ABSTRACT

The protein production from DNA to protein via RNA is a very complicated process, which could be called central dogma. In this paper, we used event based simulation to model, simulate, analyze and specify the three main processes that are involved in the process of protein production: replication, transcription, and translation. The whole control flow of event-based simulation is composed of three parts (events): replication, transcription, and translation. In each part, control flow and events are depending on their final stage event. While in progress, we describe the relationship from DNA to protein in the systematic point of view and we describe each event with more details, where we will show that each event consist of number of sub-events i.e. in order for the DNA to be transcribed into mRNA, the RNA has to go through five sub-events before a mRNA produced. Although the aim of this paper is to develop an event-based simulation model for protein production rather than the molecular events and chemical reactions that occur during this process, some basic information of the molecular events and precursors/molecules that are required for each step of the protein production process are described. By modeling the whole process of central dogma, it is expected to develop the system and to simulate them in much easier ways.

2 DISCRETE-EVENT SIMULATION

2.1 Simulation modeling paradigm

There are a number of approaches to discrete event simulation; the most commonly used are event-based modeling and process based modeling. Each of these modeling styles is described but all the previous example simulation programs are developed using the process based approach. Event-based simulation is that execution of the main control loop represents a single event, also needs event queue to maintain the information to decide which event is next. This event-based simulation is complex but efficient and very accurate [1, 2].

2.2 Discrete-event Simulation Modeling and Protein Process

Event based simulation focuses the modeler’s attention on the individual events which can occur within the system. An event may generate several actions in the model-these
are grouped together in an event routine, and this is executed when the event reaches the head of the event list. At each event time as well as the processing of the event to represent the behavior of the system some processing internal to the model might be done. The process of protein production consists of events (figure 1), so an event-based simulation tools can be used to model, simulate, specify and analyze the process of protein production. The events are replication, transcription, translation. In processing, the event is concerned only with event processing with other enzymes and by-products. In each event, several small events induce the big event as a result.

3 COMPACT DISCRETE-EVENT SIMULATION

3.1 Protein Production Process

Protein synthesis begins in the cell's nucleus when the gene encoding a protein is copied into RNA. Genes, in the form of DNA, are embedded in the cell's chromosomes. The process of transferring the gene's DNA into RNA is called transcription. Transcription helps to magnify the amount of DNA by creating many copies of RNA that can act as the template for protein synthesis. The RNA copy of the gene is called the mRNA. 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 [1, 2]. 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. DNA can either replicate to produce new double helix DNA or can be translated into mRNAs, where these mRNAs would consequently undergo the process of translation to produce secreted protein. Under some special circumstances, mRNAs can undergo the process of reverse transcription and produce double helix DNA.

3.2 Model of the Protein Production in Compact Notation

Based on the figure 2, the flow of genetic information can be modeled by using event-based simulation method. As we mentioned before, understanding and graphical description of the whole process even more simulation is very beneficial for people who are not experts in this field and can be used as a basic material for understanding and visualizing the process and relationship between products and enzymes and so on. Event-based simulation modeling is composed of only with event (here, represented as a box) and data flows.

4 DETAIL MODEL OF THE PROTEIN PROCESS

4.1 The Protein Production Process in Detail

In order to construct an adequate model of the protein production, it is necessary to give same basic information about the molecular events, and precursors/molecules that are required for protein production process in same details. Without these details it is very difficult to construct a detailed and adequate model of the protein production. The genetic information of cells is stored in the form of DNA. This information is used to direct the production/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 [1, 2].

4.2 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. The enzyme called DNA helicases helps the two parent strands unwind and replication begins at specific points called the origin of replication, and involves the separation of the two DNA strands over a short length, and the binding of short RNA (RNA primer), and enzymes (e.g. DNA polymerases).

Then DNA polymerase catalyses the synthesis of new DNA strand, using the short RNA as a primer, and the four deoxynucleoside triphosphates (dATP, dCTP, dGTP, dTTP) as nucleotide precursors. Synthesis of a DNA strand occurs only in a 5’ to 3’ direction, but, since the parental DNA duplex are anti parallel, only one strand of DNA that move the same direction as the replication fork is moving, can be continuous. The other strand is synthesized in relatively short pieces called Okazaki fragment, (see figure 4), which are subsequently ligated (sealed) by DNA ligase to give a continuous strand of DNA.

The result of the replication is that the original DNA is replaced by two new DNA double helix, each containing one old and one new strand i.e. the genetic material of cell is doubled and is ready for division. All the above information will be summarized in a detailed model of the protein production. In particular it will be shown in the part of replication.

4.3 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).

Figure 5. Model of DNA transcription to RNA

In figure 5, model of DNA transcription of RNA is shown.

Figure 6. RNA is transcribed as a single strand, which is complementary in base sequence to one strand of a gene (DNA).

At the start of the transcription unit the polymerase 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 as shown in figure 6. The initial product of transcription is pre-mRNA as it is shown in the model of figure 5. This pre-mRNA includes all of the introns (none coding sequences) and exons (coding sequences), so post-transcriptional processing is needed in
which convert the primary polypeptide (linear chain of amino acid sequences) to secondary and tertiary (three-dimensional structure).

4.4 RNA 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 (shown in figure 7), where mRNAs are read, and proteins assembled, on the ribosome, which are structures composed of tRNA 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 anticodon.

![Figure 7. Model of RNA translation to protein](image)

Figure 7 is the final step to make protein followed by replication and transcription. RNA translation to protein events are needed below.

Ribosome initially recognizes the 5’ CAP via the participation of proteins that specifically bind to the cap. It then scans along the mRNA until it encounters the initiation codon (AUG, in few cases ACG, CUG or GUG are used instead). Translation continues until a termination codon is encountered (UAA, UAG, or UGA), and then the polypeptide chain is realized. However this is not the end of the story, since a linear chain of polypeptide is not active; therefore it should undergo post-translational processing, which convert the primary polypeptide (linear chain of amino acid sequences) to secondary and tertiary (three-dimensional structure).

5 CONCLUSION

Event-based modeling helps formalization, modeling and simulation of the production of proteins. 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. 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 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 to molecular-biology processes. In its turn it is another step towards bringing computer science and molecular biology closer and calling it bioinformatics.

REFERENCES