ISSN: 2319 – 8753
International Journal of Innovative Research in Science, Engineering and Technology
Vol. 1, Issue 2, December 2012

Protein Structure Prediction Using Stochastic Process Probabilistic Model

Subhendu Bhusan Rout[1], Sarojananda Mishra[2]
Dept of Computer Science Engineering & Application, IGIT Sarang, Odisha, India[1]
Dept of Computer science Engineering & Application, IGIT Sarang, Odisha, India[2]

Abstract: Protein Structure Prediction is the process of prediction of the three dimensional structure of a protein from its amino acid sequence. In order to develop a new drug it needs to process hug amount of data to study the behaviour of various types of genes. In recent years many techniques are being used for the protein structure prediction. Recently the Bioinformatics industry is in the fledgling condition and gaining more attention of researchers. Various Soft Computing methods like Fuzzy Logic, Artificial neural network, genetic algorithms, swarm optimization, etc are used for this purpose to distinguish, compare or process various type of data. It is always a big challenge for researchers to develop new tools and methods for the processing of data as well as development of drugs. It is also a major chapter for the recent researchers and scientists for the prediction of protein structure for the designing of drugs. Probabilistic theory can be applied to predict the structure of protein in less amount of time. This paper will proposed an idea for the prediction of protein structure using stochastic process probabilistic theory.

Keywords: Protein structure, Gene, Stochastic Process, Fuzzy Logic, Artificial neural network, Swarm Optimization, Genetic Algorithm

I. INTRODUCTION

Bioinformatics is the application of computer technologies in the field of biological information. Due to the rapid advancement in computer technology it is now easier to process various type of biological information through computer. Right now large number of researchers is working upon bioinformatics and many research works also implemented every day. Now there are much more research data and information near us which is very precious and helpful for our future research. Like the secondary data these can be reuse the previous data to modify the newer one to develop a new one. Some cases the improper result after application to various genomic body of a previously developed drug may need some better research or study, which can be fruitful in same gene structure or some other genomic body also.

Proteins are the very tiny particle of a living body. Protein Structure prediction is nothing but the prediction of the three dimensional structure of a protein from its amino acid sequence. The protein structure shape changes from time to time after applying the drugs as shown in Fig 1. From the change in its protein structure in the secondary, territory & quaternary structure of the protein a researcher can study what are the effects of a particular drug to that amino acid sequence. So many times a drug may be designed in a proper way but it may not work properly for all type of living body. In that case we are having huge amount of research data, but inefficient study of the changes in the DNA, RNA & Protein structure may cause the harmfulness of the drug or may not work properly. In that case it needs some high level of study or research which may bring out the proper knowledge from the previous data. A highly talented researcher or a high level of technology may bring out the problems or the modifications to short out the problems according to that a new product may design to fit the actual need.

Probability theory is always a good application to various fields to predict or to calculate the future happenings. Stochastic Process is the prediction of a future probability after studying a number of past happenings. In this paper the section 2 gives a brief idea about various particle of a human body like DNA, RNA & Protein Structure followed by a brief knowledge about probability theory and stochastic process probability theory in section 3. In subsection 3.1 and 3.2 we give two examples of

ISSN: 2319 – 8753
International Journal of Innovative Research in Science, Engineering and Technology
Vol. 1, Issue 1, November 2012

stochastic process. In section 4 we proposed an Idea to predict the protein structure and finally the paper is concluded with conclusion in section 5.

II. PROTEIN STRUCTURE: A BRIEF INTRODUCTION

Proteins are large biological molecules consisting of one or more chains of amino acids. Proteins perform a vast array of functions within living organisms, i.e. catalyzing metabolic reactions, replicating DNA, responding to stimuli, and transporting molecules from one location to another. The Proteins differ from one another primarily from their sequence of amino acids, which is dictated by the nucleotide sequence of their genes and which usually results in folding of the protein structure into a certain three-dimensional structure that determines its activity.

There are various biological macromolecules like polysaccharides, nucleic acids, proteins etc. A polypeptide is a single linear polymer chain of amino acids bonded together by peptide bonds between the carboxyl and amino groups of amino acid residues. The sequence of amino acids in a protein is defined by the sequence of a gene, which is encoded in a form of genetic codes. Generally, the genetic code specifies 20 standard amino acids. In certain organisms the genetic code can include selenium and certain archaea-pyrrolysine.

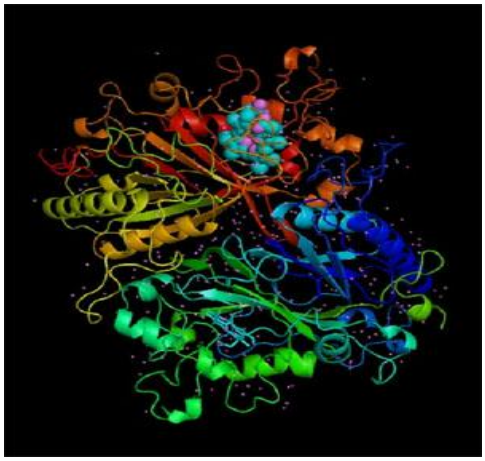

Fig. 1. 3-D Protein Structure

After the synthesis or even during synthesis, the residues in a protein are often chemically modified as shown in Fig 1. The modification may be the physical and chemical properties, folding, stability, activity, and ultimately the function of the proteins. Sometimes proteins have non-peptide groups attached, which can be called prosthetic groups orcofactors. Proteins can combine together to achieve a particular function, and also it may often associate to form stable protein complexes.

In comparison to other biological macromolecules like polysaccharides and nucleic acids, proteins are also the essential parts of organisms and participate virtually in every process within cells. Many proteins are enzymes that catalyze biochemical reactions and are vital tometabolism. Proteins also have structural or mechanical functions, such as actin and myosin in muscle and the proteins in the cytoskeleton, which form a system of scaffolding that maintains cell shape. Other proteins are also useful in various tasks like cell signaling, immune responses, cell adhesion and the cell cycle. Animals can not synthesize all amino acids, so they extract essential amino acid from foods. So in this factor Proteins are also necessary in animals' diets. Through the process of digestion, animals break down ingested protein into free amino acids that are then used in metabolism. Proteins may be purified from other cellular components using a variety of techniques such as ultracentrifugation, precipitation, electrophoresis, and chromatography. The advent of genetic engineering has made possible a number of methods to facilitate purification. There are many Methods that are commonly used to study protein structure and function like immunohistochemistry, site-directed mutagenesis, nuclear magnetic resonance and mass spectrometry.

ISSN: 2319 – 8753
International Journal of Innovative Research in Science, Engineering and Technology
Vol. 1, Issue 1, November 2012

Proteins are an important class of biological macromolecules present in all organisms. In molecular biology protein structure describes the various levels of organization of protein molecules. According to size Proteins are nanoparticles that contain polymers of amino acids. Each protein polymer (also known as a polypeptide) consists of a sequence formed from 20 possible L-α-amino acids, also referred to as residues. For chains under 40 residues the term peptide is frequently used instead of protein. To be able to perform their biological function, proteins fold into one or more specific spatial conformations, driven by a number of non-covalent interactions such as hydrogen bonding, ionic interactions, Van-Der-Waals forces, and hydrophobic packing. To understand the functions of proteins at a molecular level, it is often necessary to determine their three-dimensional structure. This is the topic of the scientific field of structural biology, which employs techniques such as X-ray crystallography, NMR spectroscopy and dual polarization interferometry to determine the structure of proteins.

Protein structures range in size from tens to several thousand residues. Very large aggregates can be formed from protein subunits. For e.g. any thousand actin molecules assemble into a microfilament. A protein may undergo reversible structural changes in performing its biological function. The alternative structures of the same protein are referred to as different conformations and transitions between them are called conformational changes.

III. STOCHASTIC PROCESS: A DETAILED STUDY

Stochastic process is basically a probabilistic theory to predict the future happenings. It is often considered as the study of behaviour of a system over some period of time. A stochastic process is defined to be an indexed collection of random variables $\{X_1\}$, where the index t runs through a given set T. The T may take as the set of nonnegative integers in certain cases and X_t represents a measurable characteristic of interest at time t. For example, X_t might represent the inventory level of a particular product at the end of week t. Stochastic processes are of interest for describing the behavior of a system operating over some period of time. A stochastic process often has the following structure.

The current status of the system can fall into any one of M+ 1 mutually exclusive category called states.

For notational convenience these states are leveled 0,1,2…….M. The random variable X_t represents the state of the system at time t, so its only possible values are 0,1,2……M. The system is observed at particular points of time, labeled t=0,1,2……. Thus the stochastic process $\{X_t\}=\{X_0, X_1, X_2…….\}$ provides a mathematical representation of how the physical system evolves over time. This kind of Process is referred to as being a discrete time stochastic process with a finite state space.

1) A weather example to Stochastic Process

The weather in the town of Chicago can change rather quickly from day to day. However the chances of being dry (No rain) tomorrow are somewhat larger if it is dry today than if it rains today. In particular the probability of being dry tomorrow are somewhat larger if it is dry tomorrow is 0.8 if it is dry today, but it is only 0.6 if it rains today. These probabilities do not change if information about the weather before to-day is also taken into account.

The evolution of the weather from day to day in Chicago is a stochastic process. Starting on some initial day (levelled as day 0), the weather is observed on each day t, for t= 0,1,2…. The states of the system on day t can be either

State 0= Day is dry
Or
State 1=Day t has rain.

Thus for t=0,1,2…… the random variable X_t takes on the values,

ISSN: 2319 – 8753
International Journal of Innovative Research in Science, Engineering and Technology
Vol. 1, Issue 1, November 2012

$$X_t = \begin{cases} 1 & \text{If dav t has rain.} \\ 0 & \text{If day t is dry.} \end{cases}$$

Here the number of states is two i.e. 0 or 1. Simply we can say there may be a rain or dry. So the matrix according to the above data will be

State	0	1
0	0.8	0.2
1	0.6	0.4

The stochastic process $\{X_t\} = \{X_0, X_1, X_2 \ldots\ldots\}$ provides a mathematical representation of how the status of the weather evolves over time. In this way a future prediction may calculate by the study of past events.

2) An application of drugs to a living cell

Suppose we want to apply a particular medicine to a living cell and let the colour of the living cell is white at the beginning & it may change in to three different color like orange Green & Black. After the end of first week the Cell color changes from white to orange. At the end of second week the color changes from Orange to Green. At the end of third week the color changes from Green to Black. If X_1 is the state in color Orange, X_2 is the state in color Green, X_3 is the state in color Black so X_4, X_5,…. can be calculated.

According to stochastic process the states are levelled as 0,1,2,3….M. The random variables in this example $\{X_t\} = X_1, X_2, X_3, X_4 \ldots$ is the state at time t= 1,2,3,4…. In this way stochastic process provides a mathematical expression for the representation of physical changes of colour.

IV. AN APPLICATION OF STOCHASTIC PROCESS TO PROTEIN STRUCTURE PREDICTION

Generally stochastic process is the prediction of a future state by studying the previous states. It is a major task of researchers to study how the behaviour of a gene or the structure of a protein changes over time after applying various drugs. Every time there should be the study of a huge amount of data in order to study the work of a particular gene. In order to develop a particular drug it needs a lot of experiments and predictions which should be upon a hug amount of genes which generally comes from various parts of a globe. Sometimes there may be some research data (that may call as secondary data) which need little bit extra effort for the application to various fields. Some times in the field of medicinal research it is often predicts the behaviour of a Genes or structure of a protein by applying various quantities of drug or minerals or vitamins. It is a lengthy process to calculate or predict the behaviour or changes upon a macromolecule that comes in a huge amount from various parts of a country or region.

Suppose a small quantity of drug that applied to various Proteins that comes from various regions. It in order to process a large amount of data or to compare the behaviour of various Proteins we may the take the help of computer. Some Proteins may react heavily or some may lightly. The three dimensional structure may change differently upon different Proteins. After a gap of time we may increase or decrease the amount of drug and again we may study the behaviour and the change in stricture of various Proteins.

According to stochastic probability theory If 't' is the observed time, X_0 is the observed behaviour of DNA at the end of week '0', X_1 is the observed behaviour at the end of weak '1', hence X_2, X_3.... can be calculate & predict according to previous change in shape and size.

V. CONCLUSION

There are various types of macromolecules like polysaccharides, nucleic acids; proteins etc are in a living body. Bioinformatics is a better way to study the behaviour of DNA, RNA and to predict the protein structure prediction. Thus application of bioinformatics provides a gateway to process huge amount of data but Probability theory is very helpful to predict the structure of Protein or study the behaviour of DNA, RNA, etc in a small amount time. So Stochastic process will helpful for designing of various drugs, after processing of huge amount data with less amount of time. Biochemistry is always a good and upcoming topic for every researcher as every day a lot of data are being processed & many changes to drugs are being developed. This Paper gives a proposed idea and our next work will focus upon the various real time effects and application of bioinformatics & probability theory upon various macromolecules. It also includes how these processes can be carried out in a beneficiary mode with less amount of time.

REFERENCES

[1] Vinicius Tragante do O, Renato Tinos, "Diversity Control in Genetic Algorithms for Protein Structure Prediction", 727-737, 2009.
[2] Gabriel, P., Lima, T., Delbem, A., Faccioli, R., Silva, I. Pure ab initio evolutionary approach to protein structure prediction. *Proceedings of the International Symposium on Mathematical and Computation Biology. BIOMAT 2007, Armação dos Búzios,* Brazil, 2007.
[3] Tinos, R., Yang, S. A self-organizing random immigrants genetic algorithm for dynamic optimization problems. *Genetic Programming and Evolvable Machines*, 8(3): 255-286, 2007.
[4] Lima, T., Gabriel, P., Delbem, A., Faccioli, R., Silva, I. Evolutionary algorithm to ab initio protein structure prediction with hydrophobic interactions. *IEEE Congress on Evolutionary Computation*, 2007 (CEC 2007), 612–619, 2007.
[5] F. S. Hiller, G.J. Liberman, Introduction to Operation Research, Tata McGraw Hill, Eighth Edition.

ACKNOWLEDGEMENT

I am very much thankful to Sumitra Kisan, VSSUT Burla, for her co-operation and encouragement regarding this research work. I am also giving a special thanks to all the faculty members and staffs of Indira Gandhi institute of technology, Sarang for their help and support during this research work. Finally I express my deep and sincere gratitude towards the Reviewer of IJIRSET for their valuable suggestion and comments which improved the quality of the paper.

Biography

Subhendu Bhusan Rout, is working as a Lecturer in the Dept of Computer Science Engineering & Application of IGIT Sarang, Odisha, India. He has completed B.Tech. from BPUT Odisha & M.Tech Degree from KIIT University, Odisha, India. Now He is also a Ph.D. scholar of Fakir Mohan University, Odisha. He is having 3 year PG teaching experience in the field of Computer Architecture, Probability & statistics, Computer Security, etc.

Dr. Sarojananda Mishra, is working as the Prof. & HOD in the Dept of Computer Science Engineering & Application of IGIT Sarang, Odisha, India. He has completed MCA from Sambalpur University, Odisha & M.Tech. from IIT Delhi. Subsequently He has completed his Ph.D. from Utkal University, Odisha. He has a lot of National & International Journals with having more than 20 years of teaching experience.