ligands (stimuli) and the pheotypes were not defined as species in the map. Therefore receptors on the plasma membrane and transcription factors were used as inputs and outputs of the system, respectively (Table 5.1).

Table 5.1. Signal transduction through the phenotypes of yeast signaling network

Stimuli	**Receptors**	**Transcription factors**	**Phenotypes**
Starvation	Sho1, Ras2*	Ste12, Tec1	Filamentous invasive growth
α-pheromone and a-pheromone	Ste2, Ste3	Far1**	Cell cycle arrest
α-pheromone and a-pheromone	Ste2, Ste3	Fus1	Mating
High osmolarity	Sho1, Sln1	Msn4, Hot1, Sko1	Osmoregulation
Hypotonic shock	Wsc3, Mid2	Rlm1	Cell wall remodeling
Hypotonic shock	Wsc3, Mid2	Swi4, Swi6	G1/S cell wall remodeling
C and N deficiency	Sps1*	Smk1**	Sporulation

The proteins indicated with (*) are intracellular signaling proteins and not receptors but they are the first known proteins in the signal transduction. The proteins indicated with (**) are not transcription factors. Far1 is a final control element regulating the gene expression to block cell cycle progression. Smk is believed to play a role in sporulation and it is the last known protein in the signal transduction. Therefore the linear paths are found

from the receptors (including Ras2 and Sps1) to these transcription factors (including Far1 and Smk1).

In a yeast cell, starvation activates the receptor Sho1 and intracellular protein Ras2 which initiate the signal transmission through the transcription factors, Ste12 and Tec1. The cell response of the signaling network is filamentous invasive growth when Ste12 and Tec1 are activated. For the second phenotype of cell cycle arrest, α-pheromone and a-pheromone bind to the receptors Ste2 and Ste3 which transmit the signal to the control element Far1. The same ligands (α-pheromone and a-pheromone) and receptors (Ste2 and Ste3) also result in a different phenotype which is mating when the transcription factor is Fus1. On the other hand, high osmolarity in the cell activates the receptors Sho1 and Sln1, and the activations of the transcription factors Msn4, Hot1 and Sko1 result in osmoregulation. Hypotonic shock to the cell affect the receptors Wsc3 and Mid2. The signals from these receptors result in cell wall remodeling when the transcription factor involved is Rlm1. The same signal results in G1/S cell wall remodeling (cell wall remodeling at the G1/S transition at the cell cycle) when the transcription factors are Swi4 and Swi6. Finally, sporulation is the response of the yeast cell in cases of carbon and nitrogen deficiencies. In these cases it is believed that the signal is transmitted from the intracellular protein Sps to another intracellular protein Smk1. With NetSearch algorithm (Steffen *et al.*, 2002) the linear paths from these starting proteins to the corresponding transcription factors or control elements were calculated.

5. 2. 1 Network Diameter

The lengths of the shortest linear paths from the receptors to the transcription factors are listed in Table 5.2 for each phenotype of yeast signaling network. The lengths of the linear paths from the ligands to the phenotypes can be obtained by adding 2 (the interactions between the ligands-receptors and transcription factors-phenotypes) to these values. The maximum of the shortest path lengths, which is 4, can be considered as the network diameter for the phenotypes of yeast signaling network. This diameter value means that the receptors can be linked to all transcription factors through 4 steps (interaction between the proteins). In other words, in 6 steps (including the interactions between ligands-receptors and transcription factors-phenotypes) all ligands can be linked

to all phenotypes. The network diameter value for the phenotypes is smaller than the value (9) obtained by the graph theoretic calculation in section 5.1 as expected. The graph theory finds the number of interactions by which any two nodes (proteins) can be linked to each other.

Table 5.2. Shortest path lengths of the phenotypes of yeast signaling network

Phenotype	**Shortest path length**
Filamentous invasive growth	3
Cell cycle arrest	3
Mating	4
Osmoregulation	3
Cell wall remodeling	2
G1/S cell wall remodeling	2
Sporulation	3

5. 2. 2 Phenotype Frequency

In analysis of EGFR signaling network, approximately same number of linear paths with different maximum path lengths for each phenotype was used for comparison of the phenotypes. However, in yeast signaling case this was not possible. When the maximum length of the paths is increased by only one the number of linear paths increases dramatically. Therefore, fixing the number of paths around any number is very difficult. Instead, another approach was used and the number of linear paths with same maximum path lengths were calculated (Table 5.3).

G1/S cell wall remodeling was found to be the most frequent phenotype as it has more linear paths than other phenotypes with same maximum path lengths. It is followed by filamentous invasive growth phenotype. Similarly, the least frequent phenotype is probably mating since the smallest number of linear paths link the ligands to this phenotype.

Table 5.3. The number of linear paths with same maximum path lengths for the phenotypes of yeast signaling network

Phenotype	# of linear paths - max path length	# of linear paths - max path length	# of linear paths - max path length	# of linear paths - max path length
Filamentous invasive growth	89 – 4	1,467 – 5	21,038 – 6	294,444 – 7
Cell cycle arrest	23 – 4	273 – 5	4,230 – 6	64,245 – 7
Mating	9 – 4	174 – 5	2,914 – 6	46,213 – 7
Osmoregulation	26 – 4	340 – 5	4,991 – 6	76,549 – 7
Cell wall remodeling	44 – 4	505 – 5	6,723 – 6	89,317 – 7
G1/S cell wall remodeling	260 – 4	3,483 – 5	46,195 – 6	608,618 – 7
Sporulation	25 – 4	426 – 5	7,443 – 6	120,562 – 7

5. 2. 3 Network Crosstalk

In order to analyze the proteins involved in the linear paths, the numbers of linear paths in Table 5.4 are chosen since these amounts of linear paths include most of the proteins in the signaling network. To get information on the crosstalking pathways, proteins specific to phenotypes and participation of proteins in the network, the proteins involved in these linear paths were analyzed comprehensively.

To determine cross-talking phenotypes , the proteins shared in linear paths leading to different phenotypes were identified using classical network crosstalk definition (Table 5.5). In this analysis the proteins involved in more than 10 % of the linear paths were considered for each phenotype.

The proteins Rvs167 (Reduced viability on starvation) and Sla1 (cytoskeletal protein binding protein) were observed to involve in the linear paths through all phenotypes except filamentous invasive growth phenotype. Since any deficiency in these proteins inhibits

the linear paths they are involved, the proteins Rvs167 and Sla1 are crucial for the signaling in yeast.

Table 5.4. The number of linear paths with the order of magitude 10^3 in yeast signaling network

Phenotype	**# of linear paths**	**max path length**
Filamentous invasive growth	1,467	5
Cell cycle arrest	4,230	6
Mating	2,914	6
Osmoregulation	4,991	6
Cell wall remodeling	6,723	6
G1/S cell wall remodeling	3,483	5
Sporulation	7,443	6

The linear paths leading to cell cycle arrest and mating phenotypes have four shared proteins that are involved in more than 10% of the linear paths. These common proteins are included in only less than 10% or in none of the linear paths leading to the remaining phenotypes. These four proteins are Ysc84 (Protein involved in the organization of the actin cytoskeleton), Akr1 (Ankyrin repeat-containing protein), Ste3 (the yeast genes whose disruption inhibited mating and caused sterility were designated as sterile (Ste) genes) and Bud6 (Bud side selection). For the phenotypes of cell wall remodeling and G1/S cell wall remodeling, there are six such common proteins, namely Smi1 (Suppresor of MAR inhibitor), Sac6 (Suppressor of actin mutations), Rom2 (GDP/GTP exchange protein (GEP) for Rho1p and Rho2p), Skt5 (activator of Chs3p (chitin synthase)), Pho85 (Phosphatase metabolism protein) and Chs5 (Chitin synthase-related protein).

Few ternary combinations of the phenotypes share some common proteins in their linear paths. The protein Cdc28 is involved in more than 10% of the linear paths leading to filamentous invasive growth, cell cycle arrest and mating phenotypes. For the phenotypes of filamentous invasive growth, mating and osmoregulation, the shared protein is Pkc1 (Protein kinase C). Similarly, for the phenotypes of cell cycle arrest, mating and osmoregulation the common protein is Las17 (actin assembly factor). The protein Bmi1 (B

lymphoma Mo-MLV insertion region) has roles in the linear paths resulting in mating, cell wall remodeling and cell wall remodeling in G1/S transition. Finally, the linear paths leading to cell wall remodeling, G1/S cell wall remodeling and sporulation share the proteins Slt2 (Mpk1, MAP kinase homolog) and Bck1 (Bypass of C kinase).

Table 5.5. Shared proteins of crosstalking pathways of yeast signaling network

Phenotypes	**Shared proteins**
Cell cycle arrest Mating	Ysc84 Akr1 Ste3 Bud6
Cell wall remodeling G1/S cell wall remodeling	Smi1 Sac6 Rom2 Skt5 Pho85 Chs5
Filamentous invasive growth Cell cycle arrest Mating	Cdc28
Filamentous invasive growth Mating Osmoregulation	Pkc1
Cell cycle arrest Mating Osmoregulation	Las17
Mating Cell wall remodeling G1/S cell wall remodeling	Bmi1
Cell wall remodeling G1/S cell wall remodeling Sporulation	Slt2 Bck1

5.2.4 Specific Proteins in Each Phenotype

The specific (non-shared) proteins were observed in the linear paths going to each phenotype. These proteins are involved in more than 10 % of the linear paths of one phenotype, but having no significant roles in others i.e. they have very little or no participation in the linear paths leading to other phenotypes. Simply it can be said that these proteins are specific to one phenotype and have a significance only for that phenotype.

Table 5.6. Specific proteins of the phenotype of yeast signaling network

Filamentous invasive growth	**Cell cycle arrest**	**Mating**	**Osmoregulation**	**Cell wall remodeling**	**G1/S cell wall remodeling**	**Sporulation**
Ste11	Cdc42	-	Vps73	-	-	MKK2
Kss1	Cdc24		Get1			Cpr6
YPR115W	Ste20		Get2			
Bem3	Sla2		Rrp14			
Dig1			Boi1			
Ste7						
Tsc11						
Bck2						

The filamentous invasive growth pathway has the largest amount of specific proteins (Table 5.6) as it has very little crosstalk with other pathways (Table 5.5). On the other hand, mating and cell wall remodeling phenotypes have no such specific proteins since they are phenotypes of highly crosstalking pathways. Especially, the cell wall remodeling and G1/S cell wall remodeling phenotypes have the largest amount of specific proteins since they are phenotypes of the same (pheromone response) pathway.

The specific proteins of filamentous invasive growth are Ste11 (Sterile; signal tranducing MEK kinase), Kss1 (Kinase suppressor of Sst2 mutations), YPR115W (Putative protein of unknown function), Bem3 (Bud emergence; Rho GTPase activating protein (RhoGAP)), Dig1 (Down-regulator of Invasive Growth), Ste7 (Sterile; signal transducing MAP kinase kinase), Tsc11 (Temperature-Sensitive Csg2 suppressor) and Bck2 (Bypass of C kinase). The proteins; Cdc42 (Cell division cycle; small rho-like GTPase), Cdc24 (Cell division cycle; Guanine nucleotide exchange factor (GEF or GDP-release factor) for Cdc42p), Ste20 (Signal transducing kinase of the PAK) and Sla2 (Synthetic lethal with Abp1) are specific to cell cycle arrest phenotype. For osmoregulation phenotype; Vps73 (Vacuolar Protein Sorting), Get1 (Golgi to ER traffic; required for the retrieval of HDEL proteins from the Golgi to the ER in), Get2 (Golgi to ER traffic; required for meiotic nuclear division and for the retrieval of HDEL proteins from the Golgi to the ER), Rrp14 (Ribosomal RNA Processing) and Boi1 (Bem1 (One) Interacting protein) are specific proteins. Sporulation phenotype has also two specific proteins that are MKK2 (Mitogen-activated Kinase Kinase) and Cpr6 (Cyclosporin-sensitive proline Rotamase).

The function of putative protein YPR115W is unkonwn and it was found specific to the pathway of filamentous invasive growth. This result can be a guide for further research to identify the role of YPR115W.

5.2.5 Participation of Proteins

The importance of each protein for each phenotype can be obtained from their participation percentages in the linear paths going to each phenotype. The most important proteins (mostly participated in the linear paths) were listed in Table 5.7.

Table 5.7. The top five proteins that participate in the linear paths of each phenotype in yeast signaling network

Filamentous invasive growth	**Cell cycle arrest**	**Mating**	**Osmoregulation**	**Cell wall remodeling**	**G1/S cell wall remodeling**	**Sporulation**
Ste11 (33.3%)	Ysc84 (78.9 %)	Ysc84 (75.6 %)	Rvs167 (33.0 %)	Smi1 (70.6 %)	Smi1 (54.4 %)	Slt2 (52.1 %)
Kss1 (24.8%)	Bem1 (63.5 %)	Bni1 (46.6 %)	Hog1 (29.4 %)	Slt2 (69.6 %)	Sac6 (33.4 %)	Abp1 (47.4 %)
YPR115W (22.5 %)	Cdc28 (23.7 %)	Cdc28 (36.9 %)	Vps73 (24.2 %)	Sac6 (43.6 %)	Slt2 (27.0 %)	Bck1 (40.2 %)
Bem3 (21.5 %)	Sla1 (23.0 %)	Bem1 (26.1 %)	Sla1 (20.8 %)	Rvs167 (21.0 %)	Cla4 (24.5 %)	Ysc84 (29.4 %)
Dig1 (16.1 %)	Rvs167 (22.2 %)	Akr1 (24.4 %)	Las17 (20.7 %)	Cla4 (20.9 %)	Rvs167 (23.5 %)	Rvs167 (28.8 %)

The most frequently observed signaling proteins in filamentous invasive growth pathway are Ste11, Kss1, YPR115W, Bem3 and Dig1 (since they are included mostly in the linear paths) which are all specific proteins of this pathway (Table 5.6). The top proteins that participate in the linear paths leading to cell cycle arrest are Ysc84, Bem1, Cdc28, Sla1 and Rvs167. For mating such proteins are Ysc84, Bni1, Cdc28, Bem1 (Bud emergence; protein containing SH3-domains) and Akr1. The proteins Rvs167, Hog1 (High osmolarity glycerol response), Vps73, Sla1 and Las17 have the greatest participation percentages in the linear paths resulting in osmoregulation. Similarly the top proteins for cell wall remodeling and G1/S cell wall remodeling are Smi1, Slt2, Sac6, Rvs167, Cla4 (Cln activity dependant) and Smi1, Sac6, Slt2, Cla4, Rvs167, respectively. Finally, the proteins Slt2, Abp1 (Actin binding protein), Bck1, Ysc84 and Rvs167 are involved in the linear paths leading to sporulation of the yeast cell with highest participation values among other known proteins in spore wall assembly pathway.

6. CONCLUSIONS AND RECOMMENDATIONS

Comprehensive structural analyses of two recently reconstructed large-scale networks of EGFR and yeast signaling were performed via graph theoretical and pathway analysis techniques. From the topological properties of the networks observed, information about connectivity distribution of the signaling molecules, phenotype frequency, crosstalking pathways, specific molecules of the phenotypes and participation of molecules in the networks were obtained.

6. 1. Conclusions

- The undirected interaction graph of EGFR signaling network with 329 nodes (molecules) and 1795 edges (interactions between the molecules) shows a small-world characteristics having a network diameter of 11 and a mean path length of 4.7.

- The connectivity distribution of the nodes follows nearly a power law with γ=1.86 ($P(k) \approx k^{-\gamma}$) due to the scale-free nature of the EGFR signaling network.

- From the shortest path lengths of the linear paths in EGFR network, the network diameter of the directed graph was obtained as 14 which is the maximum value of the shortest path lengths from the ligands through the phenotypes.

- Observing approximately same number of linear paths with different maximum path lengths allows to conclude that apoptosis is the most frequent phenotype of the EGFR signaling network since it has a smaller path length than the other phenotypes. At the opposite case, cell cycle progression, tumorigenesis and mitogenesis are the least frequent phenotypes of the EGFR signaling network.

- Investigation of the molecules involved in the linear paths from the ligands to the phenotypes showed that all pathways in EGFR network are crosstalking with each other via the cofactor GTP and the complex Gβ/Gγ since they are involved in more than 70 % of the linear paths through all phenotypes in the network. On the other hand a high crosstalk

was observed between pathways resulting with apoptosis and cell cycle progression; actin reorganization and target genes transcription; target genes transcription and tumorigenesis-mitogenesis.

- With the help of the linear paths the specific molecules of the phenotypes in the EGFR signaling were identified. The phenotype actin reorganization has no such specific molecule as its pathway shares its signaling molecules with other pathways by crosstalking. On the other hand six specific molecules were observed to be involved all linear paths through the phenotype tumorigenesis-mitogenesis. Any deficiency in these six molecules inhibits the cell response of tumorigenesis and mitogenesis (low robustness).

- The undirected protein-protein interaction graph (1,388 nodes and 4,640 edges) of the signaling mechanisms in yeast (*Saccharomyces cerevisiae*) was analyzed using graph theory and a small world property was found with network diameter of 9 and mean path length of 6.81which are relatively small in the large scale signaling network.

- The connectivity distribution of the nodes (proteins) can be modeled by a power law with γ=1.66 ($P(k) \approx k^{-\gamma}$) showing the scale-free nature of the yeast signaling network.

- From the number of the linear paths (from the receptors to the transcription factors) with the same maximum path lengths, it was concluded that the phenotype G1/S cell wall remodeling is the most frequent phenotype having higher number of linear paths compared to other pheotypes. The opposite case is mating which is the least frequent phenotype with the smallest number of the linear paths.

- An important conclusion about the network crosstalk in the signaling network of yeast is that there is a very high crosstalk among the phenotypes as the most of the signaling proteins are shared among the linear paths through the phenoytpes. For instance, the proteins Rvs167 and Sla1 were observed to be involved in the linear paths leading to all phenotypes with high participation values. As another evidence to the high network crosstalk is that there is no specific proteins for three phenotypes, namely mating, cell wall remodeling and G1/S cell wall remodeling.

6.2. Recommendations

This study aims to clarify the unknown topological features of the EGFR and yeast signaling networks. The results indicate the key points (crucial signaling molecules) in the signal transduction mechanisms. These may be considered as starting points for further researchs.

- Graph theory is a useful guide for experimental studies on protein interactions. The hubs identified with graph theoretical calculation in the EGFR and yeast signaling networks are good targets for further experimental studies.

- The six molecules that have roles in all linear paths through the tumorigenesis and mitogenesis phenotypes should be investigated thoroughly in cancer studies since tumorigenesis is the production of tumor and mitogenesis is the induction of mitosis.

- On the other hand a protein with unknown function (YPR115W) is found to participate in filamentous invasive growth pathway in yeast signaling network. This result can be used as a starting point for defining the function of this protein.

REFERENCES

Agazie, Y. M. and M. J. Hayman, 2003, "Molecular mechanism for a role of SHP2 in epidermal growth factor receptor signaling", *Moll. Cell. Biol.*, Vol. 23, pp. 7875-7886.

Albert, B., 1998, "The cell as a collection of protein machines: preparing the next generation of moleculat biologists", *Cell*, Vol. 92, pp. 291-294.

Albert, R., H. Jeong, A. L. Barabasi, 1999, "Diameter of the world-wide web", *Nature*, Vol. 401, pp. 130-131.

Albert, R., H. Jeong, A. L. Barabasi, 2000, "Error and attack tolerance of complex networks", *Nature*, Vol. 406, pp. 378-382.

Albert, R. and A. L. Barabasi, 2002, "Statistical mechanistic of complex networks", *Review of Modern Physics*, Vol. 74, pp. 47-97.

Alberts, B., A. Johnson, J. Lewis, M. Raff, K. Roberts, P. Walter, 2002, *Molecular Biology of the Cell*, 4th edition, Ch.15, Garland Science, New York.

Alessi, D. R., Y. Saito, D. G. Campbell, P. Cohen, G. Sithanandam, U. Rapp, A. Ashworth, C. J. Marshall, S. Cowley, 1994, "Identification of the sites in MAP kinase kinase-1 phosphorylated by p74raf-1", *EMBO J.*, Vol. 13, pp. 1610-1619.

Amaral, L. A. N., A. Scala, M. Barthelemy, H. E. Stanley, 2000, "Classes of behavior of small-world networks" <http://xxx.lanl.gov/abs/cond-mat/0001458>.

Anderson, N. G., J. L. Maller, N. K. Tonks, T. W. Sturgill, 1990, "Requirements for integration of signals from two distinct phosphorylation pathways for activation of MAP kinase", *Nature*, Vol. 343, pp. 651-653.

Arga, K. Y., Z. İ. Önsan, B. Kırdar, K. Ö. Ülgen, J. Nielsen, in press, "Understanding signaling in yeast: Insights from network analysis", *Biotechnology and Bioengineering.*

Bailey, J. E., 2001, "Complex biology with no parameters", *Nature Biotechnol.*, Vol. 19, pp. 503–504.

Barabasi, A. and R. Albert, 1999, "Emergence of scaling in random networks", *Science*, Vol. 286, pp. 509-512.

Barabasi, A. L. and Z. N. Oltvai, 2004, "Network biology: Understanding the cell's functional organization", *Nature Reviews Genetics*, Vol. 5, pp. 101-113, February.

Barabasi A. L., H. Jeong, Z. Neda, E. Revasz, A. Schubert, T. Vicsek, 2002, "Evolution of the social network of scientific collaborations" *Physica A.* Vol. 311, pp. 590-614.

Barthelemy, M. and L. A. N. Amaral, 1999, "Small-world networks: Evidence for a crossover picture", *Phys. Rev. Lett.*, Vol. 82, pp. 3180-3183.

Bhalla, U. S. and R. Iyengar, 1999, "Emergent properties of networks of biological signaling pathways" *Science*, Vol. 283, No. 5400, pp. 381-387, January.

Bollobas, B., 1985, *Random Graphs*, Academic, London.

Bomsztyk, K., T. H. Stanton, L. L. Smith, N. A. Rachie, S. K Dower, 1989, "Properties of interleukin-1 and interferon-γ receptors in B lymphoid cell line", *J. Biol. Chem.*, Vol. 264, 6052–6057.

Brightman, F.A. and D. A. Fell, 2000, "Differential feedback regulation of the MAPK cascade underlies the quantitative differences in EGF and NGF signalling in PC12 cells", *FEBS Letters*, Vol. 482, pp. 169-174.

Bu, D., Y. Zhao, L. Cai, H. Xue, X. Zhu, H. Lu, J. Zhang, S. Sun, L. Ling, N. Zhang, G. Li, R. Chen, 2003, "Topological structure analysis of the protein–protein interaction network in budding yeast", *Nucleic Acids Res.*, Vol. 31, pp. 2443–2450.

Buerger, C., K. Nagel-Wolfrum, C. Kunz, I. Wittig, K. Butz, F. Hoppe-Seyler, B. Groner, 2003, "Sequence-specific peptide aptamers, interacting with the intracellular domain of the epidermal growth factor receptor, interfere with Stat3 activation and inhibit the growth of tumor cells", *The Journal of Biological Chemistry*, Vol.278, pp. 37610-37621, September.

Canals, F., 1992, " Signal transmission by epidermal growth factor receptor: coincidence of activation and dimerization", *Biochemistry*, Vol. 31, pp. 4493-4501.

Carpenter, G., 2000, "The EGF receptor: a nexus for trafficking and signaling", *BioEssays*, Vol. 22, pp. 697–707.

Chang, D. Z., Z. Wu, T. L. Ciardelli, 1996, "A point mutation in interleukin-2 that alters ligand internalization", *J. Biol. Chem.*, Vol. 271, pp. 13349–13355.

Chang, C. P., J. P. Y. Kao, C. S. Lazar, B. J. Walsh, A. Wells, H. S. Wiley, G. N. Gill, M. G. Rosenfeld, 1991, "Ligand-induced internalization and increased cell calcium are mediated via distinct structural elements in the carboxyl terminus of the epidermal growth factor receptor" *J. Biol. Chem.*, Vol. 266, pp. 23467–23470.

Chang, C. P., C. S. Lazar, B. J. Walsh, M. Komuro, J. F. Collawn, L. A. Kuhn, J. A. Tainer, I. S. Trowbridge, M. G. Farquhar, M. G. Rosenfeld, H. S. Wiley, G. N. Gill, 1993, "Ligand-induced internalization of the epidermal growth factor receptor is mediated by multiple endocytic codes analogous to the tyrosine motif found in constitutively internalized receptors" *J. Biol. Chem.*, Vol. 268, pp. 19312–19320.

Crespo, J. L. and M. N. Hall, 2002, "Elucidating TOR signaling and rapamycin action lessons from *Saccharomyces cerevisiae*", *Microbiology and Molecular Biology Reviews*, Vol. 66, No. 4, pp. 579-591, December.

Çakır, T., 2006, *Stoichiometric models in metabolic systems biology of yeast*, Ph.D Thesis, Boğaziçi University.

Deng, T. and M. Karin, 1992, "Construction and expression of a monomeric c-Jun protein that binds and activates transcription of Ap-1-responsive genes", *PNAS*, Vol. 89, pp. 8572-8576.

Dikic, I., G. Tokiwa, S. Lev, S. A. Courtneidge, J. Schlessinger, 1996, "A role for Pyk2 and Src in linking G-protein-coupled receptors with MAP kinase activation" *Nature*, Vol. 383, pp. 547–550.

Douville, E. and J. Downward, 1997, "EGF induced SOS phosphorylation in PC12 cells involves P90 RSK-2", *Oncogene*, Vol. 15, pp. 373–383.

Dorogovtsev, S. N. and J. F. F. Mendes, 2000, *Evolution of reference networks with aging*, http://xxx.lanl.gov/abs/cond-mat/0001419

Downward, J., 2001, "The ins and outs of signaling", *Nature*, Vol. 411, No. 14, pp. 759-762, June.

Downward, J., 1994, "The GRB2/Sem-5 adaptor protein", FEBS Lett., Vol. 338, pp. 113-117.

Downward, J., 1996, "Control of Ras activation", *Cancer Surv.*, Vol. 27, pp. 87-100.

Erdös, P. and A. Renyi, 1960, "On the evolution of random graphs", *Publ. Math. Inst. Hung. Acad. Sci.*, Vol. 5, pp. 17-61.

Exton, J. H., 2002, "Regulation of phospholipase", D. *FEBS Lett.*, Vol. 531, pp. 58–61.

Faloutsos, M., P. Faloutsos, C. Faloutsos, 1999, "On power-law relationships of the interner topology", *Comp. Comm. Rev.*, Vol. 29, pp. 251.

Fell, D. A., 1990, "Substrates cycles. Theoretical aspects of their role in metabolism" *Comments Theor. Biol.*, Vol. 6, pp. 1–14.

Finkel, T. and J. S. Gutkind, 2003, *Signal transduction and human disease*, Wiley–Liss, Hoboken, New Jersey, USA.

Förster, J., 2003, *Pathway analysis of the metabolic network of Saccharomyces cerevisiae*, Ph.D Thesis, BioCentrum-DTU, Denmark.

Funahashi A., M. Morohashi, N. Tanimura, H. Kitano, 2003, "Cell-Designer: a process diagram editor for gene-regulatory and biochemical networks", *BioSilico*, Vol. 1, pp. 159-162.

Kitano H., 2003, "A graphical notation for biochemical networks", *BioSilico*, Vol. 1, pp. 169–176.

Gex-Fabry, M. and C. DeLisi, 1984, "Receptor-mediated endocytosis: a model and its implications for experimental analysis", *Am. J. Physiol.*, Vol. 247, R768–R779.

Goh K. L., J. Hiller, J. L. Haston, D. F. Holmes, K. E. Kadler, A. Murdoch, J. R. Meakin, T. J. Wess, 2005, "Analysis of collagen fibril diameter distribution in connective tissues using small-angle x-ray scattering" *Biochim. Biophys. Acta* Vol. 1722, pp. 183–188.

Goodman, O. B. Jr., J. G. Krupnick, F. Santini, V. V. Gurevich, R. B. Penn, A. W. Gagnon, J. H. Keen, J. L. Benovic, 1998, "Role of arrestins in G-protein-coupled receptor endocytosis", *Adv. Pharmacol.*, Vol. 42, pp. 429–433.

Han, L., D. Wong, A. Dhaka, D. Afar, M. White, W. Xie, H. Herschman, O. Witte, J. Colicelli, 1997, "Protein binding and signaling properties of RIN1 suggest a unique effector function", *Proc Natl Acad Sci USA*, Vol. 94, pp. 4954–4959.

Hartwell, L. H., J. J. Hopfield, S. Leibler, A. W. Murray, 1999, "From molecular to modular cell biology", *Nature*, Vol. 402, C47-52.

Heinrich, R. and T. A. Rapoport, 1974, "A linear steady-state treatment of enzymatic chains, General properties, control and effector strength", *European Journal of Biochemistry*, Vol. 42, pp. 89-95.

Higgins, J., 1963, "Analysis of sequential reactions", *Ann. N. Y. Acad. Sci.*, Vol. 108, pp. 305-321.

Higgins, J., 1965, "Dynamics and control in cellular systems", pp. 13–46 in Control of Energy Metabolism (edited by B. Chance, R. W. Estabrook and J. R. Williamson), Academic Press, New York.

Hoffmann, A., A. Levchenko, M. L. Scott, D. Baltimore, 2002, "The IêB–NF-êB signaling module: temporal control and selective gene activation" *Science*, Vol. 298, pp. 1241–1245.

Hucka M, A. Finney, H. M. Sauro, H. Bolouri, J. C. Doyle, H. Kitano, A. P. Arkin, B. J. Bornstein, D. Bray, A. Cornish-Bowden, A. A. Cuellar, S. Dronov, E. D. Gilles, M. Ginkel, V. Gor, I. I. Goryanin, W. J. Hedley, T. C. Hodgman, J. H. Hofmeyr, P. J. Hunter, N. S. Juty, J. L. Kasberger, A. Kremling, U. Kummer, N. Le Novere, L. M. Loew, D. Lucio, P. Mendes, E. Minch, E. D. Mjolsness, Y. Nakayama, M. R. Nelson, P. F. Nielsen, T. Sakurada, J. C. Schaff, B. E. Shapiro, T. S. Shimizu, H. D. Spence, J. Stelling, K. Takahashi, M. Tomita, J. Wagner, J. Wang, 2003, "The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models", *Bioinformatics*, Vol. 19, pp. 524–531.

Jaiswal, R. K., E. Weissinger, W. Kolch, G. E. Landreth, 1996, "Nerve growth factor mediated activation of the mitogen-activated protein (MAP) kinase cascade involves a signaling complex containing B-Raf and HSP90", *J. Biol. Chem.*, Vol. 271, pp. 23626-23629.

Jeong, H., B. Tombor, R. Albert, Z. N. Oltvai, A. L. Barabasi, 2000, "The large-scale organization of metabolic networks", *Nature*, Vol. 407, pp. 651-654.

Jeong, H., S. P. Mason, A. L. Barabasi, Z. N. Oltvai, 2001, "Lethality and centrality in protein networks", *Nature*, Vol. 411, pp. 41–42.

Jones, J. T, R. W. Akita, M. X. Sliwkowski, 1999, "Binding specificities and affinities of egf domains for ErbB receptors", *FEBS Lett.*, Vol. 447, pp. 227–231.

Jullien, J., V. Guili, L. F. Reichardt, B. B. Rudkin, 2002, "Molecular kinetics of nerve growth factor receptor trafficking and activation", *J. Biol. Chem.*, Vol. 277, pp. 38700–38708, June.

Kacser, H. and J. A. Burns, 1973, "The control of flux", *Symp. Soc. Exp. Biol.,* Vol. 27, pp. 65-104.

Kaplan, D.R., 1998, "Studying signal transduction in neuronal cells: the Trk/NGF system" Prog. *Brain Res.*, Vol. 117, pp. 35-46.

Kholodenko B. N., O. V. Demin, G. Moehren, J. B. Hoek, 1999, "Quantification of short term signaling by the epidermal growth factor receptor", *J. Biol. Chem.*, Vol. 274, pp. 30169–30181.

Kholodenko, B. N., 2003, "Four-dimensional organization of protein kinase signaling network signaling cascades: the roles of diffusion, endocytosis and molecular motors", *J. Exp. Biol.* Vol. 206, pp. 2073–2082.

Kitano, H., 2003, "A graphical notation for biochemical networks ", *BioSilico*, Vol. 1, pp. 169–176.

Kitano, H., 2004, "Biological robustness", *Nat. Rev. Genet.*, Vol. 5, pp. 826–837.

Lanzetti, L., V. Rybin, M. G. Malabarba, S. Christoforidis, G. Scita, M. Zerial, P. P. Di Fiore, 2000, "The Eps8 protein coordinates EGF receptor signalling through Rac and trafficking through Rab5", *Nature*, Vol. 408, pp. 374–377.

Lee, J. C., J. T. Laydon, P. C. Mcdonnell, T. F. Gallagher, S. Kumar, D. Green, D. Mcnulty, M. J. Blumenthal, J. R. Keys, S. W. L: Vatter, J. E. Strickler, M. M. Mclaughlin, I. V. Siemens, S. M. Fisher, G. P. Livi, J. R. White, J. L. Adams, P. R. Young, 1994, "A protein kinase involved in the regulation of inflammatory cytokine biosynthesis", *Nature*, Vol. 372, pp.739-746, December.

Lee, E., A. Salic, R. Kruger, R. Heinrich, M. W. Kirschner, 2003, "The roles of APC and Axin derived from experimental and theoretical analysis of the Wnt pathway", *PLoS Biol.*, Vol. 1, pp. 116–132.

Lemmon, M. A., Z. Bu, J. E. Ladbury, M. Zhou, D. Pinchasi, I. Lax, 1997, "Two EGF molecules contribute additively to stabilization of the EGFR dimer", *EMBO J.*, Vol. 16, pp. 281-294.

Levchenko, A., 2003, "Dynamical and integrative cell signaling: challenges for the new biology", *Biotechnol. Bioeng.*, Vol. 84, pp. 773–782.

Lehninger, *Principles of Biochemistry*, 3rd ed., Ch.13

Levkowitz G., H. Waterman, S. A. Ettenberg, M. Katz, A. Y. Tsygankov, I. Alroy, S. Lavi, K. Iwai, Y. Reiss, A. Ciechanover, S. Lipkowitz, Y. Yarden, 1999, "Ubiquitin ligase activity and tyrosine phosphorylation underlie suppression of growth factor signaling by c-Cbl/Sli-1", *Mol Cell*, Vol. 4, pp. 1029–1040.

Lund, K. A., L. K. Opresko, C. Starbuck, B. J. Walsh, H. S. Wiley, 1990, "Quantitative analysis of the endocytic system involved in hormone-induced receptor internalization", *J. Biol. Chem.*, Vol. 265, No. 26, pp. 15713-15723.

Marais, R., Y. Light, H. F. Paterson, C. S. Mason, C. J. Marshall, 1997, "Differential regulation of Raf-1, A-Raf, and B-Raf by oncogenic ras and tyrosine kinases", *J. Biol. Chem.*, Vol.272, pp. 4378-4383.

Maslov, S. and K. Sneppen, 2002, "Specificity and stability in topology of protein networks", *Science*, Vol. 296, pp. 910-913.

Mavrovouniotis, M. L., G. Stephanopoulos, G. Stephanopoulos, 1990, "Computer-aided synthesis of biochemical pathways", *Biotechnol. Bioeng.* Vol. 36, pp. 1119–1132.

Mavrovouniotis, M. L., 1992, "Synthesis of reaction mechanisms consisting of reversible and irreversible steps. Formalization and analysis of the synthesis algorithm", *Ind. Eng. Chem. Res.*, Vol. 31, pp. 1637–1653.

McAdams, H. H. and A. Arkin, 1998, "Simulation of prokaryotic genetic circuits", *Annu. Rev. Biophys. Biomol. Struct.*, Vol. 27, pp. 199–224.

McCormick, F., 1989, "Ras GTPase activating protein: signal transmitter and signal terminator", *Cell*, Vol. 56, pp. 5-8.

Mellor, H., P. J. Parker, 1998, "The extended protein kinase C superfamily", *Biochem. J.*, Vol. 332, pp. 281–292.

Millward, T.A., S. Zolnierowicz, B. A. Hemmings, 1999, "Regulation of protein kinase cascades by protein phosphatese 2A", *Trends Biochem. Sci.*, Vol. 24, pp. 186-191.

Moghal, N. and P. W. Sternberg, 1999, "Multiple positive and negative regulators of signaling by the EGF-receptor", *Curr. Opin. Cell Biol.*, Vol. 11, pp. 190–196.

Moritz, A., P. N. De Graan, W. H. Gispen, K. W. Wirtz, 1992, "Phosphatidic acid is a specific activator of phosphatidylinositol-4-phosphate kinase", *J. Biol. Chem.*, Vol. 267, pp. 7207–7210.

Neves, S. R., P. T. Ram, R. G. Iyengar, 2002, "Protein pathways", *Science*, Vol. 296, pp. 1636–1639.

Oda, K., Y. Matsuoka, A. Funashashi, H. Kitano, 2005a, "Comprehensive pathway map of epidermal growth factor receptor signaling", *Molecular Systems Biology*, 4100014-E1-E17, May.

Oda, K., H. Moriya, Y. Matsuoka, H. Kitano, 2005b, *A Comprehensive Molecular Interaction Map of Budding Yeast*, http://celldesigner.org/download/YeastMapPoster.pdf

Oda, K. and H. Kitano, 2006, "A comprehensive map of the toll-like receptor signaling network", *Molecular Systems Biology*, 4100057.

Olayioye, M. A., R. M. Neve, H. A. Lane, N. E. Hynes, 2000, "The ErbB signaling network: receptor heterodimerization in development and cancer", *EMBO J.*, Vol. 19, pp. 3159–3167.

Opresko, L. K., C. P. Chang, 1995, "Endocytosis and lysosomal targeting of epidermal growth factor receptors are mediated by distinct sequences independent of the tyrosine kinase domain" *J. Biol. Chem.*, Vol. 270, pp. 4325–4333.

Orton, R. J., O. E. Sturm, V. Vyshemirsky, M. Calder, D. R. Gilbert, W. Kolch, 2005, "Computational modelling of the receptor-tyrosine-activated MAPK pathway", *Biochem. J.* Vol. 392, pp. 249-261.

Pandey, A., A. V. Podtelejnikov, B. Blagoev, X. R. Bustelo, M. Mann, H. F. Lodish, 2000, "Analysis of receptor signaling pathways by mass spectrometry: identification of

vav-2 as a substrate of the epidermal and platelet-derived growth factor receptors", *Proc. Natl. Acad. Sci. U S A.*, Vol. 97, No. 1, pp. 179-184, January.

Papin, J. A., N. D. Price, S. W. Wiback, D. A. Fell and B. O. Palsson, 2003, "Metabolic pathways in the post-genome era", *Trends in Biochemical Sciences*, Vol. 28**,** pp. 250-258.

Papin, J. A. and B. O. Palsson, 2004a, "Topological analysis of mass-balanced signaling networks: a framework to obtain network properties including crosstalk", *Journal of Theoretical Biology*, Vol. 227, No. 2, pp. 283-297, March.

Papin, J. A. and B. O. Palsson, 2004b, "The JAK–STAT signaling network in the human B-cell: an extreme signaling pathway analysis", *Biophys. J.*, Vol. 87, No. 1, pp. 37–46, July.

Papin, J. A., T. Hunter, B. O. Palsson, S. Subramaniam, 2005, "Reconstruction of signaling networks and analysis of their properties", *Nature Rewievs Molecular Cell Biology*, Vol. 6, No. 2, pp. 99-111, February.

Pelicci, G., L. Lanfrancone, F. Grignani, J. McGlade, F. Cavallo, G. Forni, I. Nicoletti, F. Grignani, T. Pawson, P. G. Pelicci, 1992, "A novel transforming protein (SHC) with an SH2 domain is implicated in mitogenic signal transduction", *Cell*, Vol. 70, pp. 93-104.

Pierce, M., M. Wagner, J. Xie, V. Gaulis-Durner, J. Six, A. K. Vershon, E. Winter, 1998, *Mol. Cell. Biol.*, Vol.18, No. 10, pp. 5970-5980, October.

Rives, A. W. and T. Galitski, 2003, "Modular organization of cellular networks", *Proc. Natl Acad. Sci.*, Vol. 100, pp. 1128–1133.

Palecek, S. P., A. S. Parikh, S. J. Kron, 2002, "Sensing signalling and integrating physical processes during Saccharomyces cerevisiae invasive and filamentous growth", M*icrobiology*, Vol. 148, No. 4, pp. 893-907, April.

Poghosyan, Z., S. M. Robbins, M. D. Houslay, A. Webster, G. Murphy, D. R. Edwards, 2002, "Phosphorylation-dependent interactions between ADAM15 cytoplasmic domain and Src family proteintyrosine kinases", *J. Biol. Chem.*, Vol. 277, pp. 4999–5007.

Prenzel, N., E. Zwick, H. Daub, M. Leserer, R. Abraham, C. Wallasch, A. Ullrich, 1999, "EGF receptor transactivation by G-protein-coupled receptors requires metalloproteinase cleavage of proHB-EGF", *Nature*, Vol. 402, pp. 884–888.

Przulj, N., D. A. Wigle, I. Jurisica, 2004, "Functional topology in a network of protein interactions", *Bioinformatics*, Vol. 20, pp. 340-348.

Ravid, T., J. M. Heidinger, P. Gee, E. M. Khan, T. Goldkorn, 2004, "c-Cbl-mediated ubiquitinylation is required for epidermal growth factor receptor exit from the early endosomes", *J. Biol. Chem.*, Vol. 279, pp. 37153–37162.

Rozakis-Adcock, M., R. Fernley, J. Wade, T. Pawson, D. Bowtell, 1993, "The SH2 and SH3 domains of mammalian Grb2 couple the EGF receptor to the Ras activator mSos1", *Nature*, Vol. 363, pp. 83–85.

Schafer, B., B. Marg, A. Gschwind, A. Ullrich, 2004, "Distinct ADAM metalloproteinases regulate G protein-coupled receptorinduced cell proliferation and survival", *J. Biol. Chem.*, Vol. 279, pp. 47929–47938.

Schilling, C. H. and B. O. Palsson, 1998, "The underlying pathway structure of biochemical reaction networks" *Proc. Natl. Acad. Sci.U. S. A.*, Vol. 95, pp. 4193–4198.

Schilling, C. H., S. Schuster, B. O. Palsson and R. Heinrich, 1999, "Metabolic pathway analysis: Basic concepts and scientific applications in the post-genomic era", *Biotechnology Progress*, Vol. 15, pp. 296-303.

Schoeberl B, C. Eichler-Jonsson, E. D. Gilles, G. Muller, 2002, "Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors", *Nat. Biotechnol.*, Vol. 20, pp. 370–375.

Schuster, S., B. N. Kholodenko, H. V. Westerhoff, H.,2001, "Cellular information transfer regarded from a stoichiometry and control analysis perspective", *Biosystems*, Vol. 55, pp. 73–81.

Schuster, S. and C. Hilgetag, 1994, "On elementary flux modes in biochemical reaction systems at steady state" *J. Biol. Syst.* Vol. 2, pp. 165–182.

Schuster, S., C. Hilgetag, J. H. Woods, and D. A. Fell. "Elementary modes of functioning in biochemical reaction networks" In Cuthbertson, R., Holcome, M., & Paton, R., eds., *Computation in Cellular and Molecular Biological Systems*, pp.151-165, World Scientific, Singapore.

Schwartz, M. A. and V. Baron, 1999, "Interactions between mitogenic stimuli, or, a thousand and one connections" *Current Opinion in Cell Biology*, Vol. 11, pp. 197-202.

Seressiotis, A. and J. E. Bailey, 1988, "An artificially intelligent software system for the analysis and synthesis of metabolic pathways" *Biotechnol. Bioeng.*, Vol. 31, pp. 587–602

Sherrill, J. M. and J. Kyte, 1996, "Activation of epidermal growth factor receptor by epidermal growth factor", *Biochemistry*, Vol. 35, pp. 5705-5718.

Shi, C. S., S. Sinnarajah, H. Cho, T. Kozasa, J. H. Kehrl, 2000, "G13alphamediated PYK2 activation. PYK2 is a mediator of G13alpha-induced serum response element-dependent transcription", *J. Biol. Chem.*, Vol. 275, pp. 24470–24476.

Shvartsman S. Y, C. B. Muratov, D. A. Lauffenburger, 2002, "Modeling and computational analysis of EGFreceptor-mediated cell communication in Drosophila oogenesis", *Development*, Vol. 129, pp. 2577–2589.

Starbuck, C. and D. A. Lauffenburger, 1992, "Mathematical model for the effects of epidermal growth factor receptor trafficking dynamics on fibroblast proliferation responses", *Biotechnol. Prog.*, Vol. 8, pp. 132-143.

Steffen, H., A. Petti, J. Aach, P. D'haeseleer, G. Church, 2002, "Automated modeling of signal transduction networks", BMC Bioinformatics, Vol. 3, pp. 34.

Stephanopoulos, G. N. and A. A. Aristidou, 1998, *Metabolic Engineering*, Ch. 11, Academic Press.

Strogatz, S. H., 2001, "Exploring complex networks", *Nature*, Vol. 410, pp. 268-276, March.

Stryer, L., 1995, *Biochemistry*, W. H. Freeman and Company, New York.

Tall, G. G., M. A. Barbieri, P. D. Stahl, B. F. Horazdovsky, 2001, "Ras-activated endocytosis is mediated by the Rab5 guanine nucleotide exchange activity of RIN1", *Dev Cell*, Vol. 1, pp. 73–82.

Tatebayashi, K., M. Takekawa, H. Saito, 2003 "A docking site determining specificity of Pbs2 MAPKK for Ssk2/Ssk22 MAPKKKs in the yeast HOG pathway", *EMBO J.*, Vol. 22, No. 14, pp. 3624-3634, July.

Teruel, M. N. and T. Meyer, 2000, "Translocation and reversible localization of signaling proteins: a dynamic future for signal transduction", *Cell*, Vol. 103, pp. 181–184.

Theurkauf, W. E., 1994, "Premature microtubule-dependent cytoplasmic streaming in cappuccino and spire mutant oocytes", *Science*, Vol. 265, pp. 2093–2096.

Van Drogen, F., V. M. Stucke, G. Jorritsma, M. Peter, 2001, "MAP kinase dynamics in response to pheromones in budding yeast", *Nat. Cell Biol.*, Vol. 3, No. 12, pp. 1051-1059, December.

Verna J., A. Lodder, K. Lee, A. Vagts, R. Ballester, 1997, "A family of genes required for maintenance of cell wall integrity and for the stress response in Saccharomyces cerevisiae", *Proc. Natl. Acad. Sci. USA*, Vol. 94, No. 25, pp. 13804-13809, December.

Vuong, T. M. and M. Chabre, 1991, "Deactivation kinetics of the transduction cascade of vision", *Proc. Natl Acad. Sci.*, Vol. 88, pp. 9813–9817.

Wagner, A., 2001, "The yeast protein interaction network evolves rapidly and contains few redundant duplicate genes", *Mol. Biol. Evol.*, Vol. 18, pp. 1283-1292.

Wang, X. F., 2002, "Complex networks: topology, dynamics and synchronization", *Int. J. Bifurcation and Chaos*, Vol.12, No.5, pp. 885-916, May.

Watts, D. J. and S. H. Strogatz, 1998, "Collective dynamics of small-world networks", *Nature*, Vol. 393, pp. 440-442.

Wells, A., 1999, "EGF receptor", *Int. J. Biochem. Cell Biol.*, Vol. 31, pp. 637–643.

Weng, G., U. S. Bhalla, R. Iyengar, 1999, "Complexity in biological signaling systems", *Science*, Vol. 284, pp. 92–96.

Widmann, C., S. Gibson, M. B. Jarpe, G. L. Johnson, 1999, "Mitogen-activated protein kinase: conservation of a three-kinase module from yeast to human", *Physiol. Rev.*, Vol. 79, No. 1, pp. 143-180, January.

Wiley, H. S., S. Y. Shvartsman, D. A. Lauffenburger, 2003, "Computational modeling of the EGF-receptor system: a paradigm for systems biology", *Trends in Cell Biology*, Vol.13, No.1, pp. 43-50, January.

Wiley, H.S. and D. D. Cunningham, 1981, "A steady state model for analyzing the cellular binding, internalization and degradation of polypeptide ligands", *Cell*, Vol. 25, pp. 433–440.

Wiley, H.S. and D. D. Cunningham, 1982, "The endocytotic rate constant. A cellular parameter for quantitating receptor-mediated endocytosis", *J. Biol. Chem.*, Vol. 257, pp. 4222–4229.

Yarden, Y. and J. Schlessinger, 1987, "Epidermal growth factor induces rapid, reversible aggregation of the purified epidermal growth factor receptor", *Biochemistry*, Vol. 26, pp. 1443–1451.

Yarden, Y and M. X. Sliwkowski, 2001, "Untangling the ErbB signalling network", *Nat. Rev. Mol. Cell Biol.*, Vol. 2, pp. 127–137.

Yokouchi, M., T. Kondo, A. Houghton, M. Bartkiewicz, W. C. Horne, H. Zhang, A. Yoshimura, R. Baron, 1999, "Ligand-induced ubiquitination of the epidermal growth factor receptor involves the interaction of the c-Cbl RING finger and UbcH7", *J. Biol. Chem.*, Vol. 274, pp. 31707–31712.

Zubay, G., 1973, "In vitro synthesis of protein in microbial systems", *Annu. Rev. Genet.*, Vol. 7, pp. 267–287.