Boolean Networks (BN) based Molecular Biology Modeling: Protein Process

Joe W. Yeol[*], I. Barjis[**], J. Barjis[***], S. Berri[****], and Y. Ryu[****]

[*]Polytechnic University, Department of Mechanical Engineering, wjoe01@utpia.poly.edu
[**]New York City College of Technology, Department of Physics and Biological Science
[***]Georgia Southern University, Department of Information Technology, jbarjis@georgiasouthern.edu
[****]New York City College of Technology, Department of Mechanical Engineering Technology

ABSTRACT

Use of modeling and simulation tool for biological studies explores augmented analysis capability for the behavior and specification of the events in the process under study. This paper introduces applications of Boolean Networks (BN) as a modeling tool. In particular, it shows how, by means of Boolean Networks (BN), the process of mRNA transcription and translation can be modeled, simulated and analyzed. In order to develop the model of mRNA transcription and translation by means Boolean Networks, first we construct a model of the whole process. Then based on biological information available each event (Transcription, Translation) will be modeled with some details. The aim of this paper is to model, simulate and analyze the process of protein production (Transcription, Translation). However some basic information about the molecular events, and precursors/molecules that are required for each step of the protein production process are described in the first chapter, then based on the biological information a model is constructed. All the enzymes and molecular precursors that are required for the process of protein production is not included in the models.

1 INTRODUCTION

Bioinformatics is the study of the inherent structure of biological information and biological systems. It brings together the avalanche of systematic biological data (e.g. genomes) with the analytic theory and practical tools of mathematics and computer science. Bioinformatics integrates courses and research in biology (molecular biology) with computer modeling and information sciences [5]. In this paper efforts of authors from two different disciplines (one from the Biomedical Sciences another from the Information Systems), are put together to study the protein production process from computing point of view. Consequently, this paper includes both biological and analytical approaches towards the protein production process. From the computer science we apply a method to model the protein production process for the mathematical representation of the system. Analysis of the modeling can reveal important information about the structure and dynamic behavior of the system. This information can then be used to evaluate the modeled system and suggest improvements or changes or simply understand the nature of the system under study. The content of this paper is divided into two sections (a compact model and a detailed model). In section one of this paper, a compact model of the protein production is developed; and in part two, a detailed model of the protein production is presented. Each section consists of two subsections, where the first subsection of each section describes same reactions and molecular events that occur during the process of protein production. While, in the second subsections models in compact and detailed notations are developed. Conclusions that are derived from the results of this paper are represented in the conclusion part of this paper. Finally, suggestions for future works are given at the end of this paper.

2 COMPACT MODELING OF THE PROTEIN PRODUCTION

In this and next sections we will show how model of the protein production process can be built in both compact and detailed notations. Firstly we discuss the protein production process as sequence of DNA, RNA and protein without the molecular events and chemical reactions that occur during this process. Based on this description we will build a model of the protein production in compact notation. A compact model of the protein production covers only core operations (activities) such as *transcription*, *translation*, reverse *transcription* and *replication* (see figure 1). But a detailed model of the protein production, in addition to those operations, covers some molecular events, and precursors/molecules that are required for each step of the protein production process.

2.1 The Protein Production Process

The production of proteins does not proceed directly from the DNA. For the information to be translated from the DNA sequences of the genes into amino acid sequences of proteins, a special class of RNA molecules is used as intermediates [3, 4]. Complementary copies of the genes to be expressed are transcribed from the DNA in the form of messenger RNA (mRNA) molecules. The mRNAs are used by the protein-synthesizing machinery of the cell to make the appropriate proteins. This process, which takes place on sub-cellular particles called ribosome, is referred to as

translation. The flow of genetic information in the cell can be summarized by the simple schematic diagram shown in figure 1. This figure shows the flow of genetic information. As shown in the figure 1, DNA can either replicate to produce new double helix DNA or can be translated into m-RNAs, where these mRNAs would consequently undergo the process of translation to produce secreted protein. mRNAs can undergo the process of reverse transcription and produce double helix DNA.

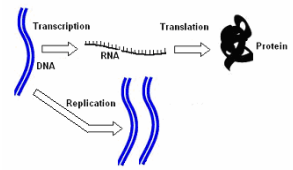

Figure 1: Flow of generic information.

From the above brief description of the protein production process the following conclusions can be made. Firstly, the protein production process is dynamic process, which changes its states after each operation. Secondly, there some conditional and optional processes take place. For example, DNA can be either replicated or translated into mRNA. Under some special circumstances, mRNA can be, in reverse, transcribed to produce DNA.

2.2 Boolean Networks

In Boolean Networks, each element has K inputs and one output. The signals at inputs and outputs take binary (0 or 1) values. The Boolean elements of the network and the connections between elements are chosen in a certain manner. There are no external inputs to the network. An automaton operates in discrete time. The set of the output signals of the Boolean elements at a given moment of time characterizes a current state of an automaton. During an automaton operation, the sequence of states converges to a cyclic attractor. The states of an attractor can be considered as a "program" and the typical attractor length L and important characteristics of Boolean Networks. According to this model, the network is represented as an oriented graph, $G = (V, F)$, whose nodes V represent element of the network, and F defines a topology of edges between the nodes and a set of Boolean functions. A node may represent either a gene or a biological stimulus, where a stimulus is any relevant physical or chemical factor which influences the network and is itself not a gene or product. A node has an associated steady-state expression level, representing the amount of product or the amount of stimulus present in the cell. This level is approximated as high or low and presented by the binary value 1 or 0, respectively. The ssignment of values to nodes fully describes the state of the model at any given time. The change of model state over time is fully defined by the functions in F. Initial assignment of values uniquely defines the model state at the next step, and, consequently, on all the future steps. Thus, the network evolution is realized as a sequence of consecutive states, and it is uniquely defined by the initial state. Such a sequence of states is called trajectory [6-8].

2.3 Boolean Networks of the Protein Production in Compact Notation

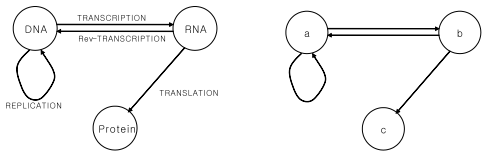

Figure 2: Boolean Network of DNA to protein via RNA.

The transition of genetic information in figure 2 is described using a Boolean Networks modelling method shown in figure 2. The beginning of the process with double helix DNA regarded as an initial product can take two different paths of processes; one is replication, the other is transcription. The process called transcription makes DNA into RNA, which is another product. After that product RNA becomes protein through the process: translation. The arrows in the figure above show that the processes replication, transcription, and translation can not be re-directed or undone, which means that the direction of the state is sequential by a certain order. Rev-transcription is a process from RNA to DNA, which is similar with transcription. The difference between (a) and (b) in figure 2 is DNA is a, RNA is b, and Protein is C. Also, all the processes replication, transcription, rev-transcription, and translation are not mentioned.

A	a'
0	0
1	1

b	c
0	0
1	1

Table 1: Boolean Networks of DNA to Protein via RNA.

Boolean Networks table is another representation of *Boolean Networks* model in figure 2. In table 1, the relationship between *a* and *b*, DNA and RNA is when a is 1, RNA is 1, if DNA is 0, RNA is 0.

a'	b	a
0	0	0
0	1	0
1	0	1
1	1	0

Table 2: Boolean Networks of DNA to Protein via RNA.

Table 2 shows us that *a'* is for a symbol of a replicated DNA strand, *b* for mRNA, and *a* is the original DNA. Different from table 1, influence of *a'* and *b* to *a* is a little

more complex. The value of *a* is only 1 when *a'* is 1 and *b* is 0.

3 DETAILED MODEL OF THE PROTEIN PRODUCTION

With reference to a compact model of the protein production it is much easier to go towards further details of the protein production process. In the following two subsections, the protein production process will be described in more detail and the corresponding model will illustrate more detailed information than the previous model (figure 2). These details concern some molecular events and precursors/molecules that are necessary for each step of the protein production process.

3.1 The Protein Production Process in Detail

Chromosomal DNA must be replicated at a rate that will at least keep up with the rate of cell division, and this process is a semi conservative process, i.e. the two strands of the parental DNA duplex act individually as templates for the synthesis of a complementary daughter strand (new strand of DNA) as shown in figure 3.

3.2 DNA Replication

Chromosomal DNA must be replicated at a rate that will at least keep up with the rate of cell division, and this process is a semi conservative process, i.e. the two strands of the parental DNA duplex act individually as templates for the synthesis of a complementary daughter strand (new strand of DNA) as shown in figure 3.

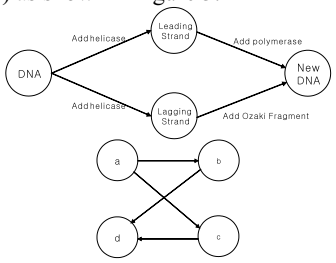

Figure 3: Boolean Network of DNA Replication.

A	b
0	0
1	1

Table 3: DNA to Leading Strand (A)

A	C
0	0
1	1

Table 3: DNA to Lagging Strand (B)

B	d
0	0

1	1

Table 3: Leading Strand to a New DNA (C)

b	c	d
0	0	0
0	1	0
1	0	1
1	1	0

Table 3: Leading and Lagging Strand to a New DNA (D)

3.3 DNA Transcription to mRNA

From a mechanistic standpoint transcription is quite similar to DNA replication apart from that where in replication only one DNA template strand is transcribed, and only a fraction of DNA strand in a genome is being expressed, and undergoes the process of transcription, in which an RNA molecule complementary to a fraction of DNA strand is synthesized.

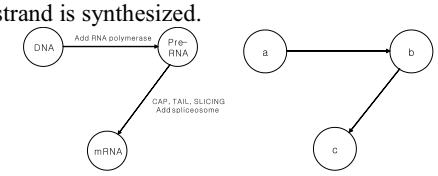

Figure 4. Boolean Network of DNA transcription to mRNA.

a	B
0	0
1	1

DNA TO PRE-mRNA (A)

b	c
0	0
1	1

PRE-mRNA to mRNA (b)

Table 4: Boolean Networks Table of DNA Replication

3.4 mRNA Translation to Protein

Each mRNA codes for the primary amino acid sequence of a protein, using a triplet of nucleotides (called codon) to represent each of the amino acids. In this process mRNA is decoded on ribosome to specify the synthesis of polypeptides (proteins). Following post-transcriptional processing, mRNA transcribed from DNA (gene) in the nucleus, migrates to the cytoplasm (as it is shown in figure 4), where mRNAs are read, and proteins assembled, on the ribosome, which are structures composed of rRNA and proteins. Transfer RNA (tRNA) is also needed for translation. Each of these tRNAs can be covalently linked to a specific amino acid, forming an aminocyl tRNA

(charged tRNA), and each has a triplet of bases called anti-codon.

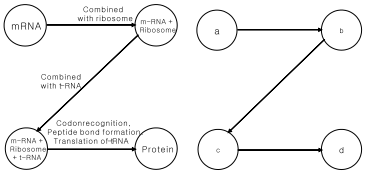

Figure 5. Boolean Networks of mRNA translation to protein.

a	B
0	0
1	1

m-RNA TO m-RNA * RIBOSOME (A)

a	B
0	0
1	1

m-RNA*RIBOSOME to m-RNA*RIBOSOME*tRNA (B)

C	d
0	0
1	1

m-RNA*RIBOSOME*tRNA TO PROTEIN (C)

Table 5: Boolean Networks Table of RNA Translation

3.5 A Whole Process of Protein Production

As it has been mentioned earlier from subsection 3.1-3.4, we can have the model of whole process DNA → RNA → protein. Even though we didn't cover everything, principal products, enzymes, process and data flows could be represented simply and successfully.

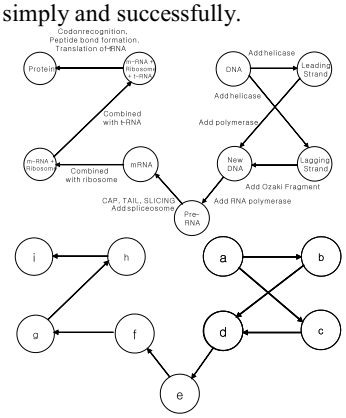

Figure 6. Boolean Networks of DNA replication, transcription, and translation.

Figure 6 just shows a whole process from DNA to protein, which is from figure 2-4. In the *replication*, a parent DNA becomes two DNAs, then in the figure 5, we used NEW DNA1, for two copied DNAs are exactly same.

4 CONCLUSION

The mean purpose of this paper was to develop modelling methodology for the production of proteins. This methodology helps formalization, modelling and simulation of the production of proteins. Therefore the first conclusion is that dynamic processes of molecular and biological systems in general, the protein production process in particular can be modelled as a discrete dynamic system.

Two areas can benefit from such a methodology that has been presented in this paper: to stimulate research and to assist teaching. For the teaching purposes, this can assist to visualize the protein production processes model from state to state and to explain how all molecular events, reactions and operations together provide production of proteins from DNA. It can show how the precursors and substrates, which are required for each step of the protein production processes, are bound to their targets. This paper can be also useful for the training program offering molecular biology with modelling and information sciences integrated into the individual courses, to train students in the use of computational techniques in the study of molecular and biological science. For the research purposes, one can use this methodology for the protein production modelling and simulation. It is also useful for protein and DNA sequence analysis. Finally, it seems that the results of this paper are one of the first efforts to apply discrete systems modelling technique to molecular-biology processes. In its turn it is another one step towards bringing computer science and molecular biology closer and calling it bioinformatics.

REFERENCES

[1] Alberts,B., Bray,D., Lewis,J., Raff,M., Roberts,K., and Watson,J.D., 1994. Molecular Biology of the Cell, (3rd edition). Garland.

[2] Karp,G. 1996. Cell and Molecular Biology. Wiley.

[3] Stryer,L., 1995. Biochemistry, (4th edition). Freeman, USA.

[4] Hawkins,J.D., 1997. Gene Structure and Expression. Cambridge University Press.

[5] Baxevanis,A.D. and Ouellette,B.F.F., 1998. Bioinformatics - A Practical Guide to the Analysis of Gene and Proteins.

[6] S.A. Kauffman., 1991, Scientific American.August. pp 64.

[7] S.A. Kauffman., 1993, The Origins of Order: Self-Organization and Selection in Evolution, Oxford University Press, New York.

[8] S.A. Kauffman., 1995, At Home in the Universe: The Search for Laws of Self-Organization and Complexity, Oxford University Press, Oxford.