GAME THEORIC ANALYSIS OF MAPK SIGNALING PATHWAYS IN *Saccharomyces cerevisiae*

by

Can Çavlı

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 work subject to this thesis has been performed at the Department of Chemical Engineering, Boğaziçi University between September 2005 and May 2007.

I am more than grateful to my supervisor Prof. Kutlu Ö. Ülgen and research assistant Kazım Yalçın Arga for their contribution and support throughout the whole procedure. Either during the course of formation or the development of the thesis they have guided me in the best way possible and helped me with their critics. Furthermore they always welcomed me with a smile, which has been a great source of motivation. Without these two special people this work could not be done.

I am indebted to my family for their understanding and financial support that allowed me to undergo this process and step on to the next desired level in my life.

I also want to thank to my friends at the labs for making the time more enjoyable, creating a peaceful environment and motivating each other to work, to Uğur Emekli, Mehmet Hikmet Üçışık, and Saliha Durmuş for helping me to overcome the mental process most easily based on their experience, to all my friends whose names are not mentioned that wanted to help me in whatever way they could, to those that connived my absence at periods of research and to all the people that made my life worth living.

ABSTRACT

GAME THEORIC ANALYSIS OF MAPK SIGNALING PATHWAYS IN *Saccharomyces cerevisiae*

Understanding cellular signaling is central for gaining insight into the molecular mechanisms behind diseases as well as adaptation of living cells to changes in the environment. Signaling pathways are often branched in an interconnected fashion and are therefore integrated into signaling networks that are quite complex with many levels of interconnectivity of different molecular components. Recently, it became apparent that each MAPK pathway is a part of a network in which there is extensive sharing of signaling elements among the MAPK signaling pathways. Understanding the design principles that bridge the topology to the function of the network is a major challenge in systems biology since almost all known diseases exhibit dysfunctional aspects in these signaling networks. In the present study, using a probabilistic graph model (Bayesian Network) the feasibility of alternative signaling mechanisms was tested in the MAPK network in *Saccharomyces cerevisiae*. As a result of the large cross-talks between MAPK pathways, several signal transmission mechanisms that are biologically inactive were observed to be feasible. On the other hand, adaptation of a game theoretical formulation, in which the optimum strategy of a player was determined by considering the possible strategies of other players, resulted in a Nash Equilibrium (i.e., the set of optimum strategies) which eliminated the false-positives due to the crosstalk and represented the biology successfully. This mathematical framework has shown that logical reasoning is in accordance with real biology and thus provides an opportunity to model complex systems. The proposed methodology with further improvements in biological data is expected to provide more insight about the underlying principle in evolutionary construction of network topology.

ÖZET

MAYA HÜCRESİNDEKİ MAPK SİNYAL YOLİZLERİNİN OYUN TEORİSİ YAKLAŞIMIYLA ANALİZİ

Hücreiçi sinyal ağlarının anlaşılması, gerek hastalıkların arkasında yatan moleküler metabolizmaların ortaya çıkarılmasında gerekse de canlı hücrelerin çevrelerindeki değişimlere uyumunun açıklanmasında önem taşımaktadır. Sinyal yolizleri dallanarak aralarında iletişim halinde olan bir yapıya dönüşürler ve bu şekilde birden çok molekülün birbiriyle bağlantılı olduğu karmaşık sinyal ağları oluştururlar. Yakın zamanda her MAPK yolizinin, bünyesindeki sinyalleme elemanlarının bu yolizlerince ortak kullanıldığı bir ağın parçası olduğu anlaşılmıştır. Bilinen neredeyse tüm hastalıklar bu sinyal ağlarında anormallikler gösterdikleri için bu ağların topolojisi ile fonksiyonları arasındaki bağlantıların ardında yatan prensiplerin ortaya çıkarılması sistem biyolojisindeki ana hedeflerden biridir. Bu çalışmada, olasılıklı grafik modeli aracılığıyla (Bayes Ağı), mayanın MAPK ağları için alternatif sinyal mekanizmalarının uygunluğu test edilmiştir. MAPK yolizleri arasında büyük oranda ortak kullanılan elemanlar dolayısıyla oluşan karışıklık sebebiyle biolojik olarak aktif olması beklenmeyen çeşitli sinyal iletim mekanizmalarının olası olduğu gözlemlenmiştir. Bunun yanı sıra, bir oyuncunun en uygun strajesinin diğer oyuncuların olası stratejileri de gözönünde bulundurularak belirlendiği bir oyun teorisi formülasyonu uygulanması sonucu ortak kullanım dolayısıyla ortaya çıkan hatalı beklentileri yok eden, ideal strateji kararlarını ifade eden bir Nash dengesi ortaya çıkmıştır ve biyolojiyi başarılı bir şekilde yansıtmıştır. Bu matematiksel yaklaşım mantıksal adımlarla komplike biyolojik sistemlerin de gerçeği yansıtacak şekilde modellenebilmesinin mümkün olduğunu göstermiştir. Önerilen metod, gelişmekte olan biyolojik verilerin ışığında ağ topolojisinin evrimsel yapılandırılmasının altında yatan prensiplerin ortaya çıkarılmasını olası kılacaktır.

TABLE OF CONTENTS

LIST OF FIGURES

1. INTRODUCTION

Cellular mechanism is a complex interplay of enzymatic and genetic regulatory events. In a particular environment, the cell chooses to express and regulate a specific set of molecules. These molecules are organized in a modular manner, and these modules (i.e. protein complexes, pathways, etc.) act in concert to produce a network state which may be thought as the "phenotype" of that cell at those conditions. Understanding the conditional selection of the modules by the cell is a major concept in systems biology and several studies have been performed within the concept of "design principles" (Milo *et al*., 2002; Milo *et al*., 2004; Wolf and Arkin, 2003; Zhang *et al*., 2004). However, the underlying principle in evolutionary construction of network topology is still unknown.

The evolutionary selection process forces microorganisms to make the choice such that the cellular network acts in concert to attain the "optimal" behavior that enhance the survivability of the cell (Varma and Palsson, 1994). In traditional optimality theory, it is assumed that the best behavior for a particular species can be predicted irrespective of what others are doing. However, in biology, in almost all cases, the selection among alternative behaviors depends to a large extend on what others are doing. In the development of game theory, the understanding of the most rational way for humans to make decisions between alternative courses of action was its original purpose (Von Neumann and Morgenstern, 1953) where the outcome of a contest to a particular player was shaped by both the actions of the central player and the opponents. One of the most important consequences of game theory is that it enables the prediction of situations where one behavior or a specific mix of behaviors is more feasible than all known alternatives. The theory of rational choice states that the action chosen by a rational player is at least as good as all other available actions. In the subsequent years, game theory has been adapted by a number of scientists to examine biological dilemmas, especially the evolutionary problems (Axelrod and Hamilton, 1981; Bishop and Cannings, 1978; Hamilton, 1967; Lewontin, 1961; Maynard Smith, 1974; Slobodkin and Rapoport, 1974; Trivers, 1971). More recently, the evolutionary game theory was applied to understand the

host-parasite interactions (Frank, 1996), the properties of chromosomes (Maynard Smith and Szathmary, 1993), the RNA virus (Turner and Chao, 2003), bacteriocin diversity (Riley and Wertz, 2002), bacterial parasites (West and Buckling, 2003; Griffin *et al*., 2004), and ATP-producing pathways (Frick and Schuster, 2003; Pfeiffer *et al*., 2001; Pfeiffer and Schuster, 2005). Its ability to predict the optimal selection among alternative behaviors makes game theory an alternative viewpoint to study the structural concepts (i.e. design principles) underlying the network function.

Here, game theoretical concepts were adapted into bioinformatics to understand the underlying design principles of biological pathways. A network of four interconnected MAPK signaling pathways in *Saccharomyces cerevisiae* is chosen as the model system because these pathways have been studied well, and hence, most of the network structure and structural properties such as cross-talk relationships have been reported in literature (Gustin *et al*., 1998; Palecek *et al*., 2002; Tatebayashi *et al*., 2003; Verna *et al*., 1997; Van Drogen *et al*., 2001; Widmann *et al*., 1999). However, little is known about the design principles of MAPK networks. The balance between available information and unknown design principles makes MAPK networks a good candidate for a game theoretical study. In the present work, using a probabilistic graph model (Bayesian Network), the feasibility of alternative signaling mechanisms was tested in the MAPK network. As a result of the large cross-talks between MAPK pathways, several signal transmission mechanisms that are biologically inactive were observed to be feasible, when the network is represented as a protein-protein interaction graph. However, the Bayesian network model coupled with a game theory based solution algorithm yielded accurate results in representing the real biology by eliminating incorrect signaling routes resulting from crosstalks since it considers the preferences of other players in determining the optimum strategies.

2. THEORY

2.1. Cell Signaling

No cell lives in isolation. All organisms, whether single-celled or multicellular, have to sense the environment surrounding them and make decisions based on the information they retrieve in order to survive, which depends on an elaborate intercellular communication network that coordinates the growth, differentiation, and metabolism of the multitude of cells in diverse tissues and organs. Even the simplest bacteria sense and swim toward high concentrations of nutrients, such as glucose or amino acids. Independent of the organism type, every cell has to communicate with its surroundings, with other cells and within itself. This communication is provided by extra-, inter-, or intracellular signalling, which occurs through the transportation of specific molecules. Cells as those in nerve- or immunesystem get highly specialized in signaling. Consequently signaling controls important aspects of the cell, and possible malfunctions will lead to threatining the life of the organism. In depth characterization of signaling pathways will lead eventually to an ability to intervene in diseases in which those pathways are defective (Downward, 2001; Lodish *et al*., 2000; Cooper, 2000).

In higher animals, cells communicate using hundreds of kinds 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, whereas yeast cells communicate with one another by secreting only several kinds of small peptides. 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. Target cells, generally, respond by means of a specific protein called a receptor, which specifically binds the signaling molecule and then initiates a response in the target cell (Figure 2.1). These cell surface receptor proteins act as signal transducers: they bind the signaling molecule with

high affinity and convert this extracellular event into one or more intracellular signals that alter the behavior of the target cell (Downward, 2001; Alberts *et al*., 2002).

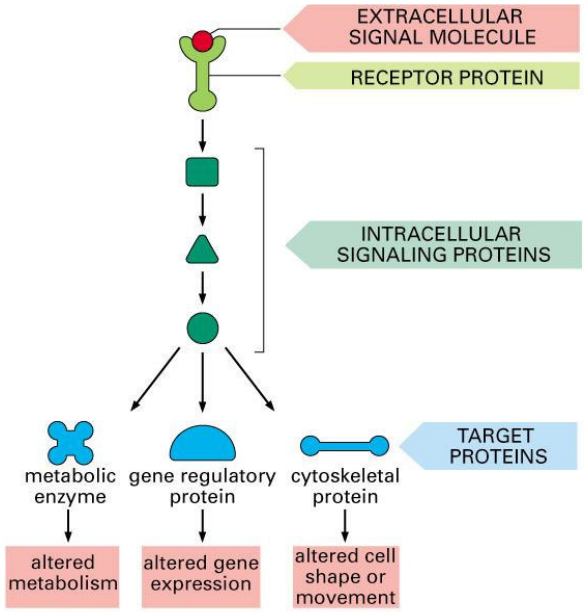


Figure 2.1. Simplified view of an intracellular signaling pathway (Alberts *et al*., 2002).

2.1.1. Ligand

A ligand is any molecule, other than an enzyme substrate, that binds tightly and specifically to a macromolecule, usually a protein, forming a macromolecule-ligand complex (Lodish *et al*., 2000). Yeast cells communicate with one another 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).

2.1.2. Receptor

A receptor is any protein that binds a specific extracellular signaling molecule (ligand) and then initiates a cellular response. Receptors for steroid hormones, which diffuse across the plasma membrane, are located within the cell; receptors for water-soluble hormones, peptide growth factors, and neurotransmitters are located in the plasma membrane with their ligand-binding domain exposed to the external medium (Lodish *et al*., 2000).

Cell signaling requires not only extracellular signal molecules, but also a complementary set of receptor proteins in each cell that enable it to bind and respond to the signal molecules in a characteristic way. Some small hydrophobic signal molecules, including steroid and thyroid hormones, diffuse across the plasma membrane of the target cell and activate intracellular receptor proteins that directly regulate the transcription of specific genes. Most extracellular signal molecules can activate receptor proteins only on the surface of the target cell; these receptors act as signal transducers, converting the extracellular binding event into intracellular signals that alter the behavior of the target cell (Alberts *et al*., 2002).

There are three main families of cell-surface receptors, each of which transduces extracellular signals in a different way. Ion-channel-linked receptors are transmitter-gated ion channels that open or close briefly in response to the binding of a neurotransmitter. G-protein-linked receptors indirectly activate or inactivate plasma-membrane-bound enzymes or ion channels via trimeric GTP-binding proteins (G proteins). Enzyme-linked receptors either act directly as enzymes or are associated with enzymes; these enzymes are usually protein kinases that phosphorylate specific proteins in the target cell.

Once activated, enzyme- and G-protein-linked receptors relay a signal into the cell interior by activating chains of intracellular signaling proteins; some transduce, amplify, or spread the signal as they relay it, while others integrate signals from different signaling pathways. Many of these signaling proteins function as switches that are transiently activated by phosphorylation or GTP binding. Functional signaling complexes are often formed by means

of modular binding domains in the signaling proteins; these domains allow complicated protein assemblies to function in signaling networks.

Target cells can use a variety of intracellular mechanisms to respond abruptly to a gradually increasing concentration of an extracellular signal or to convert a short-lasting signal into a long-lasting response (Figure 2.2). In addition, through adaptation, they can often reversibly adjust their sensitivity to a signal to allow the cells to respond to changes in the concentration of a particular signal molecule over a large range of concentrations (Alberts *et al*., 2002).

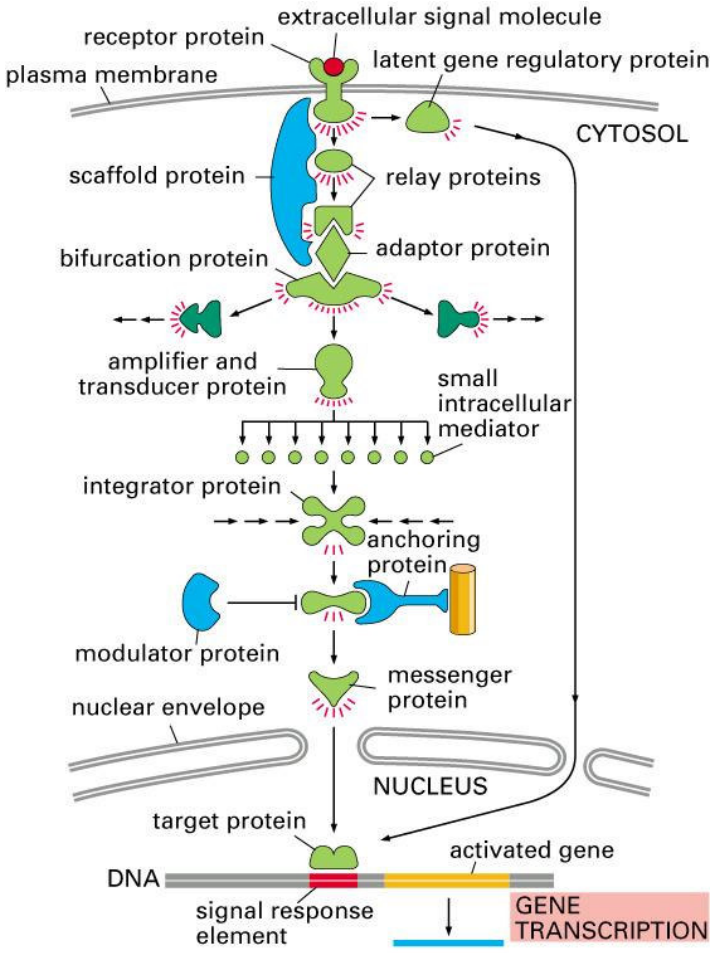


Figure 2.2. A more detailed intracellular signaling pathway (Alberts *et al*., 2002).

2.1.3. Transcription Factors

Transcription factor is the general term for any protein, other than RNA polymerase, required to initiate or regulate transcription in eukaryotic cells. General factors, required for transcription of all genes, participate in formation of the transcription-initiation complex near the start site. Specific factors stimulate (or repress) transcription of particular genes by binding to their regulatory sequences (Lodish *et al*., 2000).

More specifically, transcription factors regulate the binding of RNA polymerase to DNA and as a result control the subsequent transcription of DNA into messenger RNA and eventually protein. Transcription factors bind to specific sequences of DNA upstream or downstream to the gene they regulate and then either enhance or repress transcription of these genes by assisting or blocking RNA polymerase binding respectively. A cluster of transcription factors is the preinitiation complex (PIC) that recruits and activates RNA polymerase. Conversely, repressor transcription factors inhibit transcription by blocking the attachment of activator proteins (Brivanlou and Darnell, 2002; Karin, 1990).

A defining characteristic of transcription factors is that they contain a DNA binding domain (DBD) which bind to gene specific regulatory sites (*e.g.*, promoter sequences). In addition, transcription factors often contain a second domain that sense external signals and in response transmit these signals to the rest of the transcription complex resulting in up or down regulation of gene expression. In some cases the DBD and signal sensing domains reside on separate proteins that associate within the transcription complex to regulate gene expression (Brivanlou and Darnell, 2002; Karin, 1990).

2.1.4. Signal Transduction Pathways

Elucidation of the mechanisms that connect extracellular signal inputs to the control of transcription factors was until recently restricted to small-scale biochemical, genetic and pharmacological techniques. Signal transduction pathways have traditionally been viewed as linear chains of biochemical reactions and protein-protein interactions, starting from signal-

sensing molecules and reaching intracellular targets; however, the increasingly recognized abundance of components shared by several pathways indicates that an interconnected signaling network exists (Albert, 2005).

2.2. Network Modeling

Complex networks are currently being studied across many fields of science like biotechnology. Undoubtedly, many systems in nature can be described by models of complex networks, which are structures consisting of nodes or vertices connected by links or edges. The ubiquity of complex networks in science and technology has naturally led to a set of common and important research problems concerning how the network structure facilitates and constrains the network dynamical behaviours (Wang and Chen, 2003).

2.2.1. Reconstruction

Reconstruction of biochemical pathways is a complex task. In metabolism, databases like KEGG and EcoCyc serve as valuable resources for metabolic networks. Such extensive and well-curated databases are not yet available for cellular signaling. The role of each protein in a signaling network is to communicate the signal from one node to the next, and to accomplish this the protein has to be in a defined signaling 'state'. The state of a signaling molecule is characterized by covalent modifications of the native polypeptide, the substrates and/or ligands bound to the protein, its state of association with other protein partners, and its location in the cell. A signaling molecule may be a receptor, a channel, an enzyme or several other functionally defined species, depending on its state. In the process of passing a signal, a molecule may undergo a transition from one functional state to another. Interactions within and between functional states of molecules, as well as transitions between functional states, provide the building blocks for the reconstruction of a signaling network (Papin and Subramaniam, 2004)

The process of construction of signaling pathway models requires the assembly of a network of interacting proteins in a given context of the cell. Much of the knowledge on the pathways

comes from interrogation of cells by specific perturbations followed by assays and systematic biochemical analysis of protein complexes. Reconciliation of a large body of cellular data provides validation strategies for reconstructed networks. Even though the standard representations of biochemical pathways have been incomplete, they serve as useful models for constructing and testing specific biological hypotheses. Several efforts are underway currently to build databases of biochemical signaling pathways and networks of pathways. These databases are also combined into larger infrastructures containing graphical user interfaces and some rudimentary analysis tools (Papin and Subramaniam, 2004).

2.2.2. Mathematical Modeling and Simulation

The biochemical processes in biological cells are complex and interwoven. Pathway cross-talk by shared metabolites or enzymes, regulation, and positive and negative feedback all contribute to the complexity. Although not a traditional experimental "method," mathematical modeling can provide a powerful approach for investigating complex cell signaling networks, such as those that regulate the eukaryotic cell division cycle. Mathematical models are essential for understanding biochemical networks and predicting their behaviour under perturbation. Mathematical models permit the time dependence of metabolite concentrations to be simulated (Stein *et al*., 2007, Sible and Tyson, 2006).

Models of biochemical networks are typically constructed as sets of differential equations describing the time dependence of compound concentrations. Ordinary differential equations (ODEs), dependent on only one variable (e.g. time), can be used to describe the change of states (e.g. compound concentrations). The equations can describe changes due to reaction and diffusion or other transport processes. Partial differential equations (PDEs) depend on more variables and enable the investigation of spatial and temporal changes of molecular concentrations. The differential equations contain kinetic parameters, which need to be determined experimentally or estimated computationally.

When molecular concentrations fall below a threshold such that they can no longer be treated uniformly and deterministically, then the discrete nature of the molecules and their stochastic

behaviour has to be taken into account in simulations with random fluctuations (Stein *et al*., 2007).

Quantitative models of large systems are considerably harder to develop because they require large sets of relatively precise data that are yet to be obtained. The required experimental data fall into several categories. These include concentrations of cellular components, the rates of their interactions, including rates of enzyme and binding activities, their locations within cells, and rates of regulated movement between cellular compartments. Despite these limitations, the number of quantitative differential equation–based models has been growing steadily, and high-throughput experiments that measure rate constants are becoming a reality. Such models will be essential for drug development from network models (Ma'ayan *et al*., 2006).

Because a universal mechanism controlling DNA synthesis, mitosis, and cell division underlies the growth, development, and reproduction of all eukaryotes, an understanding of this molecular regulatory system is one of the most important goals of modern cell biology. As the complex network of cell cycle controls is uncovered, it becomes increasingly difficult to make reliable predictions about how modification of one component affects the system as a whole. However, such predictions are needed if the host of mutations contributing to cancer is to be identified or found within the molecular network novel targets for therapeutic intervention. Mathematical models provide powerful tools for managing the complexity of the cell cycle control system and of other signaling networks. Models organize a large body of experimental data, describe the fundamental behaviors of the system as a whole, bridge gaps where experimental data are missing, and drive hypothesis-building for the next round of experimentation. The value of mathematical modeling in describing and predicting the behavior of complex systems has been well established in fields such as chemical engineering and meteorology, but its power has been underappreciated until recently in molecular cell biology (Sible and Tyson, 2006).

Although mathematical models can be built to describe any signaling network, application of modeling tools to cell cycle regulation is particularly well suited and timely. First, the data in this field are vast, both providing a large body of information to build comprehensive models and creating the need for a tool to understand how these data fit together. Second, cell cycle

signaling networks are modular, allowing models to be constructed in parts and then assembled and reassembled in various ways. Furthermore, many models of the network are comparable between different organisms (e.g., budding yeast and mammals) so that it is feasible to make relatively small changes to a model describing one particular system in order to apply it to another. Thus, each new model need not be constructed from scratch. Third, a reasonable amount of quantitative or semi-quantitative information can be extracted from the literature, facilitating the early phases of model building. By modifying established protocols (described in the accompanying articles), additional quantitative data can be generated to improve parameter estimation and experimental validation of models. Finally, despite the wealth of detailed information on cell cycle molecules and their specific interactions, there is a lack of a systems-level perspective of this complex control network. Modeling can provide this perspective by helping to identify underlying regulatory principles. Where a specific experimental detail is missing, modeling can serve as bridge, enabling progress in building a systems-level view, and guiding the design and execution of future experiments (Sible and Tyson, 2006).

Although it is possible to attempt whole-cell simulations with all relevant metabolites included, it is more common to model individual pathways or groups of pathways. A modular approach can then be taken to investigate cross-talk between pathways and reconstitute cell-like simulations from independent module. Pathways that have been extensively modelled and simulated recently include the mitogen-activated protein kinase (MAPK) pathway and glycolysis (Stein *et al*., 2007).

2.3. MAPK Signaling in Yeast

Yeast is used as a model organism, since it is easy to handle and many data are already available. Signaling pathways are evolutionarily highly conserved. From an experimental as well as from a modeling viewpoint, the developed techniques, the specific results and the art of interaction of modeling and experimental research can be applied to higher organisms, in order to arrive at a better understanding of the dynamic operation of those pathways and to offer new opportunities for drug discovery.

Signal transduction networks permit cells to receive external stimuli and respond to the signals in an appropriate manner. The Mitogen-Activated Protein Kinase (MAPK) signaling pathways play an important role in signal transduction in eukaryotic cells, where they modulate many cellular events including: mitogen-induced cell cycle progression through the G1 phase, regulation of embryonic development, cell movement and apoptosis, as well as cell and neuronal differentiation. These evolutionarily conserved pathways are organized in three-kinase modules consisting of a MAP kinase, an activator of MAP kinase (MAP Kinase Kinase or MEK) and a MAP Kinase Kinase Kinase (MEK Kinase, MEKK, or MAPK Kinase Kinase). There are at least three distinct MAP kinase signal transduction pathways in mammalian cells, each named after the particular MAPK associated with it. Since the budding yeast *Saccharomyces cerevisiae* is known to have more than three distinct MAPK pathways, it is logical to expect that there might be additional MAPK pathways in mammalian cells (Promega Corporation, 2000).

The MAP kinase cascade is a highly conserved signal transduction module that propagates signals from cell surface receptors to various cytosolic and nuclear targets by way of a phosphorylation cascade and is thought to be present in all eukaryotes. The cascade typically consists of three layers, each based on a kinase that phosphorylates and activates an immediately downstream kinase. The final component in the cascade, the MAP kinase, once activated, phosphorylates various cytosolic targets and translocates to the nucleus to phosphorylate transcription factors affecting gene expression. Full activity of the MAP kinase (MAPK) generally requires phosphorylation at both a conserved tyrosine and threonine residue although evidence suggests that in some pathways partial activity is possible with phosphorylation at a single site. Activation of the cascade is often associated with cell death, environmental stress, cell proliferation, and cell differentiation responses (Schwacke and Voit, 2006).

2.3.1. Mating Pheromone Response Pathway

Saccharomyces cerevisiae exists in two haploid cell types, MATa and MATα. The α-factor released by α cells acts on a cells by binding to the G-Prote in Coupled Receptor (GPCR)

Ste2; and, a-factor released by a cells acts on α cells by binding to the GPCR Ste3. Both pheromone receptors are coupled to a common heterotrimeric G protein, Gpa1–Ste4–Ste18, where Gpa1 is Gα and Ste4–Ste18 is the Gβγ complex. Engagement of these GPCRs by their cognate pheromones leads to activation of Cdc42 and, eventually, to activation of the MAPK, Fus3. The action of Fus3 is responsible for eliciting the expression of numerous mating specific genes, imposing cell cycle arrest, promoting polarized cell growth to form copulatory projections toward the mating partner, establishing the changes in the plasma membrane and cell wall necessary for cell–cell fusion, and orienting the nucleus and modifying its envelope to permit fusion of the two haploid nuclei. Both the heterotrimeric G protein and Cdc42 also act through additional effectors to stimulate other branches of the response machinery that are necessary to produce mating-competent cells and achieve optimally efficient mating. Thus, yeast pheromone response is clearly a network of interlocking events, rather than a simple linear pathway, and is arguably one of the best understood MAPK-based signal-response systems in biology (Chen and Thorner, 2007; Van Drogen *et al*., 2001).

The binding of a pheromone to its cognate GPCR facilitates the release of GDP and the subsequent binding of GTP by Gpa1 (Gα subunit). GTP binding to Gα alters its interaction with Gβ (Ste4), dissociating Gpa1 from the Gβγ complex. Released Gβγ can interact with three known effectors: Ste20 and Ste5 and a protein weakly related to Ste5, Far1. Unlike Ste5, the C-terminus of Far1 binds to, and most likely activates Cdc24. Ste20, activated by Cdc42–GTP, serves as the MAPKKKK to phosphorylate and thereby trigger activation of the MAPKKK, Ste11, initiating activation of the remainder of the MAPK cascade, namely Ste7 (MAPKK) and Fus3 (MAPK) (Figure 2.3). Ste5 is a scaffold protein that binds all three component kinases of the cascade (Ste11, Ste7, and Fus3) (Chen and Thorner, 2007).

In addition to Fus3, pheromone stimulation also leads to transient activation of another MAPK, Kss1. Activation of Kss1 also occurs via Ste11 and Ste7, but is not dependent on the scaffold protein, Ste5. Cells lacking both Fus3 and Kss1 are sterile, whereas the presence of either one alone permits mating. Quantitative analysis shows that loss of Kss1 does not

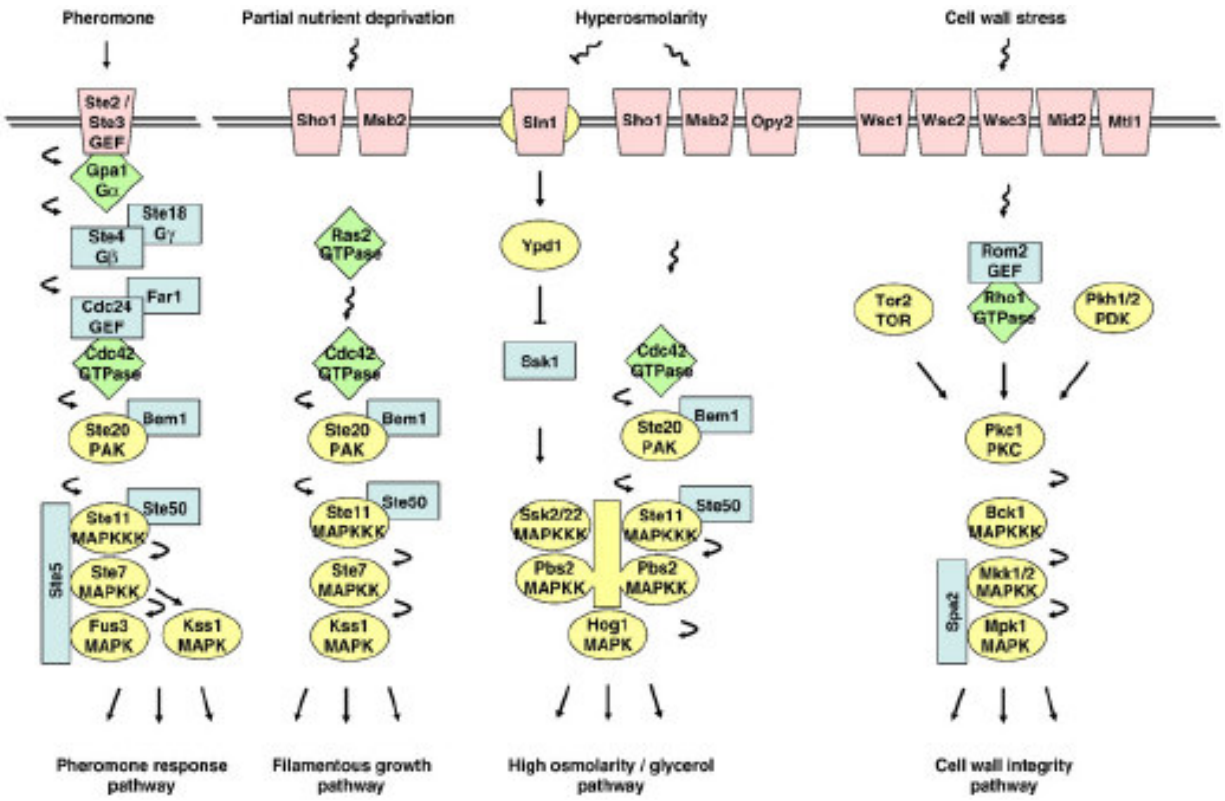


Figure 2.3. Schematic diagrams of the MAPK signaling pathways in *Saccharomyces cerevisiae*. Symbols are: protein kinases, ovals; GTP-binding proteins, diamonds; scaffold, adaptor, and activating proteins, rectangles; cell surface proteins, trapezoids; activation, arrows; inhibition, T-bars; direct action, smooth lines; indirect action (or unknown molecular mechanism), squiggly lines. For clarity, not all factors and interactions are shown, connections to other pathways and processes upstream of the MAPKs are omitted, and direct targets of the MAPKs are not included (Chen *et al*., 2007).

measurably reduce mating proficiency, whereas loss of Fus3 reduces mating efficiency to ~10 per cent of the wild-type level. Analysis of other indicators (cell cycle arrest, morphological changes, gene induction patterns) of signal throughput in cells lacking either Fus3 or Kss1 indicates that Fus3 is responsible for the majority, but not the entirety, of the MAPK-dependent pheromone response. In contrast, Kss1, but not Fus3, is essential for the invasive growth response in haploids and the pseudohyphal growth response in diploids (filamentous growth response). Fus3 is much more efficient than Kss1 at mediating pheromoneinduced cell

cycle arrest, most likely because Fus3 phosphorylates Far1 more efficiently due to a highaffinity docking site in Far1 that binds Fus3, but not Kss1. Fus3 also serves as a negative regulator of filamentous growth because, unlike Kss1, it phosphorylates and leads to the degradation of the Tec1 transcription factor necessary for induction of the genes involved in this developmental outcome (Chen and Thorner, 2007).

2.3.2. Filamentation and Invasion Pathway

Nutrient limitation induced behaviors are termed pseudohyphal growth in diploids and invasive growth in haploids. Although there are important biological and mechanistic differences between the two, many of the primary molecular components and regulatory pathways involved in these filamentous growth responses are the same. The MAPK cascade mediates signal transduction in filamentation-invasion pathway from Ras2 and Cdc42. Signaling from Ras2 requires the proteins Bmh1 and Bmh2 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. The MAPK for the filamentation-invasion pathway is Kss1. Activation of Kss1 requires Ste20 (PAK), Ste11 (MAPKKK), and Ste7 (MAPKK) (Figure 2.3). Activation of Cdc42 during filamentous growth is known to be dependent on active Ras2. Sho1 can form hetero-oligomeric complexes with Msb2, and the absence of either protein blocks Kss1 activation and prevents filamentous growth in haploids. 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 (Chen and Thorner, 2007; Palecek *et al*., 2002).

2.3.3. High Osmolarity/Glycerol Pathway

It is a treat for cell viability that the dissolved solute concentration of the extracellular medium increases to a level higher than the internal osmolarity of the cell. To restore osmotic balance the cell increases the internal osmolyte concentration by synthesis of glycerol, a highly water soluble and inert solute. This mechanism is referred to as the High-Osmolarity-Glycerol (HOG) response. Survival under hyperosmotic conditions via the HOG pathway requires activation of the Hog1 from either of two upstream inputs (Chen and Thorner, 2007) (Figure 2.3).

First route is initiated by the osmosensor Sln1. Under iso-osmotic conditions, Sln1 is active and catalyzes autophosphorylation and subsequent phospho-transfer to an intermediate protein, Ypd1, which transfers the phosphate group to a response regulator, Ssk1, preventing interaction of Ssk1 with two semi-redundant MAPKKKs, Ssk2 and Ssk22. Mild hyperosmotic stress inhibits Sln1, resulting in an increase in the amount of unphosphorylated Ssk1. Unphosphorylated Ssk1 is able to bind to and activate Ssk2 and Ssk22. These MAPKKKs phosphorylate a dedicated MAPKK, Pbs2, which in turn, is responsible for dual phosphorylation and activation of the MAPK, Hog1. The second route for Hog1 activation is via the alternative MAPKKK, Ste11, which is also encountered in both the pheromone response pathway and the filamentous growth pathway (Figure 2.3). Stimulation of Ste11 in the Sho1-dependent branch requires the function of Cdc42 and Ste20. It is necessary to steer active Ste11 toward Pbs2 and prevent it from encountering Ste7. The MAPKK Pbs2 represents a true node for both the the Sln1-dependent and the Sho1-dependent branches of the HOG pathway. Pbs2 contains a highaffinity docking site for the MAPKKKs, Ssk2 and Ssk22, of the Sln1 branch and also associates with Ste11. Pbs2 serves as both the MAPKK of the HOG pathway and also as scaffold for proper assembly of the signaling complexes necessary to propagate the signals that initiate the HOG pathway in the first place (Chen and Thorner, 2007).

Activation of Hog1 causes its rapid translocation from the cytoplasm to the nucleus. Nuclear Hog1 binds and phosphorylates several transcription factors, interacts with chromatin

modifying enzymes and RNA polymerase II, and affects the expression of hundreds of genes in response to hyperosmotic shock. Osmostress-regulated genes are implicated mainly in carbohydrate metabolism, general stress protection, protein production and signal transduction. 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 (Chen and Thorner, 2007; Tatebayashi *et al*., 2003).

2.3.4. Cell Wall Integrity Pathway

The MAPK Slt2/Mpk1 becomes activated under a number of different conditions that stress the structure and function of the yeast cell wall. It is thought that the common element sensed in all of these cases is stretching of the plasma membrane and/or alterations of its connections to the cell wall. The genes under control of this response pathway include many involved in the synthesis and modification of the major components of the yeast cell wall, and lack of an Slt2/Mpk1-dependent response causes cell lysis in the absence of an osmotic support in the medium . Hence, the Slt2/Mpk1-dependent response is referred to as the cell wall integrity (CWI) pathway (Chen and Thorner., 2007).

Five plasma membrane proteins, Wsc1, Wsc2, Wsc3, Mid2, and Mtl1, have been identified as important for activation of this pathway. The cytoplasmic Cterminal domains of Wsc1 and Mid2 interact with Rom2, thought to be specific for the small Rashomologous GTPase, Rho1. With respect to signals ensuring cellular integrity, the main effector of Rho1 is PKC1. PKC1 is an essential activator (MAPKKKK) of the MAPK cascade required for CWI signaling—Bck1 (MAPKKK), Mkk1 and Mkk2 (two semi-redundant MAPKKs), and Slt2/Mpk1 (MAPK) (Figure 2.3); the MAPKKs and MAPK in this pathway are bound by the scaffold protein Spa2 (Chen and Thorner, 2007; Verna *et al*., 1997).

Slt2/Mpk1 is responsible for stimulating expression of the genes for enzymes and other factors involved in cell wall biosynthesis and remodeling both directly and indirectly. Slt2/Mpk1 stimulates expression of cell wall biosynthesis genes directly via phosphorylation of the

transcription factors. Rlm1 and the Sbf complex (consisting of Swi4 and Swi6) have been reported as targets of the MAPK. 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 (Chen and Thorner, 2007; Verna *et al*., 1997).

2.4. Mathematical Methods

2.4.1. Graph Theory

Since the late 1990s, the approach of abstracting complex systems to networks, resulting in directed or undirected graphs made of nodes and links, is increasingly employed to analyze systems from a range of scientific fields. These applications nicely fit the molecules to nodes and interactions to links simplification used to describe intracellular mammalian interactions networks. Watts and Strogatz (1998) used two previously defined global statistical properties of graphs to characterize networks: clustering coefficient and characteristic path length. It was found that biological interaction networks have higher clustering coefficients and similar characteristic path lengths expected if the networks would be randomly rewired. An initial striking result from the analysis of biochemical interaction networks is that network nodal connectivity distribution fits a power-law. Such networks are termed scale-free. Another approach for analysis of complex systems abstracted to network maps is characterization of motifs. Network motifs are subsets of interactions involving several different components. Alon's group (Milo *et al*., 2002) was the first to propose this approach for analyzing biochemical interactions networks. They initially analyzed a gene regulatory network of bacteria. It is also possible to break up large-size biochemical interaction maps into subnetworks based on specific criteria such as limiting the number of steps from a receptor to a transcription factor and then searching for network motifs only in those subnetworks. Major attention is targeted toward the hubs: the highly connected nodes in protein–protein, ligand–

protein, and gene regulatory networks. Han and coworkers (2004) distinguished between party hubs and date hubs. Party hub proteins interact with many other proteins in the same compartment and at the same time, whereas date hubs interact with many other proteins at different times and places in the cell. When network maps include directionality of the links it is possible to separate hubs based on their in-links and out-links (Ma'ayan *et al*., 2006).

2.4.2. Game Theory

Game theory is the study of the ways in which strategic interactions among rational players produce outcomes with respect to the preferences (or utilities) of those players, none of which might have been intended by any of them (Ross, 2004). Game Theory is a misnomer for Multiperson Decision Theory, analyzing the decision making process when there are more than one decision-makers where each agent's payoff possibly depends on the actions taken by the other agents. Since an agent's preferences on his actions depend on which actions the other parties take, his action depends on his beliefs about what the others do. Of course, what the others do depends on their beliefs about what each agent does. In this way, a player's action, in principle, depends on the actions available to each agent, each agent's preferences on the outcomes, each player's beliefs about which actions are available to each player and how each player ranks the outcomes, and further his beliefs about each player's beliefs, ad infinitum (Yildiz, 2004).

One way to describe a game is by listing the players (or individuals) participating in the game, and for each player, listing the alternative choices (called actions or strategies) available to that player. In the case of a two-player game, the actions of the first player form the rows, and the actions of the second player the columns, of a matrix. The entries in the matrix are two numbers representing the utility or payoff to the first and second player respectively. A very famous game is the Prisoner's Dilemma game (Levine).

2.4.2.1. Prisoner's Dilemma in biochemistry : Two yeast strains that both used sugar as energy resource, but which may choose between two different pathways of ATP production,

were studied from a game-theory point of view. These pathways were considered as distinct strategies to which payoffs were assigned that were proportional to the expected steady-state number of individuals sustainable on the basis of these strategies. In a certain parameter range the payoffs fulfilled the conditions for the prisoner's dilemma. Therefore, cooperative behaviour was unlikely to occur, unless additional factors intervened. In fact, the yeast *Saccharomyces cerevisiae* used a competitive strategy by fermenting sugars even under aerobic conditions, thus wasting its own resource (Frick and Schuster, 2003).

2.4.2.2. Game-theoretical approaches to studying the evolution of biochemical systems : Evolutionary processes need to be considered for a detailed understanding of complex biochemical systems. Evolutionary optimization has been successfully used to increase the understanding of key properties of biochemical systems. Evolutionary processes do not simply optimize these systems because there are interactions between the evolving population and its environment. Thus, traditional optimization is often insufficient for understanding the dynamics of evolutionary processes because usually there is a mutual relationship between the properties optimized by evolution and the properties of the environment. Thus, by evolving towards optimal properties, organisms change their environment, which in turn alters the optimum. Evolutionary game theory provides a more appropriate tool with which the dynamics and outcome of evolution of biochemical systems can be studied (Pfeiffer and Schuster, 2005).

Recent studies have applied evolutionary game theory to key issues in the evolution of energy metabolism. Biochemical phenomena such as biofilms that show strong interactions between the population and its environment were analyzed by evolutionary game theory. In another application, the coexistence of three types of strain was reported to resemble the rock–scissors–paper game. Assuming that both toxin production and toxin resistance were associated with metabolic costs, each of the three strains described above could invade another type but was also susceptible to invasion by the remaining type (Pfeiffer and Schuster, 2005).

2.4.3. Bayesian Networks

Cell signaling networks typically modeled by computer simulations could also be modeled by classifier systems (Holland, 2002). Classifier systems share many aspects of cell signaling networks such as parallelism and coordination, conditional actions, modularity, and adaptation. Exploratory statistical models built using these concepts can provide insight into the operation of perturbed pathways in DS. One of the most commonly used classification methods for this purpose is Bayesian networks (BN). Construction of BN from experimental measurements of mRNA levels, as a time-series or under different perturbations, to reverse-engineer gene regulatory networks from microarray data are the most common approach so far to rebuild networks from these data. Bayesian networks are acyclic graphs in which nodes represent the experimentally measured variables and links are probabilistic influences of variables on each other. Woolf and coworkers (2005) used this approach to build a BN from a multivariant dataset of 28 signaling proteins under 16 combinations of experimental conditions applied to mouse embryonic stem cells. These cells can be driven to self-renewal or to differentiation in culture based on the extracellular media provided (i.e., stimulation by extracellular ligands). The authors searched for network topology and probabilities to connect variables to best fit the experimental results. The resulting network was validated against shuffled networks and helped the authors to hypothesize about the outcome of differentiation *versus* self-renewal under conditions not yet tested experimentally. Sachs and coworkers (2005) used the BN approach to study the relationships between proteins and phospholipids, and the directionality of their links, after T-cell activation of naïve T-cells. Sachs and coworkers (2005) used data from single cell measurements using flow cytometry to measure the phosphorylation levels of key signaling nodes (proteins and phospholipids). The authors then determined hierarchical ordering of signaling components by applying experimental perturbations, such as knocking out a protein, by either pharmacological agents or RNA interference. They were able to determine which proteins are upstream or downstream in the signaling network using statistical correlations. Dynamic Bayesian networks (DBN) attempt to solve some limitations of standard BN analysis. Dynamic Bayesian networks analysis is applied to multivariant time-series data. The idea is to identify correlations between variables at different time points. For example, if variable x is up at time point $t1$ and variable y is up at

time point *t*2, it is possible that variable *x* upregulates variable *y*. Dynamic Bayesian networks was implemented by Zou and Conzen (2005) to infer a gene regulatory network function from the yeast cell cycle dataset. The first implementation of DBN in biology is attributed to Murphy and Mian (1999). Bayesian networks do not perform well when only few time points are available, while thousands of variables are measured (i.e., time-series microarrays data). Segal and coworkers (2005) recently suggested a potential solution. Instead of treating each variable independently, variables are grouped into modules such that the BN is constructed to connect the modules. Bayesian networks application to experimental results is essentially statistical. Another elaborate statistical method was used by Janes and coworkers (2005) to study cell signaling axes of apoptosis. The author analyzed 7980 experimental measurements by constructing high-dimensional vectors from the data to understand the trajectory of cellular response to different stimuli that induce either apoptosis or promote cell survival. Such approaches may be useful in identifying where cellular dynamics trajectories are deformed in DS (Ma'ayan *et al*., 2006).

A Bayesian network is a directed acyclic graph which encodes the joint probability distribution over a set of random variables x1, . . . , xn. The joint probability distribution can be expressed as

$$P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i \mid pa[x_i]),$$

(2.1)

where pa[xi] is the set of parent variables of xi in the graph. This decomposition gives a space-efficient method of representing the joint probability distribution compared to the general case.

To study system dynamics, one can extend Bayesian networks by "unfolding" the graph structure. For each time point under consideration, a nodes of the Bayesian network are duplicated. The probability distribution for a variable in time step t + 1 is then given in terms

of its parents in time step t. The resulting graph is also a Bayesian network, and it can be used to make inferences involving dynamic behaviour of the variables (Pitkänen, 2007).

3. MATERIALS AND METHODS

3.1. MAPK Network

Like other eukaryotes, yeasts share several mitogen-activated protein kinase (MAPK) signaling pathways (Widmann *et al*., 1999). The MAPK cascades are activated by diverse stimuli (cytokines, growth factors, hormones, cellular stress, etc.) and they are regulated by a variety of extra- or intra-cellular signals (Gustin *et al*., 1998). When the cascade is activated, several transcriptional factors are regulated by MAP kinase phosphorylation. In *Saccharomyces cerevisiae*, four functionally distinct pathways including a MAPK cascade have been established and extensively investigated: (i) high-osmolarity signaling pathway (Tatebayashi *et al*., 2003), (ii) mating pheromone response pathway (Van Drogen *et al*., 2001), (iii) filamentous growth and invasion pathway (Palecek *et al*., 2002), and (iv) cell-wall integrity pathway (Verna *et al*., 1997).

In the present study a protein-protein interaction network of 49 proteins annotated to be functioning in the above-mentioned MAPK pathways in the MIPS-CYGD functional catalogue (Güldener *et al*., 2005) is reconstructed by May 12th, 2006. For the physical interactions through which the signal is transmitted, the assembly of protein-protein interactions obtained from three public databases is used, BioGRID (Breitkreutz *et al*., 2003), DIP (Salwinski *et al*., 2004) and STRING (Von Mering *et al*., 2005) by February 9th, 2006. Only the physical interactions representing the signal transduction, which is validated by specific signaling studies using small-scale experimentation, were considered to eliminate possible false-positives in the interactome assembly. So the basis of the work, namely the proteins in the network and the interactions among them, is limited to current knowledge in biology, and it will be improved with the advances in the field. The present reconstruction attempt for MAPK network resulted in an intertwined network (Figure 3.1) of four MAPK pathways.

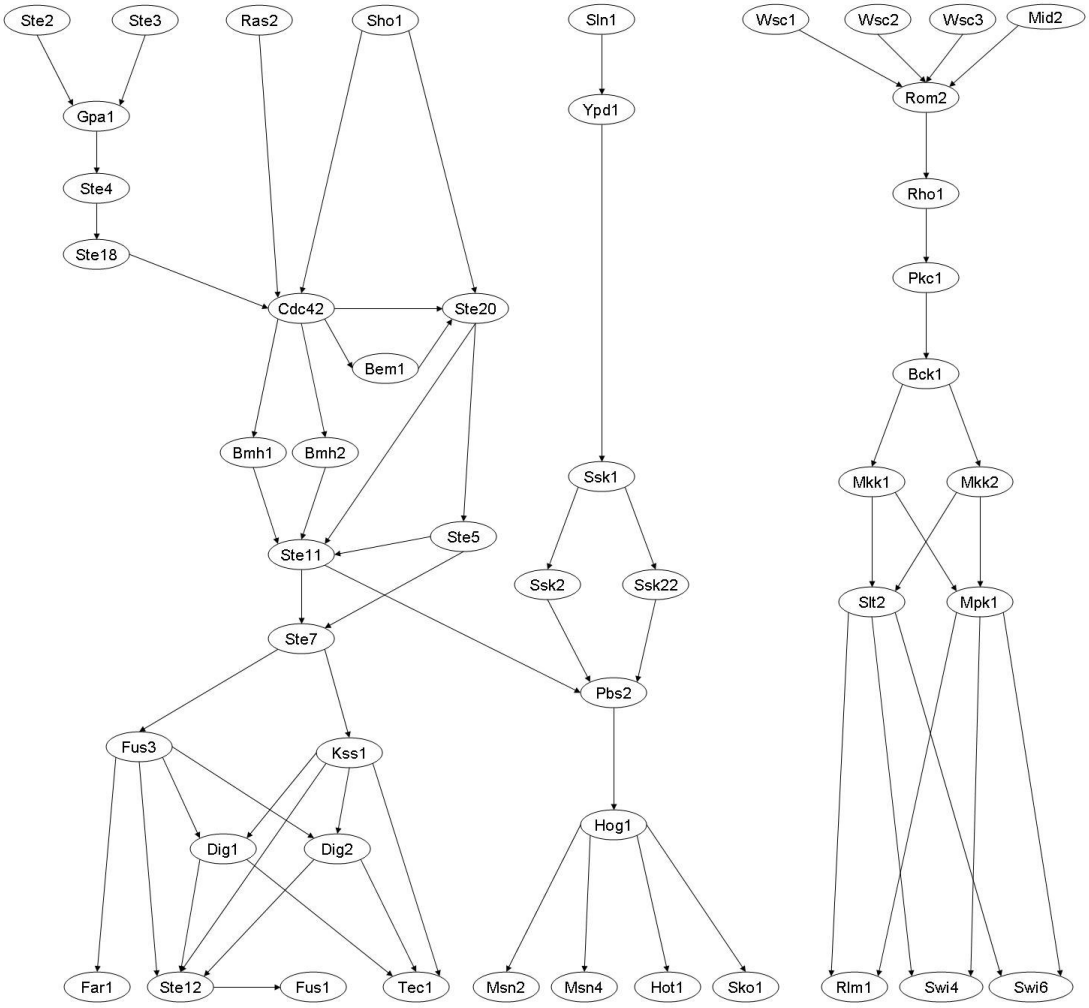


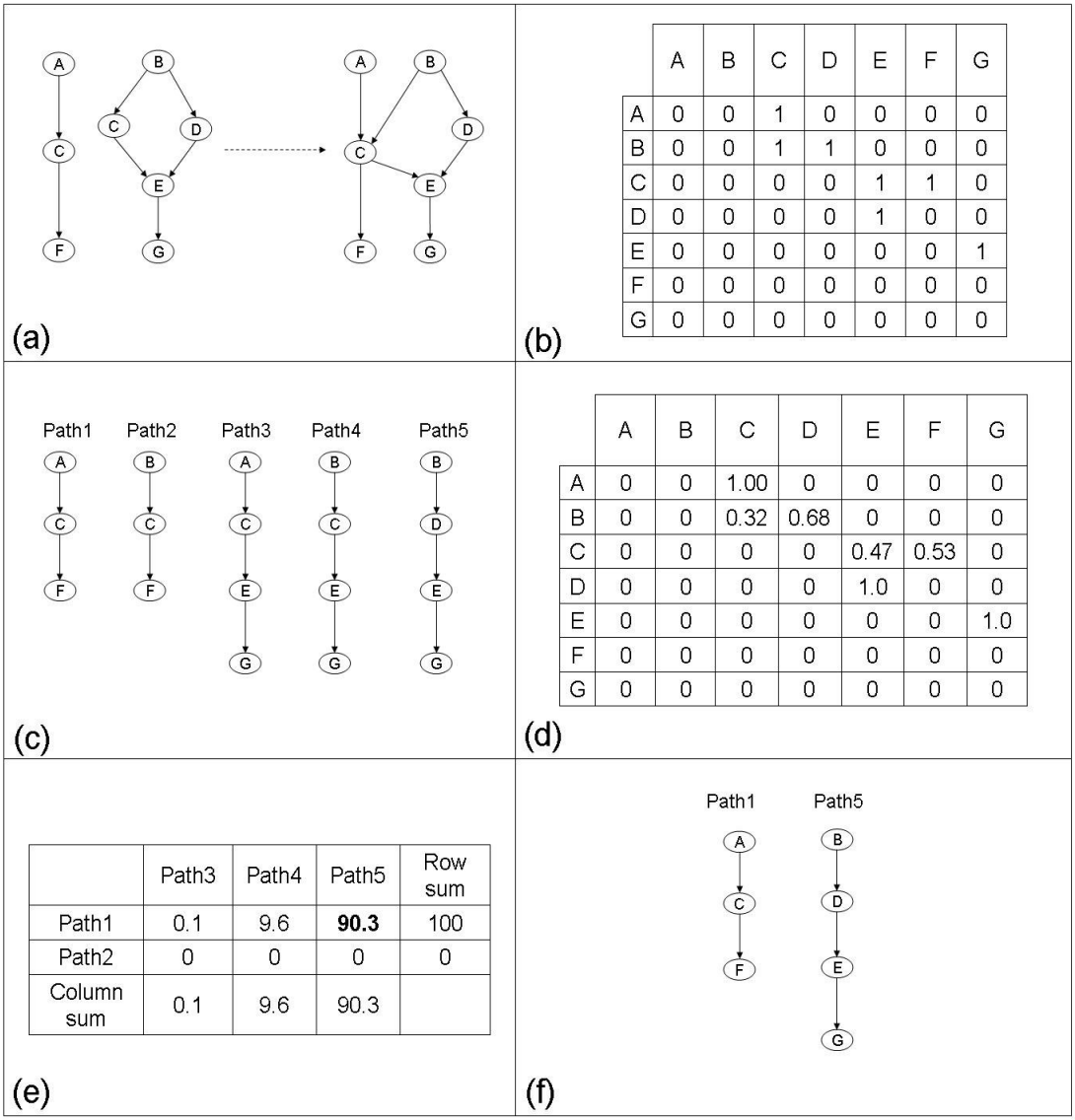
Figure 3.1. The interwined structure of the reconstructed MAPK network including the four functionally distinct MAPK pathways in *Saccharomyces cerevisiae*.

Perception of environmental conditions is maintained by existence of nine receptors (Ste2, Ste3, Ras2, Sho1, Sln1, Wsc1, Wsc2, Wsc3, and Mid2) that are able to sense several kinds of ligands including cytokines, growth factors, hormones, and cellular stress. The transduction of the perceived signal to the eleven regulatory components, i.e. transcription factors (Ste12, Tec1, Msn2, Msn4, Hot1, Sko1, Far1, Fus1, Rlm1, Swi4, and Swi6), is sustained through several signaling mechanisms, i.e. MAPK cascades. The basic assembly of a MAPK cascade

is a three-component cascade of sequentially activated kinases: a MAP kinase kinase kinase (MAPKKK or MEKK), a MAP kinase kinase (MAPKK or MEK), and a MAP kinase (MAPK). The sequential activation is achieved by tethering to scaffold proteins, e.g. Ste5, as well as direct interaction between kinases of the cascade. Organization into MAPK cascades ensures segregation of the pathways from other signaling events in the cell and also allows the use of a kinase in more than one MAPK module without affecting the specificity of the response (Widmann *et al*., 1999).

3.2. Graph Representation of the MAPK Signaling Network

The MAPK network is represented as a directed acyclic graph (DAG), which is composed of a set of nodes (signaling proteins) and a set of directed edges (protein-protein interactions indicating the direction of signal transmission) on the set of nodes (Figure 3.2.a). The graph containing 49 nodes is represented by a binary, square adjacency matrix, A, of dimensions 49x49. A(i,j)=1 if there is an edge from node i to node j; otherwise, A(i,j)=0 (Figure 3.2.b). The directed graph of the MAPK network is made up of 9 inputs (receptors) and 11 outputs (transcription factors).



(b)

	A	B	C	D	E	F	G
A	0	0	1	0	0	0	0
B	0	0	1	1	0	0	0
C	0	0	0	0	1	1	0
D	0	0	0	0	1	0	0
E	0	0	0	0	0	0	1
F	0	0	0	0	0	0	0
G	0	0	0	0	0	0	0

(d)

	A	B	C	D	E	F	G
A	0	0	1.00	0	0	0	0
B	0	0	0.32	0.68	0	0	0
C	0	0	0	0	0.47	0.53	0
D	0	0	0	0	1.0	0	0
E	0	0	0	0	0	0	1.0
F	0	0	0	0	0	0	0
G	0	0	0	0	0	0	0

(e)

	Path3	Path4	Path5	Row sum
Path1	0.1	9.6	**90.3**	100
Path2	0	0	0	0
Column sum	0.1	9.6	90.3	

Figure 3.2. Illustrative example of the methodology. (a) The pathways are represented by a directed acyclic graph (DAG), in which nodes represent signaling proteins and the directed edges represent the physical interactions. (b) The binary adjacency matrix (A) representing the network. (c) The set of linear paths connecting each input to output (conically independent basis vectors). (d) A normalized, uniformly-distributed random matrix (S) with exactly the same topology as the adjacency matrix. (e) The payoff matrix (PM) obtained by the Bayesian Network model for the output pair. The pareto-optimal NEQ (most probable strategy pair) is represented in bold. (f) The optimal strategies given by the NEQ.

3.3. MAPK Signaling Network as a System of Linear Paths

The time-invariant topological structure of the signaling network can be mathematically captured as a system of linear paths that characterize the phenotypic potential of the network. Previously, the topology of biological networks is investigated via a set of fundamental routes (pathways) in the networks (Schuster and Hilgetag, 1994). This set of routes may include all possible linear paths from inputs to outputs (Mavrovouniotis, 1995; Seressiotis and Bailey, 1988; Steffen *et al*., 2002) or conically independent basis vectors (called "elementary flux modes" or "extreme pathways") in solution space (Schilling and Palsson, 1998; Schuster and Hilgetag, 1994; Schuster *et al*., 1996). Given the proteins and the interactions among them, both sets as obtained from literature, together with the start and end terminals, the set of linear paths connecting each input to output were calculated using NetSearch algorithm (Steffen *et al*., 2002) so that any network state can be defined as a nonnegative, linear combination of them (Figure 3.2.c).

3.4. Bayesian Network Model (BN)

A probabilistic graphical model called Bayesian Network (BN) is used to represent the probabilistic relationships between interacting proteins. The most important advantage of using BN is that an edge between two nodes can be taken as an indicator of causality, i.e. how causes generate effects, on a solid mathematical basis (Pearl, 2000). Bayesian Network is represented by a directed acyclic graph (DAG), a family of conditional probability distributions (CPD) and their parameters. In BN formulation, the proteins (nodes) were represented as random variables. BN associates a conditional probability $P(X_i|U_i)$ with each variable X_i, where $U_i \subseteq X$ is the set of parents of X_i. The conditional independence between proteins was represented by the absence of edges between these proteins. In BN, the conditional independence relationship states that a node is independent of its ancestors given its parents, where the ancestor/parent relationship is with respect to some fixed topological ordering of the nodes. The joint distribution over a set of random

variables, $X = \{X_1, X_2, ..., X_n\}$, is represented as a product of conditional probabilities. Then, the joint distribution is in the following form:

$$P(X_1, X_2, ..., X_n) = \prod_i P(X_i | U_i) \quad (3.1)$$

The DAG structure defines a unique rule for expanding the joint probability in terms of simpler conditional probabilities (Jensen, 2001), i.e. the product decomposition is guarantied to be a coherent probability distribution (Friedman, 2004).

3.5. Strategic Game Representation – BN coupled with Game Theory

The strategic game is a model of competitive players consisting of (i) a set of players, (ii) a set of strategies for each player, and (iii) preferences over the set of strategies for each player.

In the present study, the targets of the MAPK network, which are the transcription factors, are considered as the players. Considering the structure of the network (Figure 3.1) consisting of four different MAPK pathways in yeast, the players may be (i) co-operating, i.e. the TFs are functioning in the same pathway, (ii) competitive, i.e. the TFs are functioning in different but cross-talking pathways, or (iii) independent, i.e. the TFs are functioning in different and non-cross-talking pathways. As the set of strategies for each player (i.e. TF), all possible alternative routes (i.e. all possible linear paths ending at the player) are used in which the signal can be transmitted to the specified player from any of the receptors. The preferences over the set of strategies for each player are defined as the likelihood of preferring a specific linear path with respect to others.

Assuming simultaneous activation of all receptors by the same amount of ligand (independent of the type of the ligand) and guaranteeing that receptor saturation does not occur (i.e., the ligand concentrations are below the binding capacity of all receptors), the BN model is used to simulate the transmission of signal from an upstream protein to any of the possible succeeding

proteins. The conditional probability distribution at each node is assumed to be uniform between the interval [0, 1]. During simulation, these probabilities are assigned by computer based on uniform distribution for each of the succesive runs. The conditional probabilities are described by a uniformly-distributed random matrix, S, with exactly the same topology as the adjacency matrix, A (Figure 3.2.d). Since each entry S_{ij} describes the probability of signal transmission from protein *i* to protein *j*, the random matrix is normalized to ensure that total probability mass is one (i.e. summation of elements in each row is restricted to be equal to 1). In order to use a matrix representation for the game, the players, i.e. transcription factors, are considered pair-wise. For each of the linear paths related to the considered transcription factor pair, the joint probability of signal termination at the target transcription factor is calculated by Equation (3.1), i.e. by multiplication of all conditional probabilities of adjacent protein-pairs on the path; and the linear path pair yielding the maximum signal transmission to both of the outputs is considered as the preferred strategy. This procedure is repeated for 1000 times, which is found to be a sufficient number in a range up to 50000 such that the results does not change significantly, i.e. the results can be easily reproduced. The payoff matrix, PM, is created for each output pair (Figure 3.2.e) where the rows of the PM correspond to the preferred strategies of the first player (transcription factor) and the columns correspond to those of the second player. The PM indicates the density of preferred strategies, i.e. PM_{ij} = (number of times the strategy pair (i,j) is preferred)/1000. The present mathematical formulation assures that the payoffs of the two players are the same.

3.6. Solution Algorithm

In order to determine the most-preferred strategies in a game consisting of two transcription factors as players, the solution algorithm uses the payoff matrix considering the basic principles of game theory:

i. *Elimination of dominated strategies:* A strategy S *dominates* the other strategy T if every outcome in S is at least as good as the corresponding outcome in T, and at least one outcome in S is strictly better than the corresponding outcome in T. A rational player should never play a strategy dominated by any of its other strategies.

ii. *Saddle point search*: An outcome in a matrix game is called a *saddle point* if the entry at that outcome is both less than or equal to any entry in its row, and greater than or equal to any entry in its column. If a matrix game has a saddle point, both players should play a strategy which contains it. The game does not necessarily contain a saddle point.

iii. *Identification of Nash Equilibrium (NEQ):* The outcome in a strategic game is called *Nash equilibrium* if the outcome of each player is at least as good as the other outcomes according to the player's preferences. A strategic game necessarily has at least one Nash equilibrium. Due to the structure of the payoff matrix, NEQ was observed at the most preferred strategy pair for every player and every outcome of these players.

iv. *Analyzing the Pareto-optimality and multiplicity of NEQ*: An outcome of a game is *Pareto-optimal* if there is no other outcome which would give both players higher payoffs, or would give one player the same payoff but the other player a higher payoff. To be acceptable as a solution to a game, an outcome should be Pareto-optimal. Due to the structure of the payoff matrix, NEQ is always Pareto-optimal. However, multiple NEQ may be present indicating same payoffs for the players.

4. RESULTS AND DISCUSSION

The topological analysis of signaling networks revealed numerous alternatives of signal transmission mechanisms from receptors to target transcription factors, most of these mechanisms are inactive in reality. In the present work, game theoretical concepts were adapted to bioinformatics to question the rationality of the "optimal" choice of pathway structure in real signaling networks. Game theory is thus here used to predict situations when one type of behavior or specific combination of them is more feasible than all known alternatives taking into account the preferences of all players.

Eleven target transcription factors (Ste12, Tec1, Msn2, Msn4, Hot1, Sko1, Far1, Fus1, Rlm1, Swi4, and Swi6) of MAPK signaling network are considered as competitive players. All possible linear paths starting from any receptor and ending at each transcription factor are used as the set of strategies for each player, i.e. transcription factor. Using the Bayesian Network (BN) model, the likelihood of preferring a specific linear path with respect to others is determined in pair-wise games.

The cell wall integrity pathway neither shares an input signal nor shows a cross-talk with the other MAPK pathways (Figure 3.1), and it seems to work independent of other pathways. Therefore, the comparison of its transcription factors (Rlm1, Swi4 and Swi6) with others was not meaningful in the game theoretical approach as the players were not competing. Among the remaining eight players, the transcription factors of the high osmolarity signaling pathway (Msn2, Msn4, Hot1, and Sko1) have the same strategy set; i.e. they are mathematically identical. Consequently, five different players (Ste12, Tec1, Far1, Fus1, and Msn2 - representing any of the Msn2/Msn4/Hot1/Sko1) are considered and all possible pair-wise games (ten games) among them are analyzed.

4.1. Crosstalk Analysis in the Protein Interaction Network

Perception of environmental conditions in cells is maintained by existence of a great variety of receptors that are able to sense several kinds of stimuli. The transduction of the perceived signals to the regulatory machinery is sustained through several signaling pathways triggering corresponding responses. Cells must be able to process multiple signals in parallel, each relaying its signal to the corresponding response specifically. The required specificity of signaling would seem to dictate that parallel cascades function independently. However, cells must also integrate the signals in order to trigger the appropriate response and this is achieved by the interaction between the signaling pathways, a phenomenon called 'crosstalk'. This was the case in the reconstruction attempt of MAPK network which resulted in an intertwined network (Figure 3.1) of four MAPK pathways, rather than parallel streaming of independent signaling pathways.

Many different definitions and measures of crosstalk have been described in the literature taking into account topological, structural and dynamical aspects (Binder and Heinrich, 2004; Cowan and Storey, 2003; Komarova *et al*., 2005; Papin and Palsson, 2004; Schwartz and Baron, 1999; Schwartz and Madhani, 2004; Somsen *et al*., 2002). Here, the classical definition is accepted in which sharing of identical signaling molecules in different signaling pathways is used as the measure of crosstalk (Schwartz and Baron, 1999).

There are several mechanisms by which the three MAPK pathways (high-osmolarity signaling pathway, mating pheromone response pathway and filamentous growth and invasion pathway) crosstalk, i.e. interact with each other :

a. The plasma membrane protein Sho1 containing an SH3 domain acts as receptor of both high-osmolarity and filamentous growth and invasion pathways.
b. The Cdc42 protein, a rho-like GTPase which is essential for establishment and maintenance of cell polarity, functions in all of the three pathways by activating the signal transducing kinase Ste20.

c. The Ste20 protein, which is a signal transducing kinase of the p21-activated kinase family, is involved in all of the three pathways by activating the MAPK cascades.
d. The signal transducing MEK kinase Ste11 functions in upstream element of all of the three MAPK cascades by phosphorylating either the Ste7 protein or the Pbs2 protein.
e. The Ste7 protein acts as the MAPKK in cascades involved in both of the mating pheromone response pathway and filamentous growth and invasion pathway, phosphorylating Fus3 and Kss1 proteins, respectively.
f. The transcription factor Ste12 activates genes involved in both the mating pheromone response pathway and filamentous growth and invasion pathway, and it is activated by the corresponding MAP kinase signaling cascades.

4.2. Traditional Optimization with Bayesian Network : Preferred Strategies of Players when the Preferences of Other Players are Not Considered

In order to simulate the transmission of signal from a set of receptors to a specific transcription factor, thousand simulations, found to be adequate for the saturation level, were performed using the BN model. For each of the linear paths related to the considered transcription factor, the joint probability of signal termination at the target transcription factor is calculated by Equation (3.1), and the linear path yielding the maximum signal transmission to the specific transcription factor is considered as the preferred strategy. Then, the vectors indicating the density of preferred strategies (number of times the strategy is preferred/total number of simulations, i.e. thousand) is calculated for each of the transcription factor.

As a result of the intertwined structure of the MAPK network, the BN simulations showed that the signal could be transduced by any nonnegative linear combination of the linear paths connecting the receptors to the target transcription factors, i.e. all strategies were probable, since for any of the transcription factors, the vectors indicating the density of preferred strategies were consisting of non-zero elements. Furthermore, for several transcription factors, a linear path indicating a signal transmission mechanism that is not biologically active was observed to be the most-probable path. For example, the most-probable path for the Msn2 transcription factor was pointing out a biologically inactive signal transmission from the

receptor of the filamentous growth and invasion pathway, Ras2, to the transcription factor of high osmolarity signaling pathway, Msn2. Additionally, in several cases, as a result of the cross-talks by other signaling pathways, the Bayesian network model produced a highly-probable signal transduction to the corresponding transcription factor even though an essential protein was not functioning in the signaling pathway. In order to test the effect of protein malfunctioning, the simulations were repeated with the exclusion of the upstream proteins of the high osmolarity signaling pathway, namely, Sln1, Ypd1, and Ssk1. It was observed that there were still highly-probable linear paths representing the ability of the osmoregulation MAPK Hog1 to transduce the signal to the targets, namely, the four transcription factors Msn2, Msn4, Hot1 and Sko1, as a result of the crosstalk through the signal transducing MEK kinase Ste11.

The above results indicated that the traditional optimization framework, in which preferred strategies of players were independent of those of other players, is not sufficient to represent real biology. To avoid false-positive signal transmission mechanisms, there is a need for a framework in which the preferences of other players could also be regarded in determining the optimum behavior.

4.3. Nash Equilibrium Represents the Real Biology : Preferred Strategies of Players when the Preferences of Other Players are Also Considered

The BN model coupled with game theory is here used to simulate the transmission of signals from an upstream protein to any of the possible succeeding proteins taking the preferences of other players into consideration. In order to use a matrix representation for the game, the players, i.e. transcription factors, are considered pair-wise. For each of the linear paths related to the considered transcription factor pair, the joint probability of signal termination at the target transcription factor is calculated by Equation (3.1), and the linear path pair yielding the maximum signal transmission to both of the outputs is considered as the preferred strategy-pair. The payoff matrix indicating the density of preferred strategies is created by thousand simulations for each output pair, and using the payoff matrix, Nash Equilibrium (NEQ) of each game is identified (see materials and methods for details).

In almost all cases (eight out of the ten games), the resultant NEQ were successfully representing the real biology. Here three possible cases will be exemplified:

(i) Co-operating TFs, i.e. the TFs are functioning in the same pathway: The Ste12 vs. Tec1 game resulted in a NEQ, which represents the Sho1-mediated branch of the filamentous growth and invasion pathway (Figure 4.1.a). Ste12 and Tec1 are co-operating players in this pathway and working in concert to cooperatively bind to filamentation response elements (FREs) in genes involved in filamentous growth (Madhani and Fink, 1997). In all the simulations, the strategies representing the real biology of the Ras2-mediated branch of the filamentous growth and invasion pathway were dominated by those of Sho1-mediated branch. Therefore, in none of the games the Ras2-mediated branch of the filamentous growth and invasion pathway was represented in the resultant NEQ. Interestingly, Sho1-mediated branch is more preferred compared to that of Ras2.

(ii) Competing TFs, i.e. the TFs are functioning in different but cross-talking pathways: In Tec1 vs. Far1 game, the full structure of the mating pheromone response pathway and Sho1-mediated branch of the filamentous growth and invasion pathway were successfully represented in the resultant NEQ (Figure 4.1.b), even though these two pathways contain almost all of the cross-talking elements (Cdc42, Ste20, Ste11, Ste7, and Ste12) present in the MAPK network.

(iii) Independent TFs, i.e. the TFs are functioning in different and non-cross-talking pathways: In the absence of any crosstalk between pathways, the game theoretical analysis proposed here also resulted in successful reconstructions. For example, the Sho1-mediated branch of filamentous growth and invasion pathway and the Sln1-mediated branch of the high osmolarity signaling pathway were successfully represented in the resultant NEQ of the Tec1 vs. Msn2 game (Figure 4.1.c).

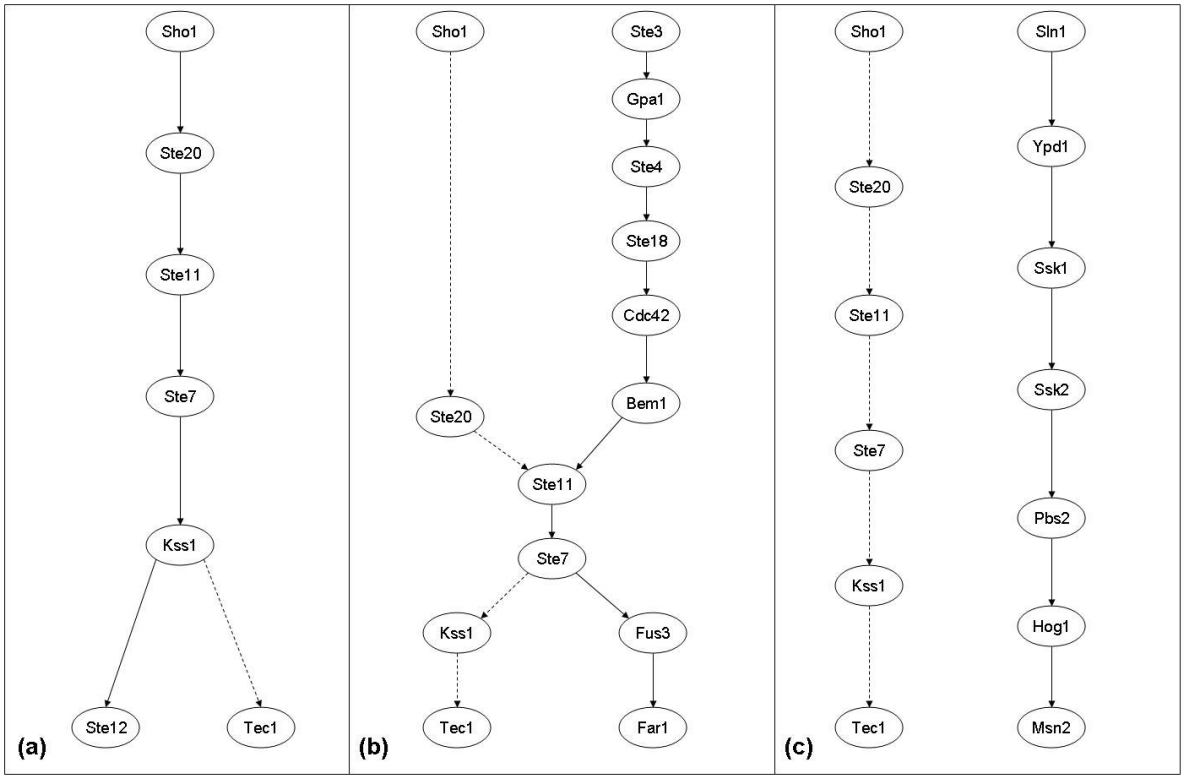


Figure 4.1. The most preferred strategies represented by the NEQ of (a) Ste12 vs. Tec1 game between co-operating players in the filamentous growth and invasion pathway, (b) Tec1 vs. Far1 game between the players which are acting in different pathways showing significant crosstalks. The Tec1-specific links in its preferred strategy are shown by dashed lines, and (c) Tec1 vs. Msn2 game between the players which are acting in branches of different pathways without crosstalk.

4.4. Bottlenecks of the Network Model and Troubleshooting

In the two games (Tec1 vs. Fus1 and Far1 vs. Msn2), the resultant NEQ included unexpected signal transduction from the receptor Ste3 to the Bmh1/2 proteins (Figure 4.2.ab) although neither Bmh1 nor Bmh2 has been reported to have a function in the mating pheromone response pathway. The presence of the unexpected link Cdc42-Bmh1/2-Ste11 in the resultant NEQ was mainly due to the crosstalk between mating pheromone response pathway and filamentous growth and invasion pathway. In fact, Cdc42 functions in all of the three

pathways, namely high-osmolarity signaling pathway, mating pheromone response pathway and filamentous growth and invasion pathway; and it activates the signal transducing kinase Ste20 either directly or indirectly through Bem1. Cdc42 also interacts with Bmh1/2 which is an essential binding protein that activates the MEK kinase Ste11 in Ras2-mediated signaling during pseudohyphal growth.

In principle, protein-protein interaction data are directly employed to reconstruct signaling networks. However, the full potential of these data cannot be utilized for discovering signal transduction networks for several reasons: (i) The incompleteness of the reconstructed network model due to lack of data and low reliability of the protein-protein interaction data due to high noise levels may lead to observations of several false-positive signal transduction mechanisms. To increase data reliability, only 49 proteins are taken into consideration, which are already annotated to have function in the above-mentioned MAPK pathways in the MIPS-CYGD functional catalogue. However, most probably, there may be several other proteins functioning in the network but not yet reported in the databanks due to the lack of experimental evidence. (ii) Lack of suitable mathematical models for representation and analysis of signaling networks is another bottleneck. The biological system is here represented as a protein-protein interaction network in which all components are homogeneously distributed over space, i.e. there is no physical separation or compartmentation; but, this system does not have the potential of representing the scaffolding and inhibitory events.

Considering the above-mentioned games (Tec1 vs. Fus1 and Far1 vs. Msn2), experimental evidence indicates that the interaction of Ste20 with Bem1 is required for association of scaffolding protein Ste5 with the Ste11-Ste7-Fus3 MAPK cascade in the mating pheromone response signaling pathway (Van Drogen *et al*., 2001). However, the reconstructed model network could not represent the effect of scaffolding. Therefore, the inclusion of the unexpected signal transductions from the receptor Ste3 to the Bmh1/2 proteins is indispensable in the resultant NEQ.

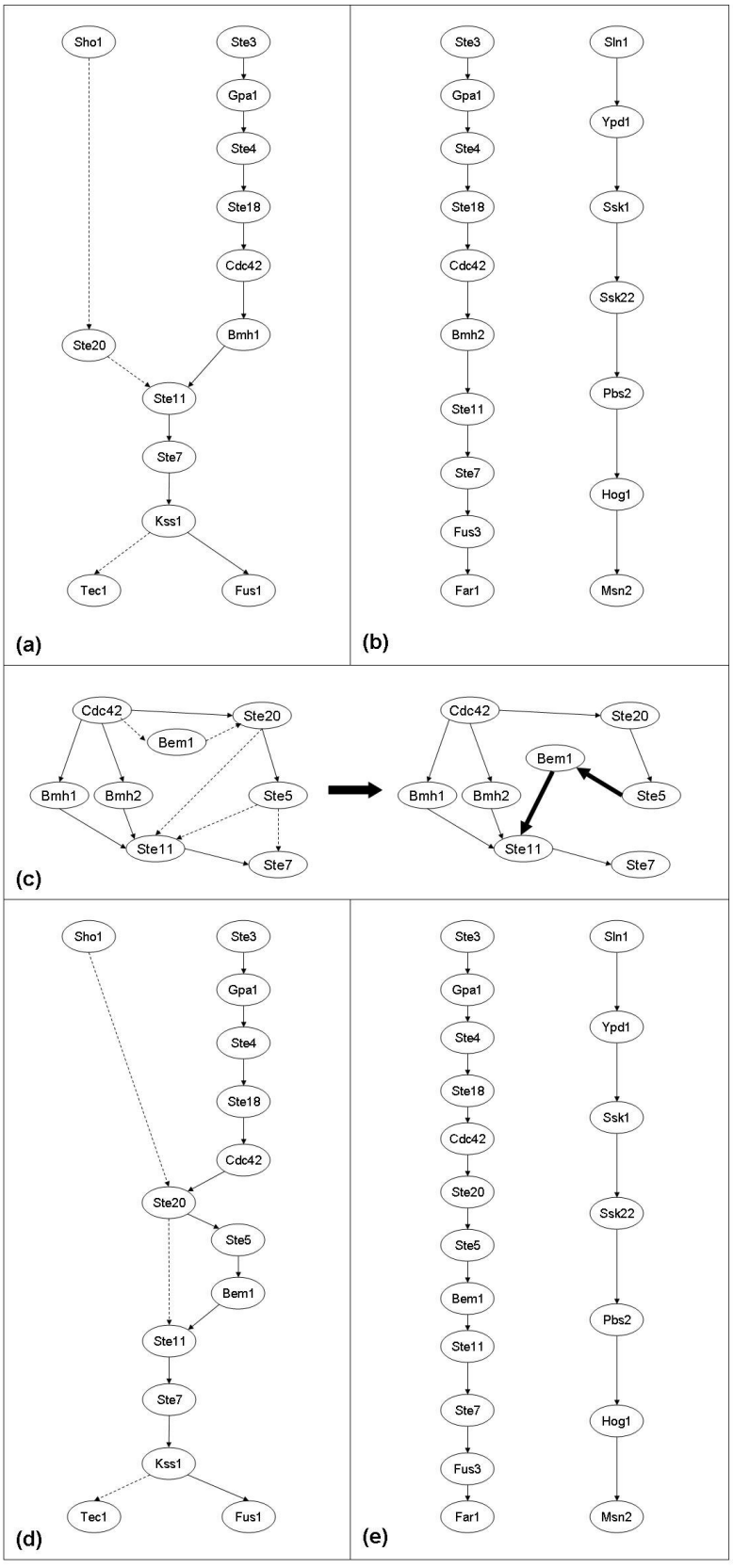
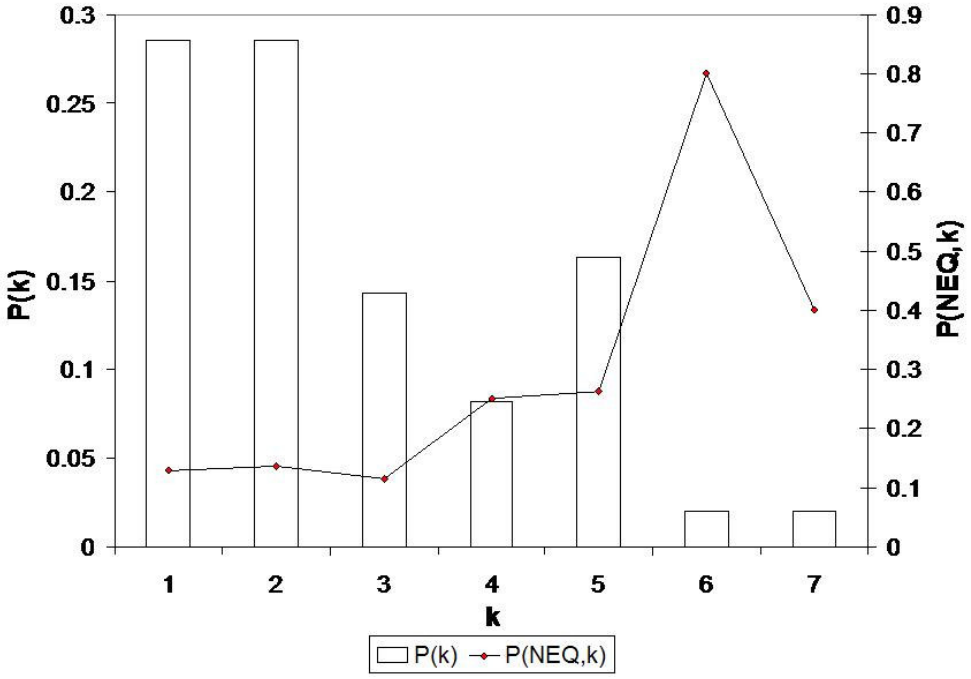

Sho1
Ste3
Gpa1
Ste4
Ste18
Cdc42
Bmh1
Ste20
Ste11
Ste7
Kss1
Tec1
Fus1
(a)
Ste3
Gpa1
Ste4
Ste18
Cdc42
Bmh2
Ste11
Ste7
Fus3
Far1
Sln1
Ypd1
Ssk1
Ssk22
Pbs2
Hog1
Msn2
(b)
Cdc42
Ste20
Bem1
Bmh1
Bmh2
Ste5
Ste11
Ste7
(c)
Sho1
Ste3
Gpa1
Ste4
Ste18
Cdc42
Ste20
Ste5
Bem1
Ste11
Ste7
Kss1
Tec1
Fus1
(d)
Ste3
Gpa1
Ste4
Ste18
Cdc42
Ste20
Ste5
Bem1
Ste11
Ste7
Fus3
Far1
Sln1
Ypd1
Ssk1
Ssk22
Pbs2
Hog1
Msn2
(e)

Figure 4.2. (a) The resultant NEQ of the Tec1 vs. Fus1 game which includes an infeasible signal transduction from the receptor Ste3 to the Bmh1 protein. (b) The NEQ of the Far1 vs. Msn2 game indicating an infeasible signal transduction from the receptor Ste3 to the Bmh2 protein. (c) The refinement in the network. Removed interactions are represented by dashed lines, whereas the newly added interactions are given in bold. (d) The resultant NEQ of the Tec1 vs. Fus1 game using the refined network. (e) The resultant NEQ of the Far1 vs. Msn2 game using the refined network.

In order to test whether the inclusion of the experimental evidence into the model can eliminate the unexpected links in the NEQ, the analysis were repeated for the corresponding games (Tec1 vs. Fus1 and Far1 vs. Msn2) by refining the network as follows: two interactions, Ste5-Bem1 and Bem1-Ste11, are included and the direct interactions Ste5-Ste11, Ste5-Ste7 and Ste20-Ste11 are removed (Figure 4.2.c), so that Bem1 was strictly required for association of Ste5 with the proteins in the MAPK cascade (Ste11 and Ste7) of the mating pheromone response signaling pathway. Since the network should be represented by a directed acyclic graph, the interactions Cdc42-Bem1 and Bem1-Ste20 were also removed in order to avoid formation of a cycle between Bem1, Ste20 and Ste5. By this way, the interaction of Ste20 with Bem1 was assumed to be achieved by sequential binding via Ste5. For both of the games, the analysis of this hypothetical network resulted in NEQ, where the false-positives are eliminated and the real biology is better represented by replacing the link of Cdc42-Bmh1/2-Ste11 protein with Cdc42-Ste20-Ste5-Bem1-Ste11 (Figure 4.2.de).

4.5. Topological Parameters as Indicators of NEQ

It is hypothesized that the resultant preferences of each player in the analysis may be due to the topological parameters of the components in the network. Therefore the existance of any correlation between the preference of the path and the connectivity of the proteins in the path or the path length is investigated.


0.3
0.25
0.2
0.15
0.1
0.05
0
P(k)
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
P(NEQ,k)
1
2
3
4
5
6
7
k
P(k)
P(NEQ,k)

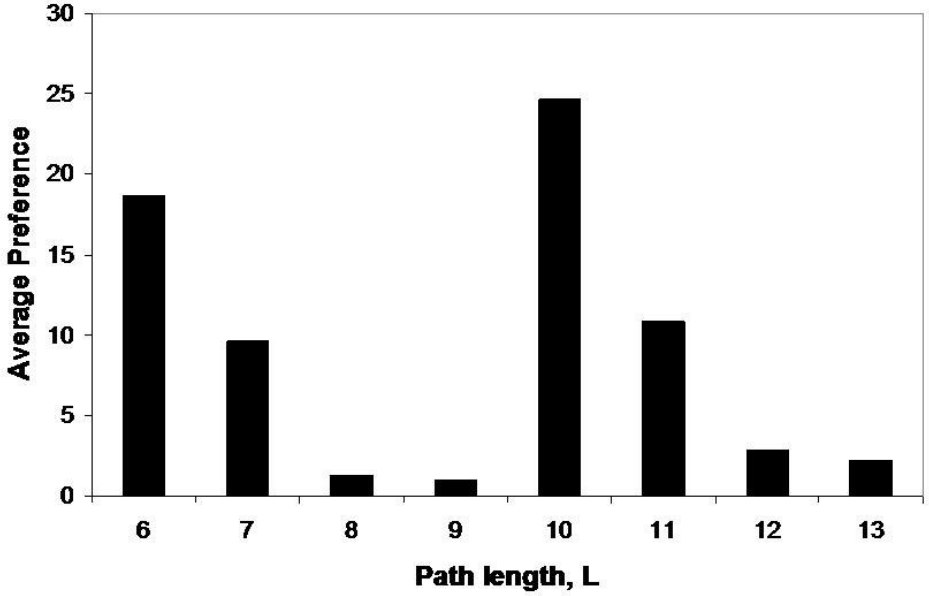

30
25
20
15
10
5
0
Average Preference
6
7
8
9
10
11
12
13
Path length, L

Figure 4.3. Analysis of network topology. (a) The trends of probability of having a connectivity k, P(k), and the probability of contribution of a protein with connectivity k to the resultant Nash equilibria, P(NEQ,k) with respect to the connectivity, k. (b) The correlation between the path length and average preference of a path with a path length of L.

Any correlation between the connectivity and the preference of a protein in the network was tested by determining the probability distribution of having a connectivity k, P(k). Although there seems to be a negative correlation between connectivity and the probability of having connectivity k, P(k); this was not statistically proved (Figure 4.3.a). The probability of contribution of a protein with connectivity k to the resultant Nash equilibria, P(NEQ,k), is calculated. Similarly, no significant correlation between connectivity, k, and probability of contribution of a protein with connectivity k to the resultant Nash equilibria, P(NEQ,k), was found although a trend can be observed in Figure 4.3.a.

Any correlation between path length and the preference of a path in the network was also investigated by calculating the average preferences of paths having a path length of L using the payoff matrices of all games. A significant correlation could not be observed between the path length and average preference of a path with a path length L (Figure 4.3.b).

These results indicated that the preferences of each protein cannot be considered as directly related to its connectivity and the preference of each strategy, i.e. linear path, is not correlated with its path length.

4.6. Effect of Scaffold Proteins

Physical segregation provided by several scaffold proteins is proposed as an explanation for the achievement of the highly specific response to available signals in spite of the uncertainties due to the crosstalk (Garrington & Johnston, 1999; Posas *et al*., 1998; Yashar *et al*., 1995). However, just looking at the available knowledge on the structure and topology of the

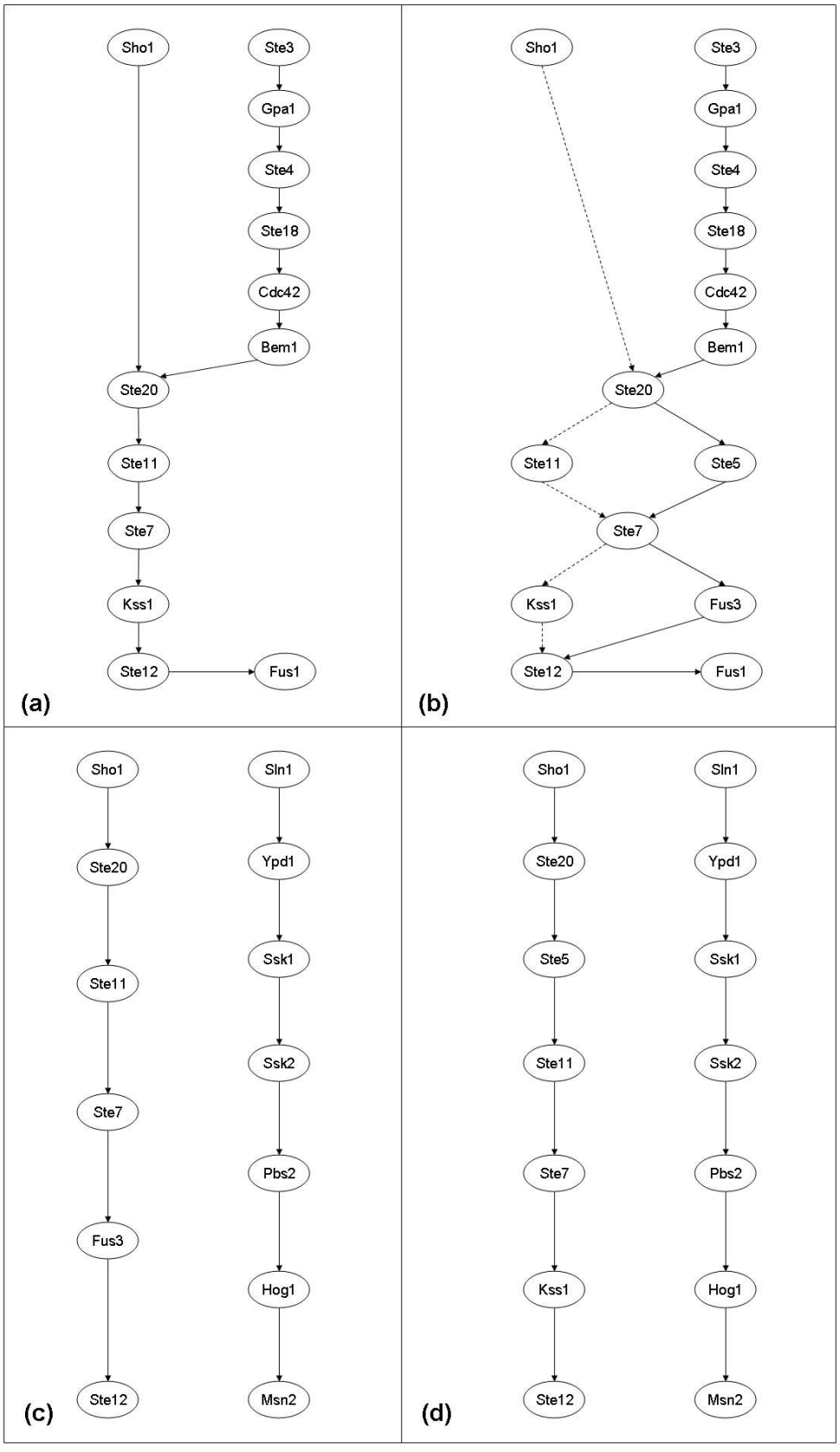

Sho1
Ste3
Gpa1
Ste4
Ste18
Cdc42
Bem1
Ste20
Ste11
Ste7
Kss1
Ste12
Fus1
(a)
Sho1
Ste3
Gpa1
Ste4
Ste18
Cdc42
Bem1
Ste20
Ste11
Ste5
Ste7
Kss1
Fus3
Ste12
Fus1
(b)
Sho1
Sln1
Ste20
Ypd1
Ste11
Ssk1
Ste7
Ssk2
Fus3
Pbs2
Hog1
Ste12
Msn2
(c)
Sho1
Sln1
Ste20
Ypd1
Ste5
Ssk1
Ste11
Ssk2
Ste7
Pbs2
Kss1
Hog1
Ste12
Msn2
(d)

Figure 4.4. The resultant NEQ of the Ste12 vs. Fus1 game in (a) the absence of the scaffold protein Ste5, (b) the presence of the scaffold protein Ste5. The resultant NEQ of the Ste12 vs. Msn2 game in (c) the absence of the scaffold protein Ste5, (d) the presence of the scaffold protein Ste5.

network, current computational frameworks based on protein interaction networks in which all components are homogeneously distributed over space, i.e. there is no physical separation or compartmentation, cannot give successful reconstruction of individual pathways. The adaptation of game theoretical concepts into bioinformatics can give an idea on the underlying design principles of this type of networks and lead to successful reconstruction of biological pathways.

Despite the presence of scaffold proteins in the protein interaction network, the preferred strategy analysis by the BN model, without considering the preferences of other players (transcription factors), gave rise to numerous alternatives of signal transmission mechanisms from receptors to target transcription factors, however most of these mechanisms are not active in reality. The biological system represented here as the protein interaction network does not have the potential of representing the scaffolding and inhibitory events. In order to test the effect of the scaffolding proteins in the network, the scaffold protein, Ste5, is removed from the network. The scaffold protein Ste5 is able to co-localize all the members of the MAPK cascade required for mating (Ste11-Ste7-Fus3) by simultaneous binding, and also enhance the activity of the Fus3 MAPK (Van Drogen *et al*., 2001). Ste7 protein activated by Ste11 phosphorylates Kss1 in the filamentous growth and invasion pathway (Widmann *et al*., 1999).

Six games (Ste12 vs. Far1, Ste12 vs. Fus1, Ste12 vs. Msn2, Tec1 vs. Fus1, Far1 vs. Fus1, and Fus1 vs. Msn2) related to at least one of these cross-talking pathways (filamentous growth and invasion pathway and the mating pheromone response pathway) were performed in the absence of the scaffold protein Ste5 in the network. They were found to yield NEQ with the unexpected signal transductions. In the Ste12 vs. Fus1 game without the scaffold protein Ste5, the NEQ points out an unfeasible strategy for Fus1 (Figure 4.4.a) including unexpected signal

transmission from the receptor Ste3 to the MAPK Kss1. The inclusion of the scaffold protein Ste5 into the network enabled the link Ste11-Ste7-Kss1 to be replaced by a linear path between Ste5-Ste7-Fus3 (Figure 4.4.b). In another game, Ste12 vs. Msn2, the absence of the scaffold protein Ste5 in the network resulted in a NEQ with unexpected signal transduction from the receptor Sho1 to the MAPK Fus3 (Figure 4.4.c). Similarly, inclusion of Ste5 into the network replaced the link Ste11-Ste7-Fus3 with that of Ste5-Ste11-Ste7-Kss1 (Figure 4.4.d), which represents the reality. Inclusion of the scaffold protein Ste5 into the network solved the problem of false-positive signals.

4.7. Network Topology from an Evolutionary Perspective

In the last decade, topological analysis of many biological networks indicated that biological networks have a "specialized" topology characterized by highly-connected proteins, scale-free degree distribution and small-world characteristics (Jeong *et al*., 2001; Wagner, 2001; Maslov and Sneppen, 2002). In addition, particularly as a result of high-throughput experimentation, highly complex protein-protein interaction graphs are obtained for several species. These findings gave rise to several new questions to be answered: What are the underlying organizational principles in this complex map of interactions and how this "specialized" topology contributes on the network function? Development of several network-growing models, such as preferential attachment (Barabasi and Albert, 1999), indicated clues on understanding of how the scale-free topology evolutionary emerges. However, the effect of the topology on the efficiency of the network function is still unknown. Since cellular systems are a result of Darwinian evolution, a promising approach should be the investigation of the answers of the above question in this view.

The linear path analysis indicated that, even with the simple MAPK network used in this study, there are a huge number of possibilities in transmitting the signal from a receptor to a transcription factor. However, previous experimental evidences showed that a very limited number of them were biologically active and traditional optimization was not sufficient to explain this fact. In the present study, by adapting a game theoretical formulation to the problem, the false-positive signaling routes were eliminated and the biology was successfully

represented. The proposed methodology with further integration of biological data will give more insight about the underlying principle in evolutionary construction of network topology.

5. CONCLUSIONS AND RECOMMENDATIONS

5.1. Conclusions

Cellular signaling is central for gaining insight into the molecular mechanisms behind diseases as well as adaptation of living cells to changes in the environment. Signaling pathways are often branched in an interconnected fashion and are therefore integrated into signaling networks that are quite complex with many levels of interconnectivity of different molecular components. Recently, it became apparent that each MAPK pathway is a part of a network in which there is extensive sharing of signaling elements among the MAPK signaling pathways. Understanding the design principles that bridge the topology to the function of the network is a major challenge in systems biology since almost all known diseases exhibit dysfunctional aspects in these signaling networks. In the present study, considering the known network topology of MAPK signaling in yeast, the traditional optimization method is found to be unsuccessful since it provides false-positive signal transmission mechanisms. The BN model coupled with a game theory based solution algorithm, however, yielded accurate results in eighty per cent of the games in terms of representing the real biology. This method eliminates the false-positive signaling routes resulting from crosstalks since it also considers the preferences of other players in determining the optimum strategies. Using a more suitable network representation describing the scaffolding events, having the ability to deal with cycling systems and a completed databank of molecular interactions, game theoretical analyses will tell more about the underlying principle in evolutionary construction of network topology and enable handling of complex networks crucial for biology.

5.2. Recommendations

The present study aimed to provide a method to obtain active pathways in an unknown network given the proteins present in the system and their possible interaction sets. To this end several choices and assumptions have been made to prove the validity for at least a system.

Further research and developments are possible to achieve a code applicable to more general systems.

- Instead of simultaneous activation of all receptors by the same amount of ligand, receptors can be activated at different times by different amounts of lignad.
- Binding capacity of receptors can be made limited, resulting in receptor saturation.
- The probabilities assigned to protein-protein interactions are based to uniform distribution, this can be checked using several other random distributions.
- For each run of the code one single pathway set has been selected to be dominating. Rather than this Boolean logic approach, fuzzy logic can be applied to sum up the probabilities for each run to obtain the final decision for the dominating pathways.
- Current approach has the advantage of considering two players simultaneously using a game-theoretical approach. This can be improved by considering more players simultaneously.
- Current protein-protein interactions are taken to be single sided, i.e. directional, and acyclic to be more target-oriented. Adding a method to deal with reverse interactions and loops will improve the applicability of the method to a great extent.
- Protein-protein interactions considered are limited with those that are known to be existent. Once the method is developed as mentioned above, the complexity of the system can be increased by taking more interactions that are available in databases.

REFERENCES

Albert, R., 2005. “Scale-free Networks in Cell Biology”, Journal of Cell Science 118, pp. 4947-4957.

Alberts, B., A. Johnson, J. Lewis, M. Raff, K. Roberts and P. Walter, 2002. Molecular Biology of the Cell, Garland Publishing, Fourth Edition, Ch. 15.

Axelrod, R. and W. D. Hamilton, 1981. “The Evolution of Cooperation”, Science 211, pp. 1390-1396.

Barabasi, A. L. and R. Albert, 1999. “Emergence of Scaling in Random Networks”, Science 286, pp. 509-512.

Binder, B. and R. Heinrich, 2004. “Interrelations Between Dyxnamical Properties and Structural Characteristics of Signal Transduction Networks”, Genome Informatics 15, pp. 164-173.

Bishop, D. T. and C. Cannings, 1978. “A Generalized War of Attrition”, J. Theor. Biol. 70, pp. 85-125.

Breitkreutz, B., C. Stark, and M. Tyers, 2003. “The Grid: The General Repository for Interaction Datasets”, Genome Biol. 4, R23.

Brivanlou, A. H. and J. E. Darnell Jr., 2002. “Signal Transduction and the Control of Gene Expression”, Science 295 (5556), pp. 813-818.

Chen, R. E. and J. Thorner, 2007. "Function and Regulation in MAPK Signaling Pathways : Lessons Learned from the Yeast *Saccharomyces cerevisiae*", Biochim. Biophys. Acta (2007), doi:10.1016/j.bbamcr.2007.05.003.

Cooper, G. M., 2000. The Cell, a Molecular Approach, Second Edition, Ch.13.

Cowan, K. J. and K. B. Storey, 2003. "Mitogen-activated Protein Kinases: New Signaling Pathways Functioning in Cellular Responses to Environmental Stress", J. Exp. Biol. 206, pp. 1107-1115.

Downward, J., 2001. "The Ins and Outs of Signaling", Nature, Vol. 411, pp. 759-762.

Frank, S. A., 1996. "Models of Parasite Virulence", Q. Rev. Biol. 71, pp. 37-78.

Frick, T. and S. Schuster, 2003. "An Example of the Prisoner's Dilemma in Biochemistry", Naturwissenschaften 90, pp. 327-331.

Friedman, N., 2004. "Inferring cellular networks using probability graphical models", Science 303, pp. 799-805.

Garrington, T. P. and G. L. Johnson, 1999. "Organization and regulation of mitogen-activated protein kinase signaling pathways", Curr Opin Cell Biol 11, pp. 211-218.

Griffin, A. S., S. A. West and A. Buckling, 2004. "Cooperation and Competition in Pathogenic Bacteria", Nature 430, pp. 1024-1027.

Güldener, U., M. Munsterkotter, G. Kastenmuller, N. Strack, J. van Helden, C. Lemer, J. Richellse, S. J. Wodak, J. Garcia-Martinez, J. E. Perez-Ortin, H. Michael, A. Kaps, E. Talla, B. Dujon, B. Andre, J. L. Souciet, J. De Montigny, E. Bon, C. Gaillardin and H. W. Mewes, 2005. "CYGD: The Comprehensive Yeast Genome Database", Nucleic Acids Res 33, D364-D368.

Gustin, M. C., J. Albertyn, M. Alexander and K. Davenport, 1998. "Map Kinase Pathways in the Yeast *Saccharomyces cerevisiae*", Microbiol Mol Biol Rev 62, pp. 1264-1300.

Han, J. D., N. Bertin, T. Hao, D. S. Goldberg, G. F. Berriz, L. V. Zhang, *et al*., 2004. "Evidence for Dynamically Organized Modularity in the Yeast Protein–Protein Interaction Network", Nature 430, pp. 88–93.

Hamilton, W. D., 1967. "Extraordinary Sex Ratios", Science 156, pp. 477-488.

Holland, J. H., 2002. "Exploring the Evolution of Complexity in Signaling Networks", *Complexity* 7, pp. 34.

Janes, K. A., J. G. Albeck, S. Gaudet, P. K. Sorger, D. A. Lauffenburger, and M. B. Yaffe, 2005. "A Systems Model of Signaling Identifies a Molecular Basis Set for Cytokine-induced Apoptosis", *Science* 310, pp. 1646–1653.

Jensen, F. V., 2001. Bayesian Networks and Decision Graphs, New York: Springer-Verlag.

Jeong, H., S. Mason, A. L. Barabási and Z. Oltvai, 2001. "Lethality and Centrality in Protein Networks", Nature 411, pp. 41-42.

Karin, M., 1990. "Too Many Transcription Factors: Positive and Negative Interactions", New Biol 2 (2), pp. 126-31.

Komarova, N. L., X. Zou, Q. Nie and L. Bardwell, 2005. "A Theoretical Framework for Specificity in Cell Signaling", Mol Syst Biol, doi:10.1038/msb41000.

Levine, D. K., What is Game Theory ?, UCLA, Department of Economics, http://levine.sscnet.ucla.edu/general/whatis.htm

Lewontin, R. C., 1961. “Evolution and the Theory of Games”, J. Theor. Biol. 1, pp. 382-403.

Lodish, H., A. Berk, L. S. Zipursky, P. Matsudaira, D. Baltimore and J. Darnell, 2000. Molecular Cell Biology, Ch.20, W. H. Freeman and Company.

Ma’ayan, A., K. Gardiner and R. Iyengar, 2006. “The Cognitive Phenotype of Down Syndrome: Insights from Intracellular Network Analysis”, NeuroRX, Volume 3, Issue 3, July 2006, pp. 396-406.

Madhani, H. D. and G. R. Fink, 1997. “Combinatorial Control Required for the Specificity of Yeast MAPK Signaling”, Science 275, pp. 1314–1317.

Maslov, S. and K. Sneppen, 2002. “Specificity and Stability in Topology of Protein Networks”, Science 296, pp. 910-913.

Mavrovouniotis, M. L., 1995. “Computational methods for complex metabolic systems: Representation of multiple levels of detail”, In: Lim H. A. and Cantor, C.R. (eds.), Bioinformatics & Genome Research, World Scientific, pp. 265-273.

Maynard Smith, J., 1974. “The Theory of Games and Animal Conflicts”, J. Theor. Biol. 47, pp. 209-221.

Maynard Smith, J. and E. Szathmary, 1993. “The Origin of Chromosomes. I. Selection for linkage”, J. Theor. Biol. 164, pp. 437-446.

Milo, R., S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii and U. Alon, 2002. “Network Motifs : Simple Building Blocks of Complex Networks”, Science 298, pp. 824-827.

Milo, R., S. Itzkovitz, N. Kashtan, R. Levitt, S. Shen-Orr, I. Ayzenshtat, M. Sheffer and U. Alon, 2004. “Superfamilies of Designed and Evolved Networks”, Science 303, pp. 1538-1542.

Murphy, K., and S. Mian, 1999. Modeling Gene Expression Data Using Dynamic Bayesian Networks Technical Report, Computer Science Division, University of California, Berkeley.

Palecek, S. P., A. S. Parikh and S. J. Kron, 2002. "Sensing, Signaling and Integrating Physical Processes during *Saccharomyces cerevisiae* Invasive and Filamentous Growth", Microbiology 148, pp. 893-907.

Papin, J. A. and B. O. Palsson, 2004. "Topological Analysis of Mass-balanced Signaling Networks: a Framework to Obtain Network Properties Including Crosstalk", J. Theor. Biol. 227, pp. 283–297.

Papin, J. and S. Subramaniam, 2004. "Bioinformatics and Cellular Signaling", Current Opinion in Biotechnology 2004, 15, pp. 78–81.

Pearl, J., 2000. Causality, Cambridge.

Pfeiffer, T. and S. Schuster, 2005. "Game-theoretical Approaches to Studying the Evolution of Biochemical Systems", TIBS 30, pp. 20-25.

Pfeiffer, T., S. Schuster and S. Bonhoeffer, 2001. "Cooperation and Competition in the Evolution of ATP-producing Pathways", Science 292, pp. 504-507.

Pitkänen, E., 2007. Modeling and Identification of Biological Networks, Seminar on Computational Systems Biology, Helsinki.

Posas, F. and H. Saito, 1998. "Activation of the Yeast *SSK2* MAP Kinase Kinase Kinase by the *SSK1* Two-Component Response Regulator", EMBO J. 17, pp. 1385-1394.

Promega Corporation, 2000. Signal Transduction Resource, Chapter 1.

Riley, M. A. and J. E. Wertz, 2002. "Bacteriocin Diversity: Ecological and Evolutionary Perspectives", Biochimie 84, pp. 357-364.

Ross, D., 2004. Game Theory, The Stanford Encyclopedia of Philosophy, Winter 2004 Edition.

Sachs, K., O. Perez, D. Pe'er, D. A. Lauffenburger and G. P. Nolan, 2005. "Causal Protein-Signaling Networks Derived from Multiparameter Single-Cell Data", *Science* 308, pp. 523–529.

Salwinski, L., C. S. Miller, A. J. Smith, F. K. Pettit, J. U. Bowie and D. Eisenberg, 2004. "The Database of Interacting Proteins", 2004 update, Nucleic Acids Res 32, D449-451.

Schilling, C.H. and B. O. Palsson, 1998. "The Underlying Pathway Structure of Biochemical Reaction Networks", Proc Natl Acad Sci U S A 95, pp. 4193-4198.

Schuster, S. and C. Hilgetag, 1994. "On Elementary Flux Modes in Biochemical Reaction Systems at Steady State", J. Biol. Syst. 2, pp. 165-182.

Schuster, S., C. Hilgetag, J. H. Woods and D. A. Fell, 1996. "In *Computation in Cellular and Molecular Biological Systems* (eds. Cuthbertson, R., Holcombe, M. & Paton, R.)", World Scientific, Singapore, pp. 151-165.

Schwacke, J. H. and E. O. Voit, 2006. The potential for signal integration and processing in interacting MAP kinase cascades, Journal of Theoretical Biology, Volume 246, Issue 4, 21 June 2007, Pages 604-620.

Schwartz, M. A. and V. Baron, 1999. "Interactions between Mitogenic Stimuli, or, a Thousand and One Connections", Curr Opin Cell Biol 11, pp. 197–202.

Schwartz, M. A. and H. D. Madhani, 2004. “Principles of MAP Kinase Signaling Specificity in *Saccharomyces cerevisiae*”, Annu Rev Genet 38, pp. 725–748.

Segal, E., D. Pe’er, A. Regev, D. Koller, and N. Friedman, 2005. “Learning Module Networks”, *J Machine Learn Res* 6, pp. 557–588.

Seressiotis, A. and J. E. Bailey, 1988. “MPS: an Artificially Intelligent Software System for the Analysis and Synthesis of Metabolic Pathways”, Biotechnol. Bioeng. 31, pp. 587–602.

Sible, J. C. and J. J. Tyson, 2006. “Mathematical Modeling as a Tool for Investigating Cell Cycle Control Networks”, Methods, Volume 41, Issue 2, February 2007, pp. 238-247.

Slobodkin, L. B. and A. Rapoport, 1974. “An Optimal Strategy of Evolution”, *Q. Rev. Biol.* 49, pp. 181-200.

Somsen, O. J., M. Siderius, F. F. Bauer, J. L. Snoep and H. V. Westerhoff, 2002. “Selectivity in Overlapping MAP Kinase Cascades”, J. Theor. Biol. 218, pp. 343–354.

Steffen, M., A. Petti, J. Aach, P. D'haeseleer and G. Church, 2002. “Automated Modelling of Signal Transduction Networks”, BMC Bioinformatics 3, pp. 34-45.

Stein, M., R. R. Gabdoulline and R. C. Wade, 2007. “Bridging from Molecular Simulation to Biochemical Networks”, Current Opinion in Structural Biology, Volume 17, Issue 2, April 2007, pp. 166-172.

Tatebayashi, K., M. Takekawa and H. Saito, 2003. “A Docking Site Determining Specificity of Pbs2 MAPKK for Ssk2/Ssk22 MAPKKKs in the Yeast HOG Pathway”, EMBO J 22, pp. 3624-3634.

Trivers, R. L., 1971. “The Evolution of Reciprocal Altruism”, Quarterly Review of Biology 46, pp. 35-57.

Turner, P. E. and L. Chao, 2003. “Escape from Prisoner’s Dilemma in RNA Phage Φ6”, American Naturalist 161, pp. 497-505.

Van Drogen, F., V. M. Stucke, G. Jorritsma and M. Peter, 2001. “MAP Kinase Dynamics in Response to Pheromones in Budding Yeast”, Nat Cell Biol 3, pp. 1051-1059.

Varma, A. and B. O. Palsson, 1994. “Stoichiometric Flux Balance Models Quantitatively Predict Growth and Metabolic By-Product Secretion in Wild Type Escherichia coli W3110”, Appl. Environ. Microbiol. 60, pp. 3724-3731.

Verna, J., A. Lodder, K. Lee, A. Vagts and R. Ballester, 1997. “A Family of Genes Required for Maintenance of Cell Wall Integrity and for the Stress Response in *Saccharomysec cerevisiae”,* Proc. Natl. Acad. Sci. USA 94, pp. 13804-13809.

Von Mering, C., L. J. Jensen, B. Snel, S. D. Hooper, M. Krupp, M. Foglierini, N. Jouffre, M. A. Huynen and P. Bork, 2005. “STRING: Known and Predicted Protein-Protein Associations, Integrated and Transferred across Organisms”, Nucleic Acids Res 33, D433-437.

Von Neumann, J. and O. Morgenstern, 1953. Theory of Games and Economic Behavior, Princeton Univ. Press.

Wagner, A., 2001. “The Yeast Protein Interaction Network Evolves Rapidly and Contains Few Redundant Duplicate Genes”, Mol. Biol. Evol. 18, pp. 1283-1292.

Wang, X. F. and G. Chen, 2003. “Complex Networks: Small-World, Scale-Free and Beyond”, IEEE Circuits and Systems Magazine, First Quarter, pp. 6-20.

Watts, D. J. and S. H. Strogatz, 1998. "Collective Dynamics of 'Small-World' Networks", *Nature* 393, pp. 440–442.

West, S. A. and A. Buckling, 2003. "Cooperation, Virulence and Siderophore Production in Bacterial Parasites", Proc. R. Soc. Lond. B. Biol. Sci. 270, pp. 37-44.

Widmann, C., S. Gibson, M. B. Jarpe and G. L. Johnson, 1999. "Mitogen-Activated Protein Kinase: Conservation of a Three-Kinase Module from Yeast to Human", Physiol. Rev. 79, pp. 143–180.

Wolf, D. M. and A. P. Arkin, 2003. "Motifs, Modules, and Games in Bacteria", Current Opinion in Microbiology 6, pp. 125-134.

Woolf, P. J., W. Prudhomme, L. Daheron, G. Q. Daley and D. A. Lauffenburger, 2005. "Bayesian Analysis of Signaling Networks Governing Embryonic Stem Cell Fate Decisions", *Bioinformatics* 21, pp. 741–753.

Yashar, B., K. Irie, J. A. Printen, B. J. Stevenson, G. F. Sprague, K. Matsumoto and B. Errede, 1995. "Yeast MEK-Dependent Signal Transduction: Response Thresholds and Parameters affecting Fidelity", Mol. Cell. Biol. 15, pp. 6545-6553.

Yildiz, M., 2004. Game Theory Lecture Notes, MIT, Department of Economics.

Zhang, L. V., O. D. King, S. L. Wong, D. S. Goldberg, A. H. Y. Tong, G. Lesage, B. Andrews, H. Bussey, C. Boone and F. P. Roth, 2004. "Motifs, Themes and Thematic Maps of an Integrated *Saccharomyces cerevisiae* Interaction Network", J. Biol. 4, pp. 6.

Zou, M. and S. D. Conzen, 2005. "A New Dynamic Bayesian Network (DBN) Approach for Identifying Gene Regulatory Networks from Time Course Microarray Data, *Bioinformatics 21*, pp. 71–79.