International Journal of ChemTech Research
CODEN(USA): IJCRGG ISSN : 0974-4290
Vol.5, No.5, pp 2623-2629, July-Sept 2013

Computational Model For The Extraction Of Human Erythropoietin

R. Selvakumar[1]*, M. Rashith Muhammad[1], G. Poornima Devi[2]

[1]School of Advanced Sciences VIT University, Vellore, Tamil Nadu – 632014, India

[2]Department of Biology, University of Heidelberg 69120 Heidelberg, Germany

***Corres.author: rselvakumar@vit.ac.in**

Abstract: A Deterministic Finite Automaton is constructed to identify the human protein called Erythropoietin from a large collection of proteins. This new technique is a computational procedure to identify erythropoietin without using *in vitro* methods. This scheme accurately reflects a DNA computer serving as a tool to analyze specified protein that can be matched with an already programmed protein sequence. Human erythropoietin has been taken as an example for this new technique to recognize the protein sequence extracted from the database. Erythropoietin is one of the highly marketed drugs for the treatment of anaemia due to chronic kidney failures. The current research is intensively concentrated for the use of erythropoietin in brain damage, neuroprotection etc., as it is involved in cell activation and proliferation. In the present study, we insist on the development of a technique for the identification of protein with erythropoietin as an example to explain the function and utility of the programmed machine constructed using Deterministic Finite Automaton.
Keywords: DNA sequence, Protein sequence, Erythropoietin (EPO), Deterministic Finite Automata (DFA).

Introduction

DNA computation has been emerged as an interdisciplinary field from mathematics, physics, chemistry, molecular biology, bio informatics, and computer science. It is one of the new area and acuity methods of computation, which uses molecules to perform calculation. It was initiated by Leonard M. Adleman [1], who is the innovator for solving an instance of the directed Hamiltonian path by using the tools of molecular biology i.e. by manipulating DNA strings. Mathematics plays a vital role in all types of computing particularly in DNA computing. It specifies that the mathematics lies at the origin of biology [2].

DNA is made up of four simple building blocks called Adenine (A), Thymine (T), Guanine (G), Cytosine (C) and it is the control center of each and every cell. It contains all of the information necessary to build an organism. DNA is a digitalized system of nature. Genes otherwise called as segments of DNA, contain specific instructions that make each individual unique. Generally the natural biological process includes DNA replication, transcription, post-transcription, translation and post translation that ends up with protein. Proteins are the essential macromolecules of all organisms and play a central role in almost all processes within cells. Proteins are the polymers of amino acids, referred to as polypeptides. The active protein is obtained after post translational modifications to transform the protein into active state and each possessing its own biological function.

In the present study, erythropoietin (EPO) is opted for developing a mathematical conduit to build a bridge between the far understood systems biology and software systems. This research effort has in-fact been

in the subject of solving two difficult issues such as an expounding theory of systems biology and an inspired computational model for gene expression in order to have an appropriate response to external stimuli. The combination of these exceptionally challenging problems of research is a very attractive and booming epistemological experiment leading to a significant amount of joint theory organization of both mathematical solution and interpretation of experimental results in biological systems. As per Chomsky hierarchy, finite-state automata are the most fundamental computational model and this is the establishing point to build worldwide DNA computers [3]. Some of the works have been stabbed to build finite automata for protein synthesis. It is the work which attempt to develop the finite automata in searching a protein by using the protein sequence.

Characteristics of Human EPO

In 1958, Jacobson and co-workers [4] proposed that the kidneys were the source of this hormone, now termed erythropoietin. Erythropoietin consists of a single chain of 165 amino acids with a polypeptide backbone to which a large number of sugars giving the whole hormone a molecular weight of 34 000 daltons [5]. The gene is located on chromosome number 7 [6] and in the kidneys is activated by hypoxia, possibly via a heme protein [7]. The principal action of erythropoietin is to serve as a differentiation factor in the blast transformation of mature erythroid progenitor cells (CFU-E) into erythrocytes and the process is termed as erythropoiesis. Thereby, it functions to increase the number of red blood cells.

Cloning of recombinant Human Erythropoietin (rHuEPO)

Due to its functional importance as a drug being used in the treatment of anaemia and other non-blood functions, it has its own importance as a commercial protein. In order to satisfy the increasing demand in the medicinal market, scientists are concentrating in developing an efficient expression systems for the production of rHuEPO. The cloning of HuEPO includes the following steps: 1. Isolation and amplification of HuEPO gene (DNA) using polymerase chain reaction (PCR), 2. Choosing the expression vector and an appropriate expression system, 3. Restriction digestion of both the amplified HuEPO gene and the expression vector separately with the appropriate restriction endonuclease enzymes, 4. Ligation of the restricted EPO gene and the expression vector with the influence of the enzyme called DNA ligase, 5. Transformation of the expression vector construct into the respective host organism, 6. Trascription and translation of the recombinant DNA to recombinant RNA inside the host, and 7. Purification of the rHuEPO from the respective host by following the methods in the downstream processes. The stages in cloning of a gene are explained briefly in the following transition figure (Fig.1).

Fig. 1: Transition diagram to get the purified recombinant Human erythropoietin

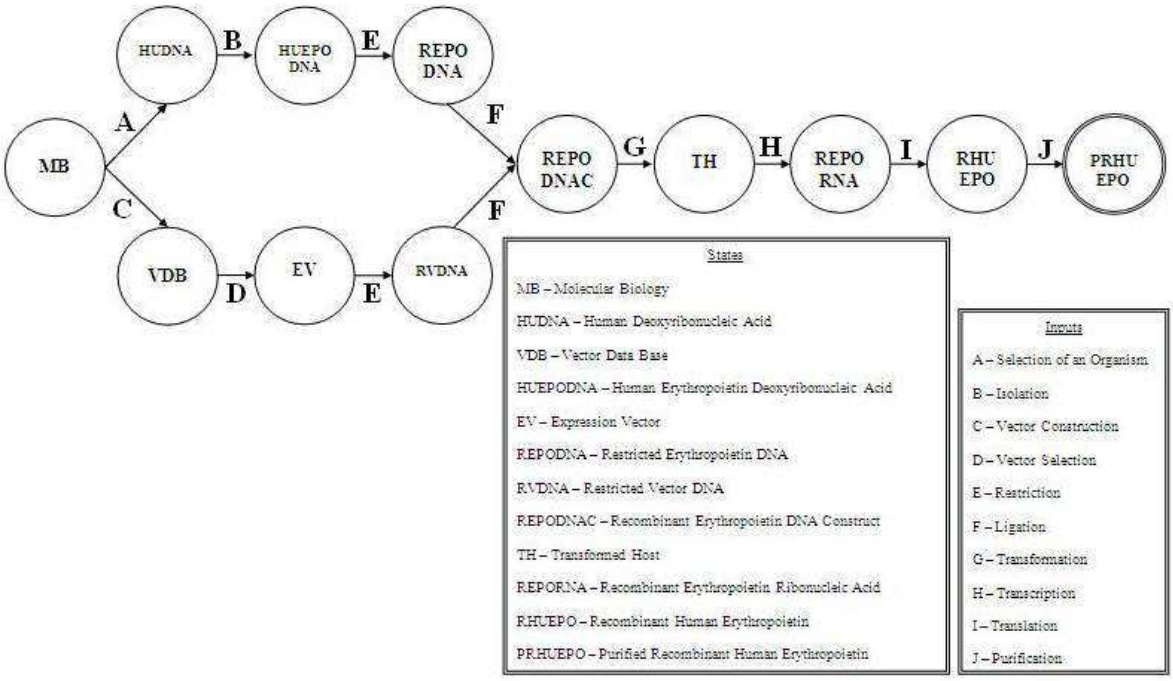


Fig. 2: Transition Table for the process of purified recombinant human erythropoietin

STATES / INPUTS	A	B	C	D	E	F	G	H	I	J
MB	HUDNA		VDB							
HUDNA		HUEPODNA								
VDB				EV						
HUEPODNA					REPODNA					
EV					RVDNA					
REPODNA and RVDNA						REPODNAC				
REPODNAC							TH			
TH								REPORNA		
REPORNA									RHUEPO	
RHUEPO										PRHUEPO

In Finite Automata, the transition table (Fig.2) is one of the ways of representation. The set of states and inputs are denoted as follows:

Set of states = {MB, HUDNA, VDB, HUEPODNA, EV, REPODNA, RVDNA, REPODNAC, TH, REPORNA, RHUEPO, PRHUEPO}

Inputs = {A, B, C, D, E, F, G, H, I, J}

Transition procedures to get the purified recombinant erythropoietin from Human DNA are explained below

1. Start from the current state as MB with input A, and then it goes to the state as HUDNA.
2. Start with the current state as MB with input C, and then it goes to the next state as VDB.
3. Start with the current state as HUDNA state with input B, and then it goes to the next state as HUEPODNA.
4. Start with the current state as VDB with input D, and then it goes to the next state as EV.
5. Start with the HUEPODNA state with input E, and then it goes to the next state as REPODNA.
6. Start with the EV state with input E, and then it goes to the next state as RVDNA.
7. Start with the two states as REPODNA and RVDNA with input F, and then it goes to the next state as REPODNAC.
8. Start with the REPODNAC state with input G, and then it goes to the next state as TH.
9. Start with the TH state with input H, and then it goes to the next state as REPORNA.
10. Start with the REPORNA state with input I, and then it goes to the next state as RHUEPO.
11. Start with the RHUEPO state with input J, and then it goes to the next state as PRHUEPO.

There are no more procedures are applicable in this Finite Automata except the above eleven procedures.

Therapeutic uses of EPO

It is a safe alternative to blood transfusion and is effective in reducing anaemia in many conditions and to boost hematocrits prior to per-surgical blood donations for autologous transfusion. It is the first hematopoietic growth factor approved for human therapy, recombinant human EPO (rHuEPO) has been used for the treatment of anemia resulting from chronic renal failure, cancers (primarily chemotherapy-associated anemia), autoimmune diseases, AIDS, surgery, bone marrow transplantation and myelodysplastic syndromes, etc., for more than 2 decades.

Interestingly, recent studies have also observed that rHuEPO has non-blood system functions and shows the potential of being used as a neuroprotective drug for cerebral ischemia, brain trauma, inflammatory disease and neural degenerative disorders [8]. EPO have a history of usage as a blood doping agent in endurance sports such as cycling, rowing, distance running, race walking, cross country skiing, biathlon, triathlons and most recently billiards [9].

Deterministic Finite Automata (DFA)

An alphabet is a finite non empty set of symbols. Symbols are denoted by small letter or digit. A string is a finite sequence of symbols. (e.g. $\Sigma = \{a,b\}$, here Σ is a alphabet; a, and b are symbols; and abbb, abab, ababab are strings) [10]. The set of strings over the given alphabet is called a formal language.

Definition: A deterministic finite automaton is a five tuple, $A=(\Sigma, Q, q_0, F, \delta)$ where Σ is the input alphabet, Q

the set of states, $q_0 \in Q$ the start state, $F \subseteq Q$ the final states and $\delta : Q \times \Sigma \rightarrow Q$ transition function.

Two major elements in finite automata are states and inputs. Here the change of state is fully governed by the input and the current state. The state moves from one place to another place due to the input and reaches the final state. The term deterministic refers to the fact that on each input there is one and only one state to which the automaton can transit from its current state. The input mechanism can move only from left to right and it can read exactly only one symbol on each step.

Existing biological methods in purification of HuEPO

The conventional method for purifying erythropoietin from a culture supernatant of erythropoietin-producing eukaryotic cells includes time consuming downstream process. The first step is to disrupt the cell for removing cell components for intracellular expression of HuEPO. The extracellular expression and secretion of proteins include dialysis and ultrafilteration based on the molecular weight. The purification is then followed by chromatography methods in the order indicated as follows: i) reversed-phase chromatography; ii) anion-exchange chromatography and iii) hydroxyapatite chromatography. The purified protein is further confirmed by performing immunoblotting in comparison with the standard erythropoietin.

Computation process of Deterministic Finite Automata

Chromatography method is one of the practical oriented, existing and difficult methods for identifying protein from the collection of proteins.

Here we design a protein based programmable computers. It means that a program is stored as a data and any computation can be accomplished by just choosing a stored program. In protein based computers, it requires that a program is encoded into a molecule different from the main and fixed units of protein computers, a molecule encoding programs can be stored and changed, and a change of molecules encoding programs accomplishes any computations [3].

In this paper, the strings are represented by DNA strands and it is denoted by capital letter. Now we extract a two set of strings of example DNA and RNA sequences each containing 30 mer oligonucleotide from the active protein human erythropoietin.

DNA Sequences:

ACGGGCTGTGCTGAACACTGCAGCTTGAAT

CCAGACACCAAAGTTAATTTCTATGCCTGG

This DNA sequence is changed into a RNA sequence. In the RNA sequence we get Uracil instead of Thymine.

RNA Sequences:

ACGGGCUGUGCUGAACACUGCAGCUUGAAU

CCAGACACCAAAGUUAAUUUCUAUGCCUGG

Let us take the DNA language D* as a set of all strings over the DNA alphabet D= {A,T,G,C} and the RNA language R* be a set of all strings over the RNA alphabet R= {A,U,G,C}. Then the protein language P* be a set of all strings over the protein alphabet

P= {F,L,I,M,V,S,P,T,A,Y,H,Q,N,K,D,E,C,W,R,G}

By using the codon table we can explain the amino acid names as a symbol [11]. The symbols are denoted as-

Symbol	Amino Acid	Symbol	Amino Acid
F	Phenylalanine	H	Histidine
L	Leucine	Q	Glutamine
I	Isoleucine	N	Asparagine
M	Methionine	K	Lysine
V	Valine	D	Asparticacid
S	Serine	E	Glutamicacid
P	Proline	C	Cysteine
T	Threonine	W	Tryptophan
A	Alanine	R	Arginine
Y	Tyrosine	G	Glycine

Through RNA sequence, we can easily get the protein sequence. The protein sequences are the group of amino acids.

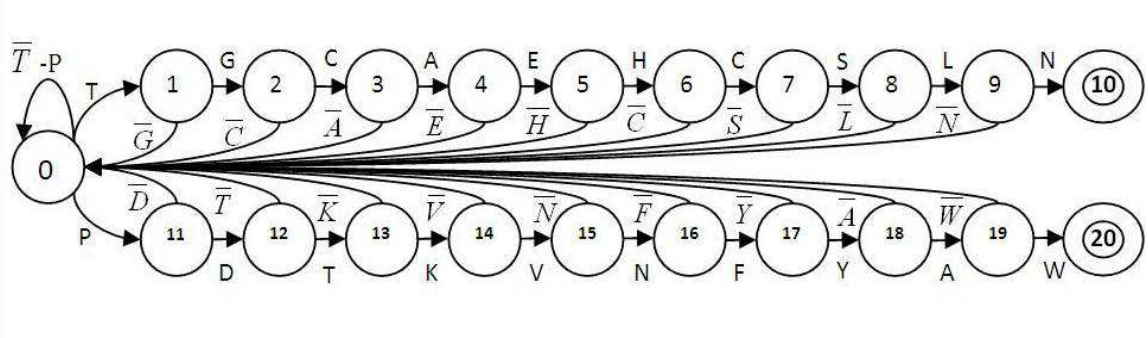
Protein Sequences:

TGCAEHCSLN
PDTKVNFYAW

For this protein we construct to accept a deterministic finite automaton. If the sequence is accepted by our machine, then it reaches the acceptance state otherwise it goes to reject state. When the end of the string is reached, the string is accepted if the automaton is in one of its final states, otherwise the string is rejected.

The input mechanism can move only from left to right and reads exactly one symbol on each step. In the following DFA diagram, $\overline{G}$, $\overline{D}$, $\overline{C}$, $\overline{T}$, etc., shows that the complement of G,D,C,T, etc., and $\overline{T}$ -P represents that the complement of T except P.

Fig. 3: Example of a Finite Automaton



The Transition functions are denoted as

δ (0, $\overline{T}$ -P) = 0, δ (0,T) = 1, δ (1,G) = 2, δ (1, $\overline{G}$) = 0, δ (2,C) = 3,

δ (2, $\overline{C}$) = 0, δ (3,A) = 4, δ (3, $\overline{A}$) = 0, δ (4,E) = 5, δ (4, $\overline{E}$) = 0,

δ (5,H) = 6, δ (5, $\overline{H}$) = 0, δ (6,C) = 7, δ (6, $\overline{C}$) = 0, δ (7,S) = 8,

δ (7, $\overline{S}$) = 0, δ (8,L) = 9, δ (8, $\overline{L}$) = 0, δ (9,N) = 10, δ (9, $\overline{N}$) = 0.

δ (0,P) = 1, δ (11,D) = 12, δ (11, $\overline{D}$) = 0, δ (12,T) = 13, δ (12, $\overline{T}$) = 0,

δ (13,K) = 14, δ (13, $\overline{K}$) = 0, δ (14,V) = 15, δ (14, $\overline{V}$) = 0, δ (15,N) = 16,

δ (15, $\overline{N}$) = 0, δ (16,F) = 17, δ (16, $\overline{F}$) = 0, δ (17,Y) = 18, δ (17, $\overline{Y}$) = 0,

δ (18,A) = 19, δ (18, $\overline{A}$) = 0, δ (19,W) = 20, δ (19, $\overline{W}$) = 0.

Conclusion

According to the knowledge of the authors, one of the finest methods for the identification of protein has been designed in this paper. A programmable machine can be constructed using DFA for the identification of protein by considering the respective DNA as a template sequence. Through this new technique we can extract the particular protein in a simple and time consuming method. This machine can be used in the medicinal research to design drugs for anaemia with less complexity.

References

[1] Adleman, L., Molecular computation of solutions to combinatorial problems. Science.1994, 266, 1021-1024.

[2] Kari, L., DNA computing: arrival of biological mathematics. The Mathematical Intelligencer.1997, 19, 9-22.

[3] Hirotaka Nakagawa., Kensaku Sakamoto., and Yasubumi Sakakibara., Development of an In Vivo Computer Based on Escherichia coli in: Carbone, A. and Pierce, N.A (1st Ed.), DNA Computing, Springer-Verlag, Berlin Heidelberg 2006, pp. 203-212.

[4] Jacobson, L.O., Goldwasser, E., Fried, W., Plzak, L., Role of the kidney in erythropoiesis. Nature.1957, 179, 633-634.

[5] Spivak, J.L., The mechanism of action of erythropoietin. The International Journal of Cell Cloning.1986, 4, 139-166.

[6] Wang, F.F., Kung, C.K.-H., Goldwasser, E., Some Chemical Properties of Human Erythropoietin. Endocrinology.1985, 116, 2286-2292.

[7] Powell, J.S., Berkner, K.L., Lebo, R.V., Adamson, J.W., Human erythropoietin gene: high level expression in stably transfected mammalian cells and chromosome localization. Proc Natl Acad Sci USA.1986, 83, 6465-6469.
[8] Tamura, K., Sudo, T., Senftleben, U., Dadak, A.M., Requirement for p38alpha in erythropoietin expression: a role for stress kinases in erythropoiesis. Cell.2000, 102, 221-231.
[9] Feldman, L., Sytkowski, A.J., Pleiotrophic actions of erythropoietin, Environmental Health and Preventive Medicine.2003, 7, 239-245.
[10] Hopcroft, J.E., Ullman, J.D., (1st Ed.), Introduction to Automata Theory, Languages, and Computation, Addison Wesley, 1979.
[11] Bozinovski, S., Skodons and Finite Automaton: New representations of the Genetic Code. Proc IEEE Int Conf Syst Man cybern.1989, 3, 1270-1271.

*****

*****