DOKUZ EYLÜL UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

ON APPLICATIONS OF CIRCUITS SYSTEMS AND SIGNAL THEORY IN SYSTEMS BIOLOGY

by

Neslihan AVCU

August, 2008

İZMİR

ON APPLICATIONS OF CIRCUITS SYSTEMS AND SIGNAL THEORY IN SYSTEMS BIOLOGY

A Thesis Submitted to the
Graduate School of Natural and Applied Sciences of Dokuz Eylül University
In Partial Fulfillment of the Requirements for the Degree of Master of Science in
Electrical and Electronics Engineering

by

Neslihan AVCU

August, 2008

İZMİR

M.Sc THESIS EXAMINATION RESULT FORM

We have read the thesis entitled **"ON APPLICATIONS OF CIRCUITS SYSTEMS AND SIGNAL THEORY IN SYSTEMS BIOLOGY"** completed by **NESLİHAN AVCU** under supervision of **Prof. Dr. CÜNEYT GÜZELİŞ** and we certify that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

……..

Prof. Dr. Cüneyt GÜZELİŞ

Supervisor

…… **……**

Prof. Dr. Alp KUT

(Jury Member)

Asst. Prof. Dr. Güleser DEMİR

(Jury Member)

……

Prof.Dr. Cahit HELVACI

Director

Graduate School of Natural and Applied Sciences

ACKNOWLEDGMENTS

This thesis would not be possible without the valuable guidance and support of my advisor, Prof. Dr. Cüneyt GÜZELİŞ. I would like to thank my advisor, for his encouragement and support not only during my education but also in every stage of my experience in assistantship.

I would like to thank my parents, sisters, grandparents and aunt for their support and understanding. I know that they are always proud of my accomplishments; I will work to continue to make them proud during my life. My grandparents, Rahime and Tevfik AVCU, always grounded me and made me better able to do anything whatever I can. I thank my grandparents for their trust and patience.

Thank you to my friend and everyone, who was supported me over the years.

I also gratefully acknowledge the support of master fellowship as well funding from TUBITAK, BIDEB.

Neslihan AVCU

ON APPLICATIONS OF CIRCUITS SYSTEMS AND SIGNAL THEORY IN SYSTEMS BIOLOGY

ABSTRACT

Nowadays, the biological researches include studies that being in cell and molecular level. All the same, the researches also consist of the system level studies instead of the studies which are related with the biological property of the organisms. The assumption of biological organism as a system and the representation of the components' functions satisfy the good understanding of organism. The changing in the study subject needs to a multidisciplinary working. The disciplinary that is present by this way is called as Systems Biology.

The analysis of the structures and the dynamical equations of the common network motifs in biological networks satisfy to understand the functions of the biological organism in system level. Because controlling of the molecular structure by genetic information is a key role in studies such as diagnosis and therapy of disease, protein synthesis, drug treatment, the understanding of the functional and structural properties of the components in protein transcription networks is very important to control the network.

In this study, the components of protein transcription networks in organisms are analyzed according to their structural and functional properties.

Keywords: Transcription networks, network motifs, negative autoregulation, feed-forward loops, dynamic equations.

DEVRELER SİSTEMLER VE SİNYAL TEORİSİNİN SİSTEMSEL BİYOLOJİ ÜZERİNE UYGULAMALARI

ÖZ

Günümüzde biyolojik yapılardaki araştırmalar hücre ve moleküler seviyede çalışmalar içermektedir. Bununla birlikte yapıların biyolojik özellikleriyle ilgili araştırmaların yerini sistem düzeyindeki araştırmalar almaktadır. Biyolojik bir yapının bir sistem olarak kabul edilmesi ve içerdiği bileşenlerin fonksiyonlarının açıklanabilmesi yapının daha iyi bir şekilde anlaşılmasını sağlamaktadır. Çalışma konusundaki değişiklik farklı disiplinlerin birlikte çalışmasını gerektirmektedir. Bu şekilde ortaya çıkan çok disiplinli çalışma alanı Sistemsel Biyoloji olarak adlandırılmaktadır.

Biyolojik ağlarda sıkça rastlanan ağ motiflerinin yapısal özelliklerinin ve dinamik denklemlerinin incelenmesi bu yapıların fonksiyonlarının ve yapının sistem düzeyinde anlaşılmasını sağlamaktadır. Organizmada genetik bilgi ile moleküler yapıdaki değişimin kontrol edilmesi hastalık tanı ve tedavisi, protein sentezi, ilaç üretimi gibi çalışmalar için anahtar rol oynadığı için protein transkripsiyon ağlarını oluşturan bileşenlerin fonksiyonlarının ve yapısal özelliklerinin bilinmesi, bu ağların kontrol edilmesi için çok önemlidir.

Yapılan bu çalışmada biyolojik yapılardaki protein sentezi ağlarının bileşenleri yapısal ve fonksiyonel olarak incelenmiştir.

Anahtar sözcükler: Transkripsiyon ağları, ağ motifleri, negatif otoregülasyon, ileri besleme döngüleri, dinamik denklemler.

CONTENTS

Page

CHAPTER ONE

INTRODUCTION

The cell is the fundamental building block of the organisms. The interactions among proteins, DNA, RNA and small molecules in cell may be defined as wireless networks. Unraveling of the dynamic processes in system-level to handle the complexity of these interactions is the goal of the systems biology. The information-processing function, which determines the rate of production of each protein, is largely carried out by transcription networks. When the scope of inspection is changed from the assembling of genes and proteins to the recurring and significant interaction patterns, the dynamic processes in organisms may be emerged and controlled.

To understand the function of all system (network), firstly the function of the building blocks of the system and the interaction between them must be analyzed. The important point is the definition of these building blocks of network which are called as network motifs. The definition criteria are the functionality and the number of appearance in a real network.

The recurring functional patterns are called as network motifs in transcription networks (Alon, 2006, Milo et al, 2002). These recurring network motifs are examples of special patterns that they have been selected for its specific dynamical information processing functions. Firstly, the network motifs were defined in Escherichia Coli as patterns that occurred in the transcription network much more often than would be expected in random networks (Lee et al, 2002). These defined motifs have been detected in different organisms' transcription networks (Eichenberger et al, 2004, Mangan et al, 2003, Lee et al, 2002).

For a network motif, some special characteristics must be defined. One of these characteristics is the response time of the network motif which is defined as the time

takes to reach halfway between the initial and final levels, is equal to half-life of the gene product (Rosenfeld et al, 2002). On the order of cell generation time for proteins that are not degraded, the slow response time of transcription network is the basic problem. The slow response is a bottleneck for networks that need to response immediately to external signals.

One of the most common network motifs is negative autoregulation (NAR) which occurs when a transcription factor represses the transcription of its own gene (Savageau, 1974, Becskei and Serrano, 2000). This network motif occurs in about half of the repressors in E. coli and many eukaryotic repressors (Thieffry et al, 1998, Rosenfeld et al, 2002, Lee et al, 2002). The dynamic analysis of this network motif explains that in contrast to the simple regulation with a production rate set to reach steady-state level, a negative regulation, with a stronger initial production rate, is suppressed to result in the desired steady-state level. The negative autoregulation has the advantage that the goal, the steady-state level, is reached faster and the fluctuations around the steady-state level due to variations in production rate are reduced in negative autoregulation (Dublanche et al, 2006, Mangan et al, 2005).

The other one of the most common network motifs is feed-forward loop (FFL) which appears in hundreds of gene systems in E. coli, yeast and as well as in other organisms (Shen-Orr et al, 2002, Milo, 2002, Milo, 2004). This network motif consists of three genes: a regulator, X, which regulates both Y and Z, which is regulated by both X and Y. FFL has eight possible types with a specific combination of positive and negative regulations. The two of FFL types, which are called coherent type-1 FFL and incoherent type-1 FFL, are much more common than other six types in transcription networks (Ma et al, 2004, Mangan and Alon, 2003, Mangan et al, 2006).

- The coherent type-1 FFL, is a sign-sensitive delay element that can protect against unwanted responses to fluctuating inputs. The magnitude of the delay in the FFL can be tuned over evolutionary timescales by varying the biochemical parameters of

regulator protein Y, such as its lifetime, maximal level and activation threshold (Mangan et al, 2003).

- The incoherent type-1 FFL, is a sign-sensitive accelerator, acts as a pulse generator. This acceleration can be used in conjunction with the other mechanisms of acceleration, such as increased degradation and negative autoregulation (Basu et al, 2004).

The transcription network can be assumed as biological circuit because the network motifs, which are the components of it, have the similar dynamic equations with the electrical circuit elements. By using this similarity, one can apply well developed circuit theory to analyze the transcription networks and also one can design effective electrical circuits. As the evolution selects the motifs, being functional, and eliminates the futile motifs to increase the functionality of transcription network like an engineering design. The dynamic equation and structure of a motif must be well-known. This master thesis reviews transcriptional networks and common network motifs to have a basis for future researches on the analysis of network motif by circuit, systems and signals theories.

CHAPTER TWO

BIOLOGY NOTES

2.1 DNA

Deoxyribonucleic acid (DNA) is described as the principal component of the genome which is hereditary information of living organism. The DNA consists of four different types of nucleotides, which are identical except that each contains a different nitrogen base. Each nucleotide contains phosphate, sugar (of the deoxy-ribose type) and one of the four bases: Adenine, Guanine, Cytosine and Thymine (denoted A, G, C, T) in Figure 2.1. (Shamir, 2002)

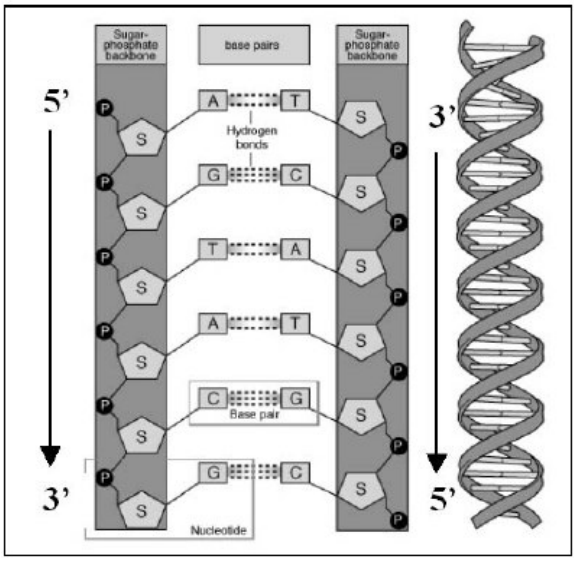

Figure 2.1 DNA structure (borrowed from Shamir, 2002)

The DNA molecule has a *double-stranded helix structure*, which looks like two interlocked bedsprings in Figure 2.1. The two strands are held together by hydrogen bonds between the bases, which gives stability but can be broken up for replication and transcription. Each strand is encoded in a complementary form (A in one strand corresponds to T in the other, and C to G) (Sontag, 2005). Because of this complementary base pairing, each one of the strand can be used as genetic code source.

2.2 RNA

Ribonucleic acid (RNA) is a nucleic acid and very similar to DNA but there are some differences between the structures of DNA and RNA. RNA is almost exclusively found in single-stranded form nevertheless DNA is generally found in double-stranded helix form. The sugar type is also different in RNA (deoxy-ribose type sugar in DNA, ribose type sugar in RNA). The RNA language is basically the same as DNA's, with the minor detail that in RNA, RNA has the pyrimidine based Uracil (abbreviated U), instead of Thymine (Sontag, 2005). Unlike DNA which is located primarily in the nucleus, RNA can also be found in the cytoplasm outside of the nucleus, e.g. messenger RNA (mRNA) - molecules that direct the synthesis of the protein in cytoplasm.

2.3 Proteins

Proteins are the primary components of living organisms, which are sequentially linked amino-acid arrays. The genetic information that is encoded in gene, which is a region of DNA that controls a discrete hereditary characteristic, usually corresponding to a single mRNA carrying the information for the production of a protein, determines the sequence of amino acids in a protein (Alberts, 1995).

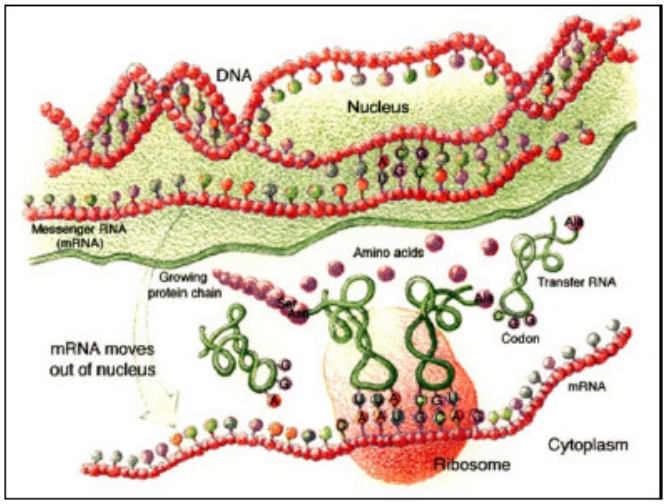

Figure 2.2 From gene to protein (borrowed from Sontag, 2005)

In a mRNA sequence, each triplet (or codon) of bases maps into a n individual amino acids. The mRNA sequence is shown in Figure 2.2. mRNA also includes a “START” and “STOP” codon for protein sequence.

2.4 The Central Dogma

During the protein synthesis, the genetic information in the DNA is translated into mRNA in transcription stage and mRNA synthesis the protein by decoding this information at ribosome in translation stage. This paradigm is explained as “DNA makes RNA, RNA makes protein, and protein makes the cell” is called the central dogma of molecular biology (Crick, 1958).

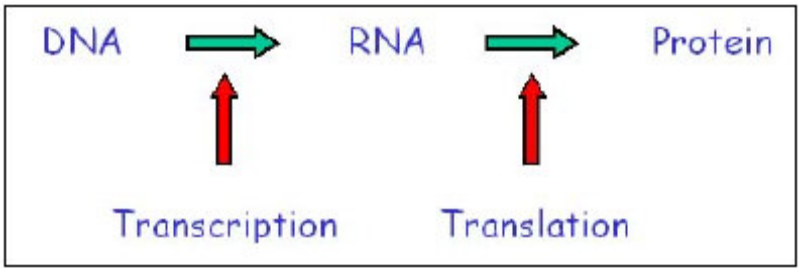

Figure 2.3 From DNA to protein (borrowed from Sontag, 2005)

CHAPTER THREE

TRANSCRIPTION NETWORKS

3.1 Networks in the Cell

There is a dynamic interaction between cell and its environment. The temperature or pressure of the environment may change and according to the new conditions, the cell can turn into a new stage in cell cycle or for demand of energy, the cell must transfer some molecules from outside of cell. These events may be given as examples of these types of interaction. The cell therefore continuously monitors its environment and calculates the amount at which each type of protein is needed. In these interactions, there are some environmental signal, which is related to the environmental conditions, and the target proteins being as the response of cell.

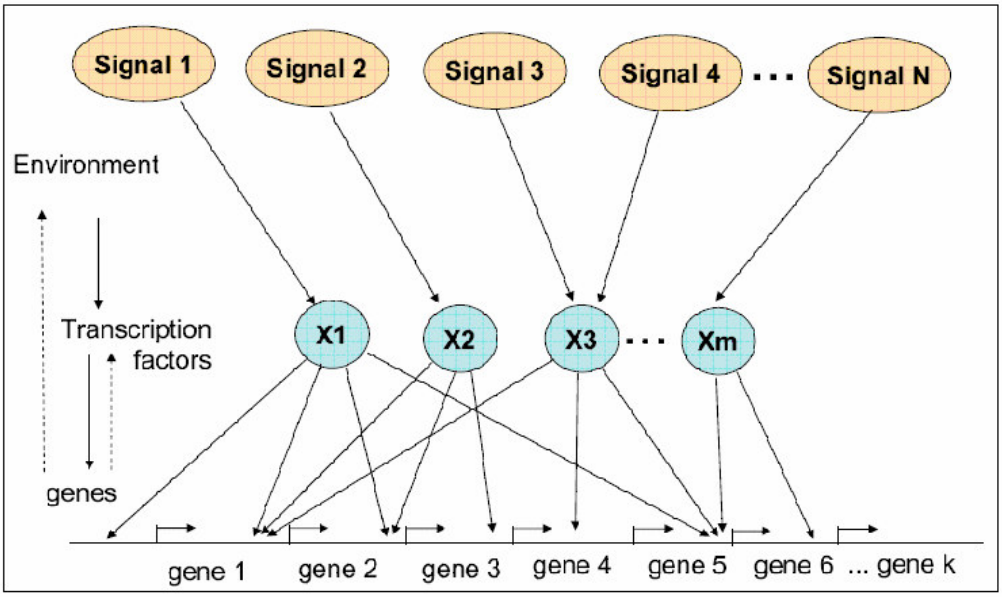

Figure 3.1 The mapping between environmental signal and transcription network (borrowed from Alon, 2007)

The dynamic interaction between environmental signals, transcription factors inside the cell and the genes regulated by them is shown in Figure 3.1. In this mapping, specific transcription factor proteins are activated by the environmental signal. When the transcription factors are activated, they bind the DNA to change the transcription rate of

specific target genes. Then the mRNA is translated into protein. Hence, transcription factor regulates the rate at which the proteins encoded by the genes are produced. These proteins affect the environment and also some proteins are themselves transcription factors that can activate or repress other genes.

Cell responds to the signals by producing appropriate proteins that act upon the internal and external environment. To represent these environmental states, the cell uses special proteins called *transcription factors* as symbols. A transcription factor is a protein that directs when- and possibly how many times- a gene is to be transcribed, by binding to DNA at a specific region. By this way, the transcription factors regulate their target genes to mobilize the appropriate protein responses in each case.

3.2 Elements of Transcription Networks

During the protein synthesis, which is a response to the environmental signals, the interaction between the transcription factors and the target genes construct networks. These networks are called transcription networks, because the transcription factors control the transcription of the genetic information from genes (DNA) to mRNA in these networks.

It is known that the gene is a specific region of DNA that controls a discrete hereditary characteristic, usually corresponding to a single mRNA carrying the information for the production of the target protein. Transcription is replication of a segment of DNA into a complementary strand mRNA. It is catalyzed by the enzyme RNA polymerase. The mRNA is then translated into a protein and called gene product.

The rate at which the gene is transcribed, the number of mRNA produced per unit time, is regulated by transcription factor. It is binding to DNA at a promoter, a regulatory region of DNA that precedes the genes. The quality of this site specifies the transcription rate of gene.

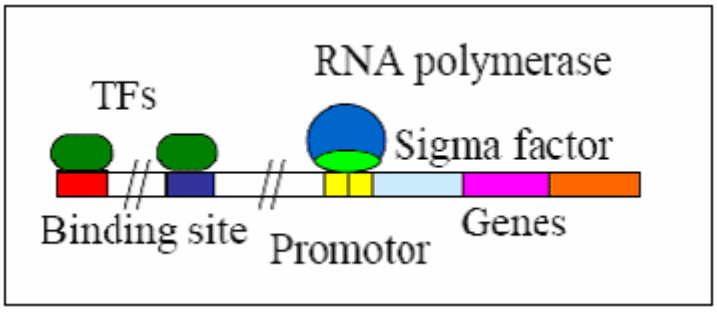

Figure 3.2 Transcription unit (borrowed from Alon, 2007)

Transcription factor affects the transcription rate by binding specific sites in the promoters of the regulated genes (Figure 3.2). When bound, they change the probability per unit time that RNAp binds the promoter and produces a mRNA molecule. The transcription factors thus affect the rate at which RNAp initiates transcription of the gene. Transcription factor can act the transcription rate by two different ways: as activators that increase the transcription rate, or as repressors that reduce the transcription rate.

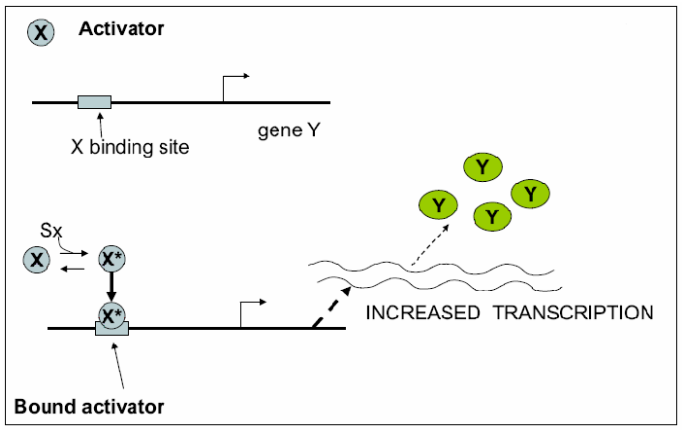

Figure 3.3 An activator (borrowed from Alon, 2007)

An activator, X, is a transcription factor protein that increases the rate of mRNA transcription by binding the promoter. The activator typically transits rapidly between active and inactive forms. In its active form, it has a high affinity to a specific site (or sites) on the promoter. The environmental (internal or external) signal, S_x, increases the probability that X is in its active form, X^*. X^* binds a specific site in the promoter of gene Y to increase transcription rate and production of protein Y (Figure 3.3).

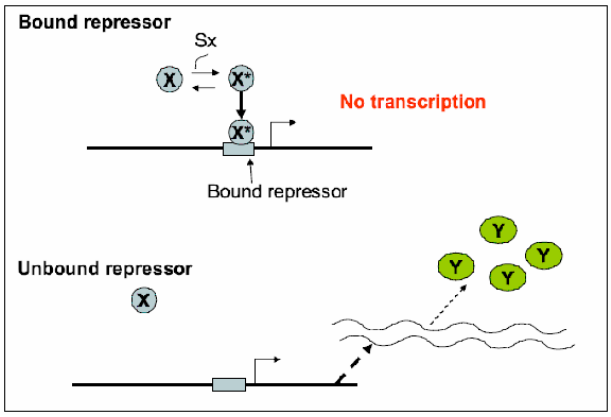

Figure 3.4 A repressor (borrowed from Alon, 2007)

A repressor, X, is a transcription factor protein that increases the rate of mRNA transcription when it binds the promoter. The signal, S_x, increases the probability that X is in its active form, X^*. X^* binds a specific site in the promoter of gene Y to decrease transcription rate and production of protein Y (Figure 3.4).

Because the transcription factor has a protein structure, they are encoded by genes, which are regulated by other transcription factors, which in turn may be regulated by yet other transcription factors, and so on. All this set of transcriptions form a transcription network describing all of the regulatory transcription interactions in a cell.

In the network, the node and edge represent the genes and transcriptional regulation of one gene by the protein product of another gene, respectively. The network can be plotted as a directed graph. A directed edge $X \rightarrow Y$ means that the transcription rate of gene Y is regulated by a transcription factor protein, which is a product of gene X, that binds the promoter of gene Y.

3.2.1 Activator or Repressor

An edge in a transcription network denotes an interaction between the target gene (the transcription rate) and the transcription factor protein controlling the target gene. There are two different types of interaction:

Activation (+ sign), or positive control, occurs when the transcription factor increases the rate of transcription when it binds the promoter.Repression (- sign), or negative control, occurs when the transcription factor decreases the rate of transcription when it binds the promoter.

Some transcription factors, called dual transcription factors, can act on a given as activators under some conditions and repressor under other conditions. Transcription networks often show comparable numbers of plus or minus edges, with more positive (activation) interactions than negative (repression) interactions.

Typically, transcription factors act primarily as either activators or repressors. In other words, the signs on the interaction edges that go out from a given node, and thus represent genes regulated by that node, are highly correlated. Some nodes send out edges with mostly minus signs. These nodes represent repressors. Other nodes, that represent activators, send out mostly plus signed edges.

Transcription factors tend to employ one mode of regulation for most of their target genes. In contrast, the signs on the edges that go into a node, which represent the transcription interactions that regulate the gene, are less correlated. Shortly, the signs on outgoing edges (edges that point out from a given node) are rather correlated, but the signs on incoming edges (edges that point into a given node) are not. A similar feature is found in neural networks, where $X \rightarrow Y$ described synaptic connections between neuron X and neuron Y. In many cases, the signs (activation or inhibition) are more highly correlated on the outgoing synapses. This feature known as Dale's rule, stems form the fact that many neurons primarily use one type of neurotransmitter, which can be either excitatory or inhibitory for most outgoing synaptic connections.

3.2.2 The Input Function

The edges can not be represented only with signs. The strength of the effect of a transcription factor on the transcription rate of its target gene is described by an input

function. When X regulates Y, represented in the network by $X \to Y$, the number of molecules of protein Y produced per unit time is a function of the concentration of X in its active form, X^*. The rate of production of

$$Y = f(X^*) \qquad \text{(Equation 3.1)}$$

Typically, the input function $f(X^*)$ is a monotonic, S-shaped function. It is an increasing function or a decreasing function, when X is an activator or a repressor, respectively. A useful function that describes many real gene input functions is called the Hill function.

The Hill input function for an activator is a curve that starts from zero and approaches to a maximal saturated level.

$$f(X^*) = \frac{\beta X^{*n}}{K^n + X^{*n}} \qquad \text{(Equation 3.2)}$$

The Hill function has three parameters, K, β and n. K is termed the *activation coefficient*, and has units of concentration. β is the *maximal expression level* of the promoter and n is the *Hill coefficient*, governs the steepness of the input function.

The value of K is related to the affinity level of the X with its binding site on promoter and when $X^* = K$, the expression level reaches the half-maximal level, $\frac{\beta}{2}$. When the activator concentration level is high enough, $X^* \rangle\rangle K$, the expression level reaches the maximal expression level, β, because X^* binds the promoter with the high probability and stimulates RNAp to produce many mRNAs per unit time.

The Hill function is a decreasing S-shaped curve for a repressor, with the same three parameters

$$f(X^*) = \frac{\beta}{1+\left(\frac{X^*}{K}\right)^n} \qquad \text{(Equation 3.3)}$$

Since a repressor allows strong transcription of a gene only when it is not bound to the promoter, the maximal production rate, β, is obtained when the repressor does not bind the promoter at all, that is when, $X^* = 0$. The half-maximal repression level is occurred when the repressor activity is equal to K, the gene's repression coefficient.

3.3 Dynamics and Response Time of Simple Gene Regulation

Let us assume that a gene that is regulated by a single regulator, with no other additional inputs. The transcription interaction between two nodes is described in the network by $X \to Y$, which means "transcription factor protein X regulates the transcription rate of gene Y". Once the activation signal, S_x, is present, the transcription factor is activated and Y concentration begins to change. Gene Y begins to be transcribed, and the mRNA is translated, resulting in accumulation of protein Y. The cell produces protein Y at a constant rate, which is denoted by β (units of concentration per unit time).

The production rate of Y is balanced by two processes;

- Protein degradation (its specific destruction by specialized proteins in the cell)
- Protein dilution (the reduction in concentration due to the increase of cell volume during growth)

The total degradation/dilution rate (in units of 1/time) is

$$\alpha = \alpha_{dil} + \alpha_{\deg} \qquad \text{(Equation 3.4)}$$

where the degradation rate and dilution rate is $\alpha_{\deg}$ and α_{dil}, respectively.

The change in the concentration of Y is due to the difference between its production and degradation/dilution, as described by a dynamic equation:

$$\frac{dY}{dt} = \beta - \alpha Y \qquad \text{(Equation 3.5)}$$

At steady state, the production of Y reaches to a constant concentration level Y_{st}. Because of the constant level of concentration, the change of Y according to time is zero, $\frac{dY}{dt} = 0$. The steady-state concentration can be calculated as:

$$Y_{st} = \frac{\beta}{\alpha} \qquad \text{(Equation 3.6)}$$

This shows that the steady-state concentration is the ratio of the production and degradation/dilution rates. According to Equation 3.6:

- The higher the production rate β, the higher the protein concentration reached, Y_{st}.
- The higher the degradation/dilution rate α, the lower Y_{st}.

When the input signal is not present any more in the environment, the production of Y stops. The solution of Equation 3.5 with $\beta = 0$ is an exponential decay of Y concentration:

$$Y(t) = Y_{st}.e^{-\alpha.t} \qquad \text{(Equation 3.7)}$$

The response time, $T_{1/2}$, is a measure of the speed at which Y levels change. It is generally defined as the time to reach the halfway between the initial and the final levels in a dynamic process. The initial level is Y_{st} and the final level is zero. By using

Equation 3.7, the response time, which is needed to reach halfway concentration, $Y(t) = Y_{st}/2$, is calculates as:

$$T_{1/2} = \frac{\log(2)}{\alpha} \qquad \text{(Equation 3.8)}$$

The degradation/dilution rate α directly determines the response time: fast degradation/dilution allows rapid changes in concentration. The production rate β affects the steady-state level but not the response time.

Consider the opposite case, in which an unstimulated cell with $Y = 0$ is provided with a signal, so that protein Y begins to accumulate. Equation 3.5 results in an approach to steady-state:

$$Y(t) = Y_{st}(1 - e^{-\alpha.t}) \qquad \text{(Equation 3.9)}$$

The response time, the time to reach the halfway concentration level, is calculated by solving for the time when $Y(t) = Y_{st}/2$. Using Equation 3.9, the response time is found as in the case of decay:

$$T_{1/2} = \frac{\log(2)}{\alpha} \qquad \text{(Equation 3.10)}$$

The response time for both increase and decrease case is the same and it depends on the degradation/dilution rate only. The larger the degradation/dilution rate α, the more rapid the changes in concentration.

CHAPTER FOUR

AUTOREGULATED GENE AS A NETWORK MOTIF

4.1 Patterns and Network Motifs

The goal is to define understandable patterns of connections that serve as building blocks of the entire network. To understand the dynamics of the entire network based on the dynamics of the individual building blocks, one approach is to look for meaningful patterns on the basis of statistical significance. To define the statistical significance, the network is compared with an ensemble of *randomized networks*.

The randomized networks are networks with the same characteristics as the real networks (the same number of nodes and edges as the real network), but where the connections between the nodes are made at random.

Patterns that occur in the real network significantly more than in randomized networks are called *network motifs* (Milo et al., 2002; Sher-Orr et al., 2002). A network motif is a recurring pattern in the network that occurs far more than at random.

The basic idea is that patterns occur in the real network much more often than in randomized networks must have been preserved over evolutionary timescales against mutations that randomly change edges. The edges are easily lost in a transcription network or new edges can be added to the network by mutations that generate a binding site for transcription factor X in the promoter region of gene Y. Edges in network motifs must be constantly selected in order to survive randomization force.

So if a network motif appears in a network much more often than in randomized networks, it must have been selected based on some advantage it gives to the organism. If the motif did not offer a selective advantage, it would be washed out and occur about as often as in randomized networks.

4.1.1 Detecting network motifs by Comparison to Randomized Networks

The network motif is detected by comparing the real network to an ensemble of randomized networks. The simplest ensemble of randomized networks is introduced by Erdos and Renyi, so called ER model (Erdos and Renyi, 1959, Bollobas 1985).

For a meaningful comparison, the randomized networks should share the basic features of the real network. The real transcription network has N nodes and E edges. To compare it to the Erdos-Renyi (ER) model, one builds a random network with the same number of nodes and edges. In random network, defined by the ER model, directed edges are signed at random between each pair of nodes.

Because of N nodes, the possibility of the connection of two nodes by an edge is $N(N-1)/2$. Since the network is a directed graph, each edge can point in one of two directions, for a total of $N(N-1)$ possible places to put a directed edge between two nodes. Additionally, an edge can begin and end at the same node, forming a self edge (total of N possible self edges). Totally, the number of possible edges is:

$$N(N-1)+N=N^2 \qquad \text{(Equation 4.1)}$$

The E edges are placed at random in the N^2 possible positions, so the each possible edge position is occupied with the probability of $p=E/N^2$, in the ER model.

4.2 Autoregulation: A Network Motif

Figure 4.1 compares a small network to a corresponding random ER network, with the same number of nodes and edges. Self-regulating genes in a network of transcription interaction are given as real network. Nodes that correspond to genes that encode transcription factor proteins that regulate their own promoters (self-regulating genes, represented by self-edges) are shown in black. The self edges are edges that originate and end at the same node. This network has $N=10$ nodes, $E=14$ edges and

$N_{self} = 4$ self-edges and its randomized Erdos-Renyi version has the same number of nodes and edges but $N_{self} = 1$ self-edge.

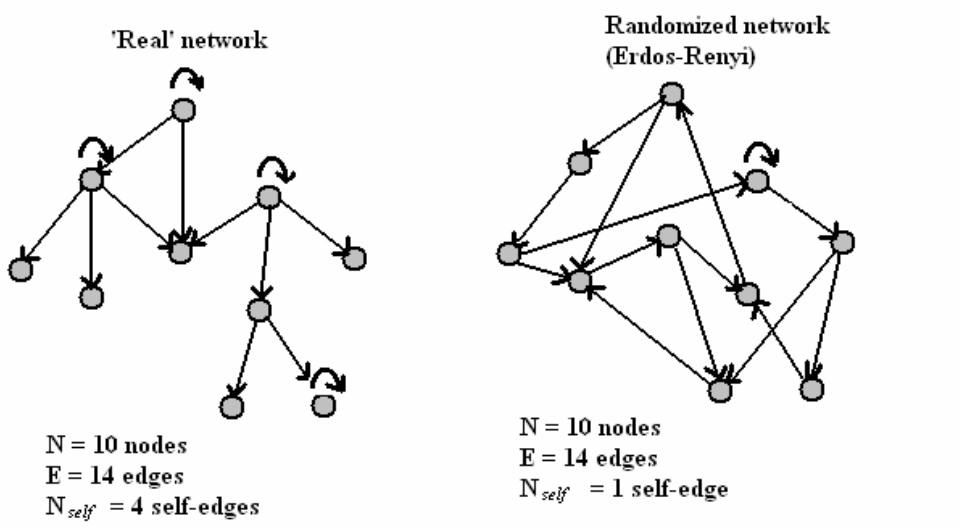

Figure 4.1 Example of a small network and its randomized Erdos-Renyi version (borrowed from Alon, 2007)

Regulation of a gene by its own gene product is known as autogenous control, or *autoregulation*. Thirty-four of the autoregulatory proteins in the real network are repressors that repress their own transcription, *negative autoregulation*.

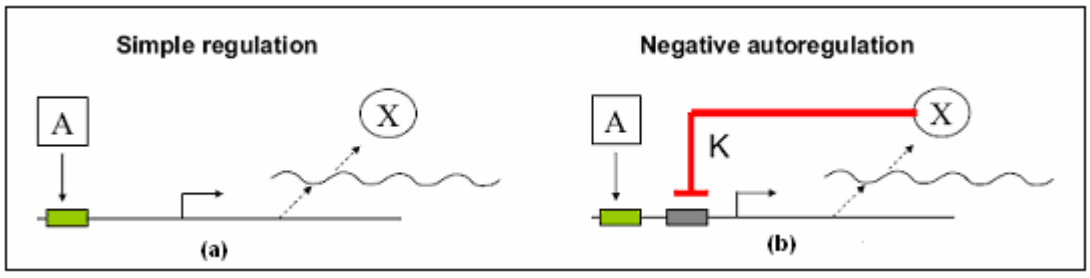

Figure 4.2 Simple regulation (a) and negative autoregulation (b) (borrowed from Alon, 2007)

Figure 4.2 shows a simple regulation means "gene X is simply regulated by gene A" and a negative autoregulation means "a gene X that is autoregulated; that is, it is repressed by its own gene product, the repressor X and is also simply regulated by gene A". Repressor X binds a site in its own promoter and thus acts to repress its own

transcription. The symbol ⊣ stands for repression. The repression threshold is K, defined as the concentration of X needed to repress the promoter activity by 50%.

To decide that the self-edge is significantly more frequent in the real network than at random, the probability of having *k* self-edges in an ER model network is calculated. To form a self-edge, an edge needs to choose its node of origin as its end node, out of the N possible target nodes. This probability is thus:

$$p_{self} = \frac{1}{N} \qquad \text{(Equation 4.2)}$$

Because E edges are placed at random to form the random network, the probability of having *k* self-edges is approximately binomial (throwing a coin E time and getting k heads):

$$p(k) = \binom{E}{k} . {p_{self}}^{k} \left(1 - p_{self}\right)^{E-k} \qquad \text{(Equation 4.3)}$$

The average number of self-edges is equal to the number of edges E times the probability that an edge is a self-edge (just as the expected number of heads is the number of times the coin is thrown multiplied by the probability of heads):

$$\left\langle N_{self} \right\rangle_{rand} \approx E.p_{self} \approx \frac{E}{N} \qquad \text{(Equation 4.4)}$$

with a standard deviation that is approximately the square root of the mean (again, similar to a coin-tossing experiment with a small probability p_{self} for heads, which approximates a Poisson process):

$$\sigma_{rand} \approx \sqrt{E/N} \qquad \text{(Equation 4.5)}$$

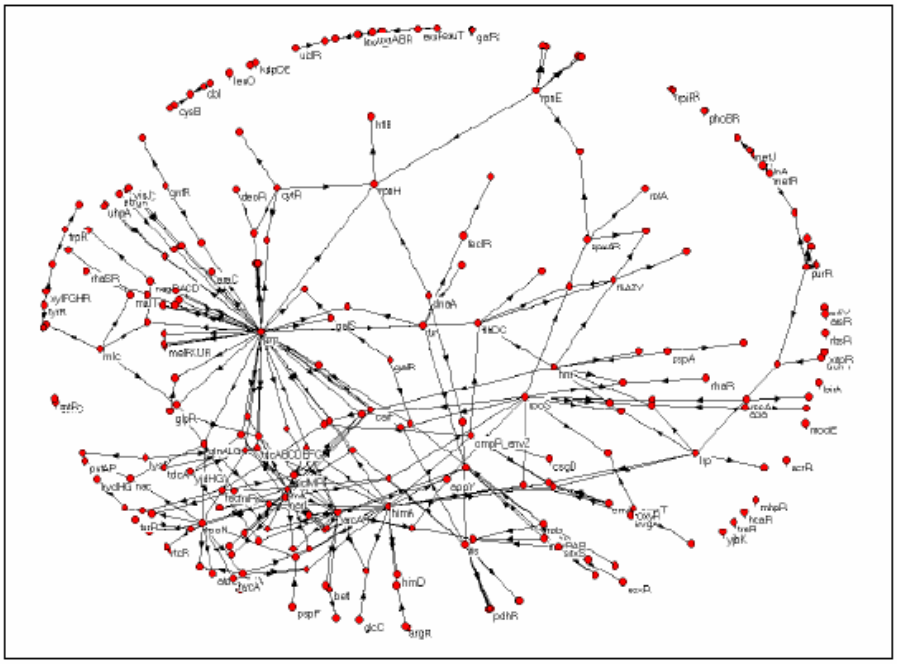

Figure 4.3 A transcription network in E. coli (borrowed from Alon, 2007)

Figure 4.3 shows the E.coli transcription network which has $N = 424$ nodes, $E = 519$ edges and $N_{self} = 40$ self-edges. According to Equation 4.5, a corresponding ER network with the same N and E would be expected to have only about one self-edge, plus minus one:

$$\langle N_{self} \rangle_{rand} \approx E/N \approx 1.2\,,\ \sigma_{rand} \approx \sqrt{1.2} \approx 1.1 \qquad \text{(Equation 4.6)}$$

But the real network has 40 self-edges, which exceeds the random networks by many standard deviations. This significant difference in the number of self-edges can be described by the number of standard deviations by which the real network exceeds the random ensemble:

$$Z = \frac{\langle N_{self} \rangle_{real} - \langle N_{self} \rangle_{rand}}{\sigma_{rand}} \qquad \text{(Equation 4.7)}$$

Self-edges show $Z \approx 32$, which means they occur far more often than at random because 32 standard deviations mark a very high statistical significance.

4.3 The Response Time of Negative Autoregulated Genes

Negative autoregulation occurs when a transcription factor X represses its own transcription. The important point is that why the negative autoregulation is a network motif. To be a network motif, it must have a useful function. For that reason a negatively autoregulated gene must be compared to a simply (non-auto) regulated gene to decide that it has or not (Figure 4.2). The comparison criteria is the response time of the system.

It is known that the response time of a simply regulated gene is governed by its degradation/dilution rate α :

$$T_{1/2} = \frac{\log(2)}{\alpha} \qquad \text{(Equation 4.8)}$$

In negative autoregulation, the self-repression occurs when X binds its own promoter to inhibit production of mRNA. As a result, the higher the concentration of X, the lower its production rate.

The dynamics of X are described by its production rate f(X) and degradation/dilution rate:

$$\frac{dX}{dt} = f(X) - \alpha X \qquad \text{(Equation 4.9)}$$

where f(X) is the input function and a decreasing Hill function:

$$f(X) = \frac{\beta}{1 + \left(\frac{X}{K}\right)^n} \qquad \text{(Equation 4.10)}$$

according to the input function equation, when X is smaller than the repression coefficient K, the denominator gets smaller. The production rate reaches its maximal value, β. On the other hand, when repressor X is at high concentration, no transcription

occurs, $f(X) \approx 0$. To solve the dynamics, the logic approximation, where the production is zero if $X \rangle K$, and production is maximal, $f(X) \approx \beta$, when X is smaller than K.

$$f(X) = \beta.\theta(X \rangle K) \qquad \text{(Equation 4.11)}$$

where $\theta(X \rangle K)$ is the step function and gets maximum value, $\theta(X \rangle K) = 1$, when X is smaller than K.

To calculate the response time, consider the case where X is initially absent, and its production starts at $t = 0$. While concentration is low, the promoter is unrepressed and production is full steam at rate β, as described by the production-degradation equation:

$$\frac{dX}{dt} = \beta - \alpha X \qquad \text{(Equation 4.12)}$$

At early time, the degradation can be neglected $(\alpha X \langle\langle \beta)$ to find linear accumulation of X with time:

$$X(t) \approx \beta.t \text{ , while } X \langle K \text{ and } X \langle\langle \beta/\alpha \qquad \text{(Equation 4.13)}$$

But, production stops when X levels reach the self-repression threshold, $X = K$ (production is zero when X exceeds K). If there are any delays in the system, small oscillations may occur around $X = K$. Delays may cause X to overshoot beyond K slightly, but then because of repression, production stops and X levels decline until they decrease below K, upon which production starts again, etc. If the input function, $f(X)$, is not strictly a logic function, but rather a smoother function like a Hill function, the oscillations are generally damped. By this way, X effectively reaches a steady-state level equal to the repression coefficient of its own promoter:

$$X_{st} = K \qquad \text{(Equation 4.14)}$$

The response time, $T_{1/2}$, can be calculated by assuming that X reaches half steady-state at $X(T_{1/2}) = X_{st}/2$. To simplify the calculation, the calculation of the response time is made by using linear accumulation of X (Equation 4.13), in which $X = \beta.t$. The response time $T_{1/2}^{(n.a.r)}$, where n.a.r stands for negative autoregulation, is given by finding the time when X reaches half of the steady-state level, $\beta.T_{1/2}^{(n.a.r)} = X_{st}/2 = K/2$, so that:

$$T_{1/2}^{(n.a.r)} = \frac{K}{2\beta}, \text{ response time for negative autoregulation} \qquad \text{(Equation 4.15)}$$

The stronger the maximal unrepressed promoter activity β, the shorter the response time. In other words, a strong promoter can be used to give an initial state and then autorepression can be used to stop production at the desired steady-state in negative autoregulation.

Now, the simply regulated gene (a gene without negative autoregulation), which is produced at rate β_{simple} and degraded at rate α_{simple}, is compared to negative autoregulated gene. To make a meaningful comparison, the two designs must have the same steady-state levels and the designs should have as many of the same biochemical parameters as possible. Such a comparison that is carried out with equivalence of as many internal and external parameters as possible between the alternative designs is termed as a *mathematically controlled comparison* (Savageau, 1976). Then, the two designs have the same protein degradation/ dilution rate $\alpha = \alpha_{simple}$.

The steady-state is balance of production and degradation, $X_{st} = \beta_{simple}/\alpha_{simple}$, in simple gene regulation. But in negative autoregulation case, the steady-state is equal to the repression threshold, $X_{st} = K$ (Equation 4.14). To make mathematically controlled comparison, K is tuned so that both designs reach the same steady-state expression level:

$$K = \beta_{simple} / \alpha_{simple} \qquad \text{(Equation 4.16)}$$

The response time of simple regulation is governed by the degradation/dilution rate as described by Equation 4.8, so that $T_{1/2} = \log(2) / \alpha_{simple}$. A much faster response can be achieved by the corresponding negative autoregulated circuit by making β large, because the response time, $T_{1/2}^{(n.a.r)} = K / 2\beta$, is inversely proportional to β. The ratio of the response times in the two designs gets smaller when β is increased.

$$T_{1/2}^{(n.a.r)} / T_{1/2}^{simple} = \frac{\beta_{simple} / \beta}{2\log(2)} \qquad \text{(Equation 4.17)}$$

So, the response time of the negative autoregulation design is faster than simple regulation.

In conclusion, negative autoregulation gets the best of both words:

- A strong promoter can give rapid production,
- A suitable repression coefficient provides the desired steady-state.

The same strong promoter on a simple-regulation circuit would reach a much higher steady-state, leading to undesirable overexpression of the gene product.

4.4 Robustness of Negative Autoregulation

Also the speedy response time, negative autoregulation has a second important benefit. The second benefit is increased robustness of the steady-state expression level with respect to fluctuations in the production rate β. This property is demonstrated using measurements of protein levels in individual cells (Becskei and Serrano, 2000).

Because of the overall fluctuations in the metabolic capacity of the cell and its regulatory systems and stochastic effects in the production of protein, the production rate of a given gene, β, fluctuates over time. So, twin cells usually have different production rates of most proteins. On the other hand, parameters such as the repression threshold K vary much less from cell to cell, because they depend on the strength of the chemical bonds between X and its DNA binding site and the position and number of the X binding sites in the promoter.

Because the steady-state level is linearly dependent on the production rate, $X_{st} = \beta / \alpha$, and a change in β leads to a proportional change in X_{st}, the simple gene regulation is affected quite strongly by fluctuations in production rate β.

On the other hand, negative autoregulation can buffer fluctuations in production rate because its steady-state level depends only on the repression threshold of X for its own promoter, $X_{st} = K$. Because the repression threshold K is determined by the chemical bonds between X and its DNA sites, this parameter varies much less from cell to cell than production rates. So, negative autoregulation is more robust in steady-state protein levels than fluctuations in production rate.

CHAPTER FIVE

THE FEED-FORWARD LOOP NETWORK MOTIF

In contrast to many possible patterns that could present in the network, only a few of them are found significantly and called as network motifs. The main point is that a pattern must have a information processing function to be a network motif. The benefit of these functions may explain why the same network motifs are discovered by evolution again and again in diverse systems. To find significant patterns, the numbers of appearances of different patterns in real and random networks must be calculated.

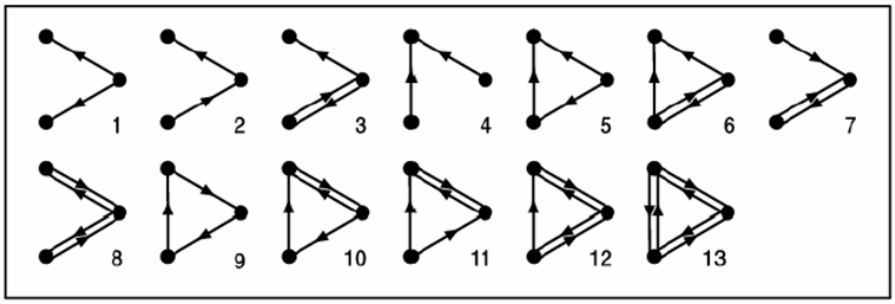

Figure 5.1 The 13 connected three-node directed subgraph (borrowed from Alon, 2007)

For patterns with three nodes (such as triangle), there are 13 possible cases. But there is only one, the feed-forward loop (FFL), is a network motif in the 13 possible three-node patterns (Figure 5.1). To understand the possible functions of the feed-forward loop, the regulation between its three edges, activation or repression interactions, must be described. There are eight possible FFL types, but six of them occur much more rarely. The common two types of FFLs can carry out interesting functions such as the filtering of noisy input signals, pulse generation and response acceleration.

5.1 The Number of The Appearances Of A Subgraph in Random Networks

The larger patterns of nodes and edges are called subgraphs. Two examples of three-node subgraphs are the three-node feedback loop and three-node feed-forward loop (Figure 5.2). The number of times that a given subgraph G appears in a random ER model network is calculated. The given subgraph G, that is feed-forward loop with n=3

nodes and e=3 edges, has n nodes and e edges. In the ER random network model, E edges are placed randomly between N nodes. Since there are N^2 possible places to put a directed edge, the probability of an edge in a given direction between an given pair of nodes is :

$$p = \frac{E}{N^2} \qquad \text{(Equation 5.1)}$$

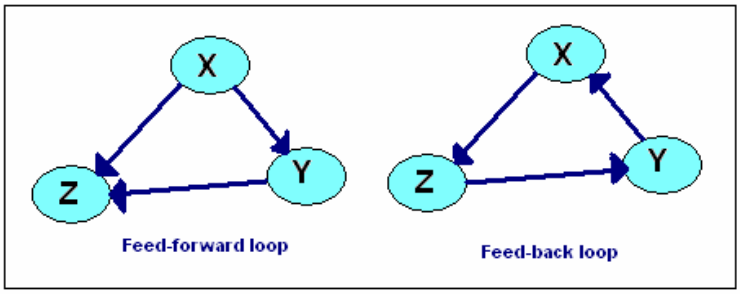

Figure 5.2 The feed-forward loop and the feedback loop (borrowed from Alon, 2007)

To calculate the average number of occurrences of subgraph G in the network, denoted $\langle N_G \rangle$, firstly n nodes and g edges in the proper places are chosen. The number of ways of choosing a set of n nodes out of N: about N^n for large number (because there are N possible ways of choosing first node, $N-1 \approx N$ possible ways of choosing the second node, etc) times the probability of g edges that are in appropriate places (each w,th probability p) is equal to the average number of occurrences of subgraph G in the network:

$$\langle N_G \rangle \approx a^{-1} N^n p^g \qquad \text{(Equation 5.2)}$$

where a is a constant that corresponds to the combinatorial factors related to the structure and symmetry of each graph. It is equal to one and three respectively for FFL and the three-node feedback loop.

The Equation 5.2 can be defined in terms of the mean connectivity of the network, defined as the average number of edges per node:

$$\lambda = E/N \qquad \text{(Equation 5.3)}$$

Also the probability p, can be recasted in terms of the mean connectivity $p = \lambda/N$. By using this probability formula, the average number of occurrences of subgraph G is a simple equation in which the higher the mean connectivity of the network λ, the higher the mean number of appearances of subgraph G:

$$\langle N_G \rangle \approx a^{-1} \lambda^g N^{n-g} \qquad \text{(Equation 5.4)}$$

The number of times that subgraph G appears in the network scales with the network size, N. assume a sense of large random networks, that have the same mean connectivity, λ. The dependency of $\langle N_G \rangle$ on the network size N is called as a scaling relation. By ignoring the prefactor in Equation 5.4, the scaling relation describes the way that the number of subgraph:

$$\langle N_G \rangle \approx N^{n-g} \qquad \text{(Equation 5.5)}$$

According to Equation 5.5, the scaling of subgraph numbers in ER networks depends only on the difference between the number of nodes and edges in the subgraph, $n - g$.

Let give some examples to realize this relation. For example, the V-shaped patterns, such as patterns 1 and 2 in Figure 5.1, have n=2 nodes and g=2 edges. Their scaling number grows linearly with network size:

$$N_{V-shaped} \approx N^{n-g} = N \qquad \text{(Equation 5.6)}$$

If the network size is doubled, the number of V-shaped subgraphs doubles. So these patterns are very common in random networks. On the other hand, the fully connected clique (the last pattern in Figure 5.1) has n=3 nodes and g=6 edges. The graph scales as $N^{n-g} = N^{-3}$, so it occurs very rarely in large random networks.

The two triangle-shaped patterns, the three-node feed-forward loop and feed-back loop, both have n=3 nodes and g=3 edges. By using Equation 5.4, the average number of appearances of subgraph is found (the appropriate symmetry factors: a=1 for feed-forward and a=3 for feed-back loop):

$$\langle N_{FFL} \rangle \approx \lambda^3 N^0 \qquad \text{(Equation 5.7)}$$

$$\langle N_{3loop} \rangle \approx \frac{1}{3} \lambda^3 N^0 \qquad \text{(Equation 5.8)}$$

The remarkable results tell us that the numbers of the triangle patterns are constant in ER networks and do not increase with network size.

While the triangle numbers do not depend on the size of the random networks, the number of V-shaped pairs of edges in the network scales linearly with the size of network N. Also there is relation between these two situations. The probability of V-shaped pattern closes to form a triangle scales as $1/N$ (because an edge that emerges from a node at one arm of the V and closes it into a triangle by pointing to the node at the other arm needs to choose the pne target node out of N possibilities). This is equal a total of $N.1/N = N^0$ triangles.

According to this result, the triangles and more complex patterns occur rarely in random networks.

5.2 The Structure of The Feed-Forward Loop Gene Circuit

The feed-forward loop consists of two transcription factors, which the first transcription factor, X, regulates a second transcription factor, Y, and both X and Y regulate gene Z. so the feed-forward loop has two parallel regulation paths that a direct path, from X to Z, consists of a single edge, and an indirect path , a cascade pf two edges, goes through Y.

Because of the activation or repression possibilities in each edge, there are $2^3 = 8$ possible types of feed-forward loops (Figure 5.3).

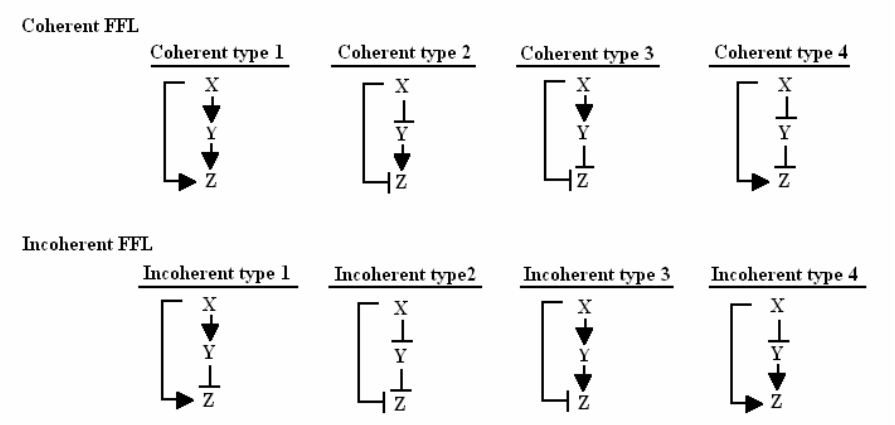

Figure 5.3 The eight sign combinations (types) of feed-forward loops. (borrowed from Alon, 2007)

The eight possible feed-forward loop types are classified into two groups: coherent and incoherent. The grouping method depends on comparing the sign of the direct path to the sign of indirect path. In coherent FFLs, the indirect path has the same overall sign, the multiplication of the sign of each arrow on the path, as the direct path. For example, in type-1 coherent FFLs, X activates Z, and also activates an activator of Z, so that both paths are positive.

In incoherent FFLs, the overall sign of the indirect path is opposite to the sign of the direct path. For example, in the type-1 incoherent FFLs, the direct path is positive and the indirect path is negative. Note that incoherent FFLs have an odd number of minus edges (one or two).

The most abundant FFL is the type-1 coherent FFL (C1-FFL) in which all three edges are activation (Mangan and Alon, 2003). The second most abundant FFL is the type-1 incoherent FFL (I1-FFL) (Ma et al., 2004). The other six FFL types appear much more rarely than C1-FFL and I1-FFL.

In addition to the signs of the edges, it must be known to calculate the type of input function of gene Z that the integration of the two regulators, X and Y, at the promoter of

gene Z. for that reason, two biologically reasonable logic functions, AND and OR, are used. In AND logic function, both X and Y activities need to be high in order to turn on Z expression. In OR logic function, either X or Y is sufficient.

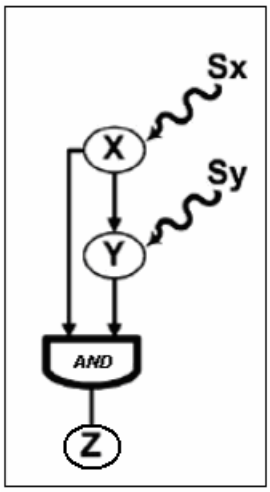

Figure 5.4 The coherent type-1 FFL with an AND input function (borrowed from Alon, 2007)

The transcription factors X and Y are activated by the external stimuli, which are represented by input signal S_X and S_Y (Figure 5.4). The signals are molecules in some systems and in other systems the signals are modifications of the transcription factor caused by signal transduction pathways activated by the external stimuli.

5.3 Dynamics of The Coherent Type-1 FFL with AND Logic

Assume that the concentration of the transcription factor protein X is high in the cell and the input to X is the signal S_X (Figure 5.5). When S_X is not present, X is in inactive state. At time $t = 0$, a strong signal S_X triggers the activation of X. This type of triggering is known as a step-like stimulation of X. the transcription factor X rapidly gets its active state, X^*. The active protein X^* binds the promoter of the second transcription factor protein Y to initiate the production of protein Y. Additionally, X also bind the promoter of Z. But, since the input function of promoter of Z is AND logic, the binding of X^* is not enough to activate Z production.

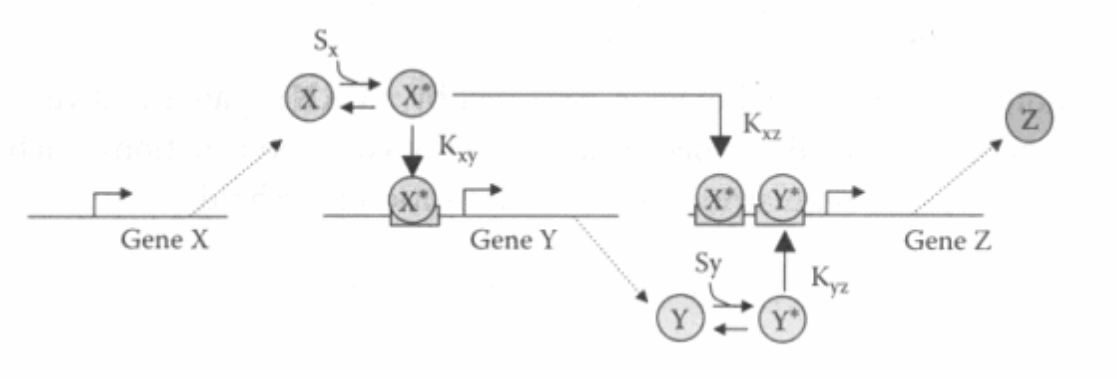

Figure 5.5 The molecular interactions in the coherent FFL of Figure 5.4 (borrowed from Alon, 2007)

To activate the production of Z, both X^* and Y^* must bind. So the concentration of Y must build up to sufficient levels to cross the *activation threshold*, K_{YZ}, for gene Z. Additionally, to activate the production of Z, the second input signal, S_Y, must be present so that Y transits to its active state Y^* (Figure 5.5). When the signal S_X is present, for activation of Z, first of all Y needs accumulate. This results in a *delay* in production of gene Z.

To understand the function of a gene circuit, the mathematical model must be constructed. To describe the FFL, let use the logic input functions. The production of Y occurs at rate β_X, when X^* exceeds the activation threshold K_{XY}, as described by the step function, θ:

production rate of $Y = \beta_Y \theta\left(X^* \rangle K_{XY}\right)$ (Equation 5.9)

When the signal S_X is present, s rapidly transits to its active state X^*. If the signal is sufficient enough, X^* exceeds the activation threshold, K_{XY}, and rapidly binds the promoter of Y to activate the production of Y. Then, Y begins to production. The accumulation of Y can be described by the dynamic equation in terms of production and degradation/dilution rate:

$$\frac{dY}{dt} = \beta_Y \theta\left(X^* \rangle K_{XY}\right) - \alpha_Y Y \qquad \text{(Equation 5.10)}$$

The production of Z can be described by a product of two step functions, each indicating whether the proper regulator crossed the activation threshold, because the promoter of Z is governed by an AND gate input function:

$$\text{production of } Z = \beta_Z \theta\left(X^* \rangle K_{XZ}\right)\theta\left(Y^* \rangle K_{YZ}\right) \qquad \text{(Equation 5.11)}$$

The C1-FFL gene circuit has three activation thresholds (Fig 4.6). When a strong step-like stimulation, X^* rapidly exceeds the two thresholds K_{XY} and K_{XZ}. Because of the time, takes that Y^* accumulates and exceeds the threshold K_{YZ}, there is a delay in production of Z. When the accumulation of Y^* exceeds the threshold level, the z production proceed at rate β_Z. The dynamic equation of the Z production is in terms of degradation/dilution rate and production with an AND input function:

$$\frac{dZ}{dt} = \beta_Z \theta\left(X^* \rangle K_{XZ}\right)\theta\left(Y^* \rangle K_{YZ}\right) - \alpha_Z Z \qquad \text{(Equation 5.12)}$$

5.4 The C1-FFL Is a Sign Sensitive Delay Element

To define the dynamics of the C1-FFL, the response to two different cases are analyzed. The first case is *ON step*, in which the signal S_X is first absent and is then suddenly appears. The second case is *OFF step*, in which S_X is at first present and is then suddenly removed. To simplify the calculations, it is assumed that the signal S_Y is present to activate the transcription factor Y.

5.4.1 Delay Following an ON Step of S_X

When S_X is in ON step, Y^* begins to be produced at rate β_Y. The concentration of y begins to exponentially converge to its steady-state level:

$$Y^*(t) = Y_{st}\left(1 - e^{-\alpha_Y t}\right) \qquad \text{(Equation 5.13)}$$

The steady-state concentration level of Y is equal to the ration of its production to degradation/dilution rate, $Y_{st} = \beta_Y / \alpha_Y$.

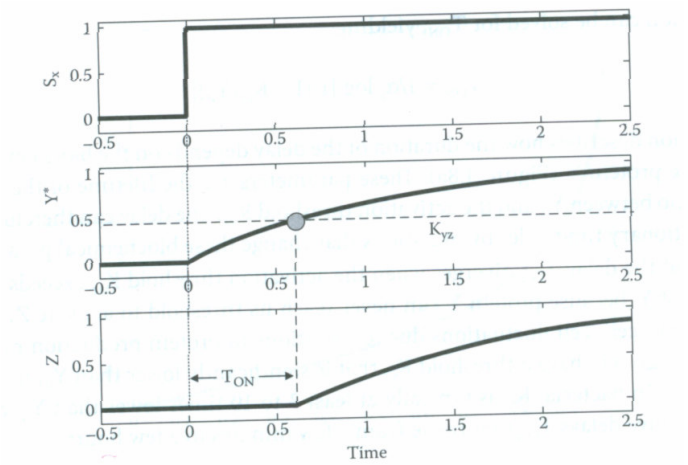

Figure 5.6 Dynamics of the coherent type-1 FFL with AND logic following an ON step of S_X at time $t = 0$ in the presence of S_Y (borrowed from Alon, 2007)

Production of Z is described by an AND input function so the presence of X^* and exceeding the threshold level of X is not enough to activate the Z production. Z begins to be expressed only after a delay because the second input, Y^*, takes some time to accumulate and to exceed the activation threshold level, K_{YZ} (Figure 5.6). The delay, which is the time needed for Y^* to reach its threshold and also time, is T_{ON} . T_{ON} , which is time when the Y concentration crosses the level at height K_{YZ} , can be calculated using Equation 5.13:

$$Y^*(T_{ON}) = Y_{st}\left(1 - e^{-\alpha_Y T_{ON}}\right) = K_{YZ} \qquad \text{(Equation 5.14)}$$

Then, T_{ON} is found as:

$$T_{ON} = \frac{1}{\alpha_Y} \log\left[\frac{1}{1 - \frac{K_{YZ}}{Y_{st}}}\right] \qquad \text{(Equation 5.15)}$$

This equation defines the relation between the duration of the delay and the biochemical parameters of the protein Y, which are the lifetime of the protein, α_Y, and the ration between Y_{st} and the activation threshold K_{Yz}.

5.4.2 No Delay Following an OFF Step of S_X

It is shown that Z has a delay following ON steps of S_X. Following an OFF step, in which S_X is suddenly removed, X rapidly becomes inactive state and does not bind the promoters of genes Y and Z. Because Z is described by an AND gate that needs binding of both X^* and Y^*, when there is only one input, the production of Z stops. So, after an ONN step of S_X, Z production stops at once without any delay.

5.4.3 The C1-FFL Is a Sign-Sensitive Delay Element

In contrast to that C1-FFL with AND input function shows a delay following ON steps of S_X, it does not have a delay following OFF steps. This type of behavior is called as sign-sensitive delay. The sign-sensitive means that the delay depends on the sign of the step, ON or OFF.

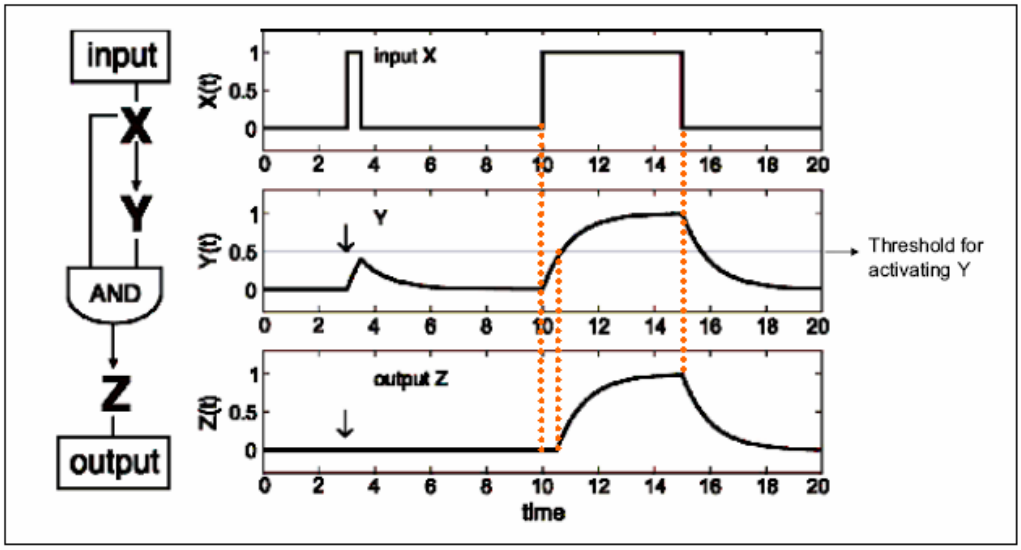

Figure 5.7 The coherent type-1FFL with AND logic as a persistence detector (borrowed from Alon, 2007)

A sign-sensitive delay element can also be considered as a kind of asymmetric filter. For example, consider a pulse of S_X that appears only briefly (an ON pulse) (Figure 5.7). an ON pulse that is shorter than the delay time, T_{ON}, does not lead to any Z expression in the C1-FFL. That is because Y does not have time to accumulate and cross its activation threshold during the pulse. Only persistent pulse (longer than T_{ON}) results in Z expression. Thus, this type of FFL is a *persistence detector* for ON pulses. On the other hand it responses immediately to OFF pulses. In contrast to the FFL, simple regulation (with no FFL) does not filter out short input pulses, but rather shows production of Z that lasts as long as the input pulse is present.

5.4.4 Sign- Sensitive Delay Can Protect against Brief Input Functions

In engineering designs, the sign-sensitive delay element is commonly used in situation where the cost of the error is not symmetric to compensate the asymmetricity. For example, in elevators: assume that the beam of light used to sense the obstruction in the elevator door. The door opens when you obstruct the light with your hand. When you remove your hand from light for a short time, shorter than T_{ON}, nothing happens (that is, a short pulse of light is filtered out). Only when you remove your hand for a sufficient length of time, the door closes (a sufficient pulse of light leads to a response, longer than or equal to T_{ON}). Put your hand back in and the door opens immediately. So the cost of the error (closing or opening time of door at the wrong time) is not symmetric: the design aims to response as quickly as possible to a person in the beam and make sure that there is no person for a sufficient period of time before closing the doors.

In transcription networks, the C1-FFLs are selected by the evolution is the systems in the cell require such a protection function. Sometimes stimuli can not be present for brief pulses that should not elicit a response like highly fluctuating environment. In fluctuating environment, the C1-FFL can offer a filtering function.

5.4.5 The OR Gates C1-FFL IS a Sign-Sensitive Delay for OFF Step of S_X

With an OR input function, the Z production begins immediately upon an ON step of S_X, because the presence of only one input is enough to activate the gene Z with an OR gate. So there is no delay following an ON step of S_X. In contrast to immediate activation, Z is deactivated at a delay following an OFF step, because both inputs need to consume for the deactivation of OR gate. Y^* can activate gene Z even though S_X is not present. So it takes time for Y^* to decay away after an OFF step of S_X. The C1-FFL shows a delay following OFF steps, whereas the AND version shows a delay after ON steps. Consequently, the C1-FFL with an OR gate is also a sign-sensitive delay element, but it has an opposite sign of the AND version. As a result, the OR gate C1-FFL can activate Z production even if the input signal is momentarily lost.

5.5 The Incoherent Type-1 FFL

5.5.1 The Structure of the Incoherent FFL

The I1-FFL consists of two parallel but antagonistic regulation paths. In the direct path, X activates Z, but in indirect path, it activates the repressor of Z, Y (Figure 5.8.a). So the two arms of the I1-FFL act in opposition: the direct path activates Z and the indirect path represses Z. The production of Z increases when the activator X^* is bound, decreases when the repressor Y^* binds (Figure 5.8.b).

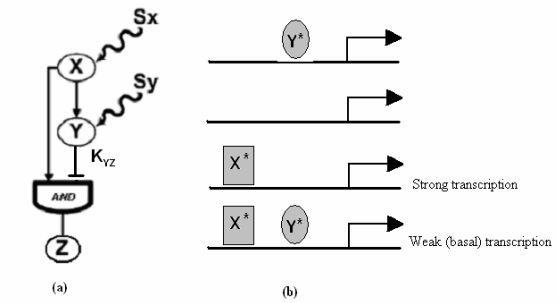

Figure 5.8 (a) The I1-FFL with an AND gate and (b) Four binding states of a simple model for promoter region of Z, regulated by activator X and repressor Y (borrowed from Alon, 2007)

5.5.2 The Dynamics of the I1-FFL: A Pulse Generator

The I1-FFL responses to the input signal S_X and S_Y. When S_X is present, X gets its active state and binds the promoter of both genes Y and Z. While Y exceeds the threshold level, X initiates the transcription and causes protein Z to begin to be produced (Figure 5.9). Because X activates the production of Y in parallel, after a delay, enough protein Y accumulates to repress Z production and the production of Z decreases. By this way, the I1-FFL can generate a pulse of Z production.

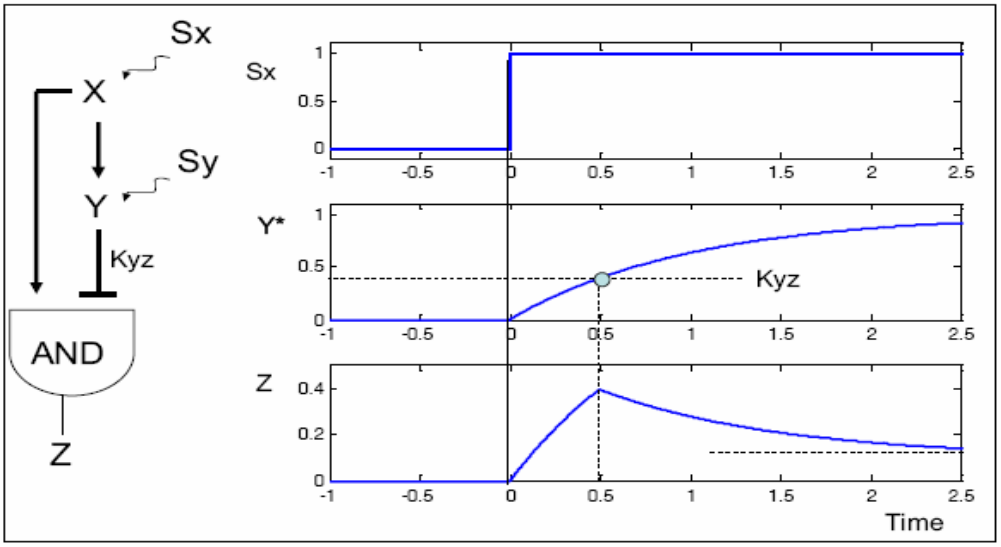

Figure 5.9 Pulse-like dynamics of the I1-FFL following an ON step S_X in the presence of S_Y (borrowed from Alon, 2007)

Consider that the response to a step function of the signal S_X, in the presence of the second signal S_Y. If the signal S_X appears, protein X rapidly becomes activated, X^*. The active transcription factor X^* binds the promoter of gene Y and Y begins to be produced. Since the second signal S_Y is present, the protein Y transits its active form Y^* and accumulates over time according to the production and degradation equation:

$$\frac{dY^*}{dt} = \beta_Y - \alpha_Y Y \qquad \text{(Equation 5.16)}$$

Then, Y shows the familiar exponential convergence to its steady-state $Y_{st} = \beta_Y / \alpha_Y$:

$$Y^*(t) = Y_{st}\left(1 - e^{-\alpha_Y t}\right) \qquad \text{(Equation 5.17)}$$

In parallel to activation of Y, X^* binds the promoter of Z to start the production of Z at a rapid rate β_Z since its promoter is occupied by the activator X^* but there are not yet enough repressor Y^* in the cell to repress the production. In this phase:

$$\frac{dZ}{dt} = \beta_Z - \alpha_Z Z \qquad \text{(Equation 5.18)}$$

So Z accumulates and converges exponentially to a high level $Z_m = \beta_Z / \alpha_Z$:

$$Z(t) = Z_m\left(1 - e^{-\alpha_Z t}\right), \text{ while } Y^* < K_{YZ} \qquad \text{(Equation 5.19)}$$

This rapid production of Z stops when the repressor Y^* exceeds its repression threshold for Z, K_{YZ}. Then, the production rate of Z rapidly decreases to a low production rate $\beta_Z^{'}$. In the extreme case of no leakiness, it drops to $\beta_Z^{'} = 0$. The repression time T_{rep}, which is the duration until the repression occurs, can be calculated by using Equation 5.13 by using the time when $Y^*(t) = K_{YZ}$, showing that T_{rep} depends on the biochemical parameters of protein Y:

$$T_{rep} = \frac{1}{\alpha_Y} \log\left[\frac{1}{1 - \frac{K_{YZ}}{Y_{st}}}\right] \qquad \text{(Equation 5.20)}$$

After T_{rep}, the repressor Y^* also binds the promoter of Z and the production rate of z is reduced. Z concentration decays exponentially to a new steady-state level $Z_{st} = \beta_Z^{'} / \alpha_Z$, which is lower than Z_m:

$$Z(t) = Z_{st} + (Z_O - Z_{st})e^{-\alpha_Z (t - T_{REP})} \quad \text{(Equation 5.21)}$$

where Z_O is the level reached at time T_{rep}, given by the Equation 5.19 at $t = T_{rep}$:

$$Z_O = Z_m (1 - e^{-\alpha_Z T_{REP}}) \quad \text{(Equation 5.22)}$$

and Z_{st} is the final steady-state level of Z, due to the low level expression when both X^* and Y^* bind the promoter of Z:

$$Z_{st} = \beta_Z^{'} \Big/ \alpha_Z \quad \text{(Equation 5.23)}$$

The shape of the dynamics of the I1-FFL depends on $\beta_Z^{'}$, the basal production rate of Z. The *repression factor* F, defined as the ratio of the maximal and basal activity of the Z promoter, also equal to the ratio of the unrepressed and repressed steady-state levels of Z:

$$F = \frac{\beta_Z}{\beta_Z^{'}} = \frac{Z_m}{Z_{st}} \quad \text{(Equation 5.24)}$$

When the repressor has a strong inhibitory effect on Z production, that is, when $F \rangle\rangle 1$, Z dynamics show a pulse-like shape. In the pulse, Z levels first increase and then decline to a low level. The I1-FFL can therefore act as a pulse generator (Mangan and Alon, 2003; Basu et al.,2004).

5.5.3 The I1-FFL Speeds the Response Time

In addition to pulse generation, the I1-FFL has another property: it accelerates the response time of the system. The response time of the I1-FFL is shorter than that of a simple-regulation circuit that exceeds the same steady-state level of Z (Figure 5.10). The response time is defined as the time which the production needs to reach its halfway to the steady-state level. In the simple regulation circuit, the response time can not be speeded by increasing its production rate, because such an increase would lead to an unwanted increase in the steady- state level.

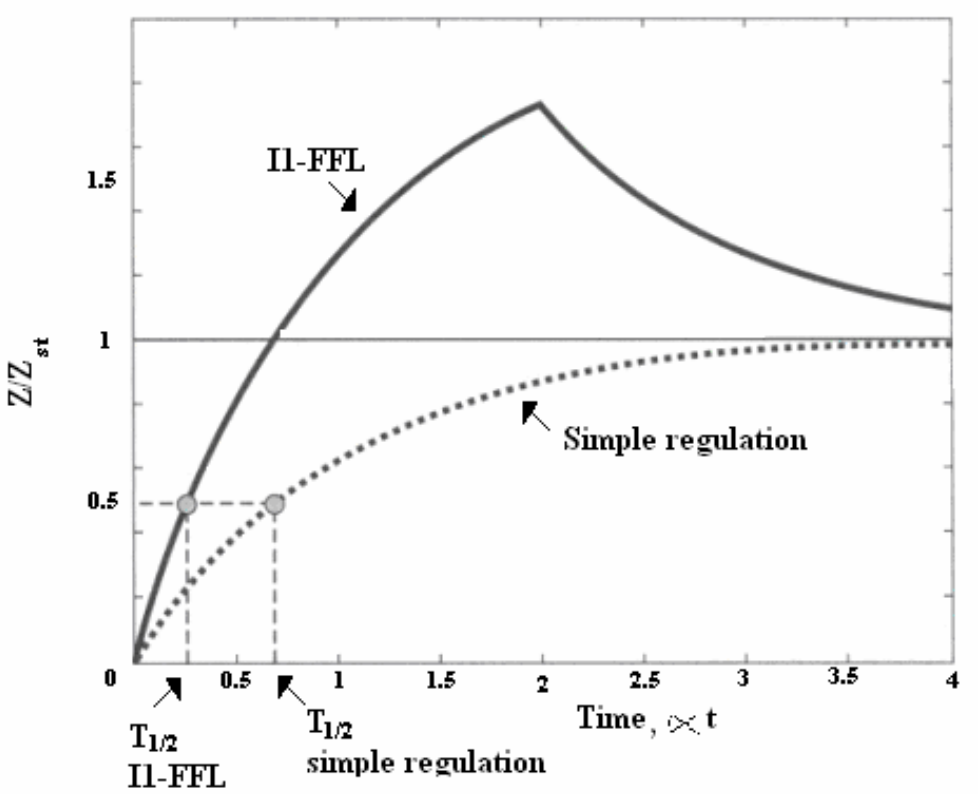

Figure 5.10 Response time of the I1-FFL is shorter than simple regulation that reaches the same steady-state level (borrowed from Alon, 2007)

To analyze the acceleration, calculate the response time $T_{1/2}$, the time to reach the half of the steady-state level. In the I1-FFL, half steady-state level is exceedded during the initial fast phase of Z production, before Y represses the gene Z. The response time $T_{1/2}$, is found by Equation 5.19 by substituting the concentration level of Z as halfway, Z_{st}:

$$Z_{1/2} = \frac{Z_{st}}{2} = Z_m\left(1 - e^{-\alpha_Z T_{rep}}\right) \qquad \text{(Equation 5.25)}$$

This expression can be solved to give an expression on the repression coefficient $F = Z_m / Z_{st}$:

$$T_{1/2} = \frac{1}{\alpha_Z}\log\left[\frac{2F}{2F-1}\right] \qquad \text{(Equation 5.26)}$$

This equation shows that the larger the repression coefficient F, the faster the response time becomes ($T_{1/2} \approx (2\alpha_Z F)^{-1}$ at large F) or it can be said, the stronger the effect of Y in repressing production of Z, the faster the performance of the I1-FFL compared to an equivalent simple-regulation circuit $X \rightarrow Z$ made to exceed the same steady-state level of Z. The response time gets very small and when $F \rangle\rangle 1$, it approaches $T_{1/2} = 0$. When $F = 1$, the limit of no repression and the response time is:

$$T_{1/2} = \frac{\log(2)}{\alpha_Z} \qquad \text{(Equation 5.27)}$$

which is the same as the response time of the simple-regulation circuit.

As a result, the I1-FFL works as a simple-regulation circuit, because the repressor Y has no effect on the gene Z when $F = 1$.

5.5.4 Response Acceleration is Sign-Sensitive

However the accelerated response after ON step, the response, when the signal S_X is rapidly absent, occurs with the same dynamics as for a simply regulated gene (no acceleration or delay). In both simple and I1-FFL circuits, OFF steps of S_X cause to rapid shut-down in Z production. The rapid response to OFF steps in the I1-FFL corresponds to the AND logic of the Z promoter, because X^* unbinds the promoter to stop production the concentration of Z production decays exponentially by its degradation/dilution rate after production of Z stops.

So the I1-FFL is a sign-sensitive response accelerator. Sign-sensitive means that response acceleration occurs only for steps of S_X in one direction (ON) and not the other (OFF). Also I1-FFLs with OR input function have the same dynamic function as those with AND input function, but accelerate OFF and not OR responses.

CHAPTER SIX

CONCLUSIONS

In this master study, the researches on transcription network and its elements, network motifs, have been reviewed.

The transcription network is defined as the mapping between the DNA and environment and its building blocks, i.e. the network motifs which are the recurring information processing functional patterns. The information-processing function, which determines the rate of production of each protein, is largely carried out by transcription networks. When the scope of inspection is changed from the assembling of genes and proteins to the recurring and significant interaction patterns, the dynamic processes in organisms may be emerged and controlled.

To understand the function of all system (network), firstly the function of the building blocks of the system and the interaction between them must be analyzed. The important point is the definition of these building blocks of network which are called as network motifs, which are recurring functional patterns in transcription networks. The definition criteria are the functionality and the number of appearance in a real network.

Detection of network motifs is crucial for understanding the functions of sub-networks of transcriptional networks and it can also be used for designing protein synthesis procedures in drug development and cancer therapy.

The transcription network-electrical circuit and network motif-subcircuit analogies can be used in possible future researches on the analysis of transcriptional network and network motifs and also on the evolutionary design of electrical circuits. Well-developed circuit theory may be useful for the analysis of transcriptional networks. On the other hand, by using evolution principle that reconstructs or cancels the connections between nodes in the transcription networks to increase the functionality of network motifs, the new circuit models which are more adaptive to environment can be designed. The

characteristics of the electrical circuit models can be modulated. For example the frequency response of a noise filter can be changed by using a similar dynamic function of network motifs.

When new network motifs are defined, the complex networks become better characterized. An important subject for a future study is to define new network motifs in transcription networks. New network motif detection algorithms can be introduced for defining new network motifs. These new algorithms may consist of either stochastic or deterministic properties. Detection algorithms can be developed in such way: Firstly a candidate motif may be replaced with a different motif which is constructed by randomly and then a utility function which is defined as main characteristic of that candidate motif or a functional sub-network containing this candidate motif is calculated. According to the calculation of the utility function, the candidate motif may be defined as network motif if it increases the utility function. To test the results of these algorithms, the needed experiments should be done in the next step for the justification of the network motif candidates. So, collaboration with molecular biologists should be provided.

Another subject for future studies is to simulate the dynamics of the network motif or a functional sub-network of a transcriptional network which is identified by a motif detection algorithm in a computational way by a system analysis tool such as PSPICE which is a widely used efficient tool for electrical circuit analysis. The simulation of the new motif or sub-network may be useful for pre-justification of the network motif candidates before a possible experimental step and also for understanding of the complex networks.

REFERENCES

Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K. and Watson, J.D. (1994). *Molecular biology of the cell*. (4th edition). New York; Garland Publishing, Inc.

Alon,U. (2007). *An introduction to systems biology: design principles of biological circuits*. (1st ed.).Boca Raton; Chapman & hall/CRC.

Basu, S., Mehreja, R., Thiberge, S., Chen, M.T. and Weiss, R. (2004). Spatiotemporal control of gene expression with pulse-generating networks. *Proc Natl Acad Sci USA*, *101*, 6355-6360.

Becskei, A., Seraphin, B. and Serrona, L. (2001). Positive feedback in eukaryotic gene networks: cell differentiation by graded to binary response conversion. *EMBO J*, *20*, 2528-2535.

Becskei, A. & Serrona, L. (2000). Engineering stability in gene networks by autoregulation. *Nature, 405,* 590-593.

Bollobas, B. (2001). *Random graphs*. (2nd edition). Cambridge; Cambridge University Press.

Dublanche, Y., Michalodimitrakis, K., Kummerer, N., Foglierini, M. & Serrano, L. (2006). Noise in transcription negative feedback loops: simulation and experimental analysis. *Mol. Syst. Biol. 2,* 41.

Eichenberger, P. et al. (2004). The program of gene transcription for a single differentiating cell type during sporulation in Bacillus subtilis. *PLoS Biology, 2,* e328.

Erdos, P. and Renyi, A. (1959). On random graphs. *Publ. Mtah. (Debrecen)*, *6*, 290.

Lee, T. et al. Transcriptional regulatory networks in Saccharomyces cerevisiae. *Science, 298,* 799-804.

Ma, H. W., Buer, J. and Zeng, A.P. (2004). Hierarchical structure and modules in the Escherichia coli transcriptional regulatory network revealed by a new top-down approach. *BMC Bioinformatic*, *5*,199.

Ma, H. W. et al. (2004). An extended regulatory network of Escherichia coli and analysis of its hierarchical structure and network motifs. *Nucleic Acids Res. 32,* 6643-6649.

Mangan, S., and Alon, U. (2003). Structure and function of the feed-forward loop network motif. *Proc Natl Acad Sci USA, 100,* 11980-11985.

Mangan, S., Zaslaver, A. & Alon, U. (2003). The coherent feed-forward loop serves as a sign-sensitive delay element in transcription networks. *J. Mol. Biol. 334,* 197-204.

Mangan, S., Zaslaver, A. and Alon, U. (2006). The incoherent feed-forward loop accelerates the response time of gal systems of Escherichia coli. *J Mol Biol, 356*, 1073-1081.

Milo,R. et al.(2004). Superfamilies of designed and evolved networks. *Science, 303,* 1538-1542

Milo,R., Itzkovitz, S., Kashtan, N., Chklovskii, D. and Alon, U. (2002). Network motifs: Simple building blocks of complex networks. *Science, 298*, 824-827.

Savageau, M.A. (1976). *Biochemical systems analysis: a study of function and design in molecular biology*.(1st edition).S Addison-Wesley

Savageau, M. A. (1974). Comparison of classical and autogenous systems of regulation in inducible operons. *Nature, 252,* 546-549.

Shamir, R. (March 14, 2002). *Analysis of gene expression data*. Retrieved July 28, 2008, from www.cs.tau.ac.il/~rshamir/ge/02/scribes/lec01.ps.gz

Shen-Orr, S.S., Milo, R., Mangan, S. and Alon, U. (2002). Network motifs in the transcriptional regulation networks of Escherichia coli. *Nature Genet, 31*, 64-68.

Sontag, E.D. (2005). Molecular systems biology and control. *Eur. J. Control, 11,* 396-435.

Rosenfeld, N., Elowitz, M. B. & Alon, U. (2002). Negative autoregulation speeds the response time of transcription networks. *J. Mol. Biol. 323,* 785-793.

Thieffry, D., Huerta, A. M., Perez-Rueda, E. & Collado-Vides, J. (1998). From specific gene regulation to genomic networks: a global analysis of transcriptional regulation in Escherichia coli. *Bioessay, 20,* 433-440.

Ziv-Ukelson, M. (2006). Computational genomics. Retrieved July 25, 2008, from www.cs.tau.ac.il/~rshamir/cg/06/lec07.pdf

Lecture 3- the cell. (n.d.). Retrieved July 31, 2008 ,from http://www.life.umd.edu/classroom/bsci124/lec3.html

APPENDIX A

Definitions and Mathematical Preliminaries:

The basic mathematical concept used to model network is a graph.

Directed and Undirected Graph:

The graphs can be divided into two broad classes:

- Directed graph: $u \rightarrow v$
- Undirected graph: $u \_\_ v$

Formally, a finite directed graph, G, consists of a set of vertices or nodes, $V(G)$:

$$V(G) = \{v_1, v_2, ..., v_n\}$$

together with an edge set, $\varepsilon(G) \subseteq V(G)xV(G)$.

Each edge $(u, v) \in \varepsilon(G)$ can be thought of as connecting the starting node u to terminal node v.

Transcriptional regulatory networks and metabolic network would usually be modeled as directed graphs. In a transcriptional regulatory network, nodes would represent genes with edges denoting the interactions between them. This would be a directed graph because, if gene A regulates gene B, then there is a natural direction associated with the edges between the corresponding nodes, starting at A and finishing at B.

Directed graph also arise in the study of neural networks, in which the nodes represent individual neurons and the edges represent synaptic connections between neurons.

An undirected graph, G, also consists of a vertex set, $V(G)$, and an edges set $\varepsilon(G)$. However, there is no direction associated with the edges in this case. Hence, the elements of $\varepsilon(G)$ are simply two-element subsets of $V(G)$, rather than ordered pairs as before. For two vertices, u and v, of an undirected graph, uv is an edge if and only if vu is also an edge.

The size or order of the graph is the number of the vertices n in a directed or undirected graph.

An edge, uv, in a directed or undirected graph G, is said to be an edges at the vertices u and v, and the two vertices are said to be adjacent to each other. In this case, it is known that u and v are neighbors.

For the undirected graph G, and a vertex, $u \in V(G)$, the set of all neighbors of u is denoted $N(u)$ and given by:

$$N(u) = \{v \in V(G) : uv \in \varepsilon(G)\}$$

Node-Degree and Adjacency Matrix

For an undirected graph G, we shall write $\deg(u)$ for the degree of node u in $V(G)$. This is simply the total number of edges at u:

$$\deg(u) = |N(u)|$$

In a directed graph G, the in-degree, $\deg_{in}(u)$, (out-degree, $\deg_{out}(u)$) of vertex u is given by the number of edges that terminates (start) at u.

Suppose that the vertices of a graph (directed or undirected) G are ordered as $v_1, v_2, ..., v_n$. Then the adjacency matrix, A, of G is given by:

$$a_{ij} = \begin{cases} 1 & if \quad v_i v_j \in \varepsilon(v) \\ 0 & if \quad v_i v_j \notin \varepsilon(v) \end{cases}$$

Thus, the adjacency matrix of an undirected graph is symmetric while this need not be the case for a directed graph.

Paths, Path Length and Diameter:

Let u, v be two vertices in a graph G. then the sequence of vertices

$$u = v_1, v_2, ..., v_k = v$$

such that for $i = 1, ..., k-1$;

$(i) \quad v_i v_{i+1} \in \varepsilon(v)$
$(ii) \quad v_i \neq v_j \quad for \quad i \neq j$

is said to be a path of length $k-1$ from u to v.

The geodesic distance, or simply distance, $\delta(u, v)$, from u to v is the length of the shortest path from u to v in G. If no such path exists, then we set $\delta(u, v) = \infty$.

If for every pair of vertices, $u, v \in V(G)$, there is some path from u to v, then we say that G is connected.

The average path length and diameter of a graph G are defined to be the average and maximal value of $\delta(u, v)$ taken over all pairs of distinct nodes, u, v in $V(G)$ which are connected by at least one path.

Clustering Coefficient:

Suppose u is a node of degree k in an undirected graph G and that there are e edges the k neighbors of u in G. Then the clustering coefficient of u in G is given by:

$$C_U = \frac{2e}{k(k-1)}$$

Thus, C_U means the ratio of the number of edges between the neighbors of u to the total possible number of such edges, which is $\frac{k(k-1)}{2}$.

APPENDIX B

Part 1: The Cell

I. Function of the cell

A. Serves as the structural building block to form tissues and organs

B. Each cell is functionally independent- it can live on its own under the right conditions

1. It can define its boundaries and protect itself from external changes causing internal changes

2. It can use sugars to derive energy for different processes which keep it alive

3. It contains all the information required for replicating itself and interacting with other cells in order to produce a multicellular organism

4. It is even possible to reproduce the entire plant from almost any single cell of the plant

II. Types of cell

A. Prokaryotic cells- eg. bacteria

1. Very simple-there are no organelles and most everything functions in the cytoplasm

B. Eukaryotic cells

1. All contain the organelles that subcompartmentalize the cell

2. Includes unicellular algae and protists (e.g. ameba) that live alone or in colonies

3. Includes multi-cellular organisms - animals, plants, fungi - where cells work together a. plant cells are unlike animal cells in that plant cells have chloroplasts and cell walls. Animal cells have neither of these. Plant cells also have relatively large vacuoles.

III. Parts of the eukaryotic cell and their function

A. Outside (boundary) of the cell

1. Cell wall

a. Protects and supports cell

b. Made from carbohydrates- cellulose and pectin- polysaccharides

c. Strong but leaky- lets water and chemicals pass through- analogous to a cardboard box

2. Cell membrane

a. Membrane is made up from lipids - made from fatty acids

b. Water-repelling nature of fatty acids makes the diglycerides form a sheet or film which keeps water from moving past sheet (think of a film of oil on water)

c. Membrane is analogous to a balloon- the spherical sheet wraps around the cell and prevents water from the outside from mixing with water on the inside

d. Membrane is not strong, but is water-tight- lets things happen inside the cell that are different than what is happening outside the cell and so defines its boundaries. Certain gate keeping proteins in the cell membrane will let things in and out.

B. Inside the cell

1. Cytosol - watery inside of cell composed of salts, proteins which act as enzymes

2. Microtubules and microfilaments - cables made out of protein which stretch around the cell

a. Provide structure to the cell, like cables and posts on a suspension bridge

b. Provide a structure for moving cell components around the cell -sort of like a moving conveyer belt.

3. Organelles - sub-compartments within the cell which provide different functions. Each organelle is surrounded by a membrane that makes it separate from the cytosol

a. Nucleus- contains the genetic information which tells the cell machinery which proteins, carbohydrates and lipids to make and how they are assembled.

i. This genetic information is coded in DNA (deoxyribonucleic acid)

ii. The DNA is seen packaged up into Chromatin/chromosomes

b. Mitochondrion - Cell powerhouse. Converts sugars into energy through aerobic respiration. (up to 100 per cell)

c. Chloroplast (50-100 per cell) - site of photosynthesis.

i. Allows production of sugars from sunlight and carbon dioxide

ii. Only found in plants and algae- other cells have to find sugar from outside the cell

d. Vacuole or tonoplast- stores compounds that may interfere with other things in the cell.

Dominates the inside of a plant cell.

i. Sugars, salts, pigments (e.g. red pigment in beets and purple onions, acids (lemon acids)

e. Ribosomes - site of protein synthesis

i. Many different proteins have to be made by the cell- the proteins that a cell makes directs the cell's function and identity

ii. Ribosomes use the information coded in the DNA of the nucleus to produce proteins

f. Endoplasmic reticulum (ER) - a network of folded membranes throughout the cytoplasm

i. Rough ER has attached ribosomes, active in protein synthesis

ii. Smooth ER lacks ribosomes and functions in the transport and packaging of proteins as well as the synthesis of lipids

g. Golgi apparatus - membranous hollow sacs arranged in a stack

i. Modifies proteins, lipids, and other substances from the ER

ii. Packets of these materials move to the edge of the golgi where the golgi membrane is pinched off to make vesicle (package); this new vesicle moves to the plasma membrane where it leaves the cell, or it goes to other sites within the cell

iii. Builds primary cell walls between newly divided nuclei

Part 2: What is a cell made from?

Four groups of biologically important molecules: lipids, carbohydrates, nucleic acids, proteins

I. Lipids

A. Composed of Carbon, Oxygen, Hydrogen atoms (COH) in building blocks of Fatty acids

B. Fats (solid) and oils (liquid at room temperature)

1. Fats associated with animals - butter, lard
2. Oils associated with plants - corn oil, olive oil

C. Main characteristic of lipids - won't dissolve in water and is repelled from water (e.g. Corn oil in water- the oil droplets separate from the water and float to the top to try and avoid the water) and form self-associating packages to exclude water. A membrane is one such self-associating structure.

D. Roles of lipids

1. Food- high energy (many C-H bonds), has more energy than any other molecule
2. Part of cell membranes
3. Also- waxes (cutin, suberin), hormones (testosterone, estrogen), certain vitamins, certain pigments (chlorophyll)
4. Basic form for energy storage- monoglycerides, diglycerides, triglycerides
   a. glycerol + 1 fatty acid -> monoglyceride + water (ex- butyric acid w/4C; 14-20 C most common)
   b. monoglyceride + 2nd fatty acid -> diglyceride
   c. diglyceride + 3rd fatty acid -> triglyceride

II. Carbohydrates

A. Composed of COH -makes building blocks of monosaccharides

B. Roles

1. Energy storage (many C-H bonds) - sugar/starch energy source
2. Structural (especially in plants- cellulose)

   Note: most important structural component in animals - protein
3. Carbon sources for making other building blocks (such as ribose and deoxyribose for nucleic acids, amino acids)

C. 3 main types

1. Monosaccharides
   a. Simple sugars
      i. example- glucose C6H12O6, energy storage - blood sugar very different structure
      ii. example- fructose C6H12O6, energy storage - fruit sugar
      iii. example- ribose C5H10O5, nucleic acids
2. Disaccharides
   a. example- sucrose (transport form for plants)

      Glucose + Fructose -> Sucrose + Water
   b. C12H22O11
   c. Sugar cane, sugar beet --> sucrose (table sugar)
3. Polysaccharides (poly = many)
   a. Polymers- composed of repeating subunits of monosaccharides -
   b. eg. all these are repeating units of glucose
      i. Starch-energy storage (plants store glucose in a compact, insoluble form)
      ii. Cellulose- cell walls in plants- component of cardboard
      iii. Glycogen- energy storage in animals

III. Nucleic acids

DNA (deoxyribonucleic acid; master information carrying molecule for the cell), RNA (ribonucleic acid; Copy of DNA molecule)

A. Function- Contains the information for entire cell-expressed through protein synthesis

B. Polymers of nucleotides- composed of:

1. Base- organic molecule with nitrogen- cytosine, guanine, thymine, adenine, and uracil (uracil is in RNA only)

2. Sugar- ribose, deoxyribose

3. Phosphate

C. Shape of DNA molecule- double helix (DNA)

D. Other important nucleotide compounds- example- ATP (universal in organisms; role-energy transfer or exchange)

1. Recall ATP <->ADP + P + energy

IV. Proteins

A. Composed of COH and Nitrogen (four main elements) -building block is amino acids (20 different)

B. Large molecular weight (10,000 - 1,000,000) - note: hydrogen atom = 1

C. Roles

1. Basic building blocks of cell - much of cell structure

2. Part of cell membranes (help control entrance and exit of materials through membranes)

3. Important in animal structure: hair, nails, connective tissue (tendons, cartilage), muscles

4. Enzymes- facilitate chemical reactions -see this avi file

D. Composed of amino acids

1. Repeating amino acids joined by the peptide bond forms a protein

2. 20 of them in proteins

3. 2 functional groups:-NH2 (amino group) & -COOH (acid group)

4. Order of amino acids is important- order determines the 3-dimensional shape of the molecule. This is significant because the function follows form: the biological activity of a protein depends largely on its 3-dimensional structure.