

DYNAMICS OF GENE REGULATORY CELL CYCLE
NETWORK IN *SACCHAROMYCES CEREVISIAE*

by

Neşe Aral

A Thesis Submitted to the
Graduate School of Sciences and Engineering
in Partial Fulfillment of the Requirements for
the Degree of

Master of Science

in

Physics

Koç University

January, 2009

Koç University
Graduate School of Sciences and Engineering

This is to certify that I have examined this copy of a master's thesis by

Neşe Aral

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Committee Members:

Asst. Prof. Dr. Alkan Kabakçioğlu

Asst. Prof. Dr. Deniz Yuret

Prof. Dr. Nihat Berker

Date: _____

*To Alexander Zabini,
who inspired me to ask the question
which set off this thesis.*

ABSTRACT

In this thesis, the genetic regulatory dynamics within the cell cycle network of the yeast *Saccharomyces cerevisiae* is examined. As the mathematical approach, an asynchronously updated Boolean network is used to model the time evolution of the expression level of genes taking part in the regulation of the cell-cycle. The attractors of the model's dynamics and their stability are investigated by means of a stochastic transition matrix. It is shown that the cell cycle network has unusual dynamical properties when compared with similar random networks. Furthermore, an entropy measure is employed to monitor the sequential evolution of the system. It is observed that the experimentally identified cell cycle phases G_1 , S , G_2 and M correspond to the stages of the network where the entropy goes through a local extremum.

ÖZETÇE

Bu tezde *Saccharomyces cerevisiae* hücresinin, hücre döngüsü içindeki gen regülasyon dinamiği incelendi. Hücre döngüsünü kontrol eden genlerin zaman içindeki ekspresyon değerlerini bulmak için matematiksel yöntem olarak asenkron biçimde güncellenen Bool tipi ağlar kullanıldı. Dinamiğin çekicileri ve bunların kararlılığı stokastik transfer matrisi kullanılarak incelendi. Rastgele ağlar ile karşılaştırıldığında hücre döngüsü ağının değişik dinamik özellikler sergilediği gösterildi. Sistemin sırayla geçtiği durumları gözlemek için bir entropi ölçüsü tanımlandı. Deneylerde görülen G_1 , S , G_2 ve M fazlarının, ağın entropisinin yerel ekstremum noktalarına karşılık geldiği gözlemlendi.

ACKNOWLEDGMENTS

I would like to thank first of all to Prof. Dr. A. Nihat Berker for his exceptional support and exhortation throughout my undergraduate and graduate years. His belief in me, not only in good days but also in my unefficacious periods of time, gave me the courage to proceed in my academic endeavor to be a physicist.

I am also grateful to my supervisors Alkan Kabakçiođlu and Deniz Yuret, who shared their invaluable time and wisdom with me, for enlightening me how to become a scientist. Their unceasing guidance and understanding are memorable.

Special thanks go to all of my office friends, from whose knowledge and help I benefited enormously, for enabling my enthusiasm for physics to increase.

I also have to thank all of my teachers in Austrian High School, especially to Peter Toplack, Hermann Quas, Hugo Dostal, Gustav Aigner and Alexander Zabini, for introducing me the world of physics, mathematics, chemistry and biology by making me believe that they are fun.

I am especially indebted to my family, because without their support, trust and understanding I could not achieve my dream of becoming a scientist. I am gratefully aware that my parents Orhan Aral and Ferihan Aral are working harder than me to provide the extraordinary opportunities I have. My sister Işıl Aral deserves a big thanks for always helping me and being (literally) there as I was writing, correcting my thesis and rehearsing its defence.

At last but not least, I send my thanks and love to A. Barış Sözen, for his understanding throughout my research.

This research is financially supported by TÜBİTAK (The Scientific and Technological Research Council of Turkey) under the project TBAG 106T553.

TABLE OF CONTENTS

List of Tables	ix
List of Figures	x
Chapter 1: PRELIMINARIES	1
1.1 Introduction	1
1.2 Motivation and Purpose of the Study	2
1.3 Methods	2
Chapter 2: Biological Background	3
2.1 The Structure of the DNA and The Templated Polymerization	3
2.2 Reading the Genome	4
2.3 Gene Regulation	6
2.4 Cell Cycle	8
2.5 The Budding Yeast <i>Saccharomyces cerevisiae</i> as a Model Organism	10
2.5.1 Cell Cycle Control System in <i>Saccharomyces cerevisiae</i>	11
2.6 Experimental Methods	14
2.6.1 Yeast two-hybrid	16
2.6.2 The Microarray Chips	17
2.6.3 Databases	18
Chapter 3: Mathematical Background	19
3.1 Graphs and Networks	19
3.2 Gene Regulatory Networks	20
3.3 Modelling Gene Regulatory Networks	20

3.3.1	Logical Models	20
3.3.2	Continuous Models	25
Chapter 4:	YEAST CELL CYCLE NETWORK	27
4.1	Network Structure	27
4.2	Asynchronously Updated Cell Cycle Network	30
4.2.1	The Simulation Procedure of The Cell Cycle Process	33
4.2.2	Transfer Matrix Description	34
4.2.3	Stability of the Attractors	42
4.2.4	Dynamics of the System	44
4.2.5	Comparison with Random Networks	47
4.2.6	Perplexities for Attractors	48
4.3	Conclusion	49
Bibliography		51
Vita		54

LIST OF TABLES

2.1	Cyclins of the budding yeast	13
3.1	Possible functions for a node with $K = 2$. The total number of functions is $2^{2^K} = 16$	22
4.1	Global states corresponding to cell cycle phases. Since G_1 and M phases are subdivided into short phases there are multiple global states corresponding to the same main phase.	29
4.2	Global states corresponding to cell cycle phases after modification of the network. The ID numbers are calculated as shown in Eq. (4.11).	32
4.3	9 fixed states of the network.	39
4.4	Probabilities of fixed states. Basin of attraction is the number of initial states preferring the attractor. P_{avg} is the average probability to reach the attractor among the initial states preferring the attractor. P_{random} is the probability to reach the attractor from a random initial state.	41
4.5	Probabilities of arriving at F_1 and F_4 when the initial state is <i>Start</i> while changing the parameter t_d	42
4.6	Behavior under perturbation. $P_b \equiv P_{back}$ Example: If Cln3 changes its value in F1, this new state return back to F1 with probability 0.577.	43
4.7	The matrix A_{ij} , where $A_{ij} = P_{Pert}(F_j, F_i)$	44

LIST OF FIGURES

2.1	The structure of the DNA [1].	4
2.2	Transfer RNA [1].	5
2.3	Gene Regulation [1].	7
2.4	Cell Cycle [1].	9
2.5	Levels of the cyclins during the cell cycle. [19]	15
2.6	The yeast two-hybrid system for detecting protein protein interactions [1].	16
2.7	Small region of microarray representing expression of 110 genes from yeast [1].	17
3.1	Trajectory of a Boolean network.	23
3.2	Illustration of a Petri net.	25
4.1	Cell Cycle Network of Li et al.	27
4.2	Gene regulatory network of the yeast cell cycle.	31
4.3	Results obtained by the average of 100 simulations.	35
4.4	The pattern of the evolution matrix for a 7-node network. The black points represent the nonzero elements.	37
4.5	The pattern of the evolution matrix, with top left corner zoomed. . .	38
4.6	The pattern of T_N for a 7-node network.	45
4.7	Flow of the system beginning from <i>Start</i> . Each line corresponds to a time step. The widths of the arrows are proportional to the transition probabilities.	46

4.8	Sequential evolution of entropy. The number of steps are greater than those in the Fig. 4.7, because the states with very low probabilities are not shown in that figure.	47
4.9	Time evolution of entropy when the original evolution matrix is used.	48
4.10	Entropy evolution for 8 random rewirings of the cell-cycle network and the yeast network for comparison.	49
4.11	Perplexities of 100 random networks with the same topology as the gene regulatory network. The perplexity value of the yeast's cell cycle network lies in the second column.	50

Chapter 1

PRELIMINARIES

1.1 Introduction

In the 20th century our knowledge about the molecular basis of life has increased enormously due to the discoveries and new techniques in genetics and molecular biology. The experimental data obtained by these new techniques are so abundant that computational and mathematical methods must be used to analyze them. Although mathematics and biology have always been in connection, in the last few decades they merged into each other due to new advances in applied mathematics, such as *theory of dynamical systems*, *population models*, and *graph theory*, which can explain biological phenomena better than ever. Additionally the geometric increase of computational power allows us to investigate large-scale problems in biology.

A main topic in this interdisciplinary field is the functional organization of the cell. The questions waiting for answers are for example: How does the behavior of a cell change when there is a change in the environment? How does it affect the cell if some part of it stops functioning? How robust is the system against mutations? Among other functions of the cell a big interest lies in the cell cycle process, and plentiful investigations have been made analyzing different cell cycle models. Understanding the concealed mechanism providing the cell to divide into two by transferring a copy of its genetic code to the next generation is a big challenge. Scientific curiosity being the most leading one, this was also needed in research of cancer, which is a consequence of errors appearing in this process. These days we have a pretty good understanding what is going on in cell cycle process. But a complete understanding of it lies far from

us. That's why we have to develop new techniques in all branches of science and use them together, so that we get a grip on the complex organization found in these natural systems. I hope, with this study, we can make a tiny contribution to this purpose.

1.2 Motivation and Purpose of the Study

In this study we want to setup a mathematical model which simulates the cell cycle process. It must be able to show how a single cell commits to division, goes through the cell cycle phases, G_1 , S , G_2 and M following each other successfully and returns back to its original steady state which it preserves until the next round of division. As our model organism we use the one celled budding yeast *Saccharomyces cerevisiae*, whose whole genome sequence, consisting of nearly 6300 genes is known. Although the whole gene regulatory network of the yeast is very big, only 800 of them have a role in the cell cycle process and a small subnetwork of these 800 genes is crucial for the progression of the cycle and this enables us to investigate cell division qualitatively and quantitatively. This model is important because the proteins taking part in the cell cycle are well conserved through the evolution and all eukaryotic cells behave similar in this process. So understanding the yeast cell cycle can give us an insight into the behavior of all eukaryotic cells, including the ones of the human.

1.3 Methods

We used an asynchronous Boolean network to simulate the time-evolution of the yeast's cell-cycle. We analyzed our model using a transfer matrix approach that completely describes the stochastic evolution of the system. We identified the attractors of the system with the eigenvectors of this matrix with eigenvalue 1. The only biologically relevant attractor is compared with the others by looking at their basin of attraction sizes and stabilities under perturbations. To demonstrate the special structure of the cell cycle network we re-wired it without changing its topology and compared it with these new random networks.

Chapter 2

BIOLOGICAL BACKGROUND

2.1 The Structure of the DNA and The Templated Polymerization

The DNA (Deoxyribonucleic acid) contains all the genetic instructions needed for the development and functioning of a living organism. It consists of two strands which twist around each other and form a double helix. Each strand is a polymer where the monomers, called nucleotides, contain a sugar (deoxyribose) with a phosphate group attached to it and one of the bases adenine (A), guanine (G), thymine (T) and cytosine (C), which are strung together in a long linear sequence that encodes the genetic information. Each base on one strand makes a hydrogen bond with the facing base on the complementary strand. This base pairing procedure gives the DNA molecule its stability. But as each base can only bind to its complementary base, just two types of base pairings are possible, namely, A-T with two hydrogen bonds and C-G with three hydrogen bonds. Since these hydrogen bonds are weaker than the sugar-phosphate links, the two DNA strands can be pulled apart and separated without breaking the backbones. This structure of the DNA is illustrated in Fig. 2.1.

However, in living cells, a DNA molecule cannot be synthesized in isolation. A new DNA molecule can only be synthesized via replication of a pre-existing DNA molecule. After the two strands of the existing DNA molecule are separated from each other they serve as templates for the new ones. The bases on the old strand bind to their complementary bases, namely A to T, C to G and vice versa. This process controls the polymerization of the new complementary strand by means of the selection of which one of the four monomers shall be added to the growing strand next. This way of copying the information in DNA replication is called templated

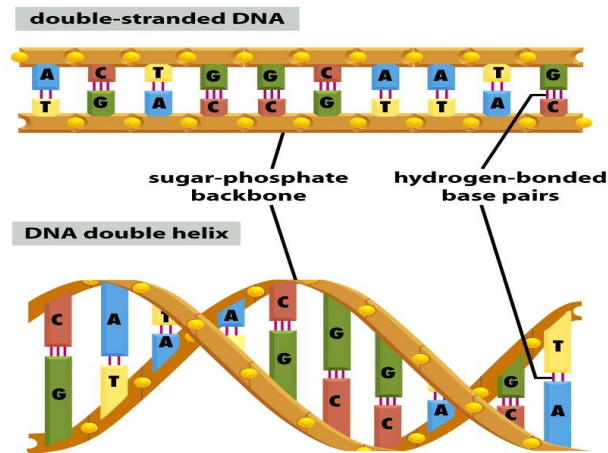


Figure 2.1: The structure of the DNA [1].

polymerization.

2.2 Reading the Genome

Since the DNA is a very long molecule it must reduce its size so that the cell can carry it. The packaging process involves many steps and at the end the organized full compact form of the DNA is called the chromosome. Usually the cells have multiple numbers of chromosomes, so that the genetic information is shared among them. The full set of the chromosomes make up the genome, the complete set of the genetic information of an organism.

It is not enough for a cell to have the genetic information stored in a long sequence and to duplicate it. The DNA must also be able to express its information, so that the cell can read it and direct the cellular processes. But not all of the information encoded in the DNA is needed at once. The DNA expresses only the portions of itself which are needed for the cell at a particular time. The expression begins with unzipping the required portion (segment) of the DNA. This time the segment serves as a template for the synthesis of a shorter polymer, RNA (ribonucleic acid). This step is called transcription. The RNA is related to DNA. Instead of deoxyribose it has ribose

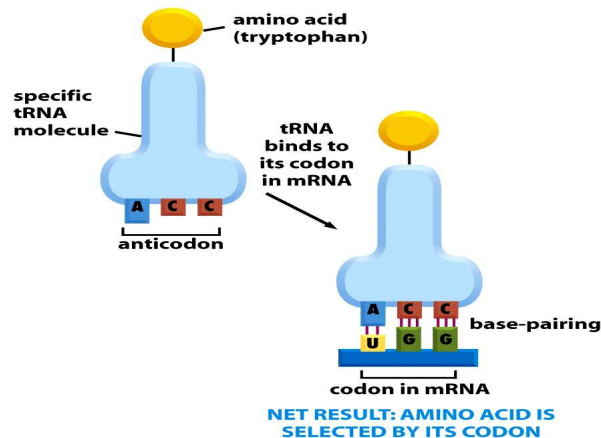


Figure 1-9a Molecular Biology of the Cell 5/e (© Garland Science 2008)

Figure 2.2: Transfer RNA [1].

as sugar and one of the bases is Uracil (U) instead of thymine (T). In addition the RNA is a single stranded molecule. During transcription, RNA monomers are lined up and selected for polymerization on a template strand of DNA, just as DNA monomers are selected during replication via templated polymerization. So the sequence of a RNA molecule represents a portion of the genetic information. As these RNAs carry the genetic information these are called the messenger RNAs (mRNA). This information will be used to synthesize the proteins, which are polymers of amino acids. There are 20 different types of amino acids but the RNA uses only 4 letters for coding. Therefore the information of the mRNA molecules is read out in groups of three nucleotides at a time. These triplets are called codons and each corresponds to a specific amino acid. Since there are 64 ($4 \times 4 \times 4$) possible codons but only 20 amino acids, multiple codons may correspond to the same amino acid. The whole code, consisting of a lot of codons, is read by another RNA molecule, the transfer RNA (tRNA). They are called so because they help to transfer the information from mRNA to the protein being synthesized. This process is enabled by the help of ribosomal RNA (rRNA) and the proteins in the ribosome. Each tRNA molecule has the anti-codon sequence at one end which can bind to the codon in the mRNA and the codon-specific amino acid on the other end (Fig. 2.2).

The mRNA molecule brings the tRNA molecules together by matching up their successive codons with the anti-codons of tRNA molecules. In this way the sequence of amino acids is being designated. In a giant multimolecular mechanism called the ribosomes, the amino acids are released from tRNA and linked together and build the protein chain. The ribosome itself consists of two main chains of ribosomal RNA (rRNA).

2.3 Gene Regulation

The gene is defined as a portion of the DNA molecule, which is transcribed on a mRNA to produce a protein or a RNA molecule. For a long time it was thought that a particular gene codes only for one type of protein. But latest discoveries showed that through a method called *RNA splicing* it is possible that several proteins can be synthesized from the same gene.

The proteins are synthesized according to the needs of the cell. So only the genes coding for the required proteins must be expressed at a particular time. The regulation of this process is extremely complex. The cell can regulate the rate of transcription and translation or the activity of the proteins. Not all but the most important regulation processes in eucaryotes for our study are presented below.

Aside from the genes, which code for the proteins, there are also noncoding nucleotide sequences in the DNA molecule. A small part of these noncoding regions is useless since they are residues of evolution and they are called *junk DNA*. But an important part of the noncoding sequences is responsible for the accurate expression of coding regions namely the genes. Some of these noncoding regions serve as binding regions for proteins especially enzymes. Of great significance are the nucleotide sequences next to the gene sequences, called the *promoter* regions, where the *RNA polymerase II*, an enzyme responsible for unzipping the DNA, is bound. But in order for this enzyme to be active and let the transcription start it must build a complex with other proteins called *transcription factors*. These transcription factors are also proteins which are products of other genes. So the products of some genes regulate

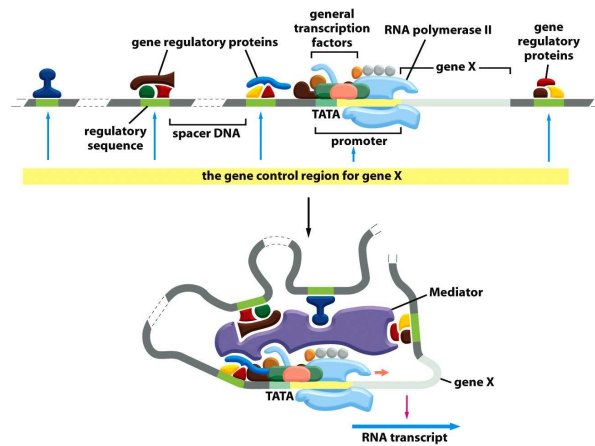


Figure 7-44 Molecular Biology of the Cell 5/e (© Garland Science 2008)

Figure 2.3: Gene Regulation [1].

the expression of other genes, a process known as *gene regulation*. Additionally this complex consisting of transcription factors and RNA polymerase II is only activated when they are bound to *gene regulatory proteins* via *mediator* proteins. These gene regulatory proteins are bound to another nucleotide sequence called *regulatory sequence*, which can be far away from the gene. An illustration of these complexes responsible for gene expression is shown in Fig. 2.3.

There are two types of regulatory proteins classified as *activators* and *repressors*. The activators once bound to the regulatory sequence activate the transcription by interacting with mediator-transcription factors-RNA polymerase II complex at the promoter. The repressors prevent the activators from interacting with the mediator protein so that the gene cannot be expressed. There are several ways in which eukaryotic gene repressor proteins can operate. They can compete with the activators for the same regulatory region and hinder them to bind. If they both are bound to the DNA next to each other, the repressor can bind to the activation domain of the activator, thereby preventing it from carrying out its activation functions. In some cases the repressor may have been bound only to the activator and not to the DNA. A repressor can also bind to the assembly of transcription factors so that this complex is not able to bind to the activator any more. In addition, it must be taken into

account that an activator (or repressor) of a specific gene can behave as a repressor (or activator) for another gene. On the other hand most regulatory proteins can only be functional by changing their shapes via building complexes with other regulatory proteins. Depending of the participants of the assembly they can act as an activator or a repressor.

After the gene is translated into proteins in the ribosomes, the activity of these proteins in the cytoplasm can also be adjusted by post-translational regulation. *Ubiquitination* is a process where a protein called *ubiquitin* binds covalently to the proteins and label them for degradation. *Phosphorylation* and *dephosphorylation* are reversible mechanisms in which a phosphate molecule is attached to or removed from the protein, and this leads to a conformational change in the structure of the proteins, causing them to become activated or deactivated.

2.4 Cell Cycle

The cell cycle process involves an orderly sequence of events which provide the cell to divide in two daughter cells. To transfer its genetic code to the next generation, the cell has to duplicate its whole genome and segregate the chromosomes equally among the new cells, producing genetically identical daughter cells.

The oversimplified eukaryotic cell cycle model has four main phases, namely G_1 (gap 1), S (synthesis), G_2 (gap 2) and M (mitosis). The G_1 phase covers the time that the cell needs to grow. In this phase, the cell also monitors the internal and external environment to make sure that the conditions are suitable and preparations are complete for the cell division. This is very critical because once the signal for the division is there, the replication of the DNA begins and this process cannot be reversed even if the external environment is not favorable anymore. This signal with which the cell commits to replicate its genome and enters the S phase is called *Start* in yeasts and *restriction point* in mammalian cells. In the S phase new chromosomes are synthesized by templated polymerization and at the end of this phase the DNA molecules in each pair of duplicated chromosomes are connected and held together

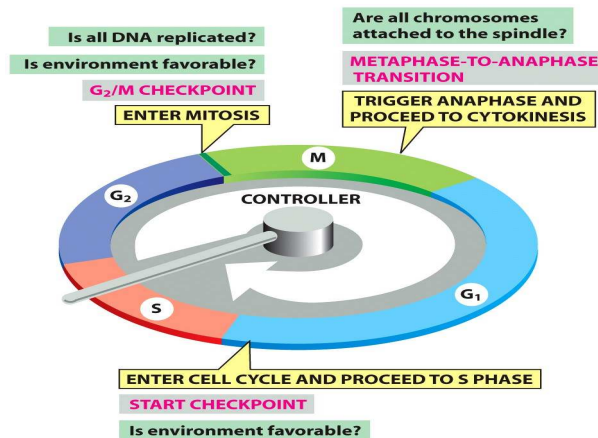


Figure 17-14 Molecular Biology of the Cell 5/e (© Garland Science 2008)

Figure 2.4: Cell Cycle [1].

by specialized protein linkages. Then it follows the G_2 phase in which the cell checks whether the duplication was successful and if it can proceed with the cycle. At the end of the G_2 phase the control system triggers the mitotic events provided that the duplication was errorless or the errors were repaired. This G_2/M checkpoint is the second checkpoint after the G_1/S checkpoint in the cell cycle process. The M phase includes the *mitosis* and the *cytokinesis* both having their own subphases. In the *prophase* of the mitosis the chromosomes are condensed in pairs of rigid and compact rods called *sister chromatids*. The *metaphase* is entered when the nuclear envelope disassembles which leads the *mitotic spindle* to take shape. The stage where all the sister chromatids are aligned at the spindle equator is called the *metaphase*. After pulling the sister chromatids to opposite poles of the spindle in the *anaphase* and destroying the spindle to package the chromosomes into separate nuclei in the *telophase*, the mitosis is completed. As a last step the cytokinesis takes place where the cytoplasm divides into two each having one copy of the genome and also one set of the duplicated organelles. For a schematic view of the cell cycle see Fig 2.4.

2.5 The Budding Yeast *Saccharomyces cerevisiae* as a Model Organism

Saccharomyces cerevisiae is a unicellular species in genus *Saccharomyces* (*sugar mold* in Greek) of kingdom *fungi* as the affix *-myces* suggests. *Cerevisiae* means "of beer" in Latin and this organism is named so because it is mostly used in brewing. It is also called the *budding yeast* since it divides into two by budding which is the formation of a new organism by the protrusion of part of another organism.

In the second half of the 20th century yeast has been introduced as an experimental system for molecular biology. In 1980's, yeast was used to produce Hepatitis B vaccine. In 1996 yeast was the first eukaryotic organism (with nucleus in the cell containing the genome), of which the complete genomic sequence could be established. In the years to follow, yeast became a useful reference against which sequences of human, animal or plant genes, and those of a multitude of unicellular organisms under study could be compared. Moreover, the ease of genetic manipulation in yeast opened the possibility to functionally dissect gene products from other eucaryotes in the yeast system.

Yeast is an ideal system to investigate cell architecture and fundamental cellular mechanisms successfully. Among all other eukaryotic model organisms, *Saccharomyces cerevisiae* combines several advantages. It is a unicellular organism which can be grown on defined media giving the investigator complete control over environmental parameters. It has a short generation time (doubling time 1.5-2 hours at 30°C). These are all positive characteristics in that they allow for the swift production and maintenance of multiple specimen lines at low cost. It can be transformed allowing for either the addition of new genes or deletion through homologous recombination and it survives a lot of mutations. Furthermore, the ability to grow *Saccharomyces cerevisiae* as a haploid simplifies the creation of gene knockouts strains. As a eucaryote, it shares the complex internal cell structure of plants and animals without the high percentage of non-coding DNA that can confound research in higher eucaryotes. *Saccharomyces cerevisiae* research had a strong economic driver, at least initially, as a result of its established use in industry (e.g. beer, bread and wine fermentation).

2.5.1 Cell Cycle Control System in *Saccharomyces cerevisiae*

The central problem in cell cycle process is the coordination of the several events, so that every event occurs in the proper order with respect to each other. For example, the segregation of chromosomes must follow the DNA duplication. Otherwise trying to segregate unreplicated chromosomes leads to chromosome breakage or aneuploidy. Additionally cytokinesis must follow the chromosome segregation since an earlier cytokinesis leads to generation of aploid and polyploid daughter cells. On the other hand, each event is restricted to take place only once in each cycle. For instance, duplicating the chromosomes more than once leads to polyploidy.

The idea of an *autonomous cell cycle clock* suggests that different cell cycle events should occur with a constant timing, even if one or more events are delayed due to environmental insults or experimental manipulation. But observations in *Saccharomyces cerevisiae* showed that this idea was incorrect. In these budding yeasts the cell cycle events are linked in dependent pathways. For example, even if the DNA replication is delayed by several hours, the chromosome segregation does not begin. These observations lead to the *cell cycle clock* idea which suggests that there are two main coordinating mechanisms. One of them is a coordinating mechanism in the form of a set of checkpoint controls such as G_1/S and G_2/M checkpoints. These checkpoint controls ensure that cell cycle events occur in the proper order. The second molecular mechanism is a biochemical oscillator, acting as a "cell cycle clock". The basis for this clock and the central components of the cell cycle control system are the cyclin-dependent kinases (CDKs). A kinase is a type of enzyme that transfers phosphate groups from high-energy donor molecules, such as ATP, to specific target molecules (substrates). Cyclins are a family of proteins involved in the progression of cells through the cell cycle. They are named cyclins because their concentration in the cell varies cyclically according to the phases of the cycle. CDKs are only active when a special type of a cyclin is bound to the proper kinase. So the activities of these kinases rise and fall as the cell progress through the cycle and this leads to cyclical changes in the Phosphorylation of proteins that initiate or regulate the major

events in the cell cycle. So we can conclude that cell cycle phases alternate due to the mechanism that one cyclin family succeeds another. Hence the systematic progress of the cell cycle depends mainly on gene regulation mechanism leading to periodically expression of cyclins.

There are four classes of cyclins, each defined by the stage of the cell cycle at which they bind CDKs and function. All eukaryotic cells require three of these classes. These are:

1. G_1/S cyclins: They activate CDKs in late G_1 and thereby help progression through Start, resulting in a commitment to cell cycle entry. Their levels fall in S phase.
2. S cyclins: They bind CDKs soon after progression through Start and help stimulate chromosome duplication. Their levels remain elevated until mitosis.
3. M cyclins: They activate CDKs that stimulate entry into mitosis at the G_2/M checkpoint. They are destroyed in mid-mitosis.

In yeast cells, the main CDK protein is Cdc28 which binds all classes of cyclins and triggers different cell cycle events by changing cyclin partners at different stages of the cycle. Its level does not fluctuate during the cell cycle. The cyclins and their partners in budding yeasts are shown in Table 2.1. The cyclin proteins do not simply activate their CDK partner but also direct them to specific target proteins. As a result, each cyclin-CDK complex phosphorylates a different set of substrate proteins.

The most important cyclins and other proteins in the cell cycle process in budding yeasts are shortly described below [3]. These proteins are products of the genes with same name.

1. Cln3: Unlike the other cyclins, its transcription is not strongly periodic with respect to the cell cycle, but there is a small rise M/G_1 border over its basal levels. Cln3-Cdc28 is the only CDK involved in activating SBF and MBF in

Cyclin-CDK complex	Cyclin	CDK partner
G_1 -CDK	Cln3	Cdc28
G_1/S -CDK	Cln1,2	Cdc28
S -CDK	Clb5,6	Cdc28
M -CDK	Clb1,2,3,4	Cdc28

Table 2.1: Cyclins of the budding yeast

normal cycling cells. Its transcription is triggered by the Start signal, when the cell reaches a critical mass.

2. Cln1,2: They stimulate DNA synthesis indirectly by accelerating the degradation of the Clb-CDK inhibitor Sic1. In this way they enable the transition from G_1 to S phase. Their expression depends on transcription factor complex SBF.
3. Clb1,2: It activates Cdc28 to promote the transition from G_2 to M phase and accumulates during G_2 and M . They negatively regulate the SBF and MBF complexes ensuring that Cln1 and Cln2 expression is kept low until the Clb2-Cdc28 are destroyed during mitosis.
4. Clb5,6: It is involved in DNA replication during S phase and activates Cdc28 to promote initiation of DNA synthesis. Their expression depends on transcription factor complex MBF.
5. Sic1: It is an inhibitor of Clb-Cdc28 complexes. One of its functions is to prevent premature DNA replication. Because of its inhibitory effect on the Clb-Cdc28 activity, it can be considered as an inhibitor of the G_1/S transition and an activator of M/G_1 transition in the cell cycle. Its transcription depends primarily on the transcription factor Swi5.
6. Cdh1: It is the activator of the *anaphase-promoting complex* (APC), which

directs ubiquitination of cyclins resulting in mitotic exit and initiates anaphase. It targets the APC to specific substrates including Cdc20. Its level is constant throughout the cell cycle.

7. Cdc14,20: They are activators of APC, which is required for metaphase/anaphase transition. They direct ubiquitination of mitotic cyclins and they are responsible for the cell to leave the mitotic phase. Their levels are strongly periodic in the cell cycle.
8. Swi5: It is a transcription factor that activates transcription of genes expressed at the M/G_1 phase boundary and in G_1 phase. Localization to the nucleus occurs during G_1 and appears to be regulated by phosphorylation by Cdc28.
9. Mcm1: It is the transcription factor of Swi5 and it requires Cdc20 as its own transcription factor to be expressed.
10. SBF: It is a complex consisting of transcription cofactor Swi6 and DNA binding protein Swi4. It is the dominant factor controlling the expression of Cln1 and Cln2. It initiates events responsible for bud formation.
11. MBF: It is a complex consisting of transcription cofactor Swi6 and transcription factor Mbp1. It is the dominant factor controlling the expression of Clb5 and Clb6. It initiates events responsible for DNA synthesis.

How the levels of the cyclins change with the cell cycle phases is illustrated in Fig. 2.5.

2.6 Experimental Methods

There are mainly two types of experiments whose results can be found in databases which show interactions between proteins and genes. Yeast two-hybrid experiments yield the protein-protein interactions and microarray chips show genes which are expressed at the same time.

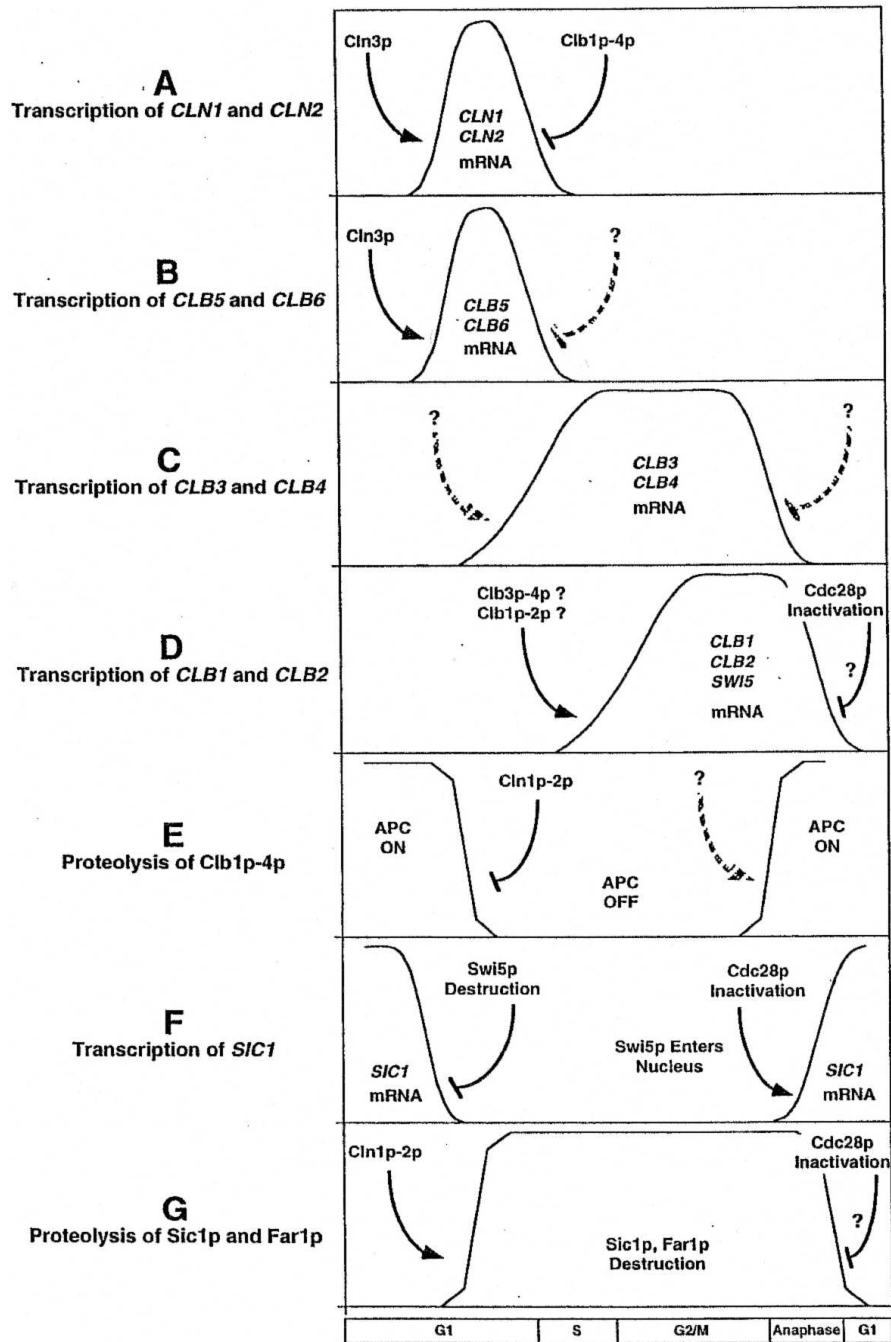


Figure 2.5: Levels of the cyclins during the cell cycle. [19]

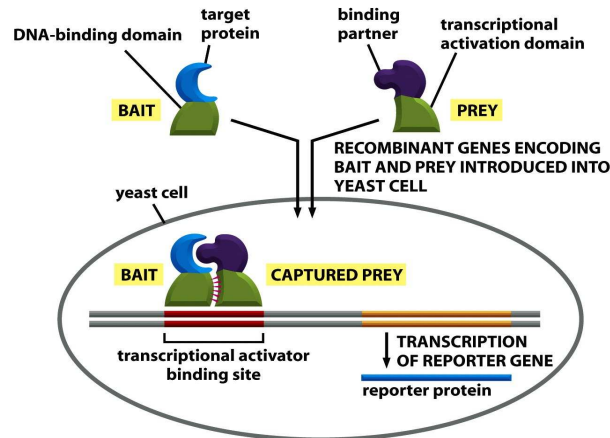


Figure 8-24 Molecular Biology of the Cell 5/e (© Garland Science 2008)

Figure 2.6: The yeast two-hybrid system for detecting protein protein interactions [1].

2.6.1 Yeast two-hybrid

The gene activator proteins have two different domains. One of the domains bind to a specific DNA sequence and the other domain activates the gene transcription. These domains are used to create separate "bait" and "prey" proteins. The "bait" is created by fusing the DNA sequence that codes for a target protein with the DNA sequence that encodes the DNA-binding domain of a gene activator protein. The whole DNA sequence is then inserted into the yeast cell and the cell produces "bait" proteins, with the target protein attached to the DNA-binding domain. This "bait" protein binds to the regulatory region of the specific gene, called the *reporter gene* in this case. The candidate proteins namely the "preys" for this protein are also created the same way. The "bait" protein binds to the regulatory region of the reporter gene. If any "prey" protein interacts with these target "bait" proteins, it is captured by the "bait" and binds to regulatory region. In this way the two halves of the activator protein are united and they can now activate the transcription of the reporter gene. An illustration of this method can be seen in Fig. 2.6.

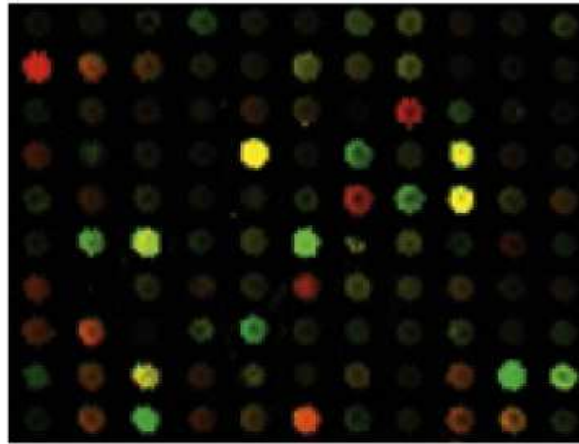


Figure 2.7: Small region of microarray representing expression of 110 genes from yeast [1].

2.6.2 The Microarray Chips

The microarray chips monitor the expression of thousands of genes at the same time. They consist of DNA fragments, each corresponding to a gene, are spotted onto a slide by a robot. mRNA is collected from two different cell samples for a direct comparison. If we want to monitor cell cycle genes we can take samples from dividing and nondividing cells. These samples are converted to cDNA and labeled, one with a red fluorochrome (sample 1), the other with a green fluorochrome (sample 2). The labeled samples are mixed and then allowed to hybridize to the microarray. After incubation, the array is washed and the fluorescence scanned with a scanning-laser microscope. Red spots indicate that the gene in sample 1 is expressed at a higher level than the corresponding gene in sample 2. Green spots indicate that expression of gene is higher in sample 2 than in sample 1. Yellow spots reveal genes that are expressed at equal levels in both cell samples. Dark spots indicate little or no expression in either sample of the gene whose fragment is located at that position in the array [1]. An example is shown in Fig. 2.7.

2.6.3 Databases

There are special databases devoted to budding yeast *Saccharomyces cerevisiae*. The data in these databases are mainly curated from papers which report results from a lot of types of experiments, yeast two-hybrid and microarray experiments being the most important ones. These are:

1. Yeasttract [4, 5]: www.yeasttract.com. Here the transcription factors of the yeast genes can be found and a regulation matrix can be constructed automatically.
2. Saccharomyces Genome Database: www.yeastgenome.org. There are tools for the analysis of gene sequences, homology comparisons between proteins. Data are available about functions, expressions and interactions of proteins.
3. Database of Interacting Proteins: <http://dip.doe-mbi.ucla.edu>. This database catalogs experimentally determined interactions between proteins. It combines information from a variety of sources to create a single, consistent set of protein-protein interactions. The data stored within the DIP database were curated, both, manually by expert curators and also automatically using computational approaches that utilize the knowledge about the protein-protein interaction networks extracted from the most reliable, core subset of the DIP data.
4. MIPS: <http://mips.gsf.de/genre/proj/yeast>. The MIPS Comprehensive Yeast Genome Database (CYGD) aims to present information on the molecular structure and functional network of the entirely sequenced, well-studied model eucaryote, the budding yeast *Saccharomyces cerevisiae*. In addition the data of various projects on related yeasts are used for comparative analysis.

Chapter 3

MATHEMATICAL BACKGROUND

3.1 Graphs and Networks

Mathematically a graph G is defined as a nonempty set V together with an irreflexive, symmetric relation R on V [2]. Although this seems to be very abstract, graphs are very powerful tools to visualize interactions between objects. A graph is a representation of a set of entities where pairs, having a connection, are linked by lines. The interconnected entities are represented usually by dots which are called *vertices* (sing. vertex) or *nodes*. The links that connect two nodes are represented by curves and these are called the *edges* of the graph. In this simplest case where the curves have no direction the graph is said to be *undirected* or *simple*. When the connections between the nodes are determined by their directions, the graph is called a *directed graph* where the edges are represented by arrows. Depending on the properties of interactions between the nodes the arrows can specify the weight or type of connections. The *indegree* of a node is the number of arriving arrows at this node and the *outdegree* is the number of the arrows leaving the node. A network is just a directed graph (*digraph*) with weighted edges. The graphs can also be classified as *connected* or *disconnected*. Defining the *path* in a graph as a sequence of nodes such that from each of these nodes there is an edge to the next node in the sequence, two nodes are said to be connected if there exists a path between these two nodes. A graph is called connected if every pair of distinct nodes in the graph is connected and disconnected otherwise.

The *connectivity matrix* of a network with N nodes is the $N \times N$ matrix C whose elements c_{ij} represent the type and the weight of the connection from node j to the node i . In our study the weights of activators and inhibitors will be equal. The type

of the connection is determined by the sign of the weight. So a connectivity matrix element $c_{ij} = 1$ (or -1) will represent the activation (inhibition) of node i by node j .

3.2 Gene Regulatory Networks

Gene regulatory networks consist of nodes which represent the genes or the products of them (proteins). The nodes are connected with arrows defining the type of the interaction. If a protein is the inhibitor of another gene there is a negative interaction or if a protein is the activator of some other gene (transcription factor) there is a positive interaction. These networks allow us to calculate the rates at which genes in the network are transcribed into mRNA and so their expression levels. As the protein levels in a cell determine its type or function, it is important to know when or how much a cell produces specific kinds of proteins.

3.3 Modelling Gene Regulatory Networks

There are different mathematical tools to analyze gene regulatory networks. These models are mainly divided into three classes: Logical models, continuous models and single-molecule models. Which model to choose depends on what we want to observe and analyze. Some of these models are shortly explained here [6].

3.3.1 Logical Models

Models that belong to this category are discrete models so they can explain the investigated network qualitatively. They allow a basic understanding of the dynamics and functions of a network under different conditions. Their advantages lie in their applicability to a wide range of systems including biological phenomena. Here we will list some of them.

Boolean Networks:

This modelling technique was introduced by Kauffman [21, 22]. These networks

are called Boolean because the nodes in these networks can only have two levels: active (1) or inactive (0). In a gene regulatory network, for example, a gene can be expressed (active) or not expressed (inactive) at any specific time. The level of one node is the *local state* and the vector whose components are the levels of each node, e.g. (0, 0, 1, 0, 1), is the *global state* of the system and a sequence of global states constitutes the *trajectory*. So the number of all possible global states of a network with N nodes is 2^N . The level of each node is updated with a Boolean function whose input variables are the levels of the nodes by which it is regulated. Each of the nodes can be assigned a different update (regulation) function. In principle at every time step all of the nodes are updated synchronously, such that the new value of a node is determined by the levels of its regulators at the previous time step. So once the global state of the system is known at one time step, the trajectory of the network can be obtained deterministically. The global state which, once reached, repeats itself in the trajectory is the *steady state*. The trajectory can also be cycle of a sequence of global states which is an important behavior in biological systems. The steady states and cycles which the network arrives at after traversing a transient trajectory are called the *attractors*.

The update functions can be in different forms [7]. Let us investigate the possible functions for a Boolean network. For a node which has K regulatory nodes from which it receives input, the input vector can have $M = 2^K$ values. These M vectors build the "input vector space". Each element of the input vector space must specify the local state of its target node at the next time step. So the number of all possible functions, whose domain is the "input vector space" with M elements and whose range is the set 0,1, can be calculated as 2^M . As an update function for any node, we can choose any of these possible functions. In Table 3.1 an example of functions for a node receiving two inputs are shown.

When we look at Table 3.1, we see that there are four types of functions. The first two functions f_1 and f_2 belong to class "frozen". These are the functions whose value does not depend on the input vector. The value is always 1 or always 0. The

Input vector	Frozen		Canalyzing I				Canalyzing II								Reversible	
	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8	f_9	f_{10}	f_{11}	f_{12}	f_{13}	f_{14}	f_{15}	f_{16}
0, 0	1	0	0	1	0	1	1	0	0	0	0	1	1	1	1	0
0, 1	1	0	0	1	1	0	0	1	0	0	1	0	1	1	0	1
1, 0	1	0	1	0	0	1	0	0	1	0	1	1	0	1	0	1
1, 1	1	0	1	0	1	0	0	0	0	1	1	1	1	0	1	0

Table 3.1: Possible functions for a node with $K = 2$. The total number of functions is $2^{2^K} = 16$

canalyzing functions of the first type ($f_{3,4,5,6}$) have the property that their value depends only on one of the inputs. f_3 and f_4 simply copy and invert the value of the first input respectively. f_5 and f_6 do the same for the second input, where the first input has no effect. The eight functions belonging to canalyzing functions of the second type have three times 1 or three times 0 in their outputs, such that for each of the two inputs there exists one value that fixes the output irrespective of the other input. f_{15} and f_{16} are the reversible functions because they change whenever an input changes.

When assigning a function to a specific node we can choose one of these possible functions according to a probability distribution. The simplest case is the constant distribution, where each function has the probability $1/2^M$. The other most frequently used probability distributions are listed below.

1. Biased functions: If a functions has the output value 1 n times and 0 $M - n$ times, it has the probability $p^n(1 - p)^{M-n}$. For the special case $p = 1/2$ all functions have the same probability $1/2^M$.
2. Weighted Classes: The classes are assigned a weight and all the functions in the same class have equal probability. The sum of the weights of the classes must

equal to 1.

3. **Canalyzing functions:** Only the functions belonging to the canalyzing class are chosen. The motivation for this is the fact that gene regulatory networks have many canalyzing functions [8, 9].
4. **Threshold functions:** If we denote the local state of each node with s_i , i being the index of the node, the update rule for a network with N nodes is in the form:

$$s_i(t+1) = \begin{cases} 1, & \text{if } \sum_{j=1}^N (c_{ij}(2s_j - 1) + h) \geq 0 \\ 0, & \text{else} \end{cases}$$

The c_{ij} coefficients are the coupling constants. If node i does not receive any input from node j , c_{ij} is zero. If node j is the activator or inhibitor of node i then $c_{ij} = 1$ or $c_{ij} = -1$ respectively.

5. If the K values of each node is the same, all nodes can be assigned the same function, so the network is then a cellular automaton with random wiring.

An illustration of the dynamics of a Boolean network with three nodes is shown in Fig. 3.1 [7].

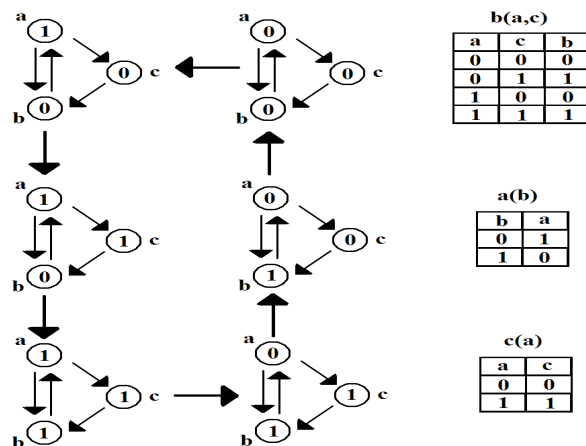


Figure 3.1: Trajectory of a Boolean network.

Probabilistic Boolean Networks:

Often, the experimental results are not enough to fully understand the system and choose the appropriate functions. This uncertainty must be involved in the system. A method for this is as follows. The functions are assigned a probability consistent with prior data. According to this probability distribution, at each time the node is updated a randomly chosen function is assigned to the node [10]. This model generates a sequence of global states constituting a *Markov chain*. A Markov chain is a stochastic process in which the next state depends only on the present state, irrespective of the past states leading the trajectory to the present states [11].

Additionally, Thomas [23] suggested that time delays must be introduced to these models. Because if the signal is fired for activation or deactivation of a specific gene it takes some time for this gene to respond. This response time depends both on the regulator and the regulated gene. So for every interaction in the system a time delay must be assigned. Such modelling reveals steady states other than those observed without time delays.

The disadvantage of the Boolean network models lies in the exponential growth of the the number of global states and possible functions. So the dynamics can efficiently be analyzed only for small networks.

Petri Nets:

Petri nets [12] are non-deterministic models which enable analysis of large metabolic networks [13]. Chaouiya *et al.* showed that Petri nets can also be used by modelling gene regulatory networks using Boolean functions [14]. Steggle *et al.* found out that the synchronous dynamics of a Boolean network can be observed using Petri nets where the uncertainties are also a part of the model [15].

In Fig. 3.2 a small Petri net is shown [6]. Here the light blue circles are the "places" which correspond to the nodes in the Boolean networks. The rectangles show the "transitions" representing the regulatory functions. Input places are connected

to transitions and transitions to their output places with arcs. Places that receive discrete values are called tokens (dark blue dots). The distribution of the tokens determine the state of the system. If a transition is fired (activated) it takes one token from the input place and puts it into the output place. Only the transitions that have enough tokens in their input places can be fired at any time. In this example, every transition takes one token from every input place and puts one token per each output place. On the arrows the fired transition is labeled. Transitions t_1 and t_3 can be fired alternately but after t_2 has fired no other transition is possible.

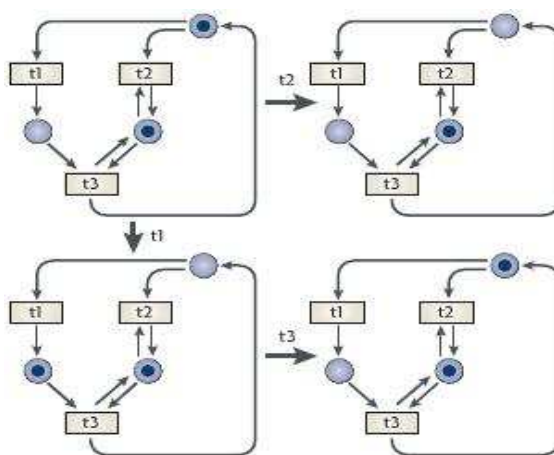


Figure 3.2: Illustration of a Petri net.

3.3.2 Continuous Models

Since the logical models are discrete valued their results are not accurate. Generally the biological experiments yield real continuous results such as reaction rates, cell mass, cell cycle length and amount of gene expression. A comparison between the experiment results and those of the model are more reliable when we use continuous models where real valued parameters are used over a continuous time-scale. Some important models belonging to this class are, *continuous linear models*, *models of transcription factor activity*, *regulated flux balance analysis* and *ordinary differential equations* [6].

Ordinary Differential Equations:

With ordinary differential equations it is possible to describe instantaneous changes in the levels of some network entities. Via a set of coupled differential equations one can relate for example the changes in the gene expression levels (S_i) with time to the levels of proteins present in the medium in the form:

$$\frac{dS_i}{dt} = f_i(S_1, S_2, \dots, S_N).$$

In the equations parameters are involved representing the reaction constants or rate of synthesis and degradation. This system can be solved analytically if the network is small. The solutions then will give the changes in the levels over time. Large networks usually require numerical solutions.

Chapter 4

YEAST CELL CYCLE NETWORK

4.1 Network Structure

Regulatory interactions among cell cycle proteins can be found in a variety of sources [17, 18]. A recent study claims that 11 genes (see Chp. 2) suffice to simulate and understand the cell cycle process in *Saccharomyces cerevisiae*. The regulatory interactions between these genes and their proteins comprise a simplified network (Fig. 4.1). Here

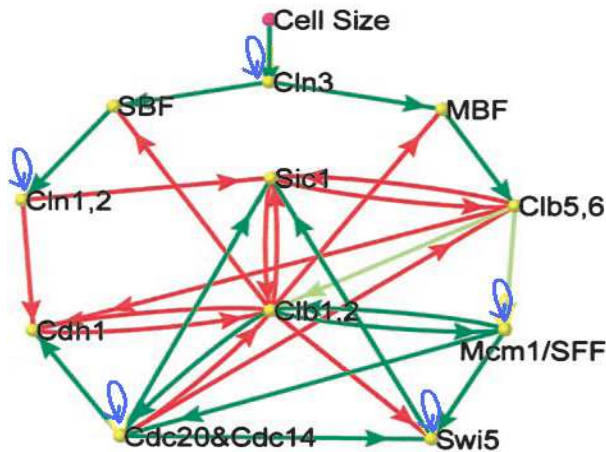


Figure 4.1: Cell Cycle Network of Li et al.

the green edges represent the activators and the red edges represent the inhibitors. The blue loops indicate a self degradation of the genes which are not down regulated by other genes in this network. So the genes inhibited by the products of genes not belonging to this network are treated as self degraded, which is a simplification of the real degradation process. To study the dynamics of this network Li et al. used Kauffman's Boolean network approach [24], where the state of gene i at time $t + 1$ is

determined by a Boolean threshold function in the form

$$S_i(t+1) = \begin{cases} 1, & \sum_j c_{ij} S_j(t) > 0 \\ 0, & \sum_j c_{ij} S_j(t) < 0 \\ S_i(t), & \sum_j c_{ij} S_j(t) = 0 \end{cases} \quad (4.1)$$

where $c_{ij} = 1$, for a green arrow from gene j to gene i and $c_{ij} = -1$, for a red arrow from j to i . Hence the inhibition and activation effects have the same strength. This function (4.1) corresponds to a majority rule. For a gene, if more activator genes are expressed at time t , itself will be activated at time $t+1$. If more inhibitor genes are expressed at time t the corresponding gene will be turned off in the next step. If there are equal numbers of activators and inhibitors which are expressed, the gene preserves its state.

However the function 4.1 is only used for the genes which are not self-regulated. For the self degraded genes, this function is slightly modified. A *time delay parameter* t_d is defined such that a gene with a blue loop will be turned off at time $t+t_d$, if its total input is zero from time $t+1$ to $t+t_d$. i.e. $S_i(t+t_d) = 0$. In the currently used model $t_d = 1$, so that a gene with zero input at time t will be off at time $t+1$.

We can specify two sets G_s and G_{ns} which contain the self degraded and the not self degraded genes respectively:

$$G_s = \{Cln3, Cln1 - 2, Cdc20 - 14, Swi5, Mcm1/SFF\} \quad (4.2)$$

$$G_{ns} = \{SBF, MBF, Sic1, Clb5 - 6, Clb1 - 2, Cdh1\} \quad (4.3)$$

The modified function (4.1) which will be used throughout the study has now the form:

$$S_i(t+1) = \begin{cases} 1, & \sum_j c_{ij} S_j(t) > 0, \forall i \\ 0, & \sum_j c_{ij} S_j(t) < 0, \forall i \\ 0, & \sum_j c_{ij} S_j(t) = 0 \wedge i \in G_s \\ S_i(t), & \sum_j c_{ij} S_j(t) = 0 \wedge i \in G_{ns} \end{cases} \quad (4.4)$$

Updating each node (gene) in this network at every timestep gives the trajectory of the states in time. Among other features of the network Li et al. were able to

observe that when they start the cell cycle process by exciting the stationary G_1 state with the cell size signal (Start), the system follows the cell cycle sequence, going from G_1 to the S phase, the M phase and finally returning back to the stationary G_1 state. Additionally they saw that this stationary G_1 phase is the attractor of the network with the biggest basin of attraction. The gene configurations corresponding to specific phases of the cell cycle are identified in their article as in Table 4.1.

	1	2	3	4	5	6	7	8	9	10	11
	Cln3	MBF	SBF	Cln1,2	Cdh1	Swi5	Cdc20	Cib5	Sic1	Cib1,2	Mcm1
Stationary G_1	0	0	0	0	1	0	0	0	1	0	0
Start	1	0	0	0	1	0	0	0	1	0	0
G_1 time 2	0	1	1	0	1	0	0	0	1	0	0
G_1 time 3	0	1	1	1	1	0	0	0	1	0	0
G_1 time 4	0	1	1	1	0	0	0	0	0	0	0
S	0	1	1	1	0	0	0	1	0	0	0
G_2	0	1	1	1	0	0	0	1	0	1	1
M time 7	0	0	0	1	0	0	1	1	0	1	1
M time 8	0	0	0	0	0	1	1	0	0	1	1
M time 9	0	0	0	0	0	1	1	0	1	1	1
M time 10	0	0	0	0	0	1	1	0	1	0	1
M time 11	0	0	0	0	1	1	1	0	1	0	0
G_1 time 12	0	0	0	0	1	1	0	0	1	0	0

Table 4.1: Global states corresponding to cell cycle phases. Since G_1 and M phases are subdivided into short phases there are multiple global states corresponding to the same main phase.

Since all of the genes are updated synchronously, given an initial state vector,

$$\vec{S}_s(t=0) = [S_1(0), S_2(0), \dots, S_i(0), \dots, S_N(0)], \quad 1 \leq i \leq N \quad (4.5)$$

with N being the total number of nodes (11 in this case), the trajectory of the system

is determined by the Boolean functions in (4.4). Here the subscript s refers to the synchronous case. Thus the global state vector of the system at every time step can be written as

$$\vec{S}_s(t+1) = [S_1(t+1), S_2(t+1), \dots, S_i(t+1), \dots, S_N(t+1)], \quad 1 \leq i \leq N, \quad (4.6)$$

where $S_i(t+1)$ is given by Eq. (4.4).

But it is more natural to think that the genes are updated at different times. Because after one gene is activated it takes some time to produce the amount of substance which interacts with another gene positively or negatively. This time interval is different for all genes, so if two genes are activated at the same time, their effect will be felt by other genes at different times. The same is the case when a gene is to be turned off. The products of different genes fall in different times under the threshold value. So although the genes are inactive their products continue to interact with other genes. Therefore it is impossible that all genes are updated at the same time. Hence, it is generally accepted that an asynchronously updated network is more realistic. We here perform an asynchronous analysis of the cell cycle network of the budding yeast.

At the end of our study we will compare our results with those obtained by Li et al.

4.2 Asynchronously Updated Cell Cycle Network

The same network discussed above is used in our study with two additions. In our model the *cell size signal* is also a node which is down regulated by Clb5 and activates Cln3. The reason for this choice is that the cell size signal does not turn off immediately after the cell cycle process has begun. Until the S phase the cell size is still big for just one set of chromosomes. But when the DNA synthesis begins at the S phase the relative cell size is reduced so that this signal should be turned off. We added a negative signal from Clb5 to cell size due to the fact that Clb5 is a gene which is active in the S phase and indicates that the G_1 phase is over and also the duplication

of the DNA has begun. So in our network Clb5 inhibits the cell size signal, which initiates all the cell cycle process. The modified network with $N = 12$ nodes and 31 edges can be seen in Fig. 4.2. By doing so the global states corresponding to cell

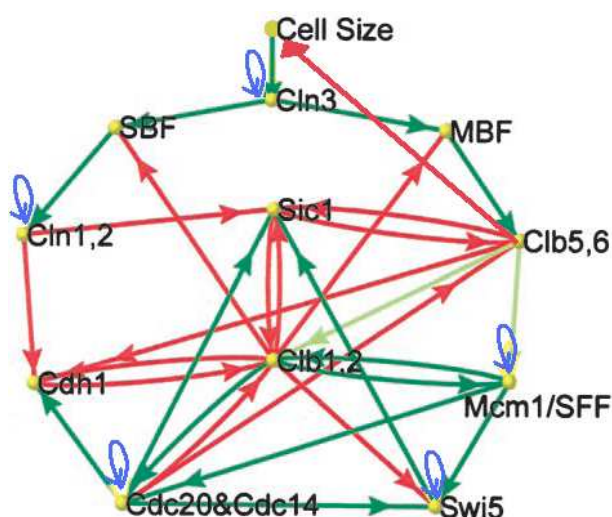


Figure 4.2: Gene regulatory network of the yeast cell cycle.

phases in Table 4.1 are also slightly modified. The new states are shown in Table 4.2.

The nodes are enumerated such that in the connectivity matrix C , the row and the column with index i represent the gene with the corresponding id number. In this matrix the elements c_{ij} are the coefficients used in Eq. (4.4). For example $c_{5,4} = -1$ shows that Cdh1 is down regulated by Cln1. The id numbers for the genes and the connectivity matrix C are as follows:

Fixed state	1	2	3	4	5	6	7	8	9	10	11	12	ID Number
	Cln3	MBF	SBF	Cln1,2	Cdh1	Swi5	Cdc20	Clb5	Sic1	Clb1,2	Mcm1	Cell Size	
Stationary G_1	0	0	0	0	1	0	0	0	1	0	0	0	137
Start	0	0	0	0	1	0	0	0	1	0	0	1	138
G_1 time 2	0	1	1	0	1	0	0	0	1	0	0	1	1674
G_1 time 3	0	1	1	1	1	0	0	0	1	0	0	1	1930
G_1 time 4	0	1	1	1	0	0	0	0	0	0	0	1	1794
S	0	1	1	1	0	0	0	1	0	0	0	0	1809
G_2	0	1	1	1	0	0	0	1	0	1	1	0	1815
M time 7	0	0	0	1	0	0	1	1	0	1	1	0	311
M time 8	0	0	0	0	0	1	1	0	0	1	1	0	103
M time 9	0	0	0	0	0	1	1	0	1	1	1	0	111
M time 10	0	0	0	0	0	1	1	0	1	0	1	0	107
M time 11	0	0	0	0	1	1	1	0	1	0	0	0	233
G_1 time 12	0	0	0	0	1	1	0	0	1	0	0	0	201

Table 4.2: Global states corresponding to cell cycle phases after modification of the network. The ID numbers are calculated as shown in Eq. (4.11).

$$C = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 1 & -1 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & -1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 1 & 1 & -1 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & -1 & 1 & -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \end{bmatrix}, \text{ with}$$

i, j	Protein
1	Cln3
2	MBF
3	SBF
4	Cln1
5	Cdh1
6	Swi5
7	Cdc14,20
8	Clb5,6
9	Sic1
10	Clb1
11	Mcm1
12	Cell Size

At each step only one node is chosen randomly and updated via the update function given in (4.4). At any time step, the probability of choosing node i out of all

nodes $\{1, 2, \dots, N\}$ is constant and given by

$$P(i) = \frac{1}{N} = \frac{1}{12}, \quad \forall i. \quad (4.7)$$

The evolution of the global state vector is now not deterministic and with subscript a referring to the asynchronous case, it can be given as

$$\vec{S}_a(t+1) = [S_1(t), S_2(t), \dots, S_{i-1}(t), S_i(t+1), S_{i+1}(t), \dots, S_N(t)], \quad (4.8)$$

where i can have any value in $\{1, 2, \dots, N\}$ with probability $P(i)$ given in (4.7).

Compared to the synchronous case, N steps of updates in asynchronous case are equivalent to a one step update in synchronous case, since we expect that after N steps all of the nodes will have been updated once on average.

4.2.1 The Simulation Procedure of The Cell Cycle Process

To investigate the time evolution of the network beginning from an initial global state, we calculated the time average values of the genes in each $N_{step} = 200$ time steps over $N_{sim} = 100$ simulations. The method of the procedure is described below:

1. Choose an initial global state:

$$\vec{S}_a(t=0) = [S_1(0), S_2(0), \dots, S_N(0)]$$

2. Pick a random node $i \in \{1, 2, \dots, N\}$ with probability $P(i)$ in Eq. (4.7).

3. Find the new local states:

$$S_j(t+1) = \begin{cases} \text{Eq. (4.4)}, & j=i \\ S_j(t), & j \neq i \end{cases}, \quad 1 \leq j \leq N$$

4. Repeat the 2. and 3. steps N_{step} times and find all

$$S_j(t), \quad 0 \leq t \leq 200, \quad 1 \leq j \leq N.$$

5. Perform N_{sim} simulations by repeating the whole process and find all

$$S_j^k(t), \quad 0 \leq k \leq 100, \quad 0 \leq t \leq 200, \quad 0 \leq j \leq N.$$

where k is the number of simulation.

6. Find average values $S_i^{avg}(t)$, $0 \leq t \leq N_{step}$ over N_{sim} simulations:

$$S_j^{avg}(t) = \langle S_j^k(t) \rangle \text{ over all } k, \quad \forall j$$

7. Plot $S_j^{avg}(t)$ vs t for each $j \in \{1, 2, \dots, N\}$.

The results of the simulations described above are shown in Fig. 4.3 with the initial global state being the *Start* state (see Table 4.2). Here each vertical line corresponds to a cell cycle phase. The intersection points of graphs with these lines show the expression levels of genes in these phases. In the stationary G_1 phase the cell is small and there are only the products of the genes Cdh1 and Sic1 are present. These two are the inhibitors of the other genes responsible for the cell cycle process. But as the size of the cell gets big enough to divide the cell size signal is activated, by a process which is not entirely understood until now [16]. So the *Start* state in Table 4.2 is the biologically relevant state for initiating the cell cycle process. We were able to observe that this coarse grained simulation help us to follow the cell cycle phases according to Li.

4.2.2 Transfer Matrix Description

The second method makes use of the advantages of the matrix algebra. With the help of the update function (Eq. (4.4)), it is possible to find the probabilities of N global states which are accessible from an initial state, when one of the N local states is changed. Totally there are

$$M = 2^N = 4096 \tag{4.9}$$

global states. We can construct a $M \times M$ *stochastic evolution matrix* T whose elements give the transition probabilities between the global states such that

$$T_{ij} = p(n(\vec{S}(t+1)) = i \mid n(\vec{S}(t)) = j), \tag{4.10}$$

where

$$n(\vec{S}) = 1 + \sum_{k=1}^N S_k \cdot 2^{N-k}. \tag{4.11}$$

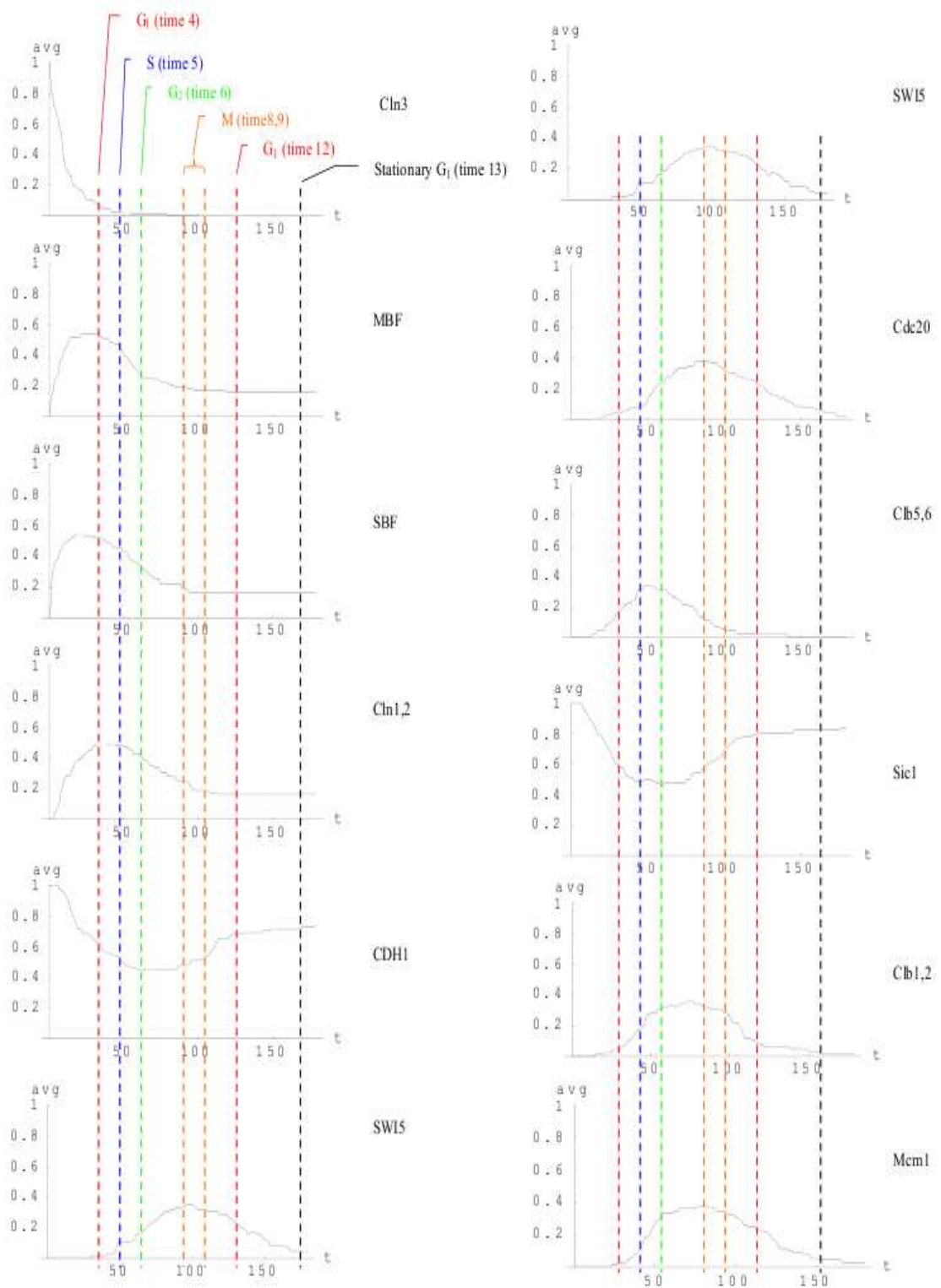


Figure 4.3: Results obtained by the average of 100 simulations.

The global state id, defined as in Eq. (4.11), is the decimal number corresponding to binary sequence \vec{S} . This ordering elucidates the internal structure of the matrix T as discussed below.

To find the elements of T, we update one of the local states of a global state i in each turn. Looking at the total numbers of different outcomes we can assign a value to each element of the evolution matrix.

Since there are 4096 global states among which only 12 are accessible from any specific state, the evolution matrix is very sparse. Using the *Hamming Distance (HD)*, the distribution of the nonzero elements can be found which gives its unique pattern to the matrix. HD between two strings of equal length is the number of positions for which the corresponding symbols are different. Hence in our case we can define HD as

$$d(\vec{S}_i, \vec{S}_j) = |\vec{S}_i - \vec{S}_j|^2 \quad (4.12)$$

whose result gives the number of different local states. Since each time at most one local state is changed, a transition between the states is only possible when their HD is at most 1. This leads to the property that

$$T_{ij} = \begin{cases} 0, & d(\vec{S}_i, \vec{S}_j) > 1 \\ 0 \leq p \leq 1, & d(\vec{S}_i, \vec{S}_j) \leq 1. \end{cases} \quad (4.13)$$

The pattern of a evolution matrix for a system with 7 nodes is presented in Fig. 4.4.

In addition to this pattern there is also one more constraint on the nonzero elements due to the Eq. (4.4). First look at the cases where the updated local state belongs to a not-self-regulating node. Even if the node changes its local state, the local states of its neighbors stay the same. So the updated node must preserve its value and cannot go back to the original state. So the number of nonzero elements is almost reduced to half such that

$$T_{ij} \neq 0 \Rightarrow T_{ji} = 0. \quad (4.14)$$

If a self-regulating node is updated, this property may not hold. But we observed that due the structure of our network, such a case does not exist and (4.14) is a general

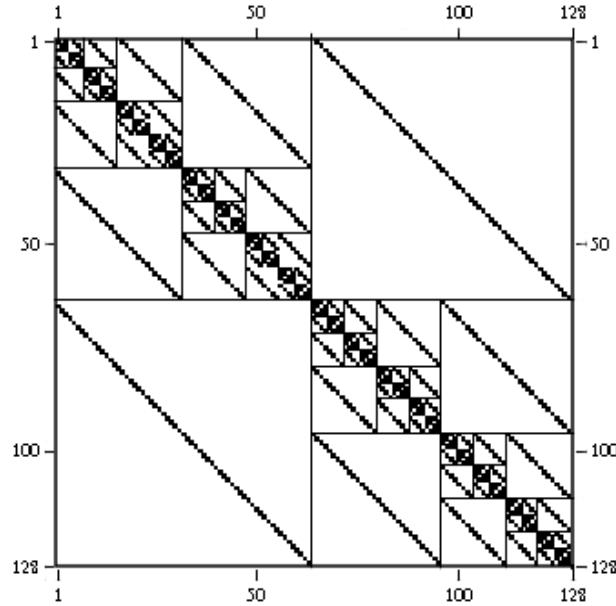


Figure 4.4: The pattern of the evolution matrix for a 7-node network. The black points represent the nonzero elements.

feature of the evolution matrix. As a result the evolution matrix obtained by our network with the majority rule function has the pattern shown in Fig. 4.5.

The evolution matrix is a stochastic matrix so that

$$\sum_{i=1}^M T_{ij} = 1, \forall j. \quad (4.15)$$

A fixed global state $\vec{S} = \vec{F}$ is characterized by the property that the diagonal element

$$T_{ii} = 1, \quad (4.16)$$

where i is given with Eq. (4.11). This means that this state has zero probability of passing to another global state other than itself and stays in this state indefinitely. In our network there are 9 fixed states compared to the 7 fixed states in Li's article. The two additionally fixed states are those with gene Cln3 active. They are shown in Table 4.3 with their corresponding state numbers.

These fixed states can also be found by looking at the eigenvectors of the evolution matrix. Since the sum of elements in every column of matrix T is unity the deter-

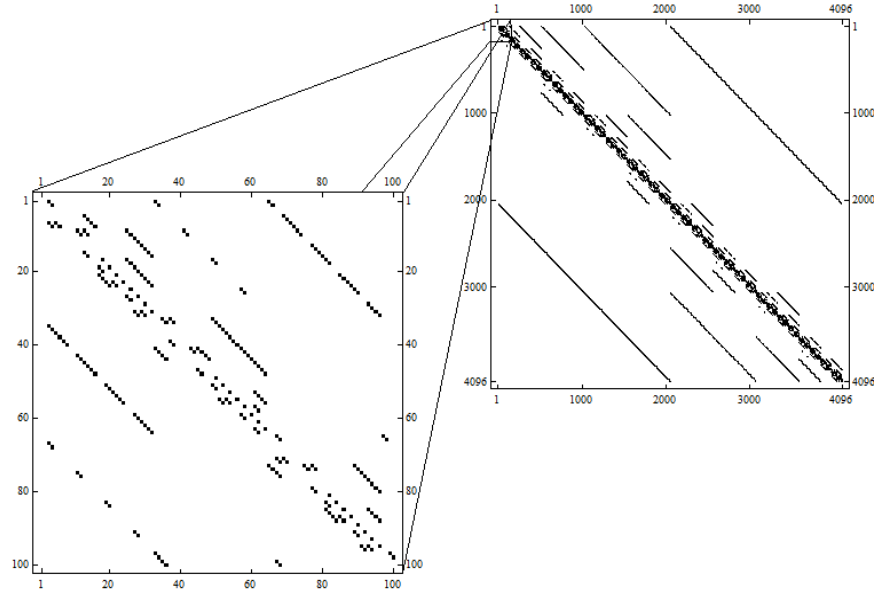


Figure 4.5: The pattern of the evolution matrix, with top left corner zoomed.

minant of the matrix $T - I$ is equal to zero and the *Perron – Frobenius* theorem ensures that one of the eigenvalues must be equal to 1 and the absolute values of other eigenvalues are lower than 1. The eigenvectors corresponding to the eigenvalue 1 have a very special meaning. Its components yield the probabilities to find the system in each of its states for $n \rightarrow \infty$. Hence the eigenvector with eigenvalue 1 in the form

$$\vec{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_i \\ \vdots \\ a_M \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \quad (4.17)$$

shows that the probability of finding the system at the global state with id number i is 1, when the system has been updated sufficiently long.

These 9 fixed states are called the point attractors of the network. Since our update function (Eq. (4.4)) does not allow a cycle between the global states, all of

Fixed state	1	2	3	4	5	6	7	8	9	10	11	12	Number
	Cln3	MBF	SBF	Cln1,2	Cdh1	Swi5	Cdc20	Clb5	Sic1	Clb1,2	Mcm1	Cell Size	
$\vec{F}_1 (G_1)$	0	0	0	0	1	0	0	0	1	0	0	0	137
\vec{F}_2	0	0	1	1	0	0	0	0	0	0	0	0	769
\vec{F}_3	0	1	0	0	1	0	0	0	1	0	0	0	1161
\vec{F}_4	0	0	0	0	0	0	0	0	1	0	0	0	9
\vec{F}_5	0	1	0	0	0	0	0	0	1	0	0	0	1033
\vec{F}_6	0	0	0	0	0	0	0	0	0	0	0	0	1
\vec{F}_7	0	0	0	0	1	0	0	0	0	0	0	0	129
\vec{F}_8	1	0	1	1	0	1	1	0	0	1	1	1	2920
\vec{F}_9	1	1	1	1	0	1	1	0	0	1	1	1	3944

Table 4.3: 9 fixed states of the network.

the global states will arrive at one of these fixed states, when the network is updated sufficiently long. The n^{th} power of the evolution matrix

$$T_{ij}^{(n)} = \underbrace{T \cdot T \cdots T \cdot T}_n \quad (4.18)$$

gives the probability of arriving at state i at $t=n$ when the initial state is j such that

$$p(\vec{S}(t=n) = \vec{S}_i \mid \vec{S}(t=0) = \vec{S}_j) = T_{ij}^{(n)}. \quad (4.19)$$

To find the probabilities with which an initial state arrives at each of the fixed states, we define the *fixed evolution matrix* as

$$T^\infty = \lim_{n \rightarrow \infty} T^{(n)}. \quad (4.20)$$

Denoting the probability of arriving at \vec{S}_i , when the initial state is \vec{S}_j , as $P(\vec{S}_i, \vec{S}_j)$, the elements of this matrix have the property that

$$P(\vec{F}_k, \vec{S}_j) = p(\lim_{n \rightarrow \infty} \vec{S}(t=n) = \vec{F}_k \mid \vec{S}(t=0) = \vec{S}_j) = T_{ij}^\infty, \quad (4.21)$$

where i is given by Eq. (4.11). We say that the fixed state $F_k, k = 1, 2, \dots, 9$ is preferred by the initial state \vec{S}_j , if the probability of arriving at this fixed state is the highest for \vec{S}_j . The preferred fixed state being \vec{F}_m , this highest probability is denoted as

$$P_m(\vec{S}_j) = \text{Max}(P(\vec{F}_k, \vec{S}_j)), k \in \{1, 2, \dots, 9\}. \quad (4.22)$$

and that \vec{F}_m is preferred by \vec{S}_j is symbolically shown as

$$\vec{S}_j \rightarrow \vec{F}_m.$$

The global state \vec{S}_j then lies in the *basin of attraction* of \vec{F}_m . The basin of attraction of a fixed state \vec{F}_m expressed as $B(\vec{F}_m)$ is the set of all initial states which prefer \vec{F}_m . Hence,

$$B(\vec{F}_m) = \{\vec{S}_1, \dots, \vec{S}_j, \dots, | \vec{S}_j \rightarrow \vec{F}_m\}. \quad (4.23)$$

Knowing the initial state, T^∞ allows us to compute the probabilities of arriving at different fixed states. With the help of this matrix we can also find the probabilities of reaching different fixed states, when the initial state is chosen randomly. Let's an ensemble \vec{V} as

$$\vec{V} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_i \\ \vdots \\ v_M \end{bmatrix}, \quad (4.24)$$

where M is given by Eq. (4.9) and v_i is the probability of finding \vec{S}_i in the ensemble. Then the elements of the column vector

$$\vec{V}^\infty = T^\infty \cdot \vec{V}, \quad (4.25)$$

have the property that

$$\vec{V}_i^\infty = P(\vec{S}_i, \vec{S}_j | P(\vec{S}(t=0) = \vec{S}_j) = v_i). \quad (4.26)$$

Now we can calculate two different average probabilities. The first one is the average probability of arriving at fixed state $F_m, m = 1, 2, \dots, 9$, among the initial states which prefer it. So we can define

$$P_{avg}(\vec{F}_m) = \langle P_m(\vec{S}_j) \rangle | \vec{S}_j \in B(\vec{F}_m). \quad (4.27)$$

Attractor	$ B(\vec{F}_m) $	P_{avg}	P_{random}
\vec{F}_1	3679	0.66	0.620224
\vec{F}_2	225	0.67	0.07402
\vec{F}_3	102	0.59	0.045991
\vec{F}_4	20	0.54	0.203688
\vec{F}_5	14	0.57	0.013335
\vec{F}_6	5	0.64	0.011774
\vec{F}_7	2	0.5	0.007585
\vec{F}_8	31	0.63	0.017256
\vec{F}_9	9	0.58	0.006129

Table 4.4: Probabilities of fixed states. Basin of attraction is the number of initial states preferring the attractor. P_{avg} is the average probability to reach the attractor among the initial states preferring the attractor. P_{random} is the probability to reach the attractor from a random initial state.

The second average probability is the probability of arriving at \vec{F}_m , when the initial state is random. These values are the 9 non-zero values of \vec{V}^∞ , corresponding to each of the fixed states. So,

$$P_{random}(\vec{F}_m) = \vec{V}_i^\infty, \quad (4.28)$$

with $i=n(\vec{F}_m)$.

These two types of average probability values for all fixed states are shown in Table 4.4, where for calculating P_{random} , all the states are assigned equal probability of $1/M$, with M given by Eq. (4.9).

We see that G_1 phase has the biggest basin of attraction and the system prefers to arrive at this global state. Beginning from a random initial state the system will return to this phase with a probability of 0.62 which is a big probability when compared with the others.

The fixed state \vec{F}_4 has also a relative high P_{random} value of 0.203, which makes

t_d	$P(\vec{F}_1)$	$P(\vec{F}_4)$
1	0.742	0.248
2	0.874	0.118
3	0.918	0.074
4	0.966	0.031
5	0.986	0.01

Table 4.5: Probabilities of arriving at F_1 and F_4 when the initial state is *Start* while changing the parameter t_d .

it also favorable as an attractor. This was unexpected, because this state has no biological meaning. We investigated how the system behaves, as the parameter t_d changes. We observed that as we increased it from 1 to 5 the probability of reaching it decreases when the initial state is *Start* and at $t_d = 5$ it reaches a value below 0.025. But meanwhile the probability of reaching \vec{F}_1 increases. The results are show in Table 4.5. This tells us that the lifetimes of the transcription factors are also relevant parameters for the time evolution of the cell-cycle regulation.

4.2.3 Stability of the Attractors

The only biologically relevant fixed state is the \vec{F}_1 . So if the cell really tries to arrive at this phase, this state must be very stable. And the other biologically irrelevant fixed states are expected to be unstable. For this reason we investigated how these attractors behave under perturbation. The perturbation process we used here includes changing the value of one of the local states. The stability of an attractor is then examined by looking at the probabilities of returning back to itself, denoted as $P_{back}(\vec{F}_m)$, under 12 possible perturbations, where

$$P_{back}(\vec{F}_m) = P(\vec{F}_m, \vec{S}_j \mid d(\vec{F}_m, \vec{S}_j) = 1). \quad (4.29)$$

All of these values can be found in fixed evolution matrix \vec{T}^∞ and the results are shown in Table 4.6. Only two attractors, \vec{F}_1 (G_1) and \vec{F}_2 , tend to return back

Perturbed gene	$P_b(\vec{F}_1)$	$P_b(\vec{F}_2)$	$P_b(\vec{F}_3)$	$P_b(\vec{F}_4)$	$P_b(\vec{F}_5)$	$P_b(\vec{F}_6)$	$P_b(\vec{F}_7)$	$P_b(\vec{F}_8)$	$P_b(\vec{F}_9)$
Cln3	0.577	0.504	0.5	0.415	0.5	0.3385	0.3351	0.5	0.3333
MBF	0	0.0093	0	0	0	0.012	0.005	0	0
SBF	0	0	0	0	0	0	0	0	0
Cln1,2	0.3333	1	0.393	0.5	0.509	1	0.5	0.5	0.5
Cdh1	0	1	0	0	0	0	0	0.5	0.5
Swi5	1	1	1	1	1	0.5	0.5	1	1
Clb5	0.771	0.143	0.34	0.53	0.065	0.025	0.032	0.5	0
Sic1	0	1	0.22	0	0.02	0	0	0	0
Clb1,2	0.79	0.11	0.34	0.46	0.25	0.07	0.34	0	0
Mcm1	1	0.667	1	0.625	0.625	0.31	0.46	0.5	0.5
Cell size	0.74	0.04	0	0.248	0	0.0055	0.0018	0	0
Average	0.5176	0.5366	0.4	0.357	0.289	0.2127	0.2162	0.375	0.2778

Table 4.6: Behavior under perturbation. $P_b \equiv P_{back}$ Example: If Cln3 changes its value in F1, this new state return back to F1 with probability 0.577.

to itself with an average probability greater than 0.5. This suggests that these fixed states are relatively stable compared to the others. It can be seen that under perturbation of SBF none of the fixed states can turn back but a perturbation in Swi5 does not have a big influence on the trajectory of the fixed states.

The average values of probabilities of coming back for 12 different perturbations, can be extended to average probabilities of choosing another attractor under perturbation. We show these probabilities as

$$P_{pert}(\vec{F}_m, \vec{F}_n) = \sum_j P(\vec{F}_m, \vec{S}_j \mid d(\vec{S}_j, \vec{F}_n) = 1) \quad (4.30)$$

In this way we can analyze if the system has any fixed states which attract other fixed states when there is a perturbation in them. In Table 4.7 we see the \vec{F}_1 can attract the perturbed fixed states with relatively big probability.

	\vec{F}_1	\vec{F}_2	\vec{F}_3	\vec{F}_4	\vec{F}_5	\vec{F}_6	\vec{F}_7	\vec{F}_8	\vec{F}_9
\vec{F}_1	0.517	0.097	0.0972	0.153	0	0.029	0.104	0	0
\vec{F}_2	0.275	0.536	0	0.09	0	0.0916	0.0018	0	0
\vec{F}_3	0.388	0.001	0.4	0.095	0.1069	0.0034	0.0038	0	0
\vec{F}_4	0.318	0.097	0	0.356	0.0972	0.128	0.0014	0	0
\vec{F}_5	0.317	0.001	0.188	0.196	0.289	0.0049	0.0017	0	0
\vec{F}_6	0.331	0.097	0.006	0.248	0.0021	0.212	0.1007	0	0
\vec{F}_7	0.464	0.097	0.022	0.0682	0.0018	0.128	0.216	0	0
\vec{F}_8	0.402	0.001	0	0.134	0	0.0022	0.00057	0.375	0.0833
\vec{F}_9	0.460	0.001	0.003	0.154	0.0013	0.0022	0.00057	0.097	0.277

Table 4.7: The matrix A_{ij} , where $A_{ij} = P_{Pert}(F_j, F_i)$

4.2.4 Dynamics of the System

The evolution matrix T_{ij} has all the information about the system. By examining this matrix, it was seen that a global state mostly tries to stay in its state. The probabilities of transitions to other states are lower, but they are equal to each other. If enough updates are done, a global state will pass to a new state at the end. We are only interested in the transitions between the global states and investigate if they occur in the proper order. The time required to arrive at a specific global state is not our main concern. Therefore we set the diagonal elements to zero (except for the fixed states) and normalize the probabilities of passing to other accessible states, so that they add up to one. We called this new evolution matrix "transition matrix (T_N)". So the elements have the property that,

$$(T_N)_{ii} = \begin{cases} 1, & \vec{S}_i \in \{\vec{F}_1, \dots, \vec{F}_9\} \\ 0, & \vec{S}_i \notin \{\vec{F}_1, \dots, \vec{F}_9\} \end{cases} \quad (4.31)$$

and,

$$\sum_{i=1}^M (T_N)_{ij} = 1, \forall j. \quad (4.32)$$

The pattern for T_N looks like in Fig. 4.6. When this new transition matrix is used a

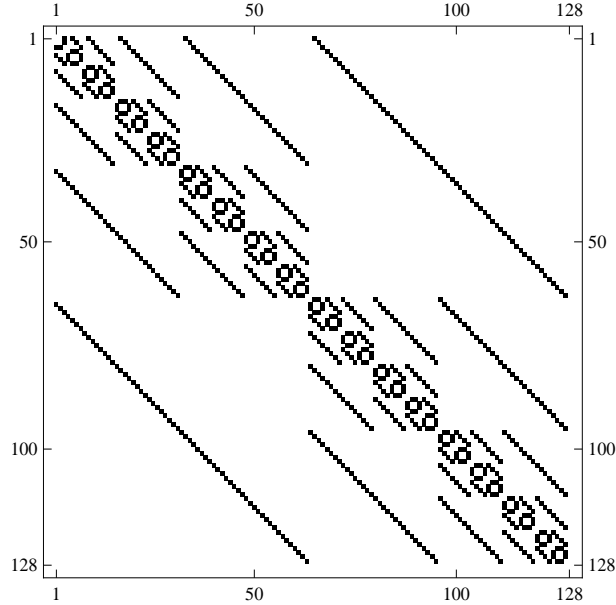


Figure 4.6: The pattern of T_N for a 7-node network.

global state cannot preserve its state at any time step and must pass to a new global state, unless it reaches a fixed one.

We begin with the *Start* state and look at the possible pathways the system can follow. In Fig. 4.7 we illustrate how the system evolves in time. Since there are a lot of different pathways it is impossible to show them all. So we only take the transitions into account which have a probability greater than 0.025. With the help of the transition matrix we can find that the probability to reach the stationary G_1 state (F1) is 74% which is 24.6% for F4 and 1.4% for other fixed states. The time steps leading to this result is shown in this figure.

As it can be seen in Fig. 4.7, between time steps 8 and 17 the system can be in different global states. Since we use the transition matrix, these steps do not correspond to the actual time steps. But as the M phase is reached the number of possible global states is decreased. To describe this behavior quantitatively the

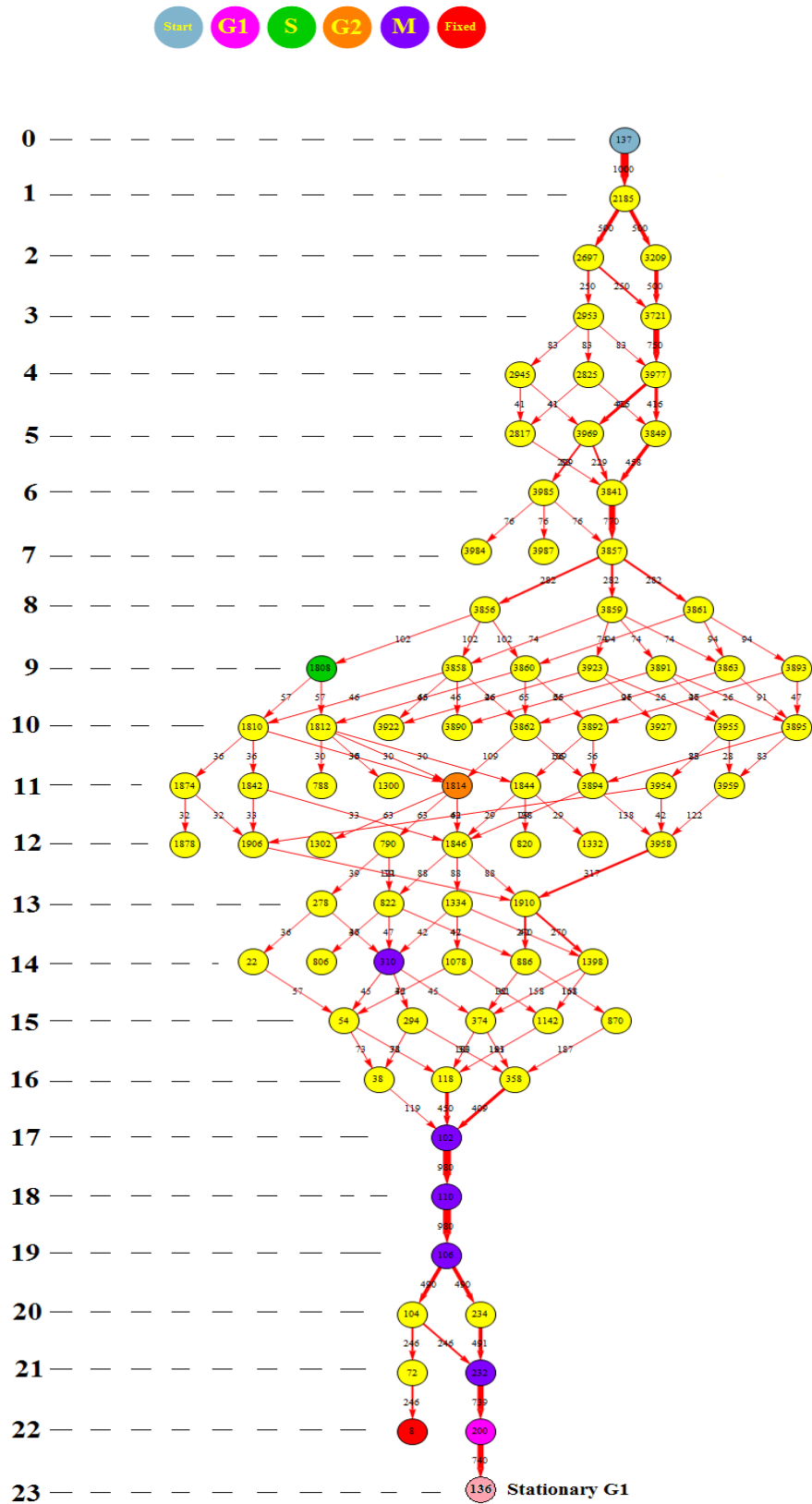


Figure 4.7: Flow of the system beginning from *Start*. Each line corresponds to a time step. The widths of the arrows are proportional to the transition probabilities.

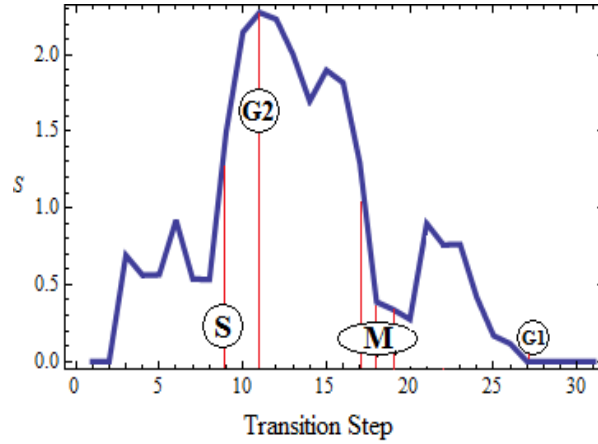


Figure 4.8: Sequential evolution of entropy. The number of steps are greater than those in the Fig. 4.7, because the states with very low probabilities are not shown in that figure.

entropy values at each time step are calculated as

$$S(t) = \sum_i p_i(t) \cdot \ln[p_i(t)], \quad (4.33)$$

where $p_i(t)$ is the probability of being at the global state i at time step t . The extreme values for the entropy correspond to the biologically defined cell phases. The evolution of the entropy is illustrated in Fig. 4.8.

If we use the original evolution matrix T in place of the transition matrix T_N we cannot observe these peaks in the behavior of entropy. Instead we have a smooth curve, which can be seen in Fig. 4.9.

4.2.5 Comparison with Random Networks

To see whether this gene regulatory network of *Saccharomyces cerevisiae* is a special network, we compared the features of this networks with the ones of random networks. Random networks are generated by reshuffling the edges of our regulatory network, such that the numbers of in degrees and out degrees of the nodes do not change. First we compared the entropy behavior in time. The random networks have a smoother behavior and the extreme values are not so ostensible. In Fig. 4.10 the results are

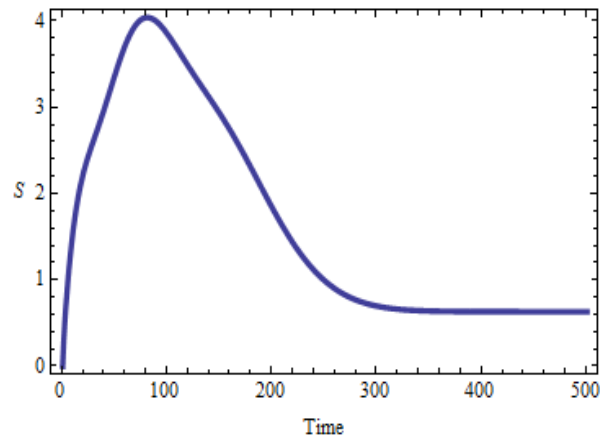


Figure 4.9: Time evolution of entropy when the original evolution matrix is used.

shown for some random networks (blue). The red curve is the average of the blue curves for 50 random networks. For comparison, the green curve belonging to gene regulatory network is also plotted.

4.2.6 Perplexities for Attractors

The number of fixed nodes have also been compared between the gene regulatory network and the random networks. 100 random networks have been investigated and in general they have more than 9 fixed states, the most being 34. Just 6 of these random networks have 8 or 9 fixed states indicating that the gene regulatory network is special in the sense of having very few fixed states. But in fact it is not so important how many fixed states there are. The significant feature is the probability distribution among the fixed states. For example consider a network having only two fixed states. If the size of the basins of attraction are the same for each of them, there won't be any preferred fixed state for the system. So neither of them characterizes the behavior. In the gene regulatory network there are 9 fixed states but one of them has the biggest basin of attraction, so there is a fixed state which is preferred by the system. In our case this is the F1 state, at which the system arrives with a probability of 62% (Table 4.4). The probabilities ending at another fixed state are very low compared to

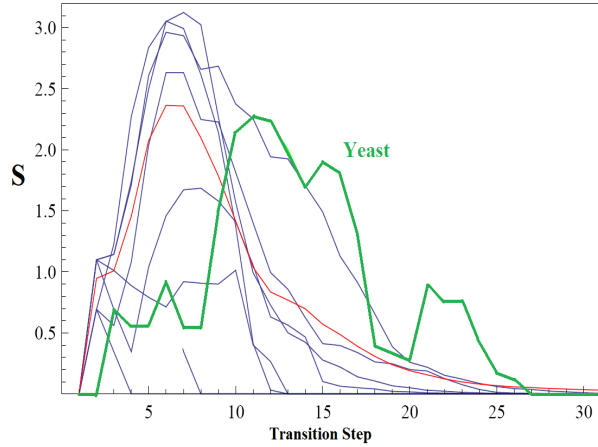


Figure 4.10: Entropy evolution for 8 random rewirings of the cell-cycle network and the yeast network for comparison.

F1. A quantitative description for this is the *perplexity*. Perplexity is calculated as

$$P = 2^{-\sum_i p_i \cdot \log_2 p_i} \quad (4.34)$$

where p_i is the probability of arriving at fixed state i . If a network has a preferred fixed state with a dominant basin of attraction, the perplexity value will be small. The perplexity increases as the probabilities for fixed states are getting closer. When compared with the 100 random graphs the gene regulatory network has a very low perplexity value of 3.32. The results for random networks are shown in Fig. 4.11.

4.3 Conclusion

We used a stochastic procedure to analyze the evolution of the gene regulatory cell cycle network of the yeast. But a low perplexity value of 3.32 tells us that the structure of this network allows the cell to act in an almost deterministic fashion. The fixed state with the biggest basin of attraction is the biologically relevant G_1 phase. When the system is stimulated to begin the cell cycle process, it tries to go back to the G_1 phase, after following the other cell cycle phases in proper order.

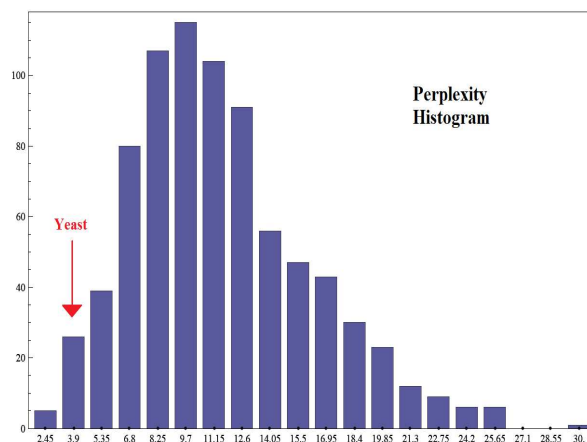


Figure 4.11: Perplexities of 100 random networks with the same topology as the gene regulatory network. The perplexity value of the yeast’s cell cycle network lies in the second column.

Our results suggest that, the cell-cycle network is a special network when compared with random networks with the same topology. The perplexity of the network (a normalized estimate of the attractor number) is significantly lower than its randomly rewired cousins. The network dynamics shows no limit cycles. The dynamical evolution of the system when perturbed by the ‘cell-size’ signal out of the G1 phase follows a path that has a non-monotonic entropy evolution, unlike the behavior of similar random networks. The extrema of the entropy evolution agree fairly well with the cell-cycle phases.

By increasing the value of t_d , we also found that, the correct choice of the lifetime of transcription factors is also an important factor for a realistic model. Although in the synchronous case Li et al. found that the behavior of the system is independent of this value, we found that it may significantly change the relative importance of the attractors.

BIBLIOGRAPHY

- [1] Alberts et al. Molecular Biology of the Cell. 5. ed. (2008).
- [2] Chartrand, G. Introduction to Graph Theory. Dover Publication New York, USA, 1985.
- [3] Mendenhall M.D., Hodge A. E., Microbiology and Molecular Biology Reviews, Dec. 1998, p.1191-1243.
- [4] Miguel C. Teixeira, Pedro Monteiro, Pooja Jain, Sandra Tenreiro, Alexandra R. Fernandes, Nuno P. Mira, Marta Alenquer, Ana T. Freitas, Arlindo L. Oliveira, and Isabel S-Correia, The YEASTRACT database: a tool for the analysis of transcription regulatory associations in *Saccharomyces cerevisiae*, Nucl. Acids Res., 2006, 34: D446-D451, Oxford University Press
- [5] Pedro T. Monteiro, Nuno Mendes, Miguel C. Teixeira, Sofia d'Orey, Sandra Tenreiro, Nuno Mira, Hlio Pais, Alexandre P. Francisco, Alexandra M. Carvalho, Artur Loureno and Isabel S-Correia, Arlindo L. Oliveira, Ana T. Freitas, YEASTRACT-DISCOVERER: new tools to improve the analysis of transcriptional regulatory associations in *Saccharomyces cerevisiae*, Nucl. Acids Res., Jan. 2008, 36: D132-D136, Oxford University Press
- [6] Kerlebach G., Shamir R. Modelling and analysis of gene regulatory networks, Nature Reviews Molecular Cell Biology, published online 17 September 2008.
- [7] Drossel, B. Random Boolean Networks, Reviews of Nonlinear Dynamics and Complexity: Volume 1, 2008..
- [8] S. E. Harris et al., Complexity 7 (2002), p.23.

-
- [9] S. A. Kauffman, C. Peterson, B. Samuelsson, and C. Troein, Proc. Nat. Acad. Sci. USA 101 (2004), p. 17102.
- [10] Shmulevich et al. Steady-state analysis of genetic regulatory networks modelled by probabilistic Boolean networks. *Comp. Funct. Genomics* 4, 601-608 (2003).
- [11] Bhattacharya, R. N., Majumdar, M. *Random Dynamical Systems: Theory and Applications*, Cambridge University Press, Cambridge, 2007.
- [12] Petri, C. A. *Kommunikation mit Automaten*. Schriften des Instituts für Instrumentelle Mathematik (1962).
- [13] Koch, I. et al. STEPP - Search tool for exploration of Petri net paths: a new tool for Petri net-based path analysis in biochemical networks. *In Silico Biol.* 5, 129-137 (2005).
- [14] Chaouiya, C. et al. *Proceedings of the 25th International Conference on Applications and Theory of Petri Nets* (eds Cortadella, J., Reisig, W.), Springer, Berlin, 2004.
- [15] Steggles, L. et al. Qualitatively modelling and analyzing genetic regulatory networks: a Petri net approach. *Bioinformatics* 23, 336-343 (2007).
- [16] Toone et al. Getting Started: Regulating the Initiation of DNA Replication in Yeast, *Annu. Rev. Microbiol.* 1997. 51:125-49.
- [17] Forsburg, S.L., Nurse, P. Cell Cycle Regulation in the yeasts *Saccharomyces cerevisiae* and *Schizosaccharomyces Pombe*. *Annu. Rev. Cell Biol.* 1991. 7:227-56.
- [18] Bähler, J. Cell-Cycle Control of Gene Expression in Budding and Fission Yeast. *Annu. Rev. Genet.* 2005. 39:69-94.

-
- [19] Dickinson, J. R., Schweizer, M. The metabolism and molecular physiology of *Saccharomyces cerevisiae*. Boca Raton, CRC Press, c2004.
- [20] Li et al. The Yeast Cell Cycle Network is Robustly Designed, PNAS April 6, 2004 vol. 101 no. 14 4781-4786.
- [21] Glass, L., Kauffman, S. A. The logical analysis of continuous, non-linear biochemical control networks. J. Theor. Biol. 39, 103-129 (1973)..
- [22] Kauffman, S. A. The Origins of Order: Self-Organization and Selection in Evolution (Oxford University Press, Oxford, 1993).
- [23] Thomas R. Regulatory networks seen as asynchronous automata: A logical description. J. Theor. Biol. 153, p. 1-23 (2001).
- [24] Kauffman, S. A. (1969) J. Theor. Biol. 22, 437-467.

VITA

Neşe Aral was born in Istanbul, Turkey on February 14, 1982. She graduated from Sankt Georg's Austrian High School, Istanbul, in 2001. She received her B. Sc. degrees in Physics and Mathematics from Istanbul Technical University in June 2006. In Koç University, she was accepted to Master program in Physics in August 2006, and to Doctorate program in Physics in January 2009 with full scholarship. In her graduate years she worked there as a teaching assistant and hold recitation in General Physics I,II and Quantum Statistical Mechanics. During her M.S. studies she received scholarships from Scientific and Technical Research Council of Turkey, TÜBİTAK. She studied renormalization group theory in statistical physics and has a research paper about hierarchical Ising lattices exhibiting chaotic behavior, published in PRB. She gave two contributed talks at the 96. Statistical Mechanics Meeting (New Brunswick, NJ, U.S.A.) and at the APS March Meeting 2009 (Pittsburgh, PA, U.S.A.). In her Ph.D. studies she wants to work on complex systems in biology.