

## **A Survey of Three Dimensional Protein Structure Prediction**

<sup>1</sup>R.Jayanthi II-M.Sc, <sup>2</sup>Dr.D.Ramyachitra Assistant Professor,  
<sup>3</sup>P.Lakshmi Ph.D Scholar.

*Department of Computer Science, Bharathiar University, Coimbatore.*  
Email: [jayanthikani27@gmail.com](mailto:jayanthikani27@gmail.com), [jaichitra1@yahoo.co.in](mailto:jaichitra1@yahoo.co.in), [visalaks@gmail.com](mailto:visalaks@gmail.com).

---

**Abstract:** Protein structure prediction is one of the most important goals pursued by bioinformatics, it is highly important in medicine (drug design) and structural bio informatics. Three dimensional protein structures are determined by its coordinates X, Y and Z. Algorithms, methods, applications and databases with various techniques addressed to predict the three dimensional protein structures. The prediction methods of 3D protein structures are categorized into comparative, folding and a b initio prediction. By using the algorithms it determines and predicts protein structure from its amino acid sequences.

**Keywords:** Protein structure prediction, Methods, Dataset, Application.

---

### **I. Introduction**

The prediction of the three dimensional structure of proteins (tertiary) is one of the most significant research problems in Structural Bioinformatics.

Tridimensional structure has the lowest potential energy. Heuristic and Meta heuristic algorithm used to utilize the minimization algorithm to find the global minimum of the potential function [1].

Tertiary structure refers to spatial relationship of all amino acids in a three dimensional structure polypeptide. The tertiary structure is described by its atomic coordinates of X, Y and Z. In three dimensional structure protein folding problem is one of the long standing goals in bio informatics and computational biology [2].

Three dimensional structures are important because they tell us how amino acids that are nearby each other on a chain interact to form higher order structures [3].

With the help of algorithms and computational resources, it is useful to design the drug and proteomics [4].

Tertiary structure is complex irregular folding polypeptide chain in three dimensions, with the combination of amino acid interactions [7].

In 3D protein structure, a membrane protein gives insight into its function and the non-membrane protein known as globular proteins. Compare with the membrane proteins, non membrane proteins are easy to obtain by experimental methods [10].

Some of the methods and algorithms are used in Structural Bio informatics to predict Three Dimensional protein structure with current algorithms and computational resources it is possible to predict the native structures of relatively small proteins.

Three dimensional protein structure prediction methods discussed in section 2, Classification of protein structures are described in section 3, in section 4 Datasets are given; Application and Performance Measurements are explained in section 5, section 6 gives conclusion.

### **II. Methods for Three Dimensional Protein Structure Prediction**

Imperialist Competitive Algorithm is used to predict the Secondary structure of the protein. Compare with previous methods it considers an effective tool and the revolution change adaptive procedure is applied to predict protein structure. It may produce errors at the time of predict secondary structure of a protein [3].

Angle Probability List Combined with Distributed Knowledge based Genetic algorithm used to predict the three Dimensional Protein structure and their topology is compared. Its need more Requirements for new strategies to extracting data [5].

By knowledge based methods best results achieved with the experiments from CASP. The predicted protein structure compared with their topology and Angle probability list combined with the knowledge based genetic algorithm to find the solution from the computational experiment test [6].

A guided macro mutation operator and 2x2 HP energy based genetic algorithm used for global search to predict three dimensional protein structures [8]. Parallel hybrid genetic algorithm with parallel computing on the grid computation helps to predict the protein structure [9].

Machine learning algorithm with local evolutionary information combined pre PSSM (Pseudo Specific Scoring Matrix) with Random forest classifier method used to predict DNA binding in protein structure prediction [11].

SSREDN (Secondary Structure Recurrent Encoder Decoder Networks) applied to get Q3 and Q8 accuracy for Secondary structure prediction with complex machine learning method [12].

### Classification of protein structure prediction methods

There are three major theoretical methods classified to predict the protein structures, they are comparative modeling, fold recognition, and ab initio prediction.

#### Comparative modeling

The term comparative modeling is the proteins with similar sequences related with each other, measured in percentage. If the similarity is very high it's called "core regions" which comprised with the elements of alpha – helices and beta – sheets secondary structure framework.

#### Fold recognition or "threading"

Threading refers to match the sequences of known three dimensional structures into unknown structure with protein folds from the database.

#### Ab initio prediction

In Ab initio prediction divided into two components, it differentiate between the correct structures with incorrect ones, and explore the search method, finally the components are join together to derive the function to find native like structures.

#### Dataset

DNA binding protein sequences can be taken from Protein Data Bank (PDB: <http://www.rcsb.org/pdb/home/home.do>). Sequences are downloaded from the server (<http://server.malab.cn/Local-DPP/Datasets.html>) [11].

Topology predictor of  $\beta$ -barrels called BOCTOPUS contains 36  $\beta$ -barrel proteins of known structures from OPM database. The position specific scoring matrix was generated with Z coordinate score with correlation status by PSI-BLAST, to search against Uniprot – SwissProt database .

A large non – homologues dataset contains 6128 and 513 protein amino acid sequences from Cull PDB and CB513 to evaluate with SSREDN (Secondary Structure Encoder Decoder Networks) to predict three dimension protein structure.

Transmembrane proteins can be used for three dimensional protein structure predictions from TOPCONS web server [10].

#### Performance Measurements

To evaluate DNA binding protein predictors LOOCV (Leave one out cross validation) test carried out for fair comparison with existing methods

Four evaluation metrics are: Sensitivity (SE), Specificity (SP), Accuracy (ACC), and Mathew's Correlation Coefficient (MCC)[11].

$$SE = (TP / (TP+TN)) * 100\% \text{ ----- (1)}$$

$$SP = (TN / (TN+FP)) * 100\% \text{ ----- (2)}$$

$$ACC = ((TP+TN) / (TP+TN+FN+FP)) * 100\% \text{ --(3)}$$

$$MCC = \frac{(TP \cdot TN - FP \cdot FN)}{\sqrt{(FP+FN)(TP+FP)(TN+FP)(TN+FN)}} \text{ ----- (4)}$$

Minimization measure of Root Mean Square Deviation (RMSD) and the Maximum measure of Global distance total score test (GDT\_TS) are used to analyze the 3D protein structures.

$$\text{RMSD}_{(a,b)} = \sqrt{(\sum_{i=1}^n ||r_{ai} - r_{bi}||) / n} \text{----- (5)}$$

Where  $r_{ai}$  and  $r_{bi}$  are the vectors representing the positions of the same atom  $I$  of the two structures  $a$  and  $b$  respectively.

$$\text{GDT}_{\text{TS}} = (\text{GDT}_{\text{P1}} + \text{GDT}_{\text{P2}} + \text{GDT}_{\text{P3}} + \text{GDT}_{\text{P8}}) / 4 \text{ ----(6)}$$

Where  $\text{GDT}_{\text{pn}}$  represents the percentage of residues under the cutoff distance  $\leq n \text{ \AA}$  [5].

### III. Conclusion

To find the native structure of a protein is a long standing goal in bioinformatics and computational biology. The proteins are classified based on the sequence of structures to be placed. The structure of protein with that applications and input sequence datasets are discussed. This paper can be use to understand concept of three dimensional protein structure prediction.

### References

- [1]. Schnell, J.R., Chou, J.J., 2008. "Structure and mechanism of the M2 proton channel of influenza A virus Nature" 451, 591–595.
- [2]. Erfan Khajia, Masoumeh Karamib, Zahra Garkani-Nejadc, n aGraduate Student, Department of Complex Adaptive Systems, University of Gothenburg, Gothenburg, Sweden Department of Bio-Chemistry,
- [3]. Iran Military University of Medical Sciences, West Fatemy St, North Karegar Ave, Tehran, Iran,"3D protein structure prediction using Imperialist Competitive algorithm and half sphere exposure prediction" E. Khaji et al. / Journal of Theoretical Biology 391 (2016) 81–87.
- [4]. BorkoBoskovic , Janez Brest , Faculty of Electrical Engineering and computer science , University of Maribor Slovenia , " Genetic algorithm with advanced mechanisms applied to the protein structure prediction in a hydrophobic – polar model and cubic lattice" . Applied Soft computing 45(2016) 61-70.
- [5]. Leonardo Correa, Bruno Borguesan, CamiloFartan, Mario Inostroza-Ponta, and Marcio Dorn, L. Corr` ea, B. Borguesan, and M. Dorn , the Institute of Informatics, Federal University of Rio Grande do Sul, Porto Alegre, Brazil. C. Farfán and M. Inostroza-Ponta are with the Dep. de IngenieriaInformática, Center for Biotechnology and Bioengineering, U de Santiago,Santiago, Chile."AMemetic Algorithm for 3-D Protein Structure Prediction Problem", Ieee/Acm Transactions On Computational Biology &Bioinformatics, Vol. 15, No. 1, April 2016.
- [6]. Alfonso E. Márquez-Chamorroa,\* , GualbertoAsencio-Cortésa,Cosme E. Santiesteban-ocab,
- [7]. Jesús S. Aguilar-Ruiza a School of Engineering, Pablo de Olavide University, Seville,SpainbCentro de Bioplasmas, University of Ciego de Ávila, Cuba "Soft computing methods for the prediction of protein tertiary structures: A survey",A.E. Márquez-Chamorro et al. Applied Soft Computing 35 (2015) 398–410 1568-4946/© 2015 Elsevier B.V. All rights reserved.
- [8]. Mahmood A. Rashida,d,\* , SumaiyaIqbalb, FirasKhatibc, MdTamjidulHoqueb, Abdul SattardaSCIMS, University of the South Pacific, Laucala Bay, Suva, Fiji bCS, University of New Orleans, LA, USA cCIS, University of Massachusetts Dartmouth, MA, USA dIIS, Griffith University, Brisbane, QLD, Australia . "Guided macro-mutation in a graded energy based genetic algorithm for protein structure prediction " ,M.A. Rashid et al. / Computational Biology and Chemistry 61 (2016) 162–177 ,1476-9271/© 2016 Elsevier Ltd. All rights reserved.
- [9]. A.-A. Tantara, N. Melaba, E.-G. Talbia, B. Parentb, D. HorvathbaLaboratoire d'Informatique Fondamentale de Lille, LIFL/CNRS UMR 8022, DOLPHIN Project - INRIA Futurs, CitéScientifique, 59655 - Villeneuve d'AscqCedex, France bCNRS UMR8576, Université des Sciences et Technologies de Lille, B` atiment C9, CitéScientifique 59655 - Villeneuve d'AscqCedex, France , "A parallel hybrid genetic algorithm for protein structure prediction on the computational grid",Future Generation Computer Systems 23 (2007) 398–409
- [10]. R.Shamima and S.Suresh , School of Computer Engineering ,Nanyang Technological University, Singapore , "Prediction of Membrane protein Structures using a projection based Meta-cognitive Radial basis function Network",International joint conference on neural networks ,978-1-5090-0620-5/16/\$31.00@2016IEEE.
- [11]. Leyiwei ,Jijin Tang ,QuanZou , School of computer science and technology,Tianjin , China,Department of Computer science and Engineering ,University of southcarolina , USA , 'Local DPP : "An improved DNA binding protein prediction method by exploring localevolutionaryinformation,"http://dx.doi.org/10.1016/j.ins.2016.06.026 0020-0255/© 2016 Elsevier Inc.
- [12]. YangxuWanga, HuaMaoa, Zhang Yia Machine Intelligence Laboratory, College of Computer Science, Sichuan University,Chengdu 610065, People's Republic of China,"Protein Secondary Structure Prediction by using Deep Learning..Method"II:S09507051(16)30471-3 DOI: 10.1016/j.knosys.2016.11.015 .