Volume 3, Issue 5, May 2013 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

A Hybrid Approach for Protein Structure Prediction

Priyanka P. Jirvankar[1], Sudipta Giri[2]
[1]*IT Department,* [2]*IT Department,Mitcoe,pune*
Pune, India

Abstract— Protein plays vital role in all living organism. Predicting structure of a protein in lab using NMR, X-ray diffraction is time consuming and critical task. There are large number of proteins whose structure is not yet discovered. There is need to develop some computational method or an algorithm to determine the structure of protein in a fast and efficient manner. To aid the Biologist, we are trying to develop a portal which will use MC simulation to predict the structure of protein. This will help biologist to understand the structure of protein and obtain structure to functional relationship for drug discovery, improve crop production.

Keywords—Monte Carlo simulation,ramchandran plot,alpha helix, beta sheet, Protein structure prediction.

I. INTRODUCTION

Protein plays a vital role in every cell of living organisms. Protein structure prediction means predicting 3D structure of protein from its amino acid. There are 20 amino acids in the universe. Any Combination of amino acids forms a protein which is connected by peptide bond. Many conformations are possible due to rotating Cα and C_β atoms.

Predicting structure of protein involves predicting secondary, tertiary and quaternary structure from sequence of amino acids.

II. LITERATURE SURVEY

A. Homology based modelling

It is known as comparative modelling of protein structure prediction. Comparative modelling predicts the 3D structure of given sequence(source) based on aligning to one or more proteins known as target protein[2].

The homology modelling can be arranged as follows:

a)Identifying an alignment between the target and relative protein sequences.b)Copy the main chain co-ordinates from the related protein from equivalent residues and inferring some side chain conformation. c)Build other structure left.

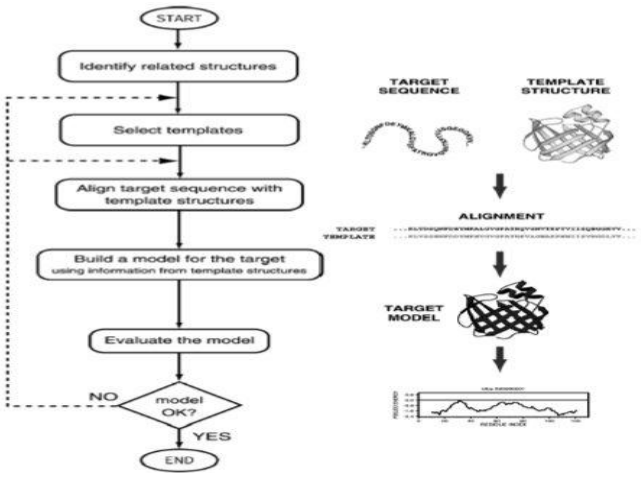

Fig. 1 .1 Homology based modelling

B. Graph theory based model

In this model, protein are treated as a graph. Alpha helix and beta sheet referred as a node, while loops and links are treated as a edge. There are some fixed protein structure refer as motif[4]. These can be treated as graph. Following are example of some well known motifs.

Fig.1.2 β-Hairpin

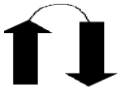

Fig.1.3. βαβ

C. Monte carlo simulation:

It is a common methodology to compute pathways and thermodynamic properties of protein.Simualtion runs in series of steps. These simulations can be long a whole day.The greatest difficulty in this case is when atom pair are closer than some cutoff distance[5] MCS of proteins can be performed in many different ways. With improved rithms and faster computers, it is becoming computationally feasible to simulate how small proteins fold to their native ates.This is an exciting development, which will lead to a better understanding not only of protein folding, but also of protein aggregation and of the interaction of proteins with other molecules and materials. *The sequences that the current version of the model is able to fold have about 20 amino acids, and there are known examples of sequences of this size that the model fails to fold.*

	Difference between various methods		
	Homology based	**Graph based**	**Monte carlo simulation**
1	Based on existing template.	Protein treated in the form of graph.	It uses metropolis algo.
2	FASTA software is used.	Need of programmin g.	Uses random number of steps
3		Used for molecular biologist	Used for biolgis as well as pharmoclogy
4		To understand how molecular m/c work.	

D. Plan Work

Protein structure prediction is very useful for biologist and pharmacologist. In current scenario the number of application for protein structure predition are linux based and need to pass command line argument.

In PROFASI, we have to pass following command at command line.Consider example of Villin headpiece,

1) $profasi_dir/bin/BasicMCRun -ac 1 " PTKLETFPLDVLVNTAAEDLPRGVDPSRKENHLSDEDFKAVFGMTRSAFANLPLWKQQNLKKEKGLF * NH2" -T "300 Kelvin" -ncyc 10000 -nrt 1000 -iavg 1000.*

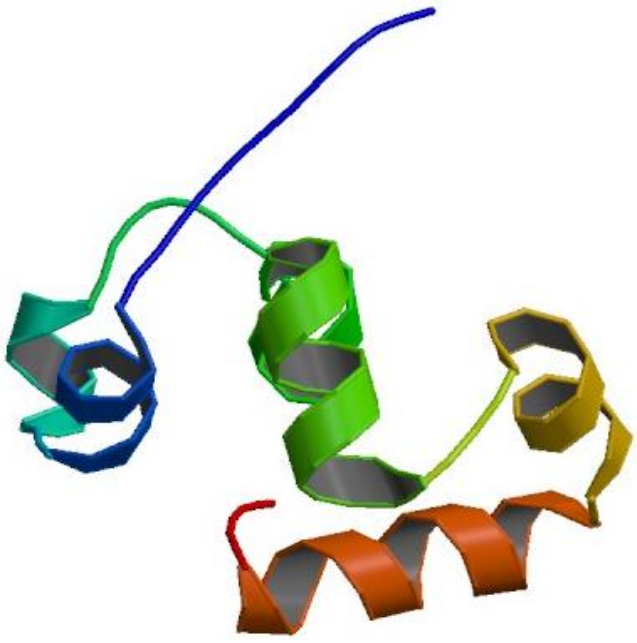

Fig: Villin headpiece structure

After applying MonteCarlo simulation to Villin Headpiece

We get the result like fig1.4

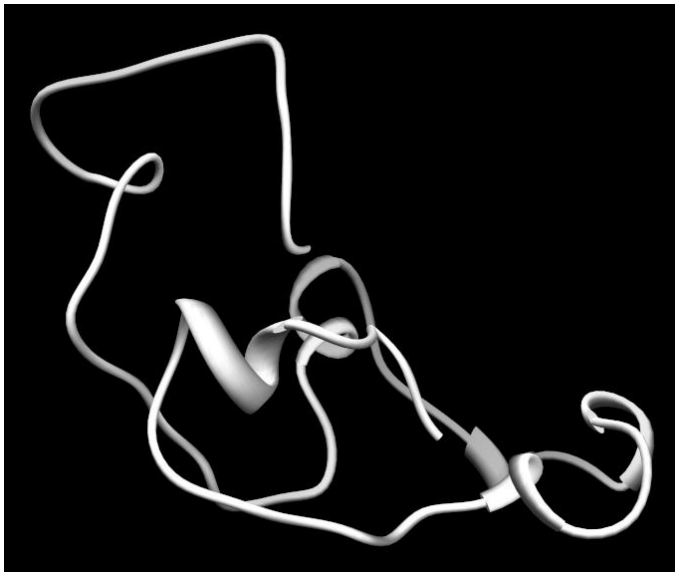

Fig.1.4:Minimum energy structure of Villin headpiece

Fig 1.5 shows how our architecture will work for protein structure prediction

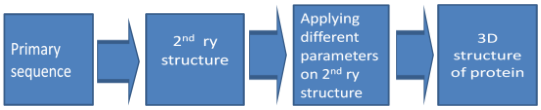

Fig1.5: Hybrid Approach for protein Structure Prediction

1)The input to system is a 1 D sequence protein.It consist of any 20 different amino acids, also referred as residues. Number of residues vary from protein to protein. The protein size vary from lower limit of 40-50 residues to several hundred residues.We can give this input either mannualy or from FASTA software.

2)The input given to FASTA program is linked with a s/w named psipred, where processing on 1D has taken place to produced a 2-D array with specifying which part is alpha helix,beta-sheet, and turns or loops.

3)Here is an actual implementation part where we are going to try number of inputs with trial and error method to check the output produced by profassi on 1D is correct with the input given by us in the 2D format.

This applies for number of inputs inorder to come on certain conclusion and based on conclusion we are going to design a program/algorithm to produced 3 dimensional structure of protein.

4)This algorithm acts as an interface between 2D and 3D.There are the drawbacks with rosetta and other s/w to produce 3-D structure for more than 100 residues.That will be overcome by using this pipelined.

III. CONCLUSIONS

We are going to implement hybrid approach for protein structure prediction that will give us Tertiary structure of protein for long sequence of input. This application can be held as service provider for large input sequence where Rosetta software and many other software are lagging. This can be held as web portal for giving GUI result of 3 dimensional.

References

[1] Georgina Mirceva, Ivana Cingovska, Zoran Dimov, and Danco Davcev," Efficient Approaches for Retrieving Protein Tertiary Structures" ieee/acm transactions on computational biology and bioinformatics, vol. 9, no. 4, july/august 2012

[2] G. Mirceva, S. Kalajdziski, K. Trivodaliev, and D. Davcev, "Comparative Analysis of Three Efficient Approaches for Retrieving Protein 3D Structures," Proc. Cairo Int'l Biomedical Eng. Conf.(CIBEC), pp. 1-4, Dec. 2008.

[3] Yan Yan1,2, Shenggui Zhang1, Fang-Xiang Wu2 ," Applications of graph theory in protein structure identification", From International Workshop on Computational Proteomics Hong Kong, China. 18-21 December 2010.

[4] saraswathi vishveshwara,_ k. v. brinda and n. kannan," protein structure: insights from graph theory", journal of theoretical and computational Chemistry, Vol. 1, No. 1 (2002) World Scientific Publishing Company.

[5] Anders irback, sandipan mohanty PROFASI: A Monte Carlo Simulation Package for Protein Folding and Aggregation , Wiley InterScience

[6] "The PSIPRED Protein Structure Prediction Server", http://bioinf.cs.ucl.ac.uk/psipred/

[7] "an efficient monte carlo simulation", shttp://robotics.stanford.edu/~itayl/mcs/

[8] " PROFASI (PROtein Folding and Aggregation SImulator [I])" http://cbbp.thep.lu.se/activities/profasi/

