IDEA GROUP PUBLISHING
701 E. Chocolate Avenue, Suite 200, Hershey PA 17033-1240, USA
Tel: 717/533-8845; Fax 717/533-8661; URL-http://www.idea-group.com
ITB12815

This paper appears in the book, *Emerging Trends and Challenges in Information Technology Management, Volume 1 and Volume 2* edited by Mehdi Khosrow-Pour © 2006, Idea Group Inc.

Modeling of DNA Transcription and Gene Regulation Using Petri Nets

Isaac Barjis, Wallied Samarrai, & Idline Augustin
Dept of Physical & Biological Sciences, New York City College of Technology

Joseph Barjis
Dept of Information Technology, Georgia Southern University

ABSTRACT

A new application area of Petri net is introduced in this paper. In particular it shows how, by means of Petri net, the protein production process can be modeled and analyzed. In order to develop a Petri net model of the protein production process, firstly the paper describes the protein production process from DNA. Although the aim of this paper is to develop a Petri net model of protein production rather then the molecular events and chemical reactions that occur during this process, some basic information about the molecular events, and precursors/ molecules that are required for each step of the protein production process are described in both part one and part two of this paper, prior to the Petri net model construction.

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 [7].

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 the Petri net modeling technique to model the protein production process. Petri net is a tool for the study of systems. Petri net theory allows a system to be modeled by a Petri net, a mathematical representation of the system. Analysis of the Petri net can then, reveal important information about the structure and dynamic behavior of the modeled 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 Petri net model of the protein production is developed; and in part two, a detailed Petri net 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 Petri net 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. A COMPACT PETRI NET MODEL OF THE PROTEIN PRODUCTION

In this and next sections we will show how Petri net 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 Petri net 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* (figure 2.1). 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 *amino acid* sequences proteins comes from a special class of RNA called messenger RNA(mRNA) transcribed from the DNA sequences of the *genes.* [5, 6].

The mRNAs are used by the protein-synthesizing machinery of the cell (ribosome) to make the appropriate proteins by translation. The flow of genetic information in the cell can be summarized by the simple schematic diagram shown in figure 2.1. As shown in the figure 2.1, DNA can either replicate to produce new double helix DNA or can be transcribed into mRNAs, where these mRNAs would consequently translated. Under some special circumstances, mRNAs can undergo the process of *reverse transcription* and produce double helix DNA.

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. In respect to the described peculiarities, inherent to the protein production process, we consider that Petri net

Figure 2.1The flow of genetic information.

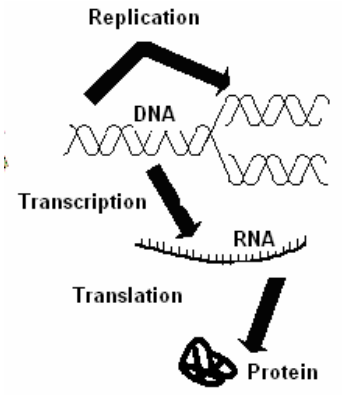
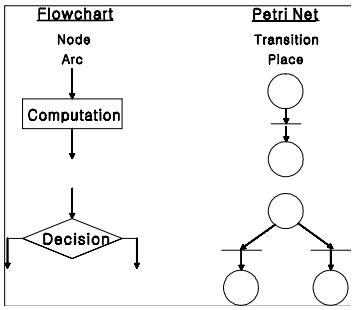


Figure 2.2 Comparison between flowchart and Petri net



can more adequately model and formalize the production of proteins. Therefore, in the following subsection we study the way in which Petri net can be applied.

2.2 A Petri net model of the protein production in compact notation

The Petri net is a formalized and graphical language for the design, specification, simulation and verification of systems [1, 2]. The Petri net structure consists of *places* and *transitions*. Corresponding to these, a Petri net graph has two types of nodes. A circle represents a *place*; a bar represents a *transition*.

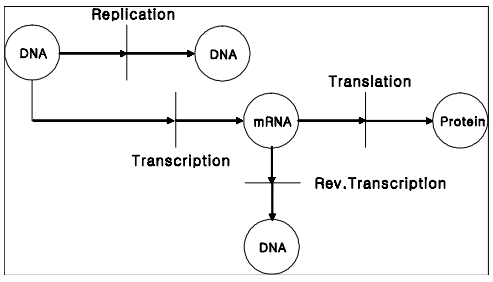
In developing a Petri net model of the protein production process, we find out operations (in terms of Petri net referred to as events or activities), which take place in the production of proteins, and states, which are reached after each of these operations. In other words we trace the sequence of operations from DNA to proteins.

Information in the Petri net model of the production of protein process (figure 2.3) is based on figure 2.1-2, and description of the protein production in the previous subsection. In this Petri net modeling, we observe three different *places* by names such as DNA, mRNA and protein *places*. All these *places*, in the Petri net model of figure 2.3, are representations of products.

Operations, which take place in this process, are *transcription, replication, translation,* and *reverse transcription (*figure 2.2) graphically represented by rectangles. Referring to figure 2.1, even visually it is obvious that a compact schema of the protein production consists of three states, as we mentioned above.

In order to develop a detailed and accurate model of protein production process, it is necessary to look at molecular events and reactions that occur in each step (i.e. process) in same details. In the following section we study a detailed model of the protein production.

Figure 2.3 Petri net model of the proteins production process



3. DETAILED PETRI NET MODEL OF THE PROTEIN PRODUCTION

With reference to a compact model of the protein production it is much easy 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 Petri net model will illustrate more detailed information than the previous model (figure 2.3). 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

The genetic information stored as DNA is used to direct the synthesis of RNA molecules or proteins. Each block of DNA that codes for a single RNA or protein is called a *gene*. Depending on the state of the cell, the genetic information could undergo *replication* or *transcription***.** If the cell undergoes *transcription*, it would consequently enter the post-transcriptional processes (*tailing* and *capping, cutting* and *splicing, transportation*), *translation*, and post-translational process before a *polypeptide* (protein) can be synthesized [3, 4].

3.3 DNA Transcription to RNA (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. *Transcription* begins when DNA dependant RNA *polymerase* binds to the *promoter* and moves along the DNA to the *transcription* unit. However RNA *polymerase* cannot initiate *transcription* by itself. Instead the binding of *transcription factors* (TF) in the *promoter* region of *gene* (e.g. TATA box, GC box) activate and guide the RNA *polymerase* (RNA *polymerase*) as shown in figure 3.1.

If we look at the first product pre-mRNA more in detail in figure 3.4, the start of the *transcription* unit the *polymeras*e begins to synthesize an RNA molecule complementary to the minus strand of DNA moving along this strand in a 3' to 5' direction, and synthesizing RNA in a 5' to 3' direction using *nucleoside triphosphates* (figure 3.2). The initial product of *transcription* is pre-mRNA (figure 3.3) which includes all of the *introns* (none coding sequences) and *exons* (coding sequences), so post-transcriptional processing is needed in *eukaryotes* (figure 3.3). In this processing a CAP sequence is added to the 5' end of the RNA, and about 150-200 *adenosine* residues are added to the 3' end, forming a *poly(A) tail*. This process also includes *cutting* and *splicing*, i.e. removal of unwanted internal segments (*introns*) and rejoining of the remaining segments (*exons*). The *capping* and *tailing* of pre-mRNA would protect

Figure 3.1 RNA is transcribed as a single strand, which is complementary in base sequence to one strand of a gene (DNA).

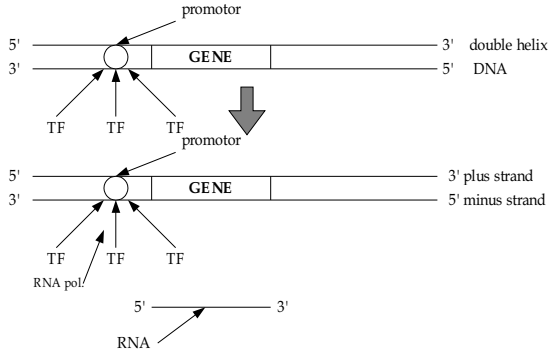
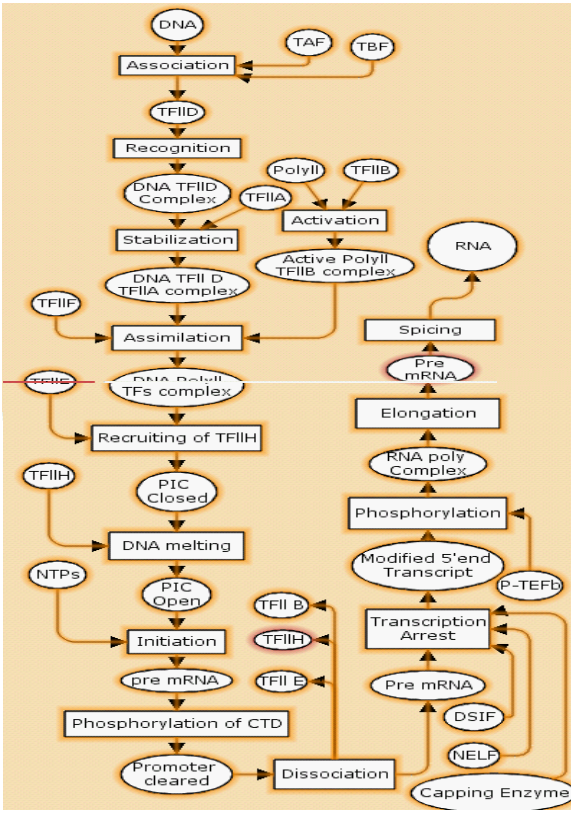


Figure 3.2 Detailed Petri net model of the protein production process.



the transcript from *endonucease* attack, facilitate transport of mRNA from nucleus to *cytoplasm*, facilitate *translation*, and facilitate RNA *splicing*

In Eukaryotes, The process of Transcription begins with the unwinding of the DNA double helix to expose the bases on both the template and the coding strand. The coding strand is used in protein synthesis where as the template strand contains the genetic information needed to produce the mRNA. The energy require for this process is called ATP. The enzyme responsible for the unwinding of the double helix is TFIIH. This exposes the active site for transcription know as the promoter. After the promoter has been exposed, the RNA polymerase attaches itself to this region with the help of Transcription factors TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIH, where "TF" stands for "transcription factor" and "II" for the RNA Pol II.

1. TFIID- recognizes the core promoter and consists of TBP (TATA-box binding protein) and TAFs (TBP associated factors). The role of TBP is to bind the core promoter and TAFs assist TBP in this process.
2. TFIIB- serves as a linkage between TFIID and TFIIF and polymeraseII. It fuction is to bind to TFIID and continue the assembly of (PIC) preinitiation complex.
3. TFIIF- reduces nonspecific interactions between polymerase and DNA.
4. TFIIH-promoters DNA repair and phosphorylate RNA polymerase II.
5. TFIIA- stabilizes TFIID and promoter binding.
6. TFIIE-recruit TFIIH and modulate TFIIH helices ATP and kinase activity.

These transcription factors do not proceed with the RNA polymerase during remainder of transcription.

During elongation of the RNA all the transcription factors dissociates itself from the promoter site with the exception of TFIIF. The RNA polymerase begins from the promoter region where it begins to string along a series of nucleotides together by covalent bonding. There are specific nucleotides involved in the growing chain of the RNA. The RNA polymerse can only attach Adenine, Cytosine, Guanine, and Uracil. Note unlike DNA synthesis the thymine has been replaced with Uracil. At each side of the DNA template the complementary RNA nucleotides form hydrogen bonds with the DNA nucleotides of the template strand. Because mRNA is copied from the DNA triplet of the template strand a threes base mRNA sequences is obtain which is complementary to the template strand and exactly like the coding strand with the exception of Uracil.

Along the template strand the RNA polymerase come in contact with a stop signal and detach itself along with the mRNA. Then the template and the coding strand of the DNA the reunites.

3.5 A Petri net model of DNA Transcription

As it has been mentioned earlier from subsection 3.2-3.4, we can have the Petri net model of whole process DNA->RNA.

Even though we didn't cover everything, principal products, enzymes, process and data flows could be represented simply and successfully using Petri net.

As it can be seen in the petrinet model above just the Transcription of the DNA can be consist of more than 100 reactions. Biological information in molecular level is modeled and expressed with *places* and *transition*s.

4. THE CONCLUSION

The mean purpose of this paper was to develop modeling methodology for the production of proteins using Petri net. This methodology helps formalization, modeling 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 modeled 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. This paper can be useful for the training program offering molecular biology with modeling 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 modeling 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 modeling technique such as Petri net to molecular-biology processes. In its turn it is another one step towards bringing computer science and molecular biology closer and calling it bioinformatics.

6. REFERENCE

1) Baxevanis,A.D. and Ouellette,B.F.F., Bioinformatics - A Practical Guide to the Analysis of Gene and Proteins 1998
2) James L. Peterson, *Petri net theory and the modeling of systems*. Prentice-Hall, Inc. 1981, Englewood Cliffs, NJ.
3) Kurt Jensen, *Colored Petri nets. Basic Concepts, Analysis Methods and Practical Use*, Volume I. 1997, Springer-Verslag.
4) Alberts,B., Bray,D., Lewis,J., Raff,M., Roberts,K., and Watson,J.D., *Molecular Biology of the Cell*, (3rd edition). Garland 1994.
5) Karp,G. 1996. *Cell and Molecular Biology*. Wiley
6) Stryer,L., 1995. *Biochemistry*, (4th edition). Freeman, USA.
7) Hawkins,J.D., 1997. *Gene Structure and Expression*. Cambridge University Press.