*IOSR Journal of Pharmacy and Biological Sciences (IOSR-JPBS)*
*e-ISSN: 2278-3008.Volume 5, Issue 6 (Mar. – Apr. 2013), PP 38-43*
*www.iosrjournals.org*

# Protein structure determination of insulin of zebra fish (Danio rerio) by homology modeling and structure analysis

Sudipta Mondal[1], Amal Kumar Bandyopadhyay[2], Rifat Nawaz Ul Islam[3]
[1(]*Department of Biotechnology, The University of Burdwan, India)*
[2(]*Department of Biotechnology, The University of Burdwan, India)*
[3(]*Department of Biotechnology, The University of Burdwan, India)*

***Abstract:*** *The protein sequence of insulin of zebra fish is obtained from UniProt. Due to lack of their structure, structure prediction is necessary, because the structure of protein plays an important role in their function. Our work is based on the production of two protein structure, from the same sequence, by computational approach and finally validates these generated structures. In this work two different widely acceptable online web tool are used for generating structure from the protein sequences of insulin of zebra fish. These are Swiss Model web server and ESyPred3D web server. After getting structure from this two web tool, the structures are passed by a series of quality tests. ProQ web software is used for checking quality of these generated structures. 3d-ss web tool is used for superimposition between two generated structures. It can compare between two structures. The Ramachandran plot is calculated by using VegaZZ software. CASTp (Computer Atlas of Surface Topology of protein) is a web tool, used to predict active sides with their respective volume and area. Finally ProFunc tool is used for analysis of two structures.*

***Key-Words:*** *CASTp, Homology Modeling, ProQ, VegaZZ, Zebra fish insulin*

## I. Introduction

Insulin is a peptide hormone, produced by beta cells of the pancreas, and is central to regulating carbohydrate and fat metabolism in the body. Insulin causes cells in the liver, skeletal muscles, and fat tissue to absorb glucose from the blood. In the liver and skeletal muscles, glucose is stored as glycogen. Protein structure plays an important role in their function. There are many sequence of this protein are present in UniProt database. But due to their lack of structure, homology modeling is necessary. But this work is based on the insulin of the zebra fish.

Homology modeling, also known as comparative modeling of protein, refers to constructing an atomic-resolution model of the "target" protein from its amino acid sequence and an experimental three-dimensional structure of a related homologous protein (the "template"). Homology modeling relies on the identification of one or more known protein structures likely to resemble the structure of the query sequence.

Evolutionarily related proteins have similar sequences and naturally occurring homologous proteins have similar protein structure. It has been shown that three-dimensional protein structure is evolutionarily more conserved than would be expected on the basis of sequence conservation alone. The sequence alignment and template structure are then used to produce a structural model of the target. Because protein structures are more conserved than DNA sequences, detectable levels of sequence similarity usually imply significant structural similarity. The quality of the homology model is dependent on the quality of the sequence alignment and template structure.

In this work two widely acceptable structure prediction tools are used, these are Swiss Model and ESyPred3D. Both are runs in windows environment. The structures are then validating using ProQ. Superimposition is done by 3d-ss.

Ramachandran plot (also known as a Ramachandran diagram or a [φ,ψ] plot), originally developed in 1963 by G. N. Ramachandran, C. Ramakrishnan, and V. Sasisekharan, is a way to visualize backbone dihedral angles ψ against φ of amino acid residues in protein structure. Ramachandran plot is calculated by VegaZZ software.

CASTp is used to predict active sides with their respective volume and area.

The ProFunc server ( http://www.ebi.ac.uk/thornton-srv/databases/profunc/index.html ) had been developed to help identifying the likely biochemical function of a protein from its three-dimensional structure.

## II. Materials And Methods

### *2.1 Obtaining Sequence*

To perform effective Bioinformatics on sequence and structure, one needs to collect sequence from highly annotated sequence database. The sequence of insulin of zebra fish (UniProt Id:-O73727) is collected from UniProt data base (http://www.uniprot.org/).

### *2.2 Sequence to 3D structure*

Two most effective software tools are used for generating 3D structure of insulin of zebra fish. These are Swiss Model (http://swissmodel.expasy.org/) and ESyPred3D (http://www.unamur.be/sciences/biologie/urbm/bioinfo/esypred/)

#### *2.2.1 Swiss Model*

Swiss Model software is one of the most reliable and effective tool in bioinformatics for generating structure from sequence. Swiss model has three modes for predicting structure from sequence (i.e. Homology Modeling). These are Automated Mode, Alignment Mode and Project Mode. Among these Automated mode is quite easier and reliable for homology modeling. Automated mode is selected for this work.

In the Swiss Model web site, the Email id, name of the protein and the sequence are entered. Then submit button is clicked. The template for this structure is selected automatically. The result is sending in the provided email id with direct link. Then the structure file (.pdb) is downloaded. QMEAN-Z score in the Swiss Model page is used for evaluation of generated 3D structure. The range of QMEAN-Z score lies between 0-1. After getting this structure, the protein is visualized in Rasmol software.

#### *2.2.2 ESyPred3D*

ESyPred3D is another valuable tool in bioinformatics for generating structure from sequence. Inside the ESyPred3D official web-site (http://www.unamur.be/sciences/biologie/urbm/bioinfo/esypred/), email address, description and the sequence in the input box, are entered. Then SUBMIT TO SERVER button is clicked. This server also sends the results in the provided email id. After getting response from ESyPred3D web server, the PDB (.pbd) and atomic coordinate (.ali) files are downloaded. Similarly the structure is visualized in Rasmol software.

### *2.3 Protein structure validation by using ProQ*

ProQ (http://www.sbc.su.se/~bjornw/ProQ/ProQ.html) is a neural network based predictor that based on a number of structural features predicts the quality of a protein model. ProQ is optimized to find correct models in contrast to other methods which are optimized to find native structures. Two quality measures are predicted, LGscore and MaxSub. It is software that checks the quality of two protein structure obtained from two different web servers, Swiss Model and ESyPred3D. If the generated structuress can satisfy the validation parameter, then it is said that the structures are good and ready for future analysis. In the PoQ- Protein quality predictor web site, ProQ web server button is clicked. In the next page the structure files are separately uploaded and finally submit button is clicked. Then we obtained Predicted LGscore and predicted maxsub values.

### *2.4 Superimposition*

In bioinformatics, structural superimposition is an effective tool of comparing two structures. 3d-ss (http://cluster.physics.iisc.ernet.in/3dss/) is used for structural superimposition of the two structures. Inside the 3d-ss web page, the option button is clicked and in the next page default parameters are keep constant. In this page, the structure obtained from Swiss Model is kept as fixed molecule and EsyPred3D structure is kept as rotated molecule.

### *2.5 Ramachandran plot calculation*

VegaZZ is the effective software for drug design, molecular dynamics simulation, and calculation of Ramachandran plot. The windows version of VegaZZ is downloaded from its official web-site (http://nova.colombo58.unimi.it/cms/index.php?Software_projects:VEGA_ZZ:Download). After downloading is over it is installed in windows environment. Then VegaZZ is opened, the two protein structures are separately opened, and calculate tab is clicked and finally Ramachandran plot option is clicked.

### *2.6 Structure Analysis by using CASTp*

CASTp (i.e. Computer Atlas of Surface Topology of protein) is a web tool is used to predict active sides with their respective volume and area. Inside the home page, (http://sts-fw.bioengr.uic.edu/castp/calculation.php) the structure files (.pdb) are separately uploaded and finally the submit button is clicked, and finally result is displayed in the next page.

***2.7 Function prediction by using ProFunc***

Inside the ProFunc homepage (http://www.ebi.ac.uk/thornton-srv/databases/ProFunc/)The structures are separately uploaded in this page. Then in the next page, name, institute/company, email address and protein name are entered and finally Run button is clicked. The result is sending in the provided email id. Finally the sequence motifs, matching folds data and 3D functional template are obtained.

## III. Results And Discussion

The protein structure generated by Swiss Model is visualized in Rasmol software is shown in Fig. 4.1 and the protein structure generated by ESyPred3D is visualized in Rasmol software is shown in Fig. 4.2. The structure generated from Swiss Model has QMEAN Z-Score = 0.93. QMEAN Z –Score is used to evaluate the generated 3D structure. The best score is ranges from 0 to 1. This QMEAN Z-Score is highly acceptable. Both structure has good LG score and Max Sub are shown in Table 4.1 obtained from ProQ. It indicates that the two structures are good model according to the ProQ validation criteria. The graphical view of structural superimposition is shown in Fig. 4.9.

The Ramachandran Plot calculation is shown in Table 4.2 and the Ramachandran plots are shown in Fig.4.3 and Fig.4.4. It indicates good result. That means both structures have good quality. The results of CASTp of the two structures are shown in the Fig. 4.5 and Fig. 4.6. From the result of CASTp server, the active site of the protein can easily identified. The results of ProFunc are shown in the Fig. 4.7 and Fig.4.8. In case of Swiss Model structure, four motifs are matched in scan against PROSITE, PRINTS, PFam-A, TIGRFAM, PROFILES and PRODOM motifs, 5869 significant structural matches are found and one significant Ligand-binding template is found in ProFunc result page. In case of EsyPred3D structure, nine motifs are matched in scan against PROSITE, PRINTS, PFam-A, TIGRFAM, PROFILES, 6710 significant structural matches are found seven significant Ligand-binding template is found in ProFunc result page. Both generated protein structure is passed a series of computational experiment. The both protein has good quality.

## IV. Figures And Tables

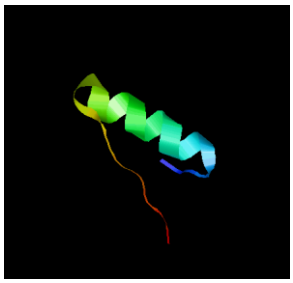

Fig.4.1 Structure of protein obtained from Swiss Model.

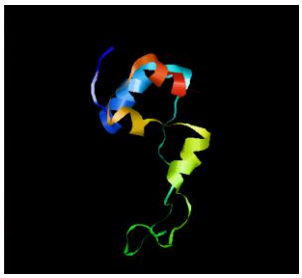

Fig.4.2 Structure of protein obtained from ESyPred3D.

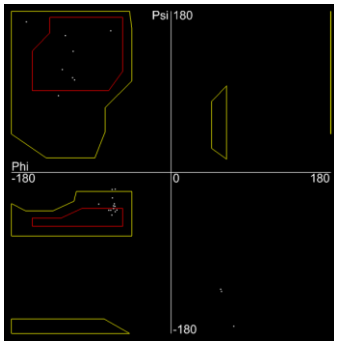


Fig.4.3 Ramachandran plot of Swiss Model protein.

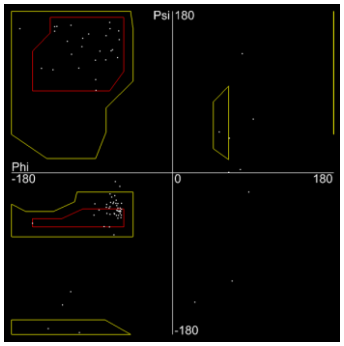


Fig.4.4 Ramachandran plot of ESyPred3D Model protein.

| SL NO. | STRUCTURE GENERATED FROM | LG SCORE | MAX SUB |
|---|---|---|---|
| 1 | Swiss Model | 0.753 | 0.120 |
| 2 | ESyPred3D | 1.134 | 0.186 |

Table 4.1 The LG SCORE and MAX SUB of two generated structure obtained from ProQ.

| SERIAL NO. | PROTEIN OBTAINED FROM | Residues inside the plot (red + yellow) | Residues without steric clashes (red) |
|---|---|---|---|
| 1 | Swiss Model | 73.08% | 38.46% |
| 2 | ESyPred3D | 82.35% | 54.12% |

Table 4.2 Ramachandran plot values of two model protein.

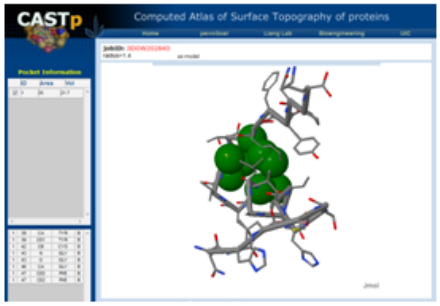


Fig.4.5 CASTp result of Swiss Model Protein. The green color indicates the active side.

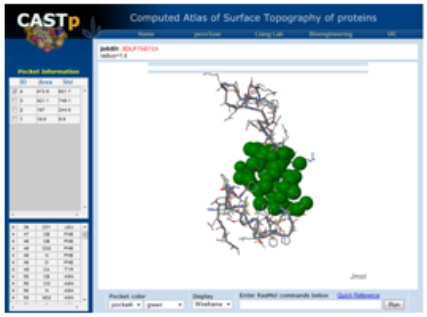


Fig.4.6 CASTp result of ESyPred3D Protein. The green color indicates the active side.


Fig. 4.7 The sequence motifs, matching fold and 3D functional template search results of Swiss Model structure obtained from ProFunc result page.

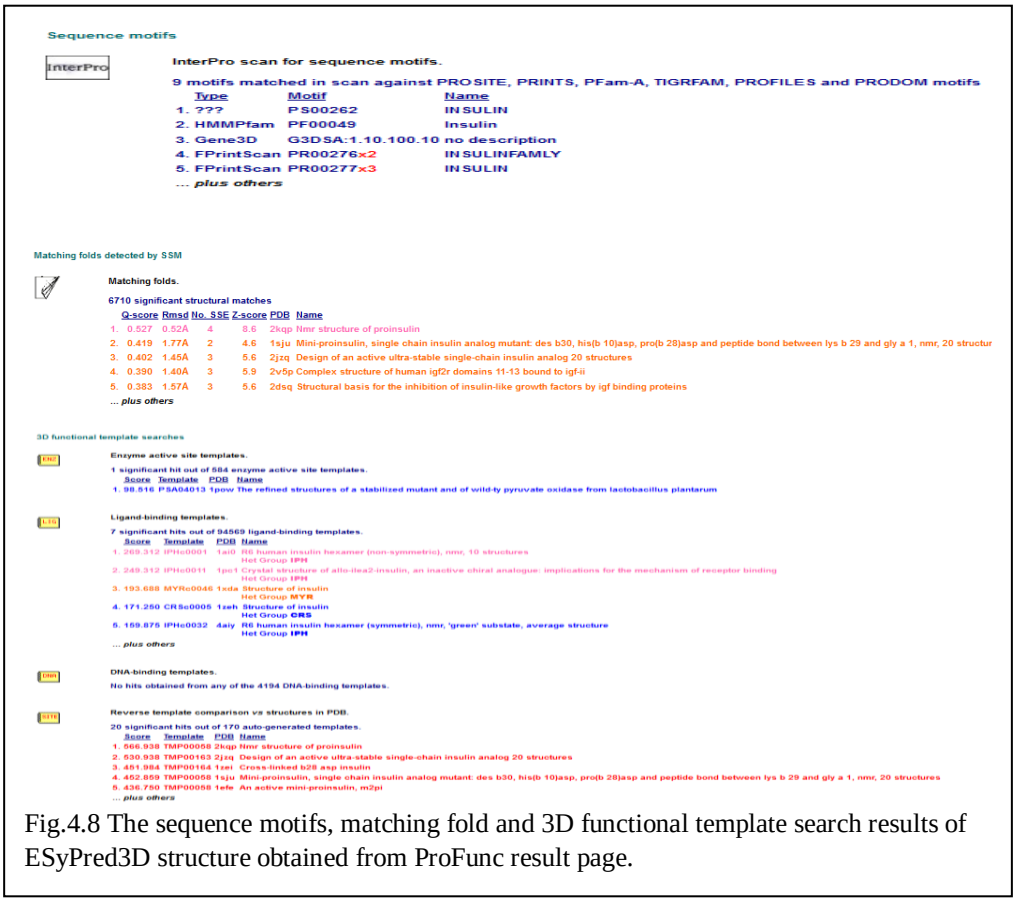


Fig.4.8 The sequence motifs, matching fold and 3D functional template search results of ESyPred3D structure obtained from ProFunc result page.

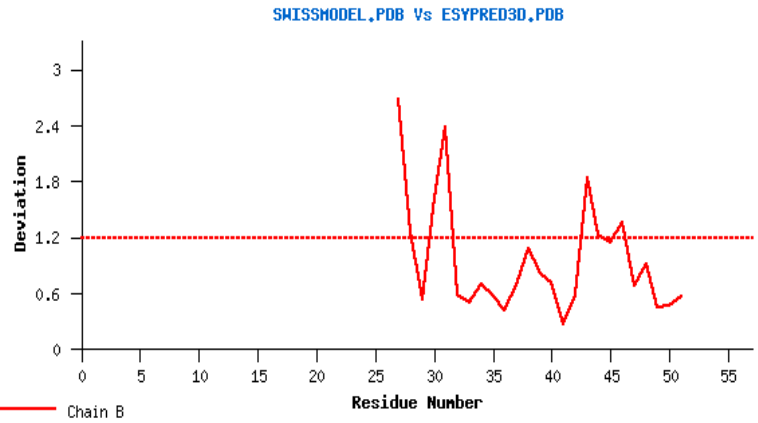


Fig.4.9 Graphical result of 3d-SS, this graph is plotted against between deviation and residue number .

## V. CONCLUSION

Now a day, Homology Modeling is very much effective tool in Bioinformatics and Biotechnology. Homology Modeling is the computer based online structure production tool. Experimentally the protein structures are generated from X-Ray Crystallography and Nuclear Magnetic Resonance (NMR). But both process are very much time consuming, laborious, and require expertise person. But the computational Homology Modeling is quite easier in this case. The insulin of zebra fish has no structure in Protein Data Bank (PDB). The sequence of zebra fish insulin is obtained from UniProt and this sequence is entered in the two most widely accepted tools for getting structure. After getting structure, it is very much essential to know the condition of quality, topology of these proteins. After completion of a series of computation experiment, both obtained proteins are passed the quality test i.e. both structures have good quality.

## Acknowledgements

Authors are grateful for the computer facility laboratory of DBT, Government of India in the department of Biotechnology, The University of Burdwan.

## REFERENCES

[1] C Lambert, N Leonard, X De Bolle, E Depiereux, ESyPred3D: Prediction of proteins 3D structures, *Bioinformatics, 18(9)*, 2002, 1250-1256

[2] R A Laskowski, J D Watson, J M Thornton, ProFunc: a server for predicting protein function from 3D structure, *Nucleic Acids Res*., *33*, 2005, W89-W93.

[3 ] R A Laskowski, J D Watson, J M Thornton, Protein function prediction using local 3D templates. *J. Mol. Biol.*, *351*, 2005, 614-626.

[4] Björn Wallner and Arne Elofsson, *Protein Sci., 15(4),* 2005, 900-913

[5] K. Arnold, L. Bordoli, J. Kopp, and T. Schwede , A web-based environment for protein structure homology modeling, *Bioinformatics, 22*, 2006, 195-201.

[6] Joe Dundas, Zheng Ouyang, Jeffery Tseng, Andrew Binkowski, Yaron Turpaz, and Jie Liang, CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated resiudes, *Nucleic Acid Research, 34*, 2006 W116-W118.

[7] F Kiefer, K Arnold, M Künzli, L Bordoli, T Schwede, The SWISS-MODEL Repository and associated resources, *Nucleic Acids Research. 37*, 2009, D387-D392.

[8] S. Brindha, Sangzuala Sailo, Liansangmawii Chhakchhuak, Pranjal Kalita, G. Gurusubramanian and N.Senthil Kumar, *Science Vision,11(3)*, 2011,125-133