Protein 3D structure determination using homology modeling and structure analysis

S. Brindha[1], Sangzuala Sailo[1], Liansangmawii Chhakchhuak[1], Pranjal Kalita[3],
G. Gurusubramanian[2] and N. Senthil Kumar[1]

[1]*Department of Biotechnology,* [2]*Department of Zoology, Mizoram University, Aizawl 796 004, India*
[3]*Department of Information Technology, ICFAI University, Aizawl, India*

ABSTRACT

The protein sequence of aspartate aminotransferase of pig was retrieved from the Swiss-Prot database. The appropriate template for homology modeling was determined using Blastp. 3D structures were determined by homology modeling softwares such as Swiss Model and EsyPred3D which passed quality test by ProQ software and set for further analysis. The pockets determined by CASTp server for the predicted structures showed a significant difference in the pocket area and volume, which were due to structural deviation between the residues 30-40 found in the 3d-ss software. Both the structures were analyzed using ProFunc tool which showed different functions as they had different structures and active sites. Thus the structure plays a vital role in determining its function.

Key words: Aspartate aminotransferase; Blastp; CASTp; 3D structure; ProQ.

INTRODUCTION

The structure of the protein plays a vital role in determining its function.[1] The proteins with same structure and pocket area (active site) have been found to have different functions.[2] The proteins that perform similar function show different structure.[3] The structures are primary factor for any disease-related protein. There are four levels of protein structure: primary, secondary, tertiary and quaternary.

Primary structure is the linear sequence of amino acids. Secondary structure is local conformation of α-helixes, β-sheets and random coils. The angle between two adjacent amino acids is called torsion angle, which determines the twists/turns of the sequences resulting in Secondary structure. Tertiary structure (3D) results from the packing of secondary structural elements at a stable conformation. Quaternary structure is combination of more than one subunits or chains.

Today protein sequencing is increasing at exponential rate, till date more than 1 million proteins are sequenced and are available in the databases (SwissProt, Protein Research Foundation and Protein Information Resource). The only protein structure database is PDB (Protein Data Bank). Determining the 3D structure from a protein sequence is a dif-

ficult and challenging problem. Basically X-ray crystallography or nuclear magnetic resonance (NMR) techniques are used, which are expensive, time consuming and complex process. Therefore, computational methods/algorithms including homology modeling, threading, and *ab initio* are developed. Homology modeling is the most accurate. It is based on alignment to a known protein structure which is derived experimentally (template). If the similarity of the sequence is greater than 30% then it can act as template.

Aspartate aminotransferase (AAT) of pig's primary sequence is functionally similar with following functions (Swiss-Prot Annotation): a) plays a vital role in amino acid metabolism, b) significant for metabolic exchange between mitochondria and cytosol, c) aids cellular uptake of long-chain free fatty acids. The present work focuses on predicting the 3D structure of AAT using two homology modeling softwares - ESyPred3D and Swiss model, which runs on Windows operating systems, validating (ProQ), analyzing their active sites (CastP), comparing predicted structures for structural similarities and dissimilarities (3d-ss) and annotating the functions from the structure (ProFunc).

Materials and Method

Sequence to 3D structure using homology modeling

The sequence of AAT of pig was obtained from Swiss-Prot (http://expasy.org/sport/). The protein sequence was given as an input

Figure 1. Swiss-Prot webpage with aatm as input.

(Fig. 1). After the result was displayed (Fig. 2), the accession number P00506 was entered to retrieve the sequence. From the fasta link (Fig. 3), fasta sequence was obtained and copied (Fig. 4), it was saved in a notepad.
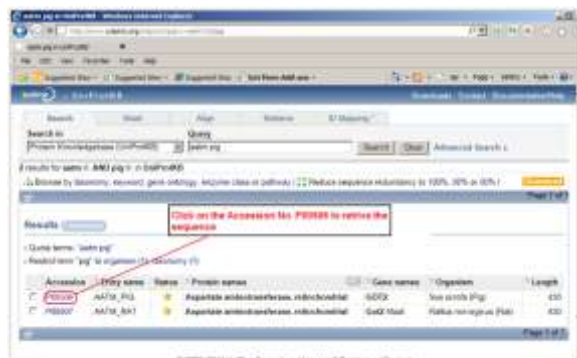

Figure 2. Result of the input aatm of pig.

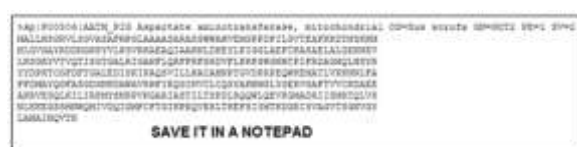
Figure 3. Complete information of accession no. P00506.

Figure 4. Sequence is displayed in the fasta format.

Selection of template

Blastp (http://blast.ncbi.nlm.nih.gov/Blast.cgi) was used to find template for homology modeling. The sequence similarity greater than 30% with known structure was selected as template. Sequence from the notepad was pasted in input box, Protein Data Bank was selected from the database and BLAST button

was entered (Fig. 5). From the Blastp result (Fig. 6) accession 3PD6_A had sequence similarity of 93% with known structure marked S in red color, which can be chosen as template for homology modeling.

Figure 5. BlastP with input and necessary parameters.

Figure 6. Results of BlastP.

Software – ESyPred3D

Software 1: ESyPred3D server predicts 3D structure based on homology modeling.[5]

Inputs: Email id, name of the protein, protein sequence and PDB ID of the template which is chosen from Blastp result (Fig. 7), after submitting, the results were sent to email id provided (Fig. 8), and the attachments of PDB (.pdb) and atomic coordinate (.ali) files were downloaded. The .pdb file was opened in the Rasmol viewer (Fig. 9), which can be downloaded from http://openrasmol.org/.

Figure 7. ESyPred3D webpage with corresponding inputs.

Figure 8. Predicted 3D structure result in the email.
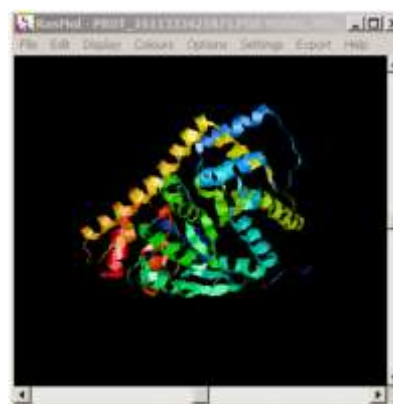
Figure 9. Rasmol viewer displaying prot_1611333421875.pdb

Software – Swiss Model Server

Software 2: Swiss model server (http://swissmodel.expasy.org/) predicts 3D structure based on homology modeling.[6,7] To input the sequence, Automated mode was selected (Fig. 10). Inputs: Email id, na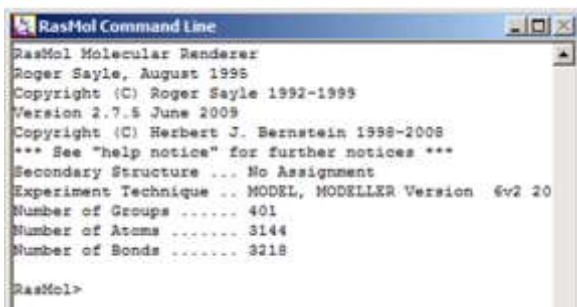me of the protein, and protein sequence. Alternatively, to speed up the process user can enter the template under the advance option and click Submit Modeling Request (Fig. 11). Significance of this model: Blastp will be done by the Swiss model server and from the result it chooses the template on its own. The quality of the model will be displayed along with the 3D structure predicted (Fig. 12). The results were sent to email with direct link the webpage. QMEAN Z-Score was used to evaluate the generated 3D structure.[8] The reliable score was between 0 and 1, i.e. 0.355, which is highly acceptable score. The predicted structure was visualized using Rasmol Software (Fig. 13).

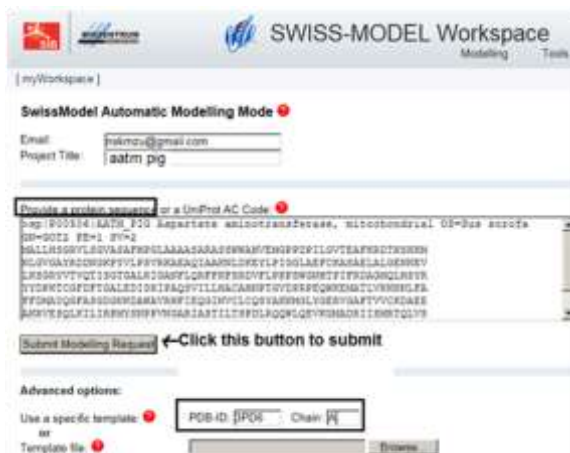

Figure 10. Swiss model webpage.

Figure 11. Inputs of the Swiss model webpage.

Figure 12. Re-directed page from the email link showing the predicted 3D structure.

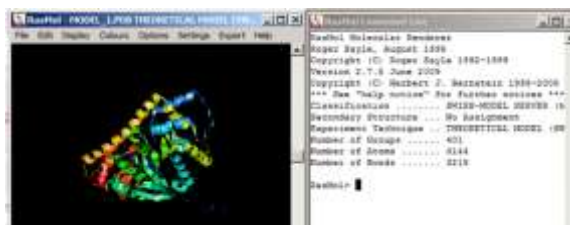Figure 13. Rasmol view of aatm_pig 3D structure modeled by Swiss model server.

Structure validation

ProQ-Protein Quality Predictor[9] is the software to check the quality of the model (http://www.sbc.su.se/~bjornw/ProQ/ProQ.html) (Fig. 14). If the predicted structure satisfies the validation parameters of ProQ process then the structure was taken for further analysis. The .pdb file predicted by

ESyPred3D uploaded (Fig. 15). The results of the structure file modeled by ESyPred3D is shown in Figure 16. Similarly the file with Swiss-Model was processed and the result is displayed in Figure 17.

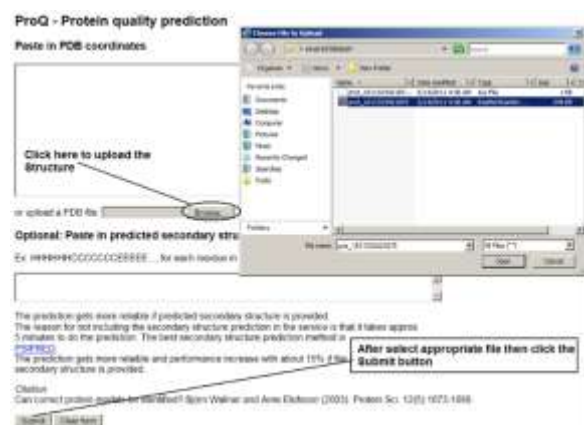
Figure 14. Webpage of the ProQ web server.

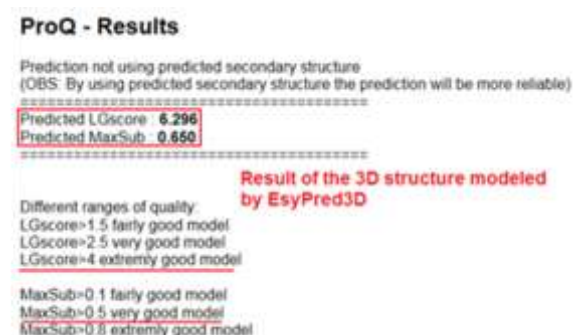Figure 15. Upload 3D structure in ProQ webpage.

Figure 16. 3D structure modeled by EsyPred3D.
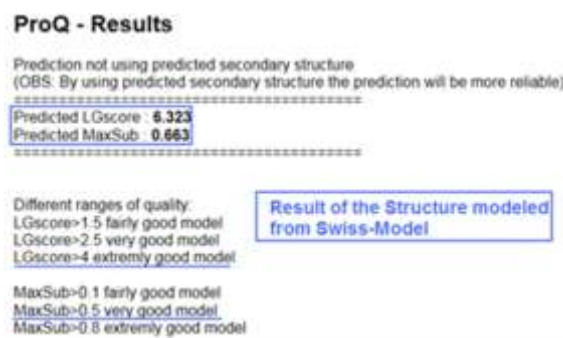
Figure 17. 3D structure modeled by Swiss model.

Structure analysis

CastP server[10] predicts the active sites with their respective volume and area (http://sts.bioengr.uic.edu/castp/calculation.php). The active pocket (active site) with more volume and area is always taken for further analysis. The .pdb file was uploaded to get the result (Fig. 18). Active site of the 3D structure predicted by EsyPred3D and Swiss-Model is shown in Figures 19&20, respectively, and the amino acids involved in pocket formation in both the models are displayed in Figure 21.

Structure comparison using superposition techniques

3d-ss (3-dimensional structure superposition) positions the 3D structure one above

Figure 18. CASTp web server.

Figure 19. Model predicted by ESyPred3D.
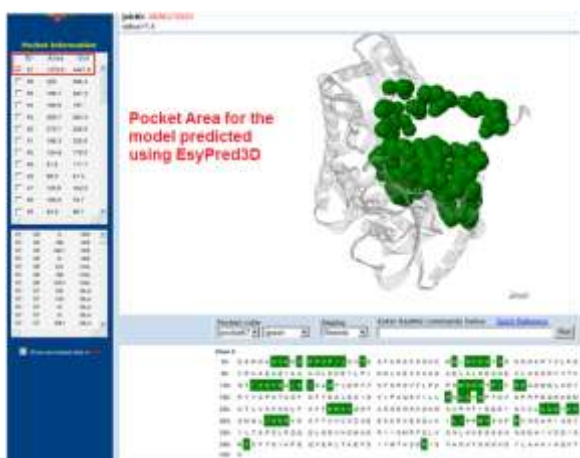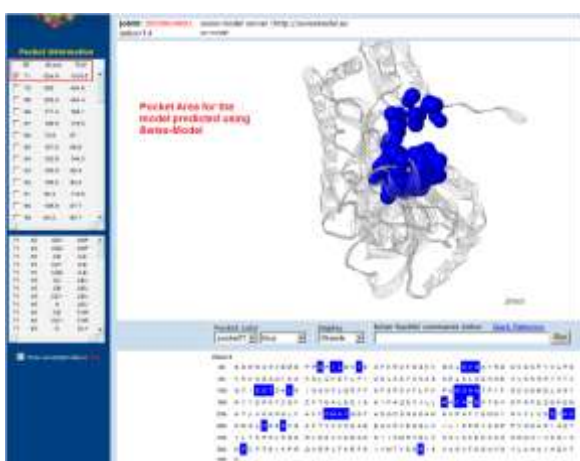

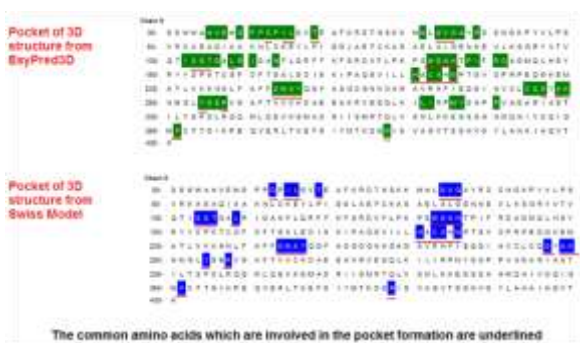Figure 20. Model predicted by ESyPred3D.

Figure 21. Pocket area of the 3D structure modeled by EsyPred3D and Swiss model.

another to find similar and deviated regions (http://cluster.physics.iisc.ernet.in/3dss/) between the structures.[8] Input was given as in Figure 22. There are only two structures for comparison. Hence, the option superimpose only two structures was selected and their respective pdb file were loaded (Fig. 23). All the options were kept in the default mode and proceeded as per Figures 24 to 27.

Here, both the protein structures modeled using two homology modeling techniques (Swiss model and EsyPred3D) were chosen, two ways of superposition were carried out as per the following criteria:

a. 3D structure by Swiss model was kept as fixed molecule and EsyPred3D as rotated molecule (Fig. 28a).

b. 3D structure by EsyPred3D model was kept as fixed molecule and Swiss model as rotated molecule (Fig. 28b).

Predicting function from structure

ProFunc identifies the likely functions of a protein pertaining to its 3D structure (http://www.ebi.ac.uk/thornton-srv/databases/

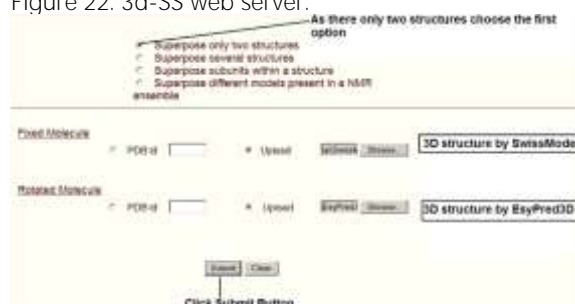Figure 22. 3d-SS web server.

Figure 23. Input of the structures.

Figure 24. Step 1 of 3d-SS result.

Figure 25. Step 2 of 3d-SS result.

Figure 26. Step 3 of 3d-SS result.

Figure 27. Step 4 of 3d-SS result.

profunc/index.html) analyzing the protein's sequence and structure by identifying functional motifs.[11] The .pdb file was loaded in the ProFunc online tool as in Figure 29. Preliminary information was entered (Fig. 30). The result was sent to the email provided. Similarly the pdb file modeled using Swiss

model was performed. The sequence motifs (Fig. 31), matching folds (Fig. 32), nest analysis (Fig. 33) and 3D functional template search results (Fig. 34a&b) were obtained.

Figure 28. Graphical result of 3d-SS showing deviation between residues 30-40.

RESULT AND DISCUSSION

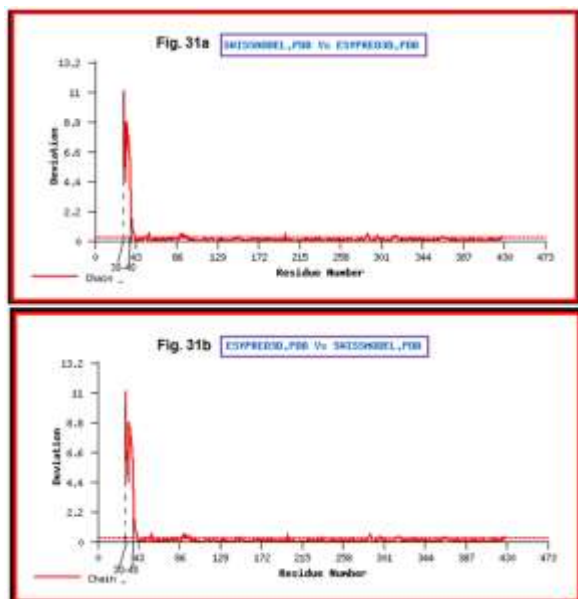Both the 3D structures modeled using ESyPred3D and Swiss model resulted with LG Score of 6.29 and 6.32, respectively, yielding to an extremely good model according to ProQ validation criteria (Fig. 16&17).

CASTp Server shows the pocket (active site) for ESyPred3D model: Area = 1379.9 and volume = 4401.9, respective, amino acids forming the packets are blocked in the sequence as highlighted in Figure 19. The
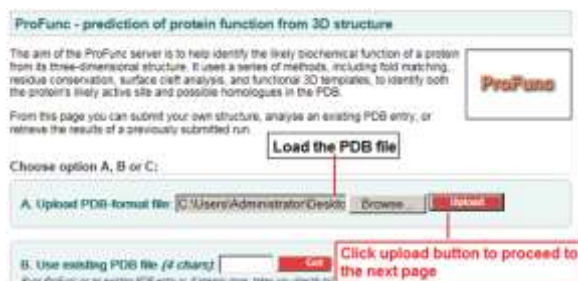

Figure 29. PDB file uploaded in the ProFunc tool.

Figure 30. ProFunc preliminary information.

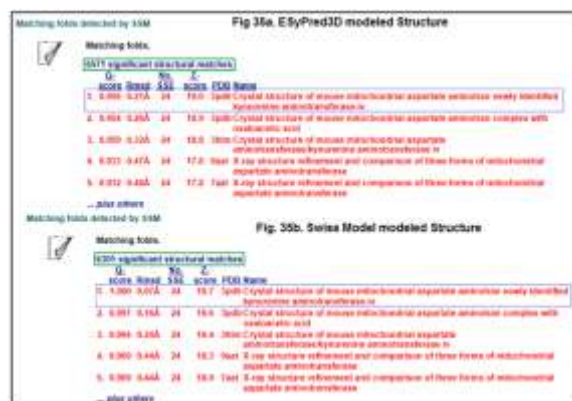Figure 31. Sequence motifs results of ProFunc.

Figure 32. Number of folds matching (predicted structure Vs PDB) - ProFunc software.
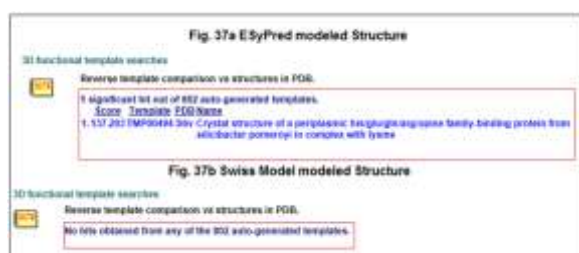
Figure 34. 3D functional template searches – ProFunc.

pocket for Swiss model: Area = 534.9 and volume = 1015.5, respective amino acids forming the packets are blocked in the sequence region (Fig. 20). The pocket size of 3D structure by Swiss model is aligning well with 3D structure by EsyPred3D (Fig. 21) because the pocket size of the 3D structure by Swiss model is smaller. The residue deviation from 30-40 (> 0.5 Armstrong) is high (Fig. 28), which plays a vital role in the pocket difference observed in the CASTp server (Fig. 21). Amino acids between 30-40 residues are involved in the pocket formation in the model determined by EsyPred3D which is not seen in structure predicted by Swiss model. But the residues from 43 to 430 remain less deviated (< 0.5 Å), which is negligible.

The results of sequence motifs of the ESyPred3D and Swiss model 3D structures reveal the presence of transaminase region in ESyPred3D model and presence of G3DSA motif in Swiss model structure (Fig. 31). There are more matching folds and nests in ESyPred3D model compared to Swiss model structure (Fig. 32&33). In reverse template comparison versus structures in PDB one hit/ function is observed only in ESyPred3D model (Fig. 34). The differences in the functions from protein structure are due to deviated site regions between two models which are observed in Figure 28.

3D structures of AAT of pig showed high deviation between residues 30-40, which affects the active site areas in the structure. The functions of the 3D structures are also in motifs, nests and reverse template comparison

versus structures. Further study is required to prove the functions in the wet lab.

ACKNOWLEDGEMENT

REFERENCES

1. Whisstock JC & Lesk AM (2003). Prediction of protein function from sequence and structure. *Q Rev Biophys*, 36, 307-340.

2. Wise E, Yew WS, Babbitt PC, Gerlt JA & Rayment I (2002). Homologous (β/α)8-barrel enzymes that catalyze unrelated reactions: orotidine 5'-monophosphate decarboxylase and 3-keto-L-gulonate 6-phosphate decarboxylase. *Biochemistry*, 41, 3861-3869.

3. Chan HS & Dill K (1993). The protein folding problem. *Physics Today*, 2, 24-32.

4. Sivakumar K (2006). Proteomics: from sequence to structure. *Advanced BioTech*, 4(11), 18-23.

5. Lambert C, Leonard N, De Bolle X & Depiereux E (2002). ESyPred3D: Prediction of proteins structures. *Bioinformatics*, 18, 1250-1256.

6. Arnold K, Bordoli L, Kopp J & Schwede T (2006). The SWISS-MODEL Workspace: A web- based environment for protein structure homology modelling. *Bioinformatics*, 22, 195-201.

7. Kiefer F, Arnold K, Künzli M, Bordoli L & Schwede T (2009). The SWISS-MODEL repository and associated resources. *Nucleic Acids Res*, 37, D387-D392.

8. Bjorn W & Arne E (2003). Can correct protein models be identified? *Protein Sci*, 12, 1073-1086.

9. Joe D, Zheng O, Jeffery T, Andrew B, Yaron T & Jie L (2006). CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic Acids Res*, 34, 116-118.

10. Sumathi K, Ananthalakshmi P, Md. Roshan MNA & Sekar K (2006). 3dSS: 3-dimensional structural superposition. *Nucleic Acids Res*, 34, 128-134.

11. Roman AL, James DW & Janet MT (2005). ProFunc: a server for predicting protein function from 3D structure. *Nucleic Acids Res*, 33, 89-93.