HIBISCUS

BIOREMEDIATION SCIENCE & TECHNOLOGY RESEARCH

Website: http://journal.hibiscuspublisher.com/index.php/BSTR

Preliminary Prediction of Lipase 3D Protein Structure and Function in *Rhodococcus* sp. NAM81 using Bioinformatics Approach

Hasdianty Abdullah*[1], Nor Suhaila Yaacob[2], Mohd Fadzli Ahmad[1], Farah Aula Mohd Fauzi[1] and Abdul Latif Ibrahim[2]

[1] Faculty of Science and Biotechnology, Universiti Selangor, Jalan Timur Tambahan, 45600 Bestari Jaya, Selangor, Malaysia

[2]Institute of Bio-IT Selangor, Universiti Selangor, Jalan Zirkon A7/A, Seksyen 7, 40000 Shah Alam, Selangor, Malaysia

*Corresponding Author: dianty@unisel.edu.my
Faculty of Science and Biotechnology,
Universiti Selangor,
Selangor, Malaysia
Tel: +603 32805123

HISTORY

Received: 29th of March 2015
Received In Revised Form: 30th of April 2015
Accepted: 30th of May 2015

KEYWORDS

Bioinformatics,
Protein structure prediction,
Lipase,
Rhodococcus sp.

ABSTRACT

Protein function depends greatly on its structure. Based on this principle, it is vital to study the protein structure in order to understand its function. This study attempts to build the predicted model of lipase gene in *Rhodococcus* sp. NAM81 using homology modelling method. The predicted structure was then used to investigate the function of protein through several bioinformatic tools. The DNA sequence of lipase gene was obtained from the *Rhodococcus* sp. NAM81 genome scaffold. Blastx analysis showed 100% identity to the target enzyme andthe appropriate template for homology modelling was determined using Blastp analysis. The 3D protein structure was built using two homology modelling software, EsyPred3D and Swiss Model Server. Both structures built obtained LGScore of greater than 4, which means they are extremely good models according to ProQ validation criteria. Both structures also satisfied the Ramachandran plot structure validation analysis. The predicted structures were 100% matched with each other when superimposed with DaliLite pairwise. This shows that both structure validation servers agreed on the same model. Structure analysis using ProFunc had found seven motifs and active sites that indicate similar function of this protein with other known proteins. Thus, this study has successfully produced a good 3D protein structure for the target enzyme.

INTRODUCTION

Lipases are ubiquitous enzymes that catalyze the breakdown of fats and oils producing free fatty acids, diacylglycerols, monoglycerols and glycerol. Lipases are also efficient esterification, transesterification and aminolysis [1]. Many microbial lipases have been commercialised, especially in detergents, papers production, cosmetics production, food flavouring and organic synthesis [2–4]. Lipases are valuable biocatalysts, because they act under mild conditions and are highly stable in organic solvents, showing broad substrates specificity [5–7].

In general, microbial lipases are 20-60 kDa proteins, with an active Ser residue of the active site structure Ser-His-Asp. Asp may be replaced by Glu in case of *Geotrichum candidum* lipases, which have the specificity for hydrolysis of fatty acids with cis-unsaturated double bonds [8]. The lipases belongs to the α/β hydrolases family [9] with a central β-sheet, containing the active Ser placed in a loop by the movement of a lid and exposure of hydrophobic pocket and the active site structure above critical micellar concentration (CMC) of the substrate [10], which is also responsible for its ability to hydrolyse fats and oils.

In this study, the main purpose is to use bioinformatic approach to predict 3D protein structure and function of lipase gene in *Rhodococcus* sp. NAM81. Bioinformatics is a novel approach in recent investigations on sequence analysis and

structure prediction of proteins [8]. The predicted model constructed from this study will be used for later analysis in understanding the lipase structural features in which will provide the knowledge of important characteristics for protein engineering. The significant structural data may be utilised to exploit the protein for industrial use and commercialisation.

MATERIALS AND METHODS

Obtaining the amino acid sequence of lipase gene

The lipase gene sequence obtained from *Rhodococcus* NAM81 genome scaffold was used throughout this study. The gene sequence was analysed using Blastx algorithm to attain the amino acid sequence. Blastx analysis was performed against non-redundant sequence database. The amino acid sequence with 100% identity was retrieved in FASTA format.

Preparation of the structure template

The amino acid sequence of lipase gene was blasted using Blastp algorithm against protein databank to search for the appropriate structure template. Blastp algorithm compared the amino acid sequence with similar sequence along with known structures in the Protein Databank. The structure with similarity that is greater than 30% with the amino acid sequence was selected as the template.

Prediction of the 3D structure using ESyPred3D and Swiss Model Server

The selected template was sent to ESyPred3D server [12] and swiss model server [13-15] to build the structure model. In this study, the template structure was sent to two different homology modelling servers to predict the structure as comparison.

Structure validation

The predicted structures were later sent to ProQ-Protein Quality Predictor [16] and the Ramachandran Plot [17] to check the quality of the model. ProQ is a neural network-based method to predict the quality of a protein model that extract structural features such as frequency of atom-atom contacts and predicts the quality of a model, which is measured by either LGscore or MaxSub. The difference between MaxSub and LGscore is the length of target protein. Correct model should have LGscore more than 1.5 and MaxSub more than 0.1 [16].

The Ramachandran plot shows the phi-psi torsion angles for all residues in the structure (except those at the chain termini). Glycine residues were separately identified by triangles, as they are not restricted to the regions of the plot appropriate to other sidechain types. The colouring/shading on the plot represents different regions described in [18]: the darkest areas (here shown in *red*) correspond to the "core" regions representing the most favourable combinations of phi-psi values. Ideally, one would hope to have over 90% of the residues in these "core" regions. The percentage of residues in the "core" regions is one of the best guides to stereochemical quality [19].

Structure analysis

Once validated, the structures were sent to DaliLite Pairwise to check for the similarities between the two structures predicted and for ProFunc server to predict their function based on the active sites and motifs. DaliLite was accessed at http://www.ebi.ac.uk/tools/structure/dalilite/. DaliLite e algorithm compared all chains in the first structure against all chains in the second structure. The superimposed coordinate file was later downloaded from the website.

RESULTS AND DISCUSSION

Obtaining the amino acid sequence of lipase gene

The contig containing lipase gene sequence obtained from the genome scaffold data of *Rhodococcus* nam81 is showed in **Fig.** 1. Meanwhile, **Fig.** 2 gives the result of Blastx. It shows the sequences from protein databases that matched with the query sequence. Blastx analysis gives good result with high percentage of similarity with the lipase gene sequences in the databases. The most similar sequence with 100% identity to the query was selected. The identical protein sequence is lipase (*Rhodococcus pyridinivorans* AK37), [Accession Number ZP_09310273.1]. The sequence was selected and used throughout the study for further analysis.

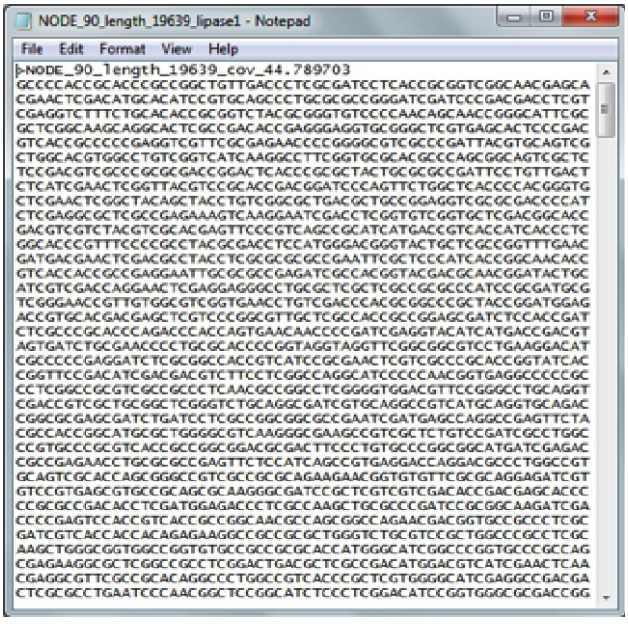

Fig. 1. Lipase gene sequence from *Rhodococcus* sp. NAM81 genome scaffold which is used as the query sequence for Blastx analysis.

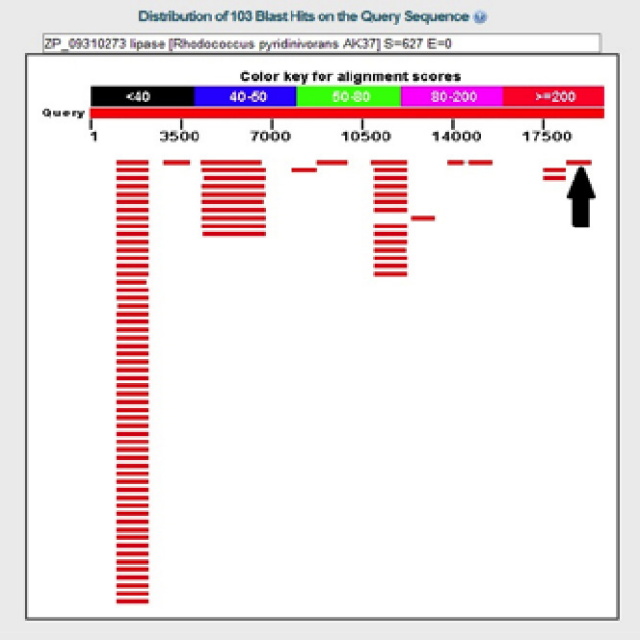

Fig. 2. Graphic display of Blastx analysis shows high score of hits. The arrow points to the region of lipase sequences from the database.

Preparation of the structure template

Using the Blastp analysis, the appropriate template for homology modelling was successfully obtained. 3DNM_A with the total score of 133 and 35% of max identity was selected as the template for the 3D protein prediction (data not shown). Blastp algorithm compares protein query with the protein sequences in the public databases and resulted with the most similar and significant sequences.

Prediction of the 3D structure using ESyPred3D and Swiss Model server

Rhodoccoccus sp. NAM81 lipase is a 46 kDA protein in a single chain monomer strand that consists of 431 amino acid residues. The 3D structures that were built using ESyPred3D and Swiss Model are given in **Fig.** 3. Generally this protein has structurally similar characteristics with other lipase, which is the existing of two types of secondary structure, known as alpha helices and beta sheets. The secondary structures of *Rhodoccoccus* sp. NAM81 lipase include 188 residues that created 16 alpha helices, shown in red, and 99 residues involved in beta sheets totalling 8 strands, shown in blue with a few random coil structures. The alpha helices were accounted for 44% of the protein, while beta sheets comprise 32%. The total number of negatively charged residues (Asp + Glu) was 35 whereas the total for positively charged (Arg + Lys) was 27 residues.

Structure validation

Structure validation analysis was performed to check the quality of the modelbuilt. The structure validation analysis was done using ProQ-Protein Quality Predictor and the Ramachandran Plot. **Fig.** 4 and 5 show the structure evaluations by ProQ-Protein Quality Predictor. The predicted LGScore for the EsyPred3D model was 7.143 (**Fig.** 4) and the predicted MaxSub was 0.731 (**Fig.** 5). Meanwhile for the Swiss Model, the predicted LGScore was 6.953 and the predicted MaxSub was 0.736. Both the predicted LGScore lies in the range of an extremely good model of predicted 3D protein structure, while their predicted MaxSub were highly acceptable score, which lies in the range of a good model.

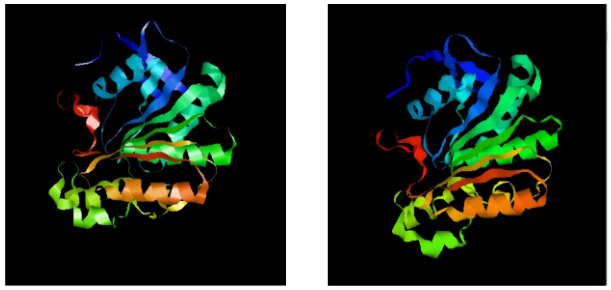

Fig. 3. Predicted structure built by a) ESyPred3D and b) swiss model server.

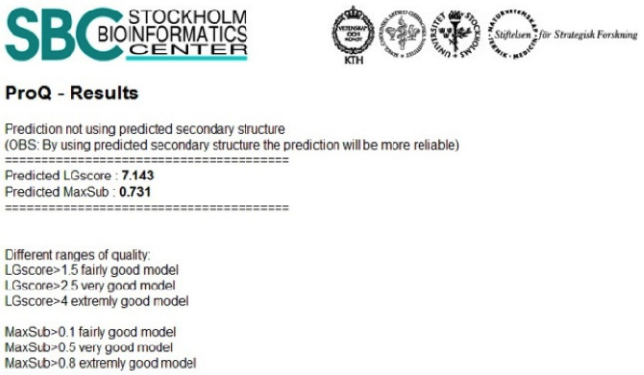

Fig. 4. ProQ Validation result for predicted structure by EsyPred3D server.

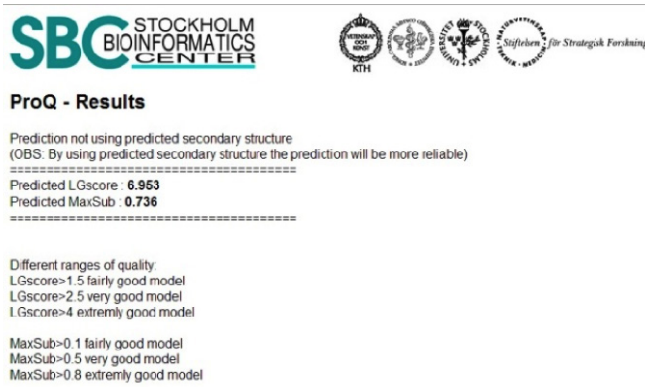

Fig. 5. ProQ Validation result for predicted structure by Swiss Model server.

Both of the 3D protein structures were also validated using Ramachandran plot. The result for the EsyPred3D model (**Fig.** 6) shows that the percentage of residue in favoured region was 96.7% with 0% of residue in outlier region. Meanwhile for the Swiss Model (**Fig.** 7), the percentage of residue in the favoured region was 87.4% with 1.3% of residue in outlier region. Results obtained from Ramachandran plot are also promising as most of the residues clustered tightly in the most favoured regions with very few at the outliers.

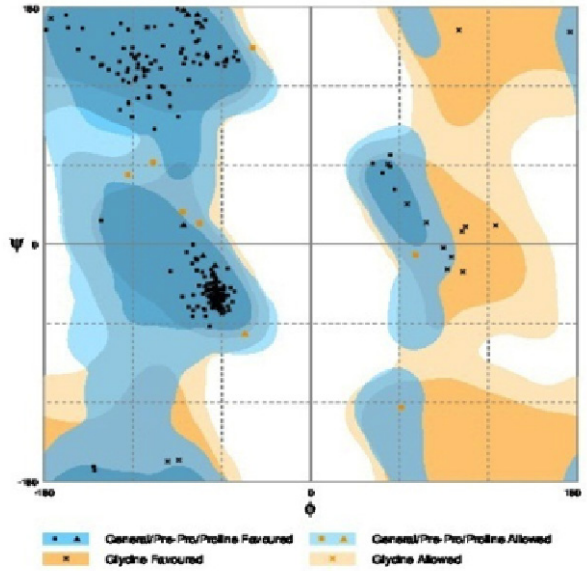

Fig. 6. Ramachandran plot for predicted structure by EsyPred3D server.

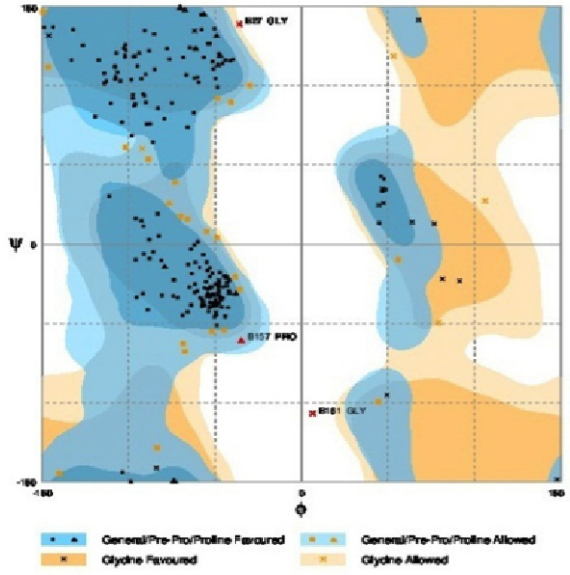

Fig. 7. Ramachandran plot for predicted structure by Swiss Model server.

Structure analysis

All the chains in the first structure against all chains in the second were compared using DaliLite pairwise. The superimposed structures in **Fig.** 8 shows exact match between the predicted model built by EsyPred3D and Swiss Model server, which indicates a good model of a predicted 3D protein structure. Although both structures were generated by two different servers, both servers agreed in the same model. This was further proved with the structure alignment result given in **Fig.** 9, which demonstrates a 100% match between both predicted structures.

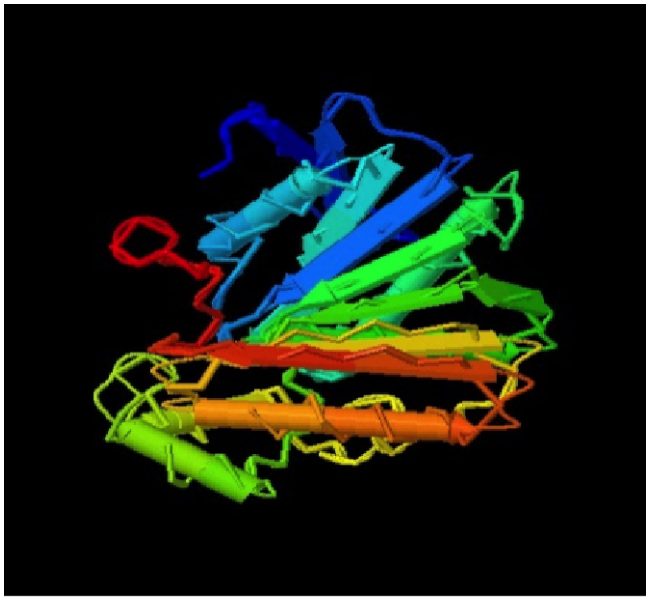

Fig. 8. The superimposed structures result from DaliLite pairwise.

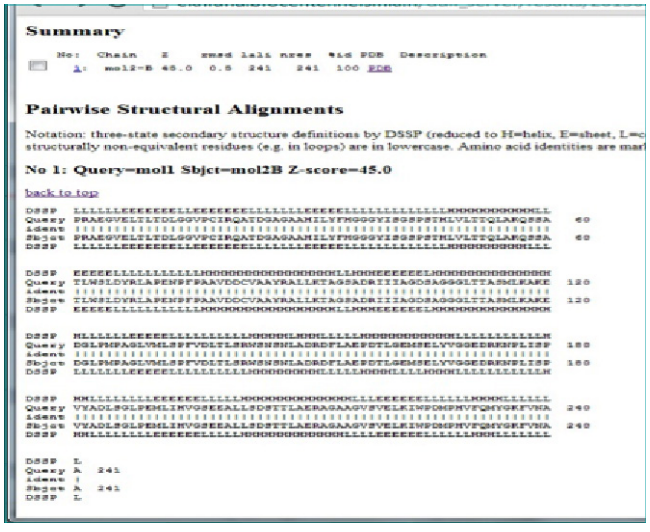

Fig. 9. Pairwise structural alignment result.

The predicted structures were then sent to ProFunc server to predict its function. ProFunc server is used to identify the likely biochemical function of a protein from its 3D structure. It uses a series of methods including fold matching, residue conservation, surface cleft analysis, and functional 3D templates to identify both the protein's likely active site and possible homologues in the pdb. The results are given in **Fig.** 10 and 11 in which ProFunc has successfully found 7 motifs and 7 active sites matched. Motif in proteins is conjectured to have biological significance. Protein sequence motifs are signatures of protein families and can often be used as tools for the prediction of protein functions [20].

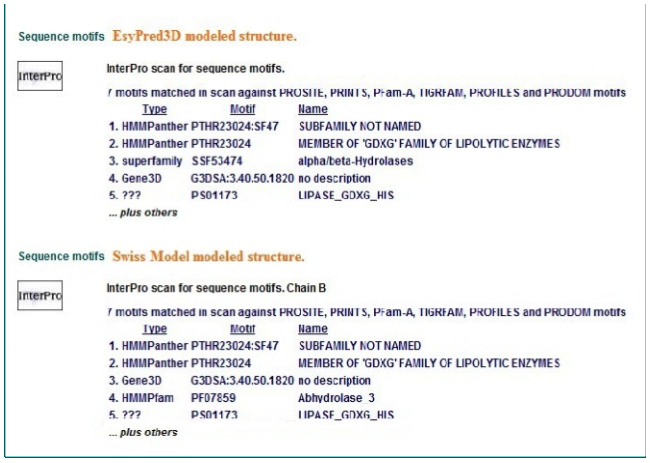

Fig. 10. ProFunc result for predicted structure by EsyPred3D server.

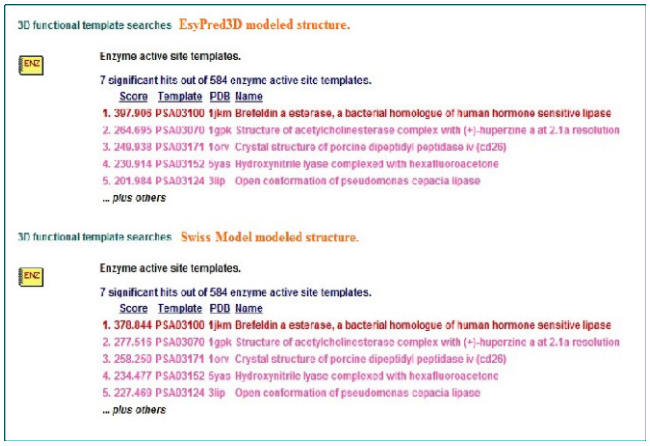

Fig. 11. ProFunc result for predicted structure by Swiss Model server.

CONCLUSION

The study has successfully produces a good predicted model for the lipase gene where the built structures had passed two validation tests, which are the ProQ-Protein Quality check and the Ramachandran plot. Nevertheless, further study is required to prove the function in the wet lab. Future work will be done in the analysis of amino acids sequence of this gene to study the uniqueness of this gene to its source organism.

REFERENCES

[1] Babu J, Pramod WR, George T. Cold active microbial lipases: Some hot issues and recent developments. Biotechnol Adv. 2008; 26:457–470.

[2] Kobayashi T, Koide O, Mori K, Shimamura S, Matsuura T, Miura T, Takaki Y, Morono Y, Nunoura T, Imachi H, Inagaki F, Takai K, Horikoshi K. Phylogenetic and enzymatic diversity of deep subseafloor aerobic microorganisms in organics- and methane-rich sediments off shimokita peninsula. Extremophiles. 2008; 12:519–527.

[3] Seiichi A, Akihiko Y, Mutsuo H. Occurrence of marine bacterial lipase hydrolyzing fish oil. Agric Biol Chem. 1991; 55:2657–2659.

[4] Chi Z, Chi Z, Zhang T, Liu G, Li J, Wang X. Production, characterization and gene cloning of the extracellular enzymes from the marine-derived yeasts and their potential applications. Biotechnol Adv. 2009; 27:236–255.

[5] Davidson MH. Mechanisms for the hypotriglyceridemic effect of marine omega-3 fatty acids. Am J Cardiol. 2006; 98:27–33.

[6] Amare G, Thomas D, Steffen BP, Per HN. Lipase and protease extraction from activated sludge. Water Res. 2003; 37:3652–3657.

[7] Pierre V, Jean MM, Jean G, Michael JH. Customizing lipases for biocatalysis: A survey of chemical, physical and molecular biological approaches. J Mol Catal B-Enzym. 2000; 9:113–148.

[8] Sandana MJG, Takeuchi S. Understanding structural features of microbial lipases – an overview. Anal Chem Insights. 2008: 3: 9–19.

[9] Ollis DL, Cheah E, Cygler M. The alpha/beta hydrolase fold. Protein Eng. 1992; 5:197–221.

[10] Svensen A. Lipase protein engineering. Biochim Biophys Acta. 2000; 1543:233–238.

[11] Lambert C, Leonard N, DE Bolle X, Depiereux E. EsyPred3D: Prediction of proteins 3d structures. Bioinformatics. 2002; 18(9):1250–1256.

[12] Arnold K, Bordoli L, Kopp J, Schwede T. The Swiss-Model workspace: a web-based environment for protein structure homology modelling. Bioinformatics. 2006; 22:195–201.

[13] Kiefer F, Arnold K, Künzli M, Bordoli L, Schwede T. The Swiss-Model repository and associated resources. Nucleic Acids Res. 2009; 37:387–392.

[14] Peitsch MC. Protein modeling by e-mail. Bio/technol. 1995; 13:658–660.

[15] Cristobal S, Zemla A, Fischer D, Rychlewski L, Elofsson A. A study of quality measures for protein threading models. BMC Bioinformatics 2001; 2(1):5.

[16] Ramachandran GN, Ramakrishnan C, Sasisekharan V. Stereochemistry of polypeptide chain configurations. J Mol Biol. 1963; 7:95–99.

[17] Morris AL, Macarthur MW, Hutchinson EG, Thornton JM. Stereochemical quality of protein structure coordinates. Proteins. 1992; 12: 345–364.

[18] Laskowski R. Main ramachandran plot. 1995. Accessed at http://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/pdbsum/getpage.pl?pdbcode=n/a&template=doc_procheck01.html on 1st December, 2014.

[19] Bork P, Koonin EV. Protein sequence motifs. Curr Opin Struct Biol. 1996; 6:366–376.