Evolutionary Bioinformatics

ORIGINAL RESEARCH

Three-Dimensional Structure of *Arabidopsis thaliana* Lipase Predicted by Homology Modeling Method

Abdelmonaem Messaoudi, Hatem Belguith and Jeannette Ben Hamida

Unité de Protéomie Fonctionnelle and Biopréservation Alimentaire, Institut Supérieur des Sciences Biologiques Appliquées de Tunis, Université Tunis El Manar, Tunisie. Corresponding author email: messaoudiabdelmonemster@gmail.com

Abstract: Triacylglycerol lipases have been thoroughly characterized in mammals and microorganisms. By contrast, very little is known about plant lipases. In this investigation, a homology model of *Arabidopsis thaliana* lipase (NP_179126) was constructed using a human gastric lipase (PDB ID: 1HLG), as a template for model building. This model was then assessed for stereochemical quality and side chain environment. Natural substrates: tributyrin, trioctanoin and triolen were docked into the model to investigate ligand-substrate interaction.

Keywords: comparative homology modeling, protein modeling, bioinformatics, lipase enzymes

Introduction

Lipases are glycerol ester hydrolases that act on triacylglycerols to release fatty acids and glycerol.[1] A major part of lipases used in today's industrial processes stems from microbial or animal sources. The oretically speaking, plant enzymes may have an advantage over animal or microbial enzymes due to their availability from natural sources, lower cost and apparent purification ease.[2] Lipases widely exist in microbe and plant species, but their structure is poorly characterized. Plant lipases are generally considered to be involved in particular in the regulation of plant growth and development. They are mainly located in seeds where triglycerides are stored in intracellular structures called oil bodies. Triglycerides are hydrolyzed by lipases to glycerol and fatty acids that provide energy needed for seed germination and seedling development.[3] Plant lipases can be classified into three major groups. The first group consists of the triacylglycerol hydrolases that are primarily present in seeds. Their study is of economic importance since they are largely responsible for seed alteration during storage. Lipases of the second group, called acylhydrolases, are present in various plant tissues. These enzymes exhibit little specificity for their substrate and are unable to hydrolyze triglycerides but can catalyze some trans-esterification reactions.[4] The main acylhydrolases are phospholipases A and B, glycolipases, sulfolipases and monoglyceride lipases. The third group consists of phosphorlipases C and D. Lipases are also involved in plant metabolism, rearrangement and degradation of chlorophyll during leaf growth and senescence as well as in fruit ripening process.[5] Recently, researches in this field became more and more attractive. Although several candidates from *Arabidopsis thaliana, Rauvolfiaserpentina, Heveabrasiliensis, Alopecurusmyosuroides* and *Brassica napus* have been extracted, cloned and characterized,[6–8] but such understanding of plant lipases is still limited.[9]

An updated and revised classification of family I bacterial "true" lipases mainly based on a comparison of their amino acid sequences and some fundamental physicochemical and biological properties, identified 11 subfamilies.[10] Although lipases belong to many different protein families, they have the same architecture, the α β-hydrolase fold as defined by Ollis et al.[11] Their activities rely mainly on a catalytic triad usually formed by Ser, His and Asp residues.[12] In amino acid sequences of α/β hydrolases, the three residues follow the order Ser-Asp-His. Also, lipases share a consensus sequence of Gly-Xaa-Ser-Xaa-Gly.[13] where X may be any amino acid residue.

Three-dimensional structure of proteins gives valuable insights into the molecular organization, function, docking simulations and also effective drug designing experiments. In the absence of an experimentally-determined crystal structure, homology modeling could provide a rational opportunity to obtain a reasonable 3D model. It is generally recognized that homology modeling of proteins is currently the most accurate method for 3D structure prediction, yielding models suitable for a wide spectrum of application, such as structure based molecular design and mechanism investigation.[30] This approach is able to provide a reasonable structure model with related template sharing more than 25% sequence identity. The present study attempts to propose a model of *Arabidopsis thaliana* lipase, a species widely used as a model organism in plant biology, due to its relatively small and genetically tractable genome.

Methods

Target and template proteins

With the aim of finding an adequate template for homology modeling of *Arabidopsis thaliana* lipase, sequence alignments of its amino acid sequence against Protein Data Bank (PDB[14]) were performed by means of the BLAST algorithm (the used default Blast parameters were: E cut-off = 10, mask low complexity = yes).[15] Sequence alignment shows that the target and the template (1HLG) share 31% of sequence identity. Because protein structures are more conserved than DNA sequences, detectable levels of sequence similarity usually imply significant structural similarity. Based on the significant e-value and alignment among the investigated templates, the human gastric lipase (PDB ID: 1HLG) was selected as a template to build a model for *Arabidopsis thaliana* lipase. The amino acid sequence of the target protein, (*Arabidopsis thaliana* lipase) was retrieved from LIPABASE database, a centralized resource database which provides taxonomic, physicochemical and molecular information about "true" lipase from different species.[16] (LIPABASE ID = 116, GenBank accession No. NP_179126) and is composed of 393 residues (Table 1). The selected template protein

Table 1. General, physicochemical and structural data of *Arabidopsis thaliana* lipase extracted from LIPABASE database.

General data	Physicochemical data	Structural data
Producer strain: *Arabidopsis thaliana* (thale cress) **Class:** True lipase **Reference:** El-kouhen K, et al (2005)	**Length:** 379 **Mass:** 44230.7 **Charge:** −15 **PI:** 5.61 **Gravy index:** 0.094 **Basic residue:** 25 **Acid residue:** 40 **Instability index:** 36.85: stable **Cystein:** 0.026 **Glycin:** 0.087 **Absent amino acid:** O, U **Common amino acid:** LSV **Aliphatic index:** 93.03 **Coefficient of extinction:** 72685 **Formule:** $C_{1867}H_{2846}N_{478}O_{527}S_{20}$	**Swiss prot Entry:** Q71DJ5 **NCBI Entry:** NP_179126 **3D structure:** Not resolved

was a human gastric lipase deposited in Protein Data Bank (PDB ID: 1HLG).[17] The template 1HLG with 371 Amino includes tow chains: A and B forming a dimmer. The resolution and R-value of the templates selected from PDB was 3°A and 0.210, respectively.

Homology modeling

Homology model was constructed using Swiss-PDB Viewer version 4.0.1.[18] This application provides a user friendly interface that allows analyzing several proteins at the same time. Homology modeling relies on the identification of one or more known protein structures likely to resemble the structure of the query sequence, and on the production of an alignment that maps residues in the query sequence to residues in the template sequence. Three different types of modeling requests can be made in the Swiss-PDB Viewer version 4.0.1program. They are automated mode, alignment mode and project mode. The automated mode is suited for cases where the target–template similarity is sufficiently high to allow for fully automated modeling. The alignment mode allows the user to test several alternative alignments and evaluate the quality of resulting models in order to achieve an optimal result. In difficult modeling situations where the correct alignment between target and template cannot be clearly determined by sequence-based methods, visual inspection and manual manipulation of the alignment can significantly help improve the quality of the resulting model. We used the project mode for obtaining the modeled

structure of *Arabidopsis thaliana* lipase (Accession No. NP_179126). Project files contain the superposed template structures and the alignment between the target and template files generated inside the program Deep View (Swiss-PdbViewer)[29] by the workspace template selection tools.

Model validation

To obtain an accurate homology model, it is very important that appropriate steps are built into the process to assess the quality of the model.[19] Therefore, accuracy of the predicted models were subjected through a series of tests. Stereochemical quality were evaluated using Ramachandran plots obtained from the RAMPAGE server[20] and amino acid environment was assessed using Verify 3D[21] and Errat[22] from the UCLA-DOE server.[23]

Docking

The lipase natural substrate was *triacylglyceride*. In order to validate the active site architecture of the *Arabidopsis thaliana* lipase model and examine its possible mode of interaction with the ligand, the triacylglyceride substrate was docked within the lipase homology model using the HEX v.5.1[24] docking environment at its default parameters. Hex is a tool for macromolecule docking and it can superpose pairs of molecules using only knowledge of their 3D shapes. Further, it is one of the few docking tools having in built graphic viewer.[25] This tool has been used in some earlier studies demonstrating

Figure 1. Ribbon representation of a three dimensional structure of *Arabidopsis thaliana* lipase enzyme in Rasmol version 2.7.5. display: cartoons, colors: structure.

ligand-protein interaction.[26] The approach was to use blind docking since it has been recommended for acquiring good results in prediction of substrate binding site.[31] Correlation type and post-processing output for receptor and ligand were kept based on shape, electrostatic potential and minimization of molecular mechanics (MM). Docking was carried out at full rotation allowing full flexibility for the ligand while keeping receptor position fixed in space.[32] Docking parameters involved Fourier transformation, steric scan, final search for ligand binding site and refinement of the complex.

Results and Discussion
Homology models and validation
Studies of Yang et al[27] demonstrated that a sequence identity higher than 25% between two proteins is indicative of similar three-dimensional structures. In our case, target and template protein share 32% of
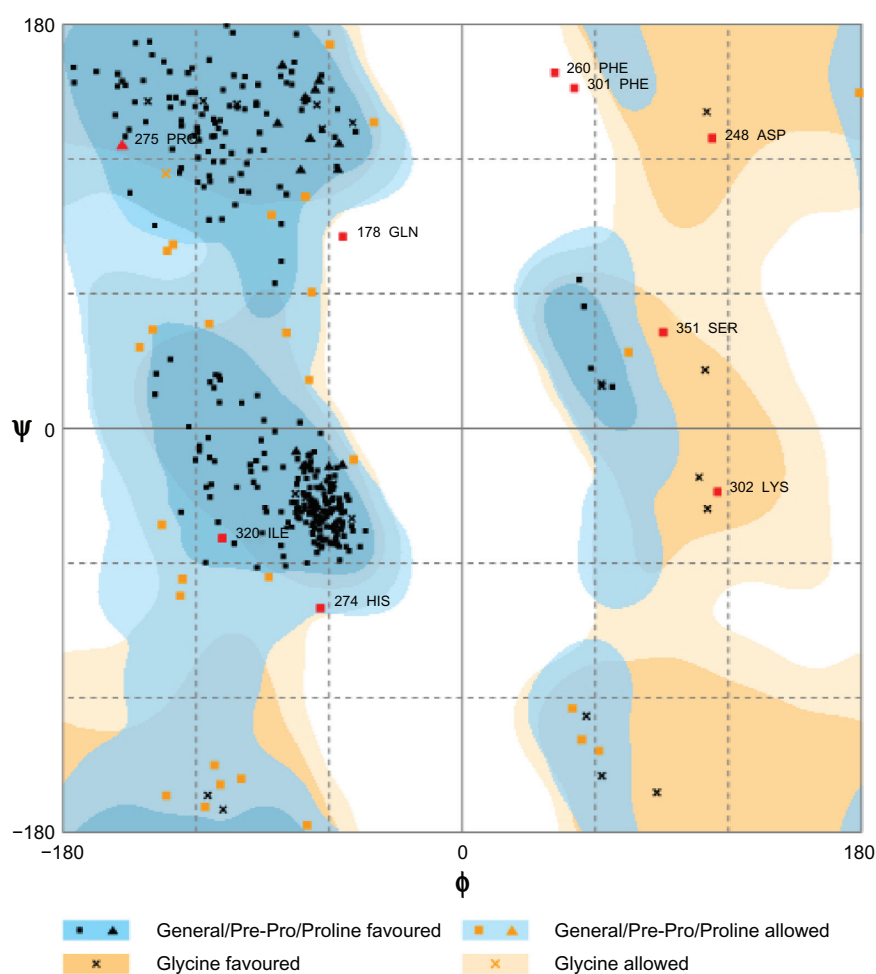
**Figure 2.** Ramachandran plot values showing number of residues in favoured, allowed and outlier region.
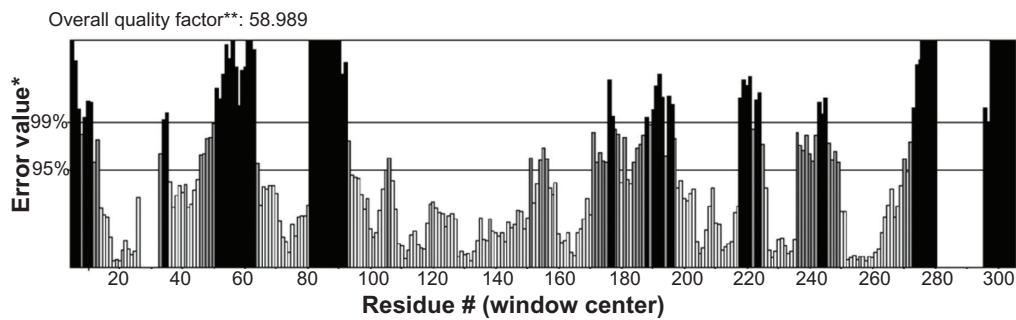
Figure 3. Errat plot for *Arabidopsis thaliana* lipase model. Black bars show the misfolded region located distantly from the active site, gray bars demonstrate the error region between 95% and 99%, white bars indicate the region having less error rate for protein folding.

sequence identity. Based on this alignment, a model of *Arabidopsis thaliana* lipase was built using a human gastric lipase (PDB ID: 1HLG) as a template for the model building (Fig. 1). The modeled enzyme is a monomer folded into α/β domain. It consists of eight central stranded β sheet flanked by twenty two α helices. The same structure is shared by other lipase enzymes from mammalian and bacterial origin, the number of α helices and β sheet differ from a species to another. The Ramachandran plot in Figure 2 indicated the region of possible angle formations by w (phi) and c (psi) angles. The conventional terms represent the torsion angles on either side of alpha carbon in peptides. The plot was divided into three regions: favored (91.5%), allowed (6%) and outlier (2.5%). This result is significant since the high percentage of residues in favored region (>90%). This indicates that the built model is of good quality. Errat plot assesses the arrangement of different types of atoms with respect to each other in protein models (Fig. 3). Errat is a sensitive technique,

which is good for identifying incorrectly-folded regions in preliminary protein models. ERRAT is a so-called "overall quality factor" for non-bonded atomic interactions, and higher scores mean higher quality. The normally-accepted range is >50 for a high quality model.[22] In the current case, the ERRAT score for the model is 58.989. The Verify 3D method assesses protein structures using three-dimensional profiles. This program analyzes the compatibility of an atomic model (3D) with its own amino acid sequence (1D). The scores range from −1 (bad score) to +1 (good score).[29] The verify score diagram presented in Figure 4 validate the *Arabidopsis thaliana* lipase model.
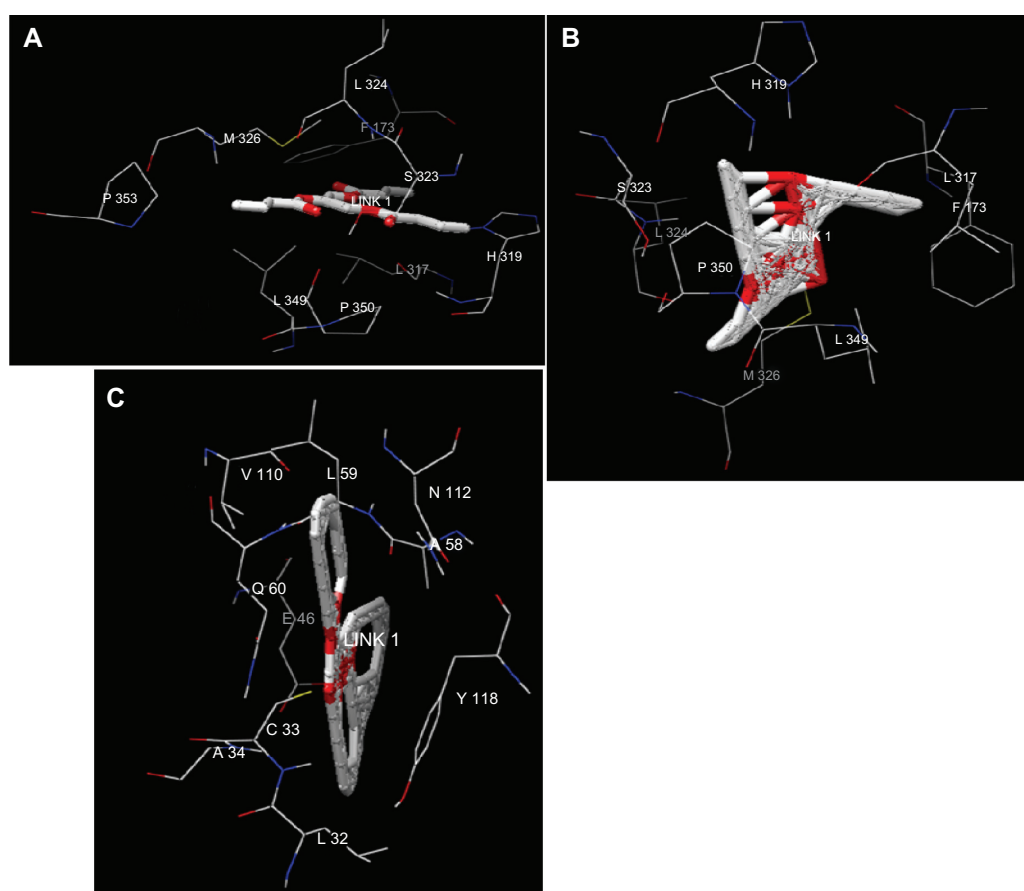
Docking

For docking, the ligand structures were obtained from the PubChem database.[33] In order to investigate the substrates binding with the enzyme, we attempted to dock A) tributyrin, B) trioctanoin and C) triolen to the lipase enzyme model. The top docking solutions of 3000 inter-
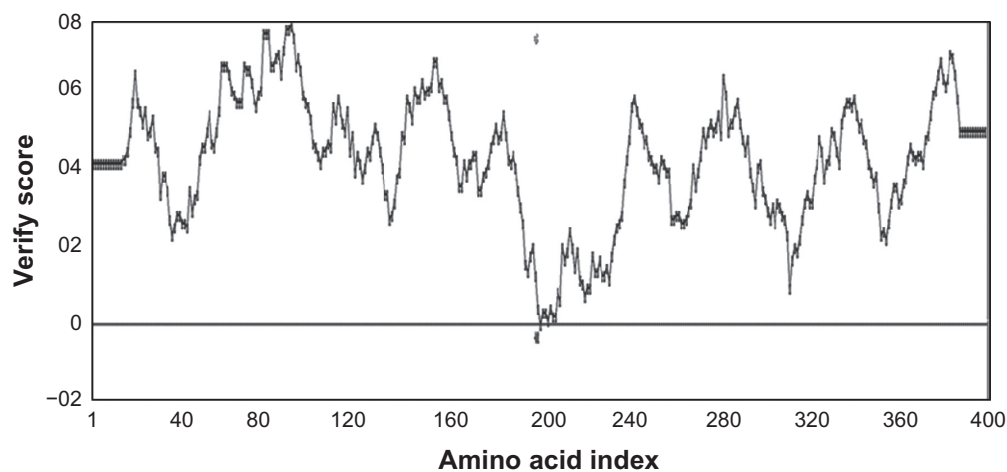
Figure 4. The verify score diagram validate the *Arabidopsis thaliana* lipase model.

Figure 5. Three-dimensional representation of substrates (**A**), tributyrin (**B**), trioctanoin and (**C**) triolen in the *Arabidopsis thaliana* lipase model active site after docking. For clarity purpose residues and ligands are shown in wireframe and stick form respectively.

action results for each ligand were selected. The total binding energies are respectively: −201.09 (kcal/mol), −191.11 (kcal/mol) and −272.83 (kcal/mol). This result confirms that the most preferred substrate for *Arabidopsis thaliana* lipase is the Triolen since its binding energy is the smallest one. The key residues involved in the ligand binding are the same for tributyrin and tri-octanoin substrate. This key is formed by the following amino acids: P353, M326, L324, P173, S323, H319, L317, P350, L349. On the other side, Triolen presents a unique binding key.

Figure 5 shows that triolein does not bind the lipase at the same site as tributyrin and trioctanoin. In addition, the catalytic active site, serine S_{166}, is not involved in the binding site, this result confirms that the lid domain which covers the active site is not accessible to solvent.

Conclusion

In summary, the constructed homology model of *Arabidopsis thaliana* lipase was validated for

stereochemical and amino acid environment quality using appropriate programs. A further validation of active site architecture was achieved by docking studies with the natural substrate to determine the key residues involved in the ligand binding.

Disclosures

This manuscript has been read and approved by all authors. This paper is unique and not under consideration by any other publication and has not been published elsewhere. The authors and peer reviewers report no conflicts of interest. The authors confirm that they have permission to reproduce any copyrighted material.

References

1. Gupta R, Gupta N, Rathi P. Bacterial lipases: An overview of production, purification and biochemical properties. *Applied Microbiology and Biotechnology*. 2004;64:763–81.
2. Villeneuve P. *Eur J Lipid Sci Technol*. 2003;35:308–17.
3. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *Journal of Molecular Biology*. 1990;215:403–10.

4. Hills MJ, Mukherjee KD. Triacylglycerol lipase from rape (*Brassica napus*) suitable for biotechnological purposes. *Appl Biochem Biotechnol*. 1990;26:1–10.
5. Tsuchiya T, Ohta H, Okawa K, et al. Cloning of chlorophyllase, the key enzyme in chlorophyll degradation: finding of a lipase motif and the induction by methyl jasmonate. *Proc Natl Acad Sci U S A*. 1999;96:15362–7.
6. Brick DJ, Brumlik MJ, Buckley JT, et al. A new family of lipolytic plant enzymes with members in rice, *Arabidopsis* and maize. *FEBS Lett*. 1995;377:475–80.
7. Oh IS, Park AR, Bae MS, et al. Secretome analysis reveals an *Arabidopsis* lipase involved in defense against *Alternariabrassicicola*. *Plant Cell*. 2005;17:2832–47.
8. Belguith H, Fattouch S, Jridi T, Ben Hamida J. Immunopurification and characterization of a rape (*Brassica napus L.*) seedling lipase. *African Journal of Biochemistry Research*. 2009;3:356–65.
9. Ling H. Sequence analysis of GDSL lipase gene family in *Arabidopsis thaliana*. *Pak J Biol Sci*. 2008;11:763–7.
10. Messaoudi A, Belguith H, Ghram I, Ben Hamida J. Classification of EC 3.1.1.3 bacterial true lipases using phylogenetic analysis. *African Journal of Biotechnology*. 2010;48:8243–7.
11. Ollis DL, Cheah E, Cygler M, et al. The a/b hydrolase fold. *Protein Eng*. 1992;5:197–211.
12. Arpigny JL, Jaeger KE. Bacterial lipolyticenzymes: classification and properties. *Biochem J*. 1999;343:177–83.
13. Kanaya S, Koyanagi T, Kanaya E. An esterase from Escherichia coli with a sequence similarity to hormone-sensitive lipase. *Biochem J*. 1998; 322:75–80.
14. Berman HM, Westbrook V, Feng Z, et al. The protein data bank. *Nucl Acids Res*. 2000;28:235–42.
15. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990;215(3):403–10.
16. Messaoudi A, Belguith H, Ghram I, Ben Hamida J. LIPABASE: a database for "true" lipase family enzymes. *International Journal of Bioinformatics Research and Applications* in press 2010.
17. Roussel A, Canaan S, Egloff MP, et al. Crystal structure of human gastric lipase and model of lysosomal acid lipase, two lipolytic enzymes of medical interest. *J Biol Chem*. 1999;274(24):16995–7002.
18. Arnold K, Bordoli L, Kopp J, Schwede T. The SWISS-MODEL Workspace: A web-based environment for protein structure homology modelling. *Bio informatics*. 2006;22:195–201.
19. Kleywegt GJ. Validation of protein crystal structures. *Acta Crystallogr D Biol Crystallogr*. 2000;56:249–65.
20. RAMPAGE Server http://ravenbioccam.ac.uk/rampage.php.
21. Bowie JU, Luthy R, Eisenberg D. A method to identify protein sequences that fold into a known three-dimensional structure. *Science*. 1991;253:164–70.
22. Colovos C, Yeates TO. Verification of protein structures: patterns of Non bonded atomic interactions. *Protein Sci*. 1993;2:1511–9.
23. UCLA-DOE Institute for Genomics & Proteomics Server http://www.doe-mbiucla.edu/Services.
24. Ritchie DW. Recent progress and future directions in protein–protein docking. *Current Protein and Peptide Science*. 2008;9:1–15.
25. Rithcie DW. Evaluation of Protein Docking Predictions using Hex 3.1 in CAPRI rounds 1–2. *Proteins*: *Struct Funct Genet*. 2003;52:98–106.
26. Daisy P, Mathew S, Suveena S, Rayan NA. A novel terpenoid from Elephantopusnscaber—antibacterial activity on Staphylococcus aureus: a substantiate computational approach. *Int J Biomed Sci*. 2008;3:3196–203.
27. Yang AS, Honig B. An integrated approach to the analysis and modeling of protein sequences and structures. *J Mol Biol*. 2000;301:665–711.
28. Eisenberg D, Lüthy R, and Bowie J. VERIFY3D: Assessment of protein models with three-dimensional profiles. *Methods in Enzymology* 1997; 277: 396–404.
29. Guex N, Peitsch MC. Swiss-Model and the Swiss-Pdb Viewer: an environment for comparative protein modeling. *Electrophoresis*. 1997;18(15): 2714–23.
30. Ginalski K. Comparative modeling for protein structure prediction. *Curr Opin Struct Biol*. 2006;16:172–7.
31. Hetenyi C, van dar Spoe LD. Blind docking of drug-sized compounds to proteins with up to a thousand residues. *FEBS Lett*. 2006;580:1447–50.
32. Nathan ST, Mathew N, Kalyanasundaram M, Balaraman K. Structure of glutathioneS-transferase of the filarial parasite Wuchereriabancrofti: a target for drug development against adult worm. *J Mol Mod*. 2005;11:194–9.
33. PubChem database http://pubchem.ncbi.nlm.nih.gov.