Trends in Bioinformatics

ISSN 1994-7941

science alert

ANSInet
an open access publisher
http://ansinet.com

Trends in Bioinformatics 8 (2): 37-51, 2015
ISSN 1994-7941 / DOI: 10.3923/tb.2015.37.51

Homology Modeling, Molecular Dynamic Simulations and Docking Studies of a New Cold Active Extracellular Lipase, *En*L A from *Emericella nidulans* NFCCI 3643

[1]Suseela Lanka, [2]Venkateswar Rao Talluri, [3]V. Ganesh and [1]J. Naveena Lavanya Latha
[1]Department of Biotechnology, Krishna University, Machilipatnam, Krishna district, A.P., India
[2]Department of Biotechnology, KL University, Vaddeswaram Guntur, A.P., India
[3]Department of Biotechnology, Research Scholar, KL University, Vaddeswaram, Guntur, A.P., India

Corresponding Author: J. Naveena Lavanya Latha, Department of Biotechnology, Krishna University, Machilipatnam, 521001, India Tel: 91-9502245034 Fax: 91-8672225960

ABSTRACT

A cold active lipase producing mesophilic fungus was isolated from Palm Oil Mill Effluent (POME) dump sites and identified by 28S rRNA molecular identification studies as *Emericella nidulans* NFCCI 3643. The BLAST P search with the sequence of the purified cold active lipase obtained by MALDI-TOF/MS analysis revealed that the protein is a hypothetical protein from *Emericella nidulans* with a gi number 67522685. Search of Lipase Engineering Database (LED) for this protein sequence revealed that this protein belongs to *Candida antarctica* lipase A like super family and to *Aspergillus* lipase like homologous family of class Y lipases. In the present study, a 3D structure of *En*L A (*Emericella nidulans* lipase A) was built using homology modeling, the model was further optimized by molecular dynamic simulations and the optimized model was then docked with natural substrates. Secondary structure analysis of *En*L A showed 37.11% of its content to be alpha helix making it stable for three dimensional structure modeling. Homology model of *En*L A was constructed using the X-ray structure of *Candida antarctica* Lip A (3 guu.1.A) as a template with which *En*L A showed 32.77% sequence identity. The stereo chemical quality and side chain environment of the model was validated by Ramachandran plot, ERRAT and Verify 3D. Natural substrates like tributyrin and trioctanoin were docked in to the optimized 3D model to further investigate the ligand-enzyme interactions.

Key words: Cold active lipase, palm oil mill effluent, *Emericella nidulans* NFCCI 3643, *En*L A, homology modeling, molecular dynamic simulations, docking, Ramachandran plot

INTRODUCTION

Lipases (triacylglycerol acylhydrolase, EC 3.1.1.3) are ubiquitous enzymes and are found in both prokaryotic and eukaryotic organisms with physiological significance. Lipases belong to the family of hydrolases acting on carboxylic ester bonds and responsible for catalyzing the hydrolysis of triglycerides to diglycerides, monoglycerides, fatty acids and glycerol (Gilbert, 1993). Among lipases, cold active lipases are gaining the attention of industrialists owing to their applications in various fields (Cavicchioli *et al*., 2002; Cavicchioli and Siddiqui, 2004) and due to their ability to catalyze the reactions at lower temperatures (Gerday *et al*., 2000; Cavicchioli *et al*., 2002; Feller and Gerday, 2003; Cavicchioli and Siddiqui, 2004). Cold active lipases are characterized by greater structural flexibility especially around the active site and this property can be best

exploited for improving the thermal stability of these enzymes (Joseph *et al*., 2008) which makes them suitable for various industrial applications. The thermal stability of cold active enzymes can be increased by chemical modification (Siddiqui and Cavicchioli, 2005) and as well by directed evolution (Zhang *et al*., 2003).

Bioinformatic applications provide insights in to sequence analysis and structure prediction of novel proteins. Structural predictions of microbial lipases have been in practice since 1980s. Ollis *et al.* (1992) for the first time reported α/ß hydrolase fold in lipases based on the 3 D folds. Most of the lipases contains 'lid', a helical segment that covers the active site and the opening and closing of this lid governs the accessibility and binding of the substrate to the active site (Derewenda *et al*., 1992; Van Tilbeurgh *et al*., 1993; Grochulski *et al*., 1994; Egloff *et al*., 1995). The active site contains the catalytic triad, usually Ser-Asp-His, in the hydrophobic environment.

Understanding the 3 dimensional structures will allow one to perform required modifications to suit various industrial applications. Molecular modeling is a valuable tool especially for medicinal chemists in the process of drug design. Computational biology is paving way to understand protein 3 dimensional structure and predict functions of novel proteins. Sequences of a number of lipases have been determined and their comparison is used for elucidating the structure-function relationship (Sinchaikul *et al*., 2001; Tyndall *et al*., 2002; Eggert *et al*., 2002; Tripathi *et al*., 2004).

In the field of molecular modeling, docking is one of the very important technique which allows one to predict the preferred orientation of one molecule relative to the other when both are binding to each other (Kumar *et al*., 2013) and this knowledge of preferred orientation in docking studies in turn predicts the binding affinity and strength of association between two molecules. In general, docking is used to predict the interactions between drug candidates (Kitchen *et al*., 2004) in drug designing process. In case of enzymes, docking can be employed to predict the substrate binding sites (active sites).

In the present study, Homology modeling was first applied to build a 3D structure of *En*L A purified and characterized from *Emericella nidulans* NFCCI 3643 screened and isolated from Palm Oil Mill Effluent (POME) dump sites. Molecular Dynamic (MD) simulations were performed to further optimize the 3D model and docking studies were then performed to study ligand-enzyme interactions.

MATERIALS AND METHODS

BLAST P, multiple sequence alignment and phylogenetic tree construction: The amino acid sequence of the cold active lipase, *En*L A, obtained by MALDI-TOF/MS analysis of the purified lipase from *Emericella nidulans* NFCCI, screened and isolated from Palm Oil Mill Effluent (POME) dump sites, Pedavegi, East Godavari District andhra Pradesh, India was used to scan the protein sequence databases using BLAST P algorithm to obtain homologous protein sequences from the available protein sequences of various organisms. The sequences so obtained were further subjected to multiple sequence alignment using CLUSTAL ω and a phylogenetic tree was then constructed using phylogeny.fr (http://www.phylogeny.fr/) to determine the evolutionary relationships.

***En*L A and Lipase Engineering Database (LED):** Lipase engineering database, a database of lipases, maintains information regarding lipase sequences including putative and hypothetical sequences from various organisms, organized into classes, super families and homologous families. Since LED (http://www.led.uni-stuttgart.de/) is a repository of lipase sequences from all possible sources it was searched for presence of *En*L A sequence.

Secondary structure prediction of *En*L A: Secondary structure of *En*L A was predicted using SOPMA (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html) tool in Expasy.

Homology modeling: The sequence of *En*L A was downloaded from the universal protein resource (Uniprot KB) (http://www.uniprot.org/) (The UniProt Consortium, 2012) (entry ID: Q5BCD1). The suitable template for homology modeling was identified through searching *En*L A on PDB using the BLAST P algorithm (Altschul *et al.,* 1990). The 3D structure of *Candida antarctica* lipase A (Cal A) was downloaded from PDB (PDB ID: 3guu.1.A) as the template structure. The 3D model of *En*L A was then built with Prime version 3.9 (Jacobson *et al.,* 2002, 2004) in Schrödinger Suite 2015-1 (Schrodinger, LLC, New York, NY). The target (*En*L A) and template (Cal A) sequences were aligned using the Clustal W method employed in Prime, followed by manual adjustment to avoid big gaps in the secondary structure domain.

Model validation: The quality of the homology model was validated by assessing the stereo chemical quality of the model using Ramachandran plot obtained from the RAMPAGE (http://mordred.bioc.cam.ac.uk/~rapper/rampage.php) server (Lovell *et al*., 2002). Verify 3D (Bowie *et al*., 1991) and ERRAT (Colovos and Yeates, 1993) were used to assess the amino acid environment from the UCLA-DOE server (http://www.doe-mbi.ucla.edu/services).

Molecular Dynamic (MD) simulations: The initial 3D structure of *En*L A obtained from homology modeling was further optimized using MD simulation. The MD simulations were performed using Desmond Molecular Dynamics module (Guo *et al*., 2010; Shivakumar *et al*., 2010) version 4.1 of Schrodinger with OPLS (Optimized Potentials for Liquid Simulations) 2005 force field. The 3D *En*L A structure was surrounded by a truncated orthorhombic box of SPC water molecules with a margin of 10.0 Å along each dimension. The charge neutrality was maintained by adding sodium ions to the system. Energy of the prepared systems was minimized for 2000 iterations using steepest descent method. The default parameters in Desmond were applied for system equilibration. The 12 nsec (nano seconds) MD simulations were then carried out with the equilibrated systems at a temperature of 300 K and at a constant pressure of 1atm, under the NPT (normal pressure and normal temperature) ensemble with a time step of 2f sec (femto seconds).

Ligand preparation: The natural substrates of lipases like Tributyrin and Trioctanoin were selected as ligands for docking studies. The 3D structures for these ligands were built using Lig prep version 3.3 in Schrödinger Suite 2015-1 with an OPLS_2005 force field (Jorgensen and Tirado-Rives, 1988; Jorgensen *et al*., 1996; Shivakumar *et al*., 2010). Epik (Shelley *et al.,* 2007; Greenwood *et al*., 2010) version 3.1 in Schrödinger Suite was used to generate their ionization states at pH7.0±2.0.

Active site identification, grid generation and molecular docking: Site Map (Halgren, 2007, 2009) version 3.4 in Schrödinger Suite was used to identify probable ligand binding sites. The optimized *En*L A structure was prepared prior to molecular docking using Protein Preparation Wizard (Sastry *et al*., 2013) in Schrödinger Suite. Bond orders were assigned and hydrogen atoms were added to the protein. The OPLS_2005 force field was then used to minimize

the structure of the protein in order to reach the converged RMSD (root mean square deviation) 0.30 Å. Hydrophobic and hydrophilic field contour maps were then generated. This was then followed by dividing of hydrophilic maps in to donar, acceptor and metal binding regions. All the sites were finally assessed by calculating various properties. Receptor Grid Generation in Schrodinger Suite was then used to define a docking grid. Glide (Friesner *et al*., 2004; Halgren *et al*., 2004; Friesner *et al*., 2006), version 6.6 in Schrödinger Suite with extra precision (XP) was used to dock the optimized 3 D structures of the two substrates (tributyrin and trioctanoin) in to docking grid in the optimized 3D structure of *En*L A.

RESULTS AND DISCUSSION

In view of growing demand for cold active lipases to use in various industrial applications, knowledge of 3 dimensional structures could pave the way to modify these enzymes as per the industrial needs. The 3D structure of proteins can be predicted by using homology modeling which uses experimentally determined protein structures as templates to predict the 3 D structure of a target protein based on target-template alignment. In the present study, homology modeling, MD simulation and docking studies were carried out with *En*L A sequence obtained by the MALDI-TOF/MS analysis of the purified lipase from *Emericella nidulans* NFCCI 3643, screened and isolated from POME dump sites.

Scanning of protein sequence databases using BLAST P (Altschul *et al*., 1990) with the sequence obtained by MALDI-TOF/MS analysis of the purified lipase revealed that the protein is a hypothetical protein from *Emericella nidulans* with a gi number 67522685 and the protein is showing up to 64% similarity with hypothetical proteins and putative secretary lipases from various organisms especially from *Aspergillus* spp. The FASTA sequence of extracellular cold active lipase (*En*L A) retrieved from Uniprot KB was shown in Fig. 1. Multiple sequence alignment of sequences obtained by BLAST P analysis using Clustal ω revealed that *En*L A also contained a conserved GXSXG sequence that is characteristic of all lipases (Ollis *et al*., 1992) (Fig. 2). A phylogram constructed based on multiple sequence alignment using phylogeny.fr (Dereeper *et al*., 2008) revealed that *En*L A was closely related to a conserved hypothetical protein from *Aspergillus terreus* NIH 2624 (Fig. 3).

tr|Q5BCD1|Q5BCD1_ EMENI Secretory lipase, putative (AFU_orthologue AFUA_7G00110) OS=*Emericella nidulans* (strain FGSC A4 / ATCC 38163 / CBS 112.46 / NRRL 194 / M139) GN=AN1799.2 PE=4 SV=1

MASLLYQLLFLLVPLLAAGLPANPVKKAGPQPPGEDPFYTPPDGWESTEPGAILRHRTPP
YPIAAFGLAEVNLDASYQILYRTTDSFGEPIATVTTILIPHNADYTKVLSYQVAQDAADP
NCSPSFAIQQFSDAGEALALVMPQLEYLFMSSALNKGFVVIVPDHLGPRSAFLANTLSGQ
AVLDNVRAALASTDITGISSQATVALWGYSGGSLASGFAAELQPSYAPELKIAGAALGG
TVPQIPPVIRASNKGIFTGLIPAGIQGLANEYPAAQQLIRDAILPDKWAEFNKTQELCLTG
NLIEYLGKDIYTYVNDPNVFESPLANSLTEPNAMGHNTPKIPILIYKGVNDQISPVKDTD
ALYDTYCSNGANVEYVRDLLAEHALMTITGAPDAFMWLTERLSGVPVKKGCRRKTQL
TGLQDPKALAALGTTVVKFLLSLLTLPVGPIGR

Q5BCD1 from UNIPROT KB. Search for the *En*L A sequence in UNIPROT KB revealed that the sequence is a putative secretary lipase from *E. nidulans*

Fig. 1: Sequence of the purified lipase from UNIPROT KB

```
gi|67522685|ref|XP_659403.1|        QAT---VALWGYSGGSLASGFAAELQPSYAPELKIAGAALGGTVPQIPPV 247
gi|115399588|ref|XP_001215422.      TAR---VTMWGYSGGSLATGFAAELQPTYAPELTIAGAALGGTVPKILSV 246
gi|238485356|ref|XP_002373916.      DPT---IALWGYSGGSLASGFAAELHPTYAPELNIVGAALGGTVPKIRPV 249
gi|169772037|ref|XP_001820488.      DPT---IALWGYSGGSLASGFAAELHPTYAPELNIVGAALGGTVPKIRPV 249
gi|391872471|gb|EIT81587.1|         DPT---IALWGYSGGSLASGFAAELHPTYAPELNIVGAALGGTVPKIRPV 249
gi|599151235|gb|EYE90686.1|         DPT---IALWGYSGGSLASGFAAELQPSYAPELNIQGAALGGTVPQILPV 243
gi|345105447|gb|AEN71554.1|         KAT---ITLWGYSGGSLASGLAAELRASYAPELNIAGAALGGTVPKIMPV 265
gi|159122840|gb|EDP47960.1|         KAT---ITLWGYSGGSLASGLAAELRASYAPELNIAGAALGGTVPKIMPV 264
gi|70982778|ref|XP_746917.1|        KAT---ITLWGYSGGSLASGLAAELRASYAPELNIAGAALGGTVPKIMPV 265
gi|119488165|ref|XP_001262626.      KAT---ITLWGYSGGSLASGFAAELRASYAPELNIAGAALGGTVPKIMPV 243
gi|212527456|ref|XP_002143885.      NAT---ITIWGYSGGSLSSGFAVELRKAYAPELKIAGAALGGTVPDLLET 248
gi|573980681|ref|XP_006668261.      DAR---VALWGYSGGSLASEFAAELHPTYAPELKIVGAALGGTVPSILSV 237
gi|729189900|emb|CEJ81243.1|        NPT---IAMWGYSGGSVASLFAAELQPSYAPELKIAGAAVGGIVPNITNV 240
gi|734656746|gb|KHN93840.1|         KPT---VTMWGYSGGSLATMFAAELQPSYAPEVEVAGAAVGGIVPSIMTV 253
gi|629735826|ref|XP_007826372.      NSSNLKIGMWGYSGGSLATNWAAELQPTYAPELRIAGAAVGGTIPNITTA 252
gi|672382699|gb|KFG84821.1|         NSSNLKIGMWGYSGGSLATNWAAELQPTYAPELRIAGAAVGGTIPNITTA 252
gi|629725822|ref|XP_007823144.      NST---ITMWGYSGGSLATNHAAELQPEYAPELQIAGAAVGGTVPNITNA 247
gi|672382640|gb|KFG84762.1|         NST---IAMWGYSGGSLATNHAAELQPEYAPELQIAGAAVGGTVPNITSA 247
gi|701767719|gb|KGQ03222.1|         DAR---VAMWGYSGGSLASQWAAELQPSYAPELHVAGAAVGGTVPDVASV 247
gi|573991003|ref|XP_006673422.      DAT---VAMWGYSAGSDATAWAAELHPTYAPELAIAGAAIGGTTPNITNV 391
gi|358394239|gb|EHK43640.1|         DAT---VALWGYSGGSLASGWAAELQPSYAPELKIAGAALGGTVPNITTV 249
gi|672382543|gb|KFG84667.1|         NAT---VALWGYSGGSVATGWAAELHASYAPELNIVGAALGGTVPSIPPV 244
gi|629734290|ref|XP_007825873.      NAT---VALWGYSGGSVATGWAAELHASYAPELNIVGAALGGTVPSIPPV 244
gi|629699887|ref|XP_007815235.      NAT---VALWGYSGGSVATGWAAELHASYAPELNIVGAALGGTVPSIPPV 244
gi|629731906|ref|XP_007825110.      DPA---IALWGYSGGSITSARAAELQPSYAPELKILGMAIGGTEPNIANV 247
gi|303319853|ref|XP_003069926.      KAR---VTMWGYSGGSLASGFAAELQPAYAPELKIAGVALGGTVPKISTV 248
gi|119183459|ref|XP_001242768.      KAR---VTMWGYSGGSLASGFAAELQPAYAPELKIAGVALGGTVPKISTV 248
gi|258570909|ref|XP_002544258.      KAR---VTMWGYSGGSLASGFAAELQPKYAPELNIAGAALGGTVPRIEPV 216
gi|302497105|ref|XP_003010553.      DAV---VTMWGYSGGSLAAGFAAELQPCYAPELKIAGAALGGTVPNITKV 248
gi|302663480|ref|XP_003023382.      DAV---VTMWGYSGGSLAAGFAAELQPCYAPELKIAGAALGGTVPNITTV 355
gi|327300066|ref|XP_003234726.      DAV---VTMWGYSGGSLASGFAAELQPCYAPELKIAGAALGGTVPNITTV 248
gi|326473534|gb|EGD97543.1|         DAT---VTMWGYSGGSLASGFAVEVQPCYAPELKIAGAALGGTVPNITSV 248
gi|326480243|gb|EGE04253.1|         DAT---VTMWGYSGGSLASGFAVEVQPCYAPELKIAGAALGGTVPNITSV 248
gi|607894905|gb|EZF33704.1|         DAT---VTMWGYSGGSLASGFAVEVQPCYAPELKIAGAALGGTVPNITSV 248
gi|315042321|ref|XP_003170537.      DAV---VTMWGYSGGSLASGFAAELQPCYAPELKIAGAALGGTIPKVKTV 248
gi|296816046|ref|XP_002848360.      DAT---ITMWGYSGGSLASGFAAELQPSYAPELKIAGAALGGTVPRVKTV 249
gi|302652744|ref|XP_003018216.      DAA---VGLWGTSGGSVASAFAADLHPTYAPELNIVGAALGGVVPSITTA 254
gi|302501833|ref|XP_003012908.      DAA---VGLWGTSGGSVASAFAADLHPKYAPELNIVGAALGGVVPSITTA 254
gi|326475508|gb|EGD99517.1|         DAA---VGLWGTSGGSVASAFAADLHPTYAPELNIVGAALGGVVPSITTA 253
gi|326483106|gb|EGE07116.1|         DAA---VGLWGTSGGSIASAFAADLHPTYAPELNIVGAALGGVVPSITTA 253
gi|607889079|gb|EZF29627.1|         DAA---VGLWGTSGGSVASAFAADLHPTYAPELNIVGAALGGVVPSITTA 253
gi|607877491|gb|EZF22622.1|         DAA---VGLWGTSGGSVASAFAADLHLTYAPELNIVGAALGGVVPSITTA 253
gi|607976730|gb|EZG05913.1|         DAA---VGLWGTSGGSVASAFAADLHLTYAPELNIVGAALGGVVPSITTA 253
gi|315048943|ref|XP_003173846.      DAV---IGLWGTSGGSVASGFAVDLHSTYAPELKIVGAALGGLVPSITTA 253
gi|682404984|gb|KFY80564.1|         SAT---YQMWGYSGGSLASEWAAELQPSYAPELNFAGVALNSLVPNISSV 252
gi|682455651|gb|KFZ13715.1|         SAT---YQMWGYSGGSLASEWAAELQPSYAPELNFAGVALNSLVPNISSV 252
gi|682471249|gb|KFZ24117.1|         SAT---YQMWGYSGGSLASEWAAELQPSYAPELNFAGVALNSLVPNVSSV 252
gi|682427570|gb|KFY94539.1|         SAT---YQMWGYSGGSLASEWAAELQPSYAPELNFAGVALNSLVPNVSSV 252
gi|682314829|gb|KFY20655.1|         SAT---YQMWGYSGGSLASEWAAELQPSYAPELSFAGVALNSLVPNVSSV 252
gi|682386063|gb|KFY67998.1|         SAT---YQMWGYSGGSLASEWAAELQPSYAPELNFAGVALNSLVPNVASV 252
                                    .      :** *.** ::  *.:::  ****: . * *:..  * :  .
```

Fig. 2: Multiple sequence Alignment using CLUSTAL ω (Omega), multiple sequence alignment using CLUSTAL ω (omega) showing a GXSXG consensus sequence characteristic of lipases

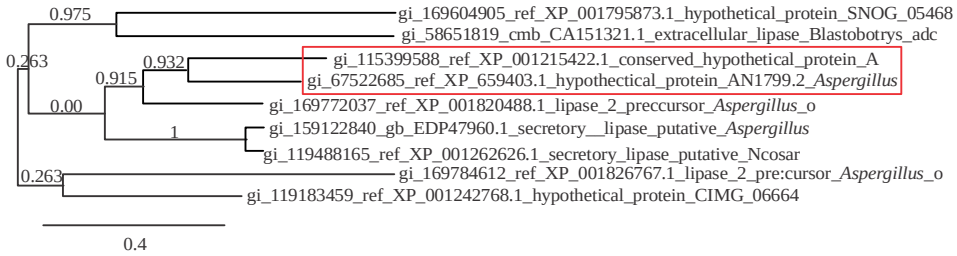

Fig. 3: Phylogenetic tree of *En*L A, A phylogenetic tree of *En*L A constructed using phylogeny.fr showing that *En*L A is closely related to conserved hypothetical protein of *Aspergillus terreus* NIH 2624

Since, LED is a repository of lipase sequences from all possible sources it was searched for the presence of *En*L A sequence. The lipase sequence under current study belongs to Class Y lipases based on oxy anion hole forming amino acid i.e., tyrosine which included 5 super families, 8 homologous families, 1905 proteins, 2553 sequences and 225 structures. Of the 5 super families and 8 homologous families, *En*L A belongs to *Candida antarctica* lipase A like super family (abH38)and to the homologous family of *Aspergillus* lipase like (abH38.03). The *Aspergillus* lipase

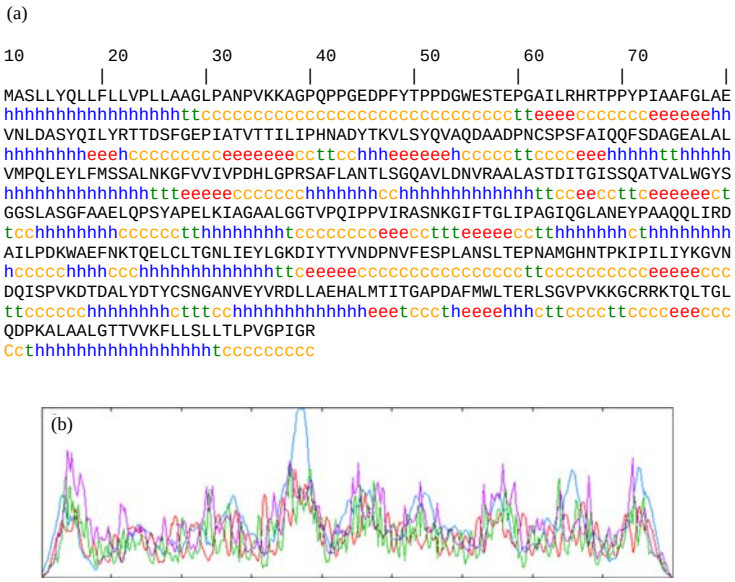

Fig. 4(a-b): Secondary structure prediction of *En*L A, (a) Sequence length: 450; Alpha helix (Hh): 167 is 37.11%, Extended strand (Ee): 70 is 15.56%, Beta turn (Tt): 45 is 10.00%, Random coil (Cc):168 is 37.33% and (b) Distribution of secondary structure elements in *En*L A. Blue line-Alpha Helix, Red-Extended strand, Green-beta turn, Orange-random coil

like homologous family contained 10 protein sequences including the *En*L A sequence and most of them were hypothetical proteins/putative secretary lipases from different *Aspergillus* spp. Till now there were no resolved 3D structures for that homologous group (*Source-Lipase Engineering database, http://www.led.uni-stuttgart.de/).

Secondary structure of the target protein was predicted by using SOPMA tool in Expasy (Fig. 4). The results indicate that *En*L A has 37.11% α-helix thus making it stable for homology modeling.

Since, there were no resolved 3D structures for the proteins of *Aspergillus* lipase like homologous group including *En*L A, an attempt was made to build the 3 D structure of *En*L A using homology modeling. The first step in homology modeling involves identification of a suitable template. This was met by performing a BLAST P search against known protein structures deposited in PDB. The studies of Rost (1999) and Yang and Honig (2000) demonstrated that 3D structures will be similar if the sequence identity between target and template proteins is higher than 25%. Generally, a target which shares a sequence similarity of 30% or more to an experimentally solved protein structure (template) can only be employed for homology modeling (Marsden and Orengo, 2008). The crystal structure of *Candida antarctica* Lip A (3 guu.1.A) with a sequence identity of 32.77% to the target sequence was selected based on BLAST P search against PDB database (Fig. 5a). The sequence alignment between the template (3guu.1.A) and the target was shown in Fig. 5b. Homology modeling was done using Prime version 3.9 of Schrodinger Suite 2015-1. The homology model so obtained was further refined by Protein Prep (Protein preparation) wizard in Schrodinger software suite. This mainly involves stabilizing the protein by hydrogen bond

(a)

Template	Seq Identity	Oligo-state	Found by	Method	Resolution	Seq Similarity	Coverage	Description
2veo.1.A	32.77	monomer	BLAST	X-ray	2.20Å	0.36	0.92	LIPASE A
3guu.1.A	32.77	monomer	BLAST	X-ray	2.10Å	0.36	0.92	Lipase A
3guu.1.A	29.36	monomer	HHblits	X-ray	2.10Å	0.34	0.93	Lipase A
2veo.1.A	28.99	monomer	HHblits	X-ray	2.20Å	0.34	0.92	LIPASE A
3zpx.1.A	30.56	homo-dimer	HHblits	X-ray	1.99Å	0.35	0.91	LIPASE
3zpx.1.A	32.34	homo-dimer	BLAST	X-ray	1.99Å	0.36	0.89	LIPASE
4ezi.1.A	14.83	monomer	HHblits	X-ray	1.15Å	0.28	0.70	Uncharacterized protein
3h2h.1.A	17.57	monomer	HHblits	X-ray	2.10Å	0.28	0.70	esterase
3h2j.1.A	16.19	monomer	HHblits	X-ray	1.89Å	0.28	0.70	esterase
3h2i.1.A	16.56	monomer	HHblits	X-ray	2.10Å	0.28	0.70	esterase
2b4k.1.A	14.19	homo-tetramer	HHblits	X-ray	3.30Å	0.26	0.64	alpha-amino acid ester hydrolase
1ryy.1.A	14.19	homo-tetramer	HHblits	X-ray	2.80Å	0.26	0.64	alpha-amino acid ester hydrolase
1lns.1.A	13.45	homo-dimer	HHblits	X-ray	2.20Å	0.25	0.64	X-PROLYL DIPEPTIDYL AMINOPEPTIDASE
1mpx.1.A	11.72	homo-tetramer	HHblits	X-ray	1.90Å	0.25	0.64	alpha-amino acid ester hydrolase
1nx9.1.A	13.78	homo-tetramer	HHblits	X-ray	2.20Å	0.26	0.63	alpha-amino acid ester hydrolase
1ju3.1.A	12.63	monomer	HHblits	X-ray	1.58Å	0.25	0.63	cocaine esterase
3i2g.1.A	12.63	homo-dimer	HHblits	X-ray	2.50Å	0.25	0.63	Cocaine esterase
1l7r.1.A	13.43	monomer	HHblits	X-ray	1.64Å	0.26	0.63	cocaine esterase
3i2h.1.A	11.93	homo-dimer	HHblits	X-ray	1.65Å	0.25	0.63	Cocaine esterase
3ida.1.A	12.32	homo-dimer	HHblits	X-ray	1.60Å	0.25	0.63	Cocaine esterase
3pui.1.A	13.88	homo-dimer	HHblits	X-ray	1.53Å	0.26	0.62	Cocaine esterase
3i2i.1.A	13.88	homo-dimer	HHblits	X-ray	2.14Å	0.25	0.62	Cocaine esterase
3puh.1.B	13.88	homo-dimer	HHblits	X-ray	2.30Å	0.25	0.62	Cocaine esterase
3i2j.1.A	13.88	homo-dimer	HHblits	X-ray	2.01Å	0.25	0.62	Cocaine esterase

(b)

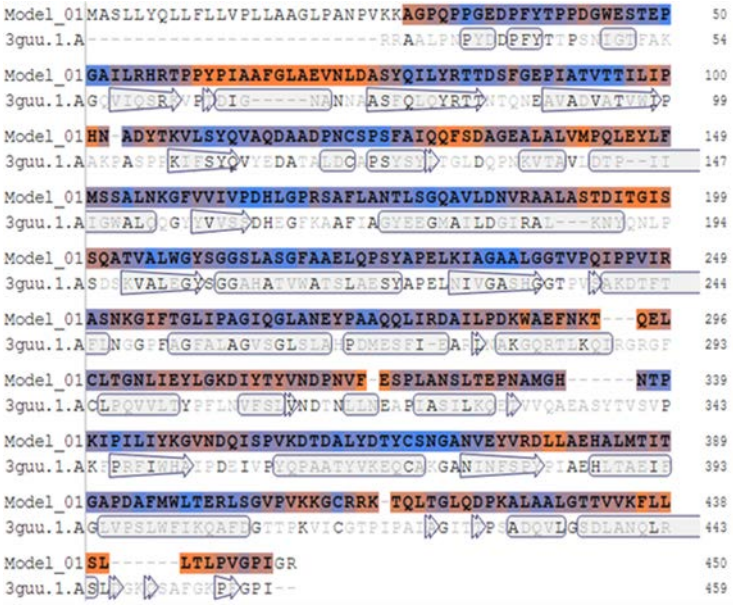

Fig. 5(a-b): BLAST P search against PDB and Target-Template alignment, (a) BLAST results of target sequence (*En*L A) against PDB for the identification of template for homology modeling and (b) Alignment between target (*En*L A) and template (*Candida antarctica* Lip A (3guu.1.A)

addition, structure optimization and energy minimization. Figure 6a and 6b shows the homology model of *En*L A and the ribbon diagram of super imposed structures of the template and target, respectively.

Built homology models can be further refined by MD simulations (Raval *et al*., 2012). In the present study, the initial 3D structure of *En*L A obtained by homology modeling was further optimized using MD simulation. In MD simulations, RMSD (Root Mean Square Deviation) serves

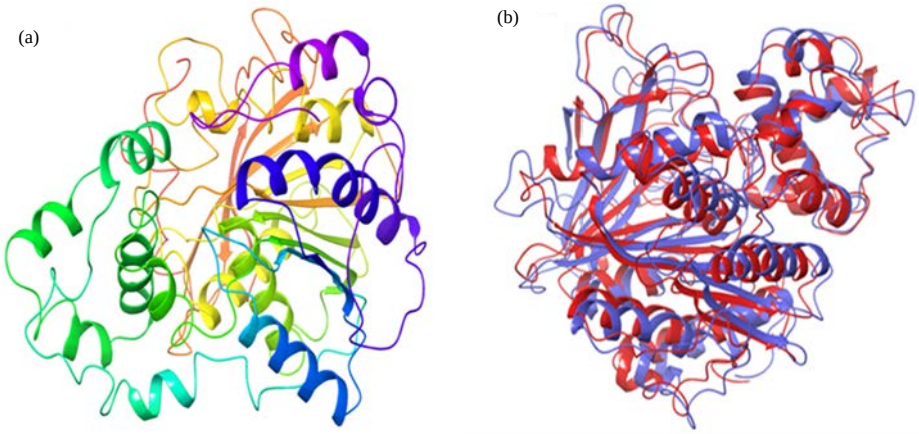

Fig. 6(a-b): Homology modeling of *En*L A, (a) 3D Ribbon representation of structure of *En*L A predicted using 3 guu.1.A as template based on target-template alignment using prime module of Schrodinger suite and (b) Ribbon representation of super imposed structures of initial homology model (blue) vs template (red)

Table 1: Ramachandran plot statistics

Amino acid residues and regions (%)	Percentage
Residues in most favored regions	85.7
Residues in the allowed	12.2
Residues in the outlier regions	2.1

Ramachandran Plot statistics for *EnL* A homology model using RAMPAGE server

as a measure to determine the stability of *En*L A structure based on its deviation from the initial structure. The optimized 3D model was shown in Fig. 7a. The RMSD values of *En*L A residues in the entire MD simulation trajectory were shown in Fig. 7b. The *En*L A structure showed deviations up to 3 Å and then gave a stable trajectory beyond that. The 3D structure of E*nL* A reached to a stable state after 6 n sec. Figure 7c shows RMSF (root mean-square fluctuations) of E*nL* A structure generated during the MD simulation and these fluctuations were calculated to characterize the mobility of individual residues. Large fluctuations were reported to the side chain residues up to 80 with regard to their high peaks in the RMSF plot and beyond 80th residue the fluctuations reported were within 3 Å indicating that the amino acid fluctuations were in acceptable range. Figure 7d shows the super imposed structures of initial homology model and MD simulated model. The stereo chemical quality of the 3D model was validated by Ramachandran plot using RAMPAGE server. Figure 8a and Table 1 shows that around 97.9% residues were present in the allowed regions (85.7% in the favored region and 12.2% residues in the allowed regions) and only 2.1% residues were present in the outlier region indicating that the quality of the model was good. The amino acid environment was assessed by using ERRAT (Fig. 8b) and Verify 3D (Fig. 8c). ERRAT determines the overall quality of the model and higher scores implies higher quality. In general a range of >50 indicates high quality model (Colovos and Yeates, 1993). In the present study, the E*nL* A structure has a score 84.677 indicating the high quality of the model. The E*nL* A structure was also assessed by Verify 3D, which assesses the structure by analyzing the compatibility of a 3D model with its own 1D model. The scores generally range from -1 (bad score) to +1 (good score) (Guex and Peitsch, 1997).

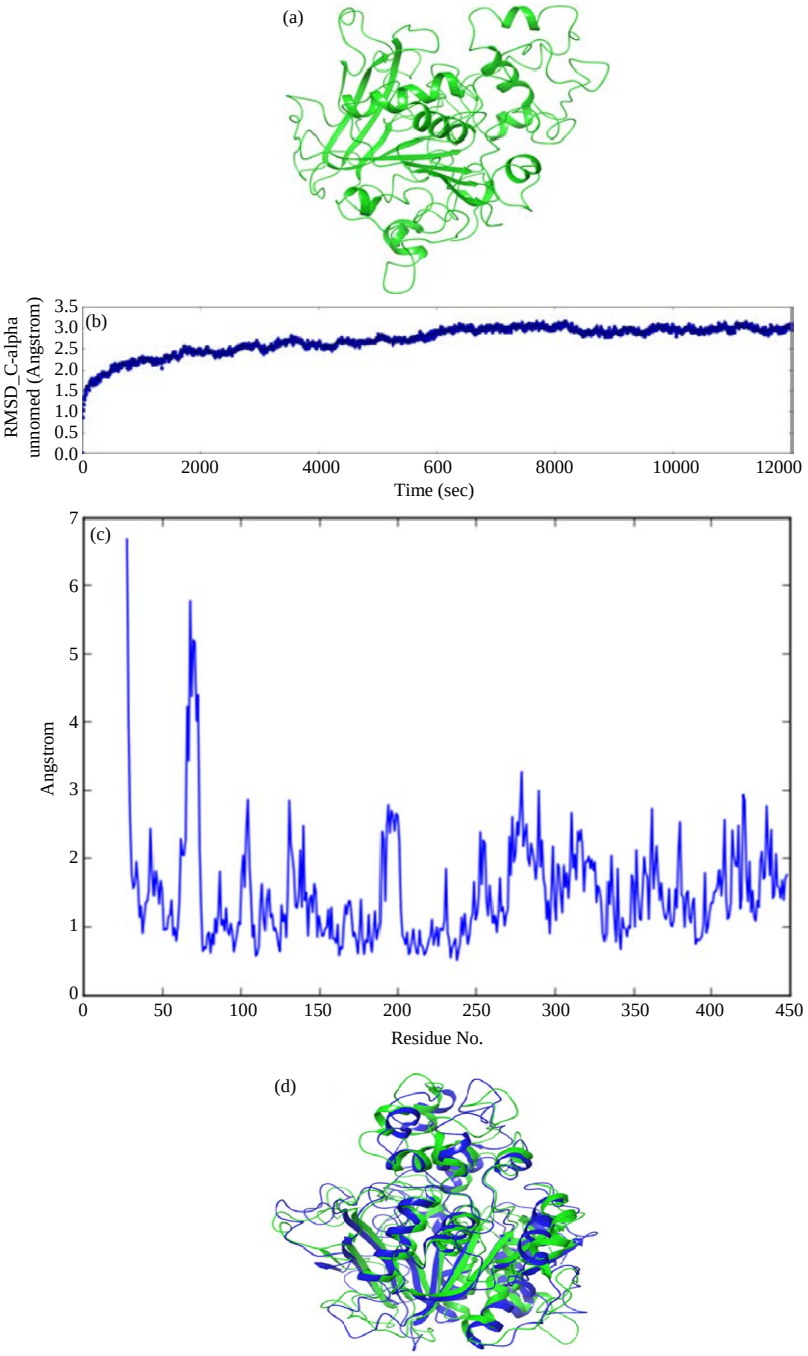

Fig. 7(a-d): MD simulations of *En*L A structure, (a) Ribbon representation of the optimized 3D model obtained by the 12 ns MD simulation run using Desmond module in Schrodimger suite, (b) RMSD (Root mean square deviations) of back bone residues of *En*L A structure, (c) RMSF (Root mean square fluctations) diagram showing fluctuations of amino acid residues and (d) Ribbon representation of the super imposed structures of initial homology model (blue) and MD simulated model (green)

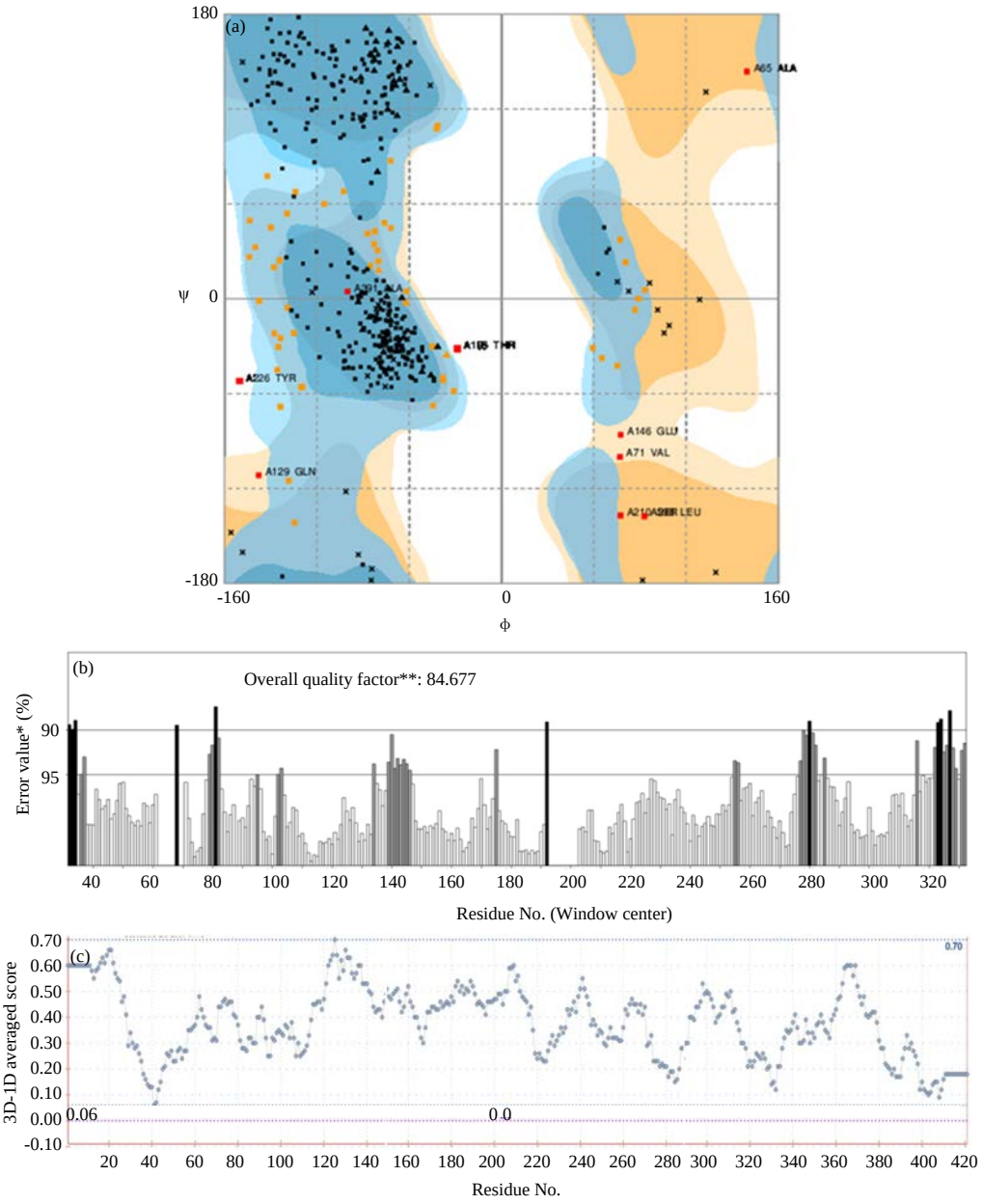

Fig. 8(a-c): Validation of *En*L A structure, (a) Stereo chemical quality of the *En*L A structure assessed by Ramachandran plot using RAMPAGE server. The plot values showing number of residues in favoured, allowed and outlier region, (b) Amino acid environment assessed by ERRAT using UCLA-DOE Institute for Genomics and Proteomics Server. Black bars show the mis folded region located distantly from the active site, gray bars demonstrate the error region and the white bars indicate the region having less error rate for protein folding and (c) Verify 3D score diagram validating the *En*L A a model

Among the 5 different active sites of *En*L A structure determined by Site Map module of Schrodinger suite, Site 2 was further selected for molecular docking studies because of its big

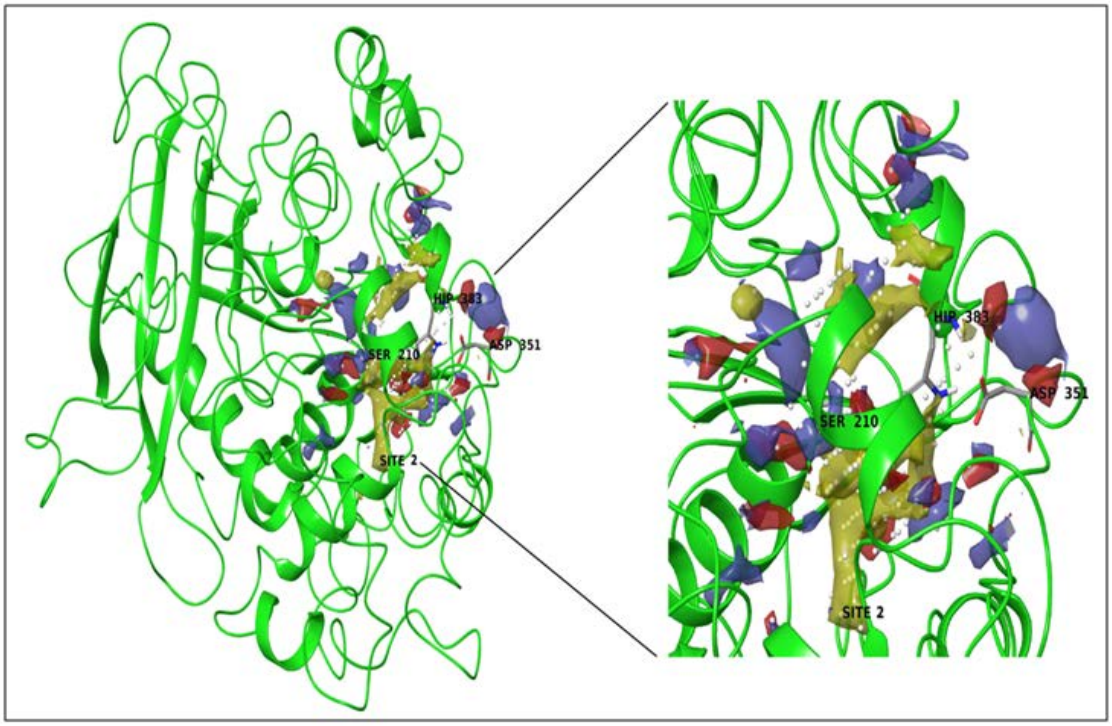

Fig. 9: Active site prediction using Site Map, Active site 2 of *En*L A structure predicted using Site Map. Closer view of site 2 displaying catalytic triad residues

Table 2: Energy values of the two substrates (Tributyrin and Trioctanoin)

Substrates	Formula	Energy value (kcal mol^{-1})
Tributyrin	$C_{15}H_{26}O_6$	-70.67
Trioctanoin	$C_{27}H_{50}O_6$	-83.14

Binding energy values of the two substrates in kcal mol^{-1} calculated based on their binding affinity to the active site of the enzyme (*En*L A)

pocket size, enclosed catalytic triad residues (Ser-210, Asp-351 and His-383) and good D score (Fig. 9). The natural substrates of lipases like tributyrin (Chem spider ID: 13849665), trioctanoin (Chem spider ID: 10393) obtained from Chem spider database were used to dock the *En*L A structure. The chemical structures and names of the substrates (ligands) used in this study were shown in Fig. 10a-b. The binding energies obtained were respectively -70.67 and -83.14 kcal mol^{-1} for tributyrin and trioctanoin (Table 2). From this result it was clear that the most preferred substrate for *En*L A is tributyrin since its binding energy is smallest. The docked substrates in the site 2 of the *En*L A structure was shown in Fig. 10c (site 2 of *En*L A docked with tributyrin) and 10 d (site 2 of *En*L A docked with trioctanoin). The key amino acid residues involved in the ligand binding were shown in Fig. 10e and f. The key residues involved in the binding of tributyrin were found to be S210, F256, T257, L302, N301, D116, L259, Y305, F172, G258, L298, L443, T299, H383, S354, I353, V247. In case of trioctanoin, the key residues were found to be same as those involved in the binding of tributyrin but some additional amino acid residues were also involved in addition to those mentioned above.

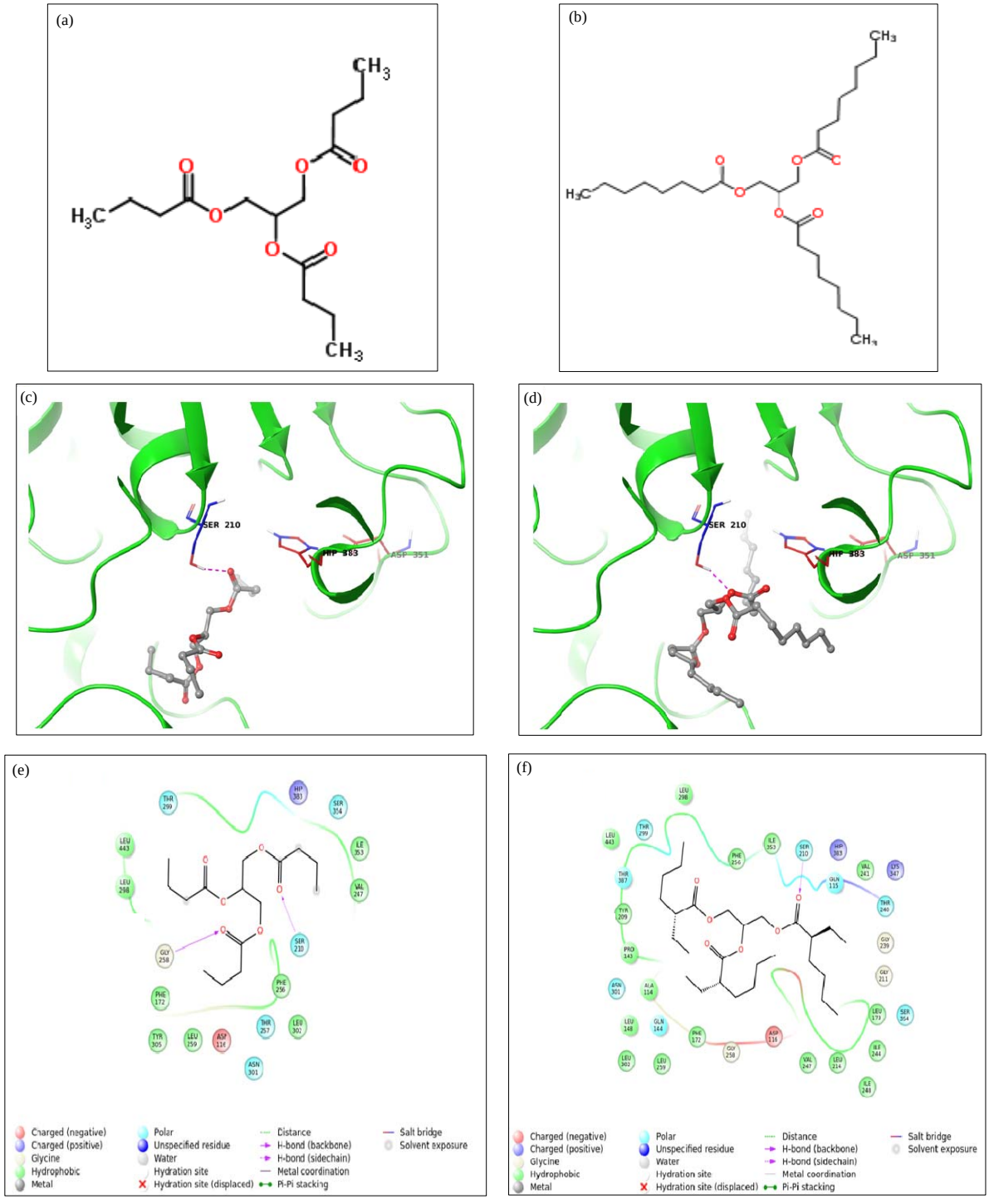

Fig. 10(a-f): Chemical structures of the substrates (ligands), (a, b) Chemical structures of Tributyrin and Trioctanoin obtained from Chem spider database, (c-d) Active site 2 of *En*L A structure docked with Tributyrin and Trioctanoin, respectively and (e-f) Active site amino acid residues of cold active lipase *En*L A as investigated by docking studies with Glide XP module in Schrodinger suite

CONCLUSION

In summary, a homology model of *En*L A was constructed, further optimized by molecular dynamic simulations and validated by assessing stereo chemical and amino acid environment quality. Docking studies were performed with the natural substrates to investigate the key residues involved in the ligand binding.

ACKNOWLEDGMENTS

We are extremely thankful to the management of K L University for providing Schrodinger software for carrying out this work.

REFERENCES

Altschul, S.F., W. Gish, W. Miller, E.W. Myers and D.J. Lipman, 1990. Basic local alignment search tool. J. Mol. Biol., 215: 403-410.

Bowie, J.U., R. Luthy and D. Eisenberg, 1991. A method to identify protein sequences that fold into a known three-dimensional structure. Science, 253: 164-170.

Cavicchioli, R., K.S. Siddiqui, D. Andrews and K.R. Sowers, 2002. Low-temperature extremophiles and their applications. Curr. Opin. Biotechnol., 13: 253-261.

Cavicchioli, R. and K.S. Siddiqui, 2004. Cold-Adapted Enzymes. In: Enzyme Technology, Pandey, A., C. Webb, C.R. Soccol and C. Larroche (Eds.). Asia Tech Publishers, New Delhi, India, pp: 608-631.

Colovos, C. and T.O. Yeates, 1993. Verification of protein structures: Patterns of nonbonded atomic interactions. Prot. Sci., 2: 1511-1519.

Dereeper, A., V. Guignon, G. Blanc, S. Audic and S. Buffet *et al*., 2008. Phylogeny.fr: Robust phylogenetic analysis for the non-specialist. Nucl. Acids Res., 36: W465-W469.

Derewenda, U., A.M. Brzozowski, D.M. Lawson and Z.S. Derewenda, 1992. Catalysis at the interface: The anatomy of a conformational change in a triglyceride lipase. Biochemistry, 31: 1532-1541.

Eggert, T., G. van Pouderoyen, G. Pencreac'h, I. Douchet, R. Verger, B.W. Dijkstra and K.E. Jaeger, 2002. Biochemical properties and three-dimensional structures of two extracellular lipolytic enzymes from *Bacillus subtilis*. Colloids Surf. B: Biointerfaces, 26: 37-46.

Egloff, M.P., F. Marguet, G. Buono, R. Verger, C. Cambillau and H. van Tilbeurgh, 1995. The 2.46. ANG. Resolution structure of the pancreatic lipase-colipase complex inhibited by a C11 alkyl phosphonate. Biochemistry, 34: 2751-2762.

Feller, G. and C. Gerday, 2003. Psychrophilic enzymes: Hot topics in cold adaptation. Nat. Rev. Mcrobiol., 1: 200-208.

Friesner, R.A., J.L. Banks, R.B. Murphy, T.A. Halgren and J.J. Klicic *et al*., 2004. Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. J. Med. Chem., 47: 1739-1749.

Friesner, R.A., R.B. Murphy, M.P. Repasky, L.L. Frye and J.R. Greenwood *et al*., 2006. Extra precision glide: Docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. J. Med. Chem., 49: 6177-6196.

Gerday, C., M. Aittaleb, M. Bentahir, J.P. Chessa and P. Claverie *et al*., 2000. Cold-adapted enzymes: From fundamentals to biotechnology. Trends Biotechnol., 18: 103-107.

Gilbert, E.J., 1993. *Pseudomonas* lipases: Biochemical properties and molecular cloning. Enzyme Microb. Technol., 15: 634-645.

Greenwood, J.R., D. Calkins, A.P. Sullivan and J.C. Shelley, 2010. Towards the comprehensive, rapid and accurate prediction of the favorable tautomeric states of drug-like molecules in aqueous solution. J. Comput. Aided Mol. Des., 24: 591-604.

Grochulski, P., Y. Li, J.D. Schrag and M. Cygler, 1994. Two conformational states of *Candida rugosa* lipase. Protein Sci., 3: 82-91.

Guex, N. and M.C. Peitsch, 1997. SWISS-MODEL and the Swiss-PDB viewer: An environment for comparative protein modeling. Electrophoresis, 18: 2714-2723.

Guo, Z., U. Mohanty, J. N oehre, T.K. Sawyer, W. Sherman and G. Krilov, 2010. Probing the α-helical structural stability of stapled p53 peptides: Molecular dynamics simulations and analysis. Chem. Biol. Drug Des., 75: 348-359.

Halgren, T.A., R.B. Murphy, R.A. Friesner, H.S. Beard, L.L. Frye, W.T. Pollard and J.L. Banks, 2004. Glide: A new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. J. Med. Chem., 47: 1750-1759.

Halgren, T., 2007. New method for fast and accurate binding-site identification and analysis. Chem. Biol. Drug Des., 69: 146-148.

Halgren, T.A., 2009. Identifying and characterizing binding sites and assessing druggability. J. Chem. Inf. Mod., 49: 377-389.

Jacobson, M.P., R.A. Friesner, Z. Xiang and B. Honig, 2002. On the role of the crystal environment in determining protein side-chain conformations. J. Mol. Biol., 320: 597-608.

Jacobson, M.P., D.L. Pincus, C.S. Rapp, T.J.F. Day, B. Honig, D.E. Shaw and R.A. Friesner, 2004. A hierarchical approach to all-atom protein loop prediction. Proteins: Struct. Funct. Bioinf., 55: 351-367.

Jorgensen, W.L. and J. Tirado-Rives, 1988. The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. J. Am. Chem. Soc., 110: 1657-1666.

Jorgensen, W.L., D.S. Maxwell and J. Tirado-Rives, 1996. Development and testing of the opls all-atom force field on conformational energetics and properties of organic liquids. J. Am. Chem. Soc., 118: 11225-11236.

Joseph, B., P.W. Ramteke and G. Thomas, 2008. Cold active microbial lipases: Some hot issues and recent developments. Biotechnol. Adv., 26: 457-470.

Kitchen, D.B., H. Decornez, J.R. Furr and J. Bajorath, 2004. Docking and scoring in virtual screening for drug discovery: Methods and applications. Nat. Rev. Drug Discovery, 3: 935-949.

Kumar, V., C.S. Kumar, G. Hari, N.K. Venugopal, P.D. Vijendra and G.B. Basappa, 2013. Homology modeling and docking studies on oxidosqualene cyclases associated with primary and secondary metabolism of *Centella asiatica*. Springer Plus., Vol. 2.

Lovell, S.C., I.W. Davis, W.B. Arendall III, P.I.W. de Bakker and J.M. Word *et al*., 2002. Structure validation by Cα geometry: φ,ψ and Cβ deviation. Proteins: Struct., Funct., Genet., 50: 437-450.

Marsden, R.L. and C.A. Orengo, 2008. Target selection for structural genomics: An overview. Methods Mol. Biol., 426: 3-25.

Ollis, D.L., E. Cheah, M. Cygler, B. Dijkstra and F. Frolow *et al*., 1992. The alpha/beta hydrolase fold. Protein Eng., 5: 197-211.

Raval, A., S. Piana, M.P. Eastwood, R.O. Dror and D.E. Shaw, 2012. Refinement of protein structure homology models via long, all-atom molecular dynamics simulations. Proteins: Struct. Funct. Bioinformatics, 80: 2071-2079.

Rost, B., 1999. Twilight zone of protein sequence alignments. Protein Eng., 12: 85-94.

Sastry, G.M., M. Adzhigirey, T. Day, R. Annabhimoju and W. Sherman, 2013. Protein and ligand preparation: Parameters, protocols and influence on virtual screening enrichments. J. Comput. Aided Mol. Des., 27: 221-234.

Shelley, J.C., A. Cholleti, L.L. Frye, J.R. Greenwood, M.R. Timlin and M. Uchimaya, 2007. Epik: A software program for pK_a prediction and protonation state generation for drug-like molecules. J. Comp. Aided Mol. Des., 21: 681-691.

Shivakumar, D., J. Williams, Y. Wu, W. Damm, J. Shelley and W. Sherman, 2010. Prediction of absolute solvation free energies using molecular dynamics free energy perturbation and the OPLS force field. J. Chem. Theory Comput., 6: 1509-1519.

Siddiqui, K.S. and R. Cavicchioli, 2005. Improved thermal stability and activity in the cold-adapted lipase B from *Candida antarctica* following chemical modification with oxidized polysaccharides. Extremophiles, 9: 471-476.

Sinchaikul, S., B. Sookkheo, S. Phutrakul, Y.T. Wu, F.M. Pan and S.T. Chen, 2001. Structural modeling and characterization of a thermostable lipase from *Bacillus stearothermophilus* P1. Biochem. Biophys. Res. Commun., 283: 868-875.

The UniProt Consortium, 2012. Reorganizing the protein space at the Universal Protein Resource (UniProt). Nucleic Acids Res., 40: D71-D75.

Tripathi, M.K, U. Roy, U.K. Jinwal, S.K. Jain and P.K. Roy, 2004. Cloning, sequencing and structural features of a novel *Streptococcus* lipase. Enzyme Microb. Technol., 34: 437-442.

Tyndall, J.D., S. Sinchaikul, L.A. Fothergill-Gilmore, P. Taylor and M.D. Walkinshaw, 2002. Crystal structure of a thermostable lipase from *Bacillus stearothermophilus* P1. J. Mol. Biol., 323: 859-869.

Van Tilbeurgh, H., M.P. Egloff, C. Martinez, N. Rugani and R. Verger, 1993. Interfacial activation of the lipase-procolipase complex by mixed micelles revealed by X-ray crystallography. Nature, 362: 814-820.

Yang, A.S. and B. Honig, 2000. An integrated approach to the analysis and modeling of protein sequences and structures. III. A comparative study of sequence conservation in protein structural families using multiple structural alignments. J. Mol. Biol., 301: 691-711.

Zhang, N., W.C. Suen, W. Windsor, L. Xiao, V. Madison and A. Zaks, 2003. Improving tolerance of *Candida antarctica* lipase B towards irreversible thermal inactivation through directed evolution. Prot. Eng., 16: 599-605.