BIOINFORMATICS APPROACH TO UNDERSTAND AND EXPLOIT THE POTENTIAL OF LIPASES

A Thesis
submitted in Partial
Fulfillment of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

By

P. SARAVANAN

DEPARTMENT OF BIOTECHNOLOGY
INDIAN INSTITUTE OF TECHNOLOGY, GUWAHATI
GUWAHATI-781039, ASSAM, INDIA
SEPTEMBER 2013

BIOINFORMATICS APPROACH TO UNDERSTAND AND EXPLOIT THE POTENTIAL OF LIPASES

A Thesis
submitted in Partial
Fulfillment of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

By

P. SARAVANAN

(Roll No. 09610625)

DEPARTMENT OF BIOTECHNOLOGY
INDIAN INSTITUTE OF TECHNOLOGY, GUWAHATI
GUWAHATI-781039, ASSAM, INDIA
SEPTEMBER 2013

DEDICATED TO

UMAKANTHAM and ARUNACHALAM

&

TEACHERS, MENTORS and NATURE

INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI

DEPARTMENT OF BIOTECHNOLOGY

STATEMENT

I do hereby declare that the matter embodied in this thesis entitled "**Bioinformatics Approach to Understand and Exploit the Potential of Lipases**" is the result of investigations carried out by me in the Department of Biotechnology, Indian Institute of Technology Guwahati, Guwahati, India, under the supervision of Dr. Sanjukta Patra and Dr. Vikash Kumar Dubey.

In keeping with the general practice of reporting scientific observations, due acknowledgements have been made wherever the referred work is based on the findings of other researchers.

Date: October, 2013. **P. Saravanan**

INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI

DEPARTMENT OF BIOTECHNOLOGY

CERTIFICATE

It is certified that the work described in this thesis entitled "**Bioinformatics Approach to Understand and Exploit the Potential of Lipases**" by P. Saravanan for the award of degree of Doctor of Philosophy is an authentic record of the results obtained from the research work carried out under our supervision in the Department of Biotechnology, Indian Institute of Technology Guwahati, India, and this work has not been submitted elsewhere for a degree.

Dr. Sanjukta Patra
Assistant Professor
(Thesis Supervisor)
Department of Biotechnology
Indian Institute of Technology Guwahati
Guwahati -781039, India

Dr. Vikash Kumar Dubey
Associate Professor
(Thesis Co-supervisor)
Department of Biotechnology
Indian Institute of Technology Guwahati
Guwahati - 781039, India

ACKNOWLEDGEMENTS

With my deepest sense of appreciation, I express my heartiest acknowledgement to my research supervisors, Dr. Sanjukta Patra and Dr. Vikash Kumar Dubey, Department of Biotechnology, for their belief on me as well as the encouragement throughout my research. I must acknowledge the freedom to think, express, plan and execute my research.

I am thankful to Indian Institute of Technology Guwahati, India and its Department of Biotechnology for providing me the opportunity and infrastructure to carry out my doctoral research. I am also thankful to Department of Information Technology, Government of India for research fellowship.

I am grateful to my doctoral committee members, Dr. Ranjan Tamuli, Dr. Vishal Trivedi and Dr. S.R.M. Prasanna for their constructive suggestions. I duly acknowledge Prof. Shyam Sundar, Banaras Hindu University, India who provided Leishmania donovani (BHU-1081) promastigote cells. I owe my gratitude to the successive Heads of Department of Biotechnology, IIT Guwahati, Prof. Arun Goyal and Dr. V. Venkata Dasu, for providing me with the departmental facilities to carry out my research work. I would also like to thank the teachers, technical and non- technical staffs of Department of Biotechnology, IIT Guwahati.

I am grateful to Dr. Pushpa Agarwal, CSIR-IMTECH, Dr. Gopi Mohan, NIPER-SAS Nagar, Dr. Mark S. Johnson, Abo Akademi University, Finland and Dr. A. T. Khan, IIT Guwahati who provided opportunity to work with them and opened the windows for my research career. I am what I am is because of the efforts and inspirations of my teachers and mentors from University of Madras especially Dr. Niranjali Devaraj, Dr. Yathindra, Dr. Vasantha Pattabhi, Dr. Anbarasu and Dr. Sivaramakrishnan as well as Dr. Subramanian, CSIR-CLRI, Dr. Kumudhini, Dr. Swarnalatha, Mrs. Savitha and Dr. Mikko Vianio. I need to specially thank Dr. Swarnalatha who questioned me ‘what you want to do after 5 years and 10 years’. The question transfixed me to pursue the meaning of my life. I specially thank Gajalakshmi who introduced me to Bioinformatics which changed the course of my life.

I am thankful to my lab members Debamitra, Bhaskar, Sonali, Nivedita, Faheem, Shilpa and ex-members for their help and cooperation. I am also thankful to Prakash for his help with *Leishmania* culture related studies. I like to extend thanks to my seniors Suhail, Saurabh, Richa, Kausik VenkataSatish and Abhay for their inspiration and guidance.

I would also like to offer special thanks to my friends Sathish, Balamurugan, Vamsi, Vijayaraj, Vidya, Shikhar, Mahendra, Archana, Vivek, Vinod, Sidick, Himangshu, Ankita and Nidhi for their support, patience, and time. I also thank the unsung friends who pat my shoulder and offered theirs in the toughest times. I wish to express my gratitude with love to my family especially Arunachalam, Umakantham, Dhanalakshmi and Kalyani who complete the circle of my life.

Date: October, 2009. **P. Saravanan**

ABSTRACT

Growing knowledge and application of lipases in industries have prioritized the necessity to develop lipases with better stability and substrate specificity. In therapeutics, these enzymes can also be good drug targets against pandemic diseases that rely on host pathogen interactions. A thorough literature review indicated that among the large number of lipases, only few species have been shown to have adequate stability and biosynthetic capabilities for routine use in industries. Thus lipases still require further attention and holistic understanding. Moreover databases emphasizing the applications dedicated to triacylglycerol acylhydrolases (E.C 3.1.1.3; a key lipase) are need of the hour. Furthermore, there is also need for a more comprehensive database with specific tool that can classify uncharacterized lipases into sub-families so that their functional characteristics along with phylogeny can be indicated and compared to known ones.

The proposed work discusses about the efforts made to understand and exploit the lipases with help of integrated bioinformatics approaches. The creation of database for triacylglycerol acylhydrolases (TLDB) implemented with non-alignment based methodology for sub-family classification distinguishes our database from existing ones. A sequence-based workflow has been implemented to identify subfamily and conserved patterns of new lipase(s). Key research areas identified from our TLDB were: (i) identification and betterment of lipases with industrial potential and (ii) exploiting lipases as drug targets to combat diseases. Bioinformatics analysis on the representative lipases chosen from our database has led to the identification of respective mutagenesis hotspots. The *in silico* mutants of representative lipases have been validated through molecular dynamics simulations and found with improved structural stability, thereby suitable for their use in industrial applications. This opens up several possibilities to rationally engineer the lipases. The other significant sub-family ascertained in TLDB were lipases from pathogenic source. This urged us to pinpoint and recognize the lipases as drug targets to combat infectious diseases. The current study also focused on targeting key lipases to combat leishmaniasis and tuberculosis, caused by lipid-scavenging pathogens that exist in several physiological states within the host. Structure-based drug discovery (SBDD) approach was implemented to identify potential inhibitors against the identified drug targets Rv0183 (monoglyceride lipase) and Rv3802c (cell wall lipase) of

Mycobacterium tuberculosis. Furthermore, potential dual inhibitors targeting mycolic acid synthesis and host lipid catabolism simultaneously were also identified. The identified drug target-specific and multi-targeting inhibitors might pave the way to combat tuberculosis which might reduce therapy time and possibility of pathogen acquiring drug resistance. In addition, lipases participate in digenetic life cycle and infection of leishmanial species, the causative agent of leishmaniasis especially in India. SBDD approach identified leishmania-specific potential inhibitors followed by its wet lab validation revealed them as novel candidates to treat leishmaniasis. Integrated SBDD approach leads to the design of new molecules which could be effective against the respective drug targets in terms of predicted free energy of binding. The thesis has thus made the utmost effort to understand and exploit the potential applications of lipases.

Key Words: Lipases, Enzymes, Bioinformatics, Databases, Sub-Family Classification, *in silico* Mutagenesis, Structure-based Drug Discovery, Drug Targets, Potential Inhibitors, Tuberculosis, Leishmaniasis

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

α/β	alpha/beta
ΔG	Predicted Free Energy of Binding
3D	Three Dimension
Å	Angstroms
AuPosSOM	Automated Analysis of Poses using Self-Organizing Map
BLAST	Basic Local Alignment Search Tool
DBMS	Database Management Systems
DMEM	Dulbecco's Modified Eagle's Medium
DOPE	Discrete Optimized Protein Energy
DOTS	Directly Observed Treatment, Short-course
DSSP	Dictionary of Secondary Structure for Proteins
EC	Enzyme Classification
ElrF	Envelope Lipids Regulation Factor
EMC	Enzyme Modified Cheese
ESTHER	ESTerases and alpha/beta Hydrolase Enzymes and Relatives
FAQ	Frequently Asked Questions
FBS	Fetal Bovine Serum
FLIP	Lipases involved in Flagellar Biosynthesis
LED	Lipase Engineering Database
LGA	Lamarckian Genetic Algorithm
LIB	Lipid Inclusion Bodies
LINCS	Linear Constraint Solver
LIPABASE	Lipase Database
mAGP	Mycolyl Arabinogalactan Peptidoglycan
MAST	Motif Alignment Search Tool

MD	Molecular Dynamics
MDR	Multi Drug Resistance
MELDB	Microbial Esterases and Lipases Database
MEME	Multiple EM for Motif Elicitation
MGL	Monoglyceride Lipase
MOLPDF	Probability Density Function
MTT	3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide
NCCS	National Centre for Cell Science
nm	nanometre
PBS	Phosphate-buffered Saline
PDB	Protein Data Bank
PME	Particle Mesh Ewald
PML	*Pseudomonas mendocina* lipase
PLRP	Pancreatic Lipase Related Protein
Rg	Radius of Gyration
RMSD	Root-Mean Square Deviation
RMSF	Root-Mean Square Fluctuations
Rm-TGL	Lipase from *Rhizomucor miehei*
SAL3	Lipase from *Staphylococcus aureus*
SASA	Solvent Accessible Surface Area
SBDD	Structure-Based Drug Discovery
SHL	Lipase from *Staphylococcus hyicus*
SPC	Simple Point Charge
SXL2	Lipase from *Staphylococcus xylosus*
TAG	Triacylglycerols
TE	Trans-esterification
THL	Tetrahydrolipstatin

WHO	World Health Organization
XDR	Extreme Drug Resistance
ZYH	reversible inhibitor of human MGL

Single/Three letter code for amino acids:

A/Ala	Alanine
C/Cys	Cysteine
D/Asp	Aspartic acid
E/Glu	Glutamic acid
F/Phe	Phenylalanine
G/Gly	Glycine
H/His	Histidine
I/Ile	Isoleucine
K/Lys	Lysine
L/Leu	Leucine
M/Met	Methionine
N/Asn	Asparagine
P/Pro	Proline
Q/Gln	Glutamine
R/Arg	Arginine
S/Ser	Serine
T/Thr	Threonine
V/Val	Valine
W/Trp	Tryptophan
Y/Tyr	Tyrosine

Chapter I

REVIEW OF LITERATURE

Chapter I

Review of Literature

1.1. Abstract

This chapter presents a review of the existing knowledge focusing on lipases. They are ubiquitous in nature carrying out lipid hydrolysis. Structurally they are composed of α/β hydrolase fold and the active site consists of a nucleophilic serine, an acid (aspartate or glutamate), and histidine residues. The active site can be identified through the conserved pentapeptide GXSXG motif [G: Glycine, S: serine, X: any amino acid]. Triacylglycerol lipases have various applications. For this purpose a comprehensive database on triacylglycerol acylhydrolases is a need and it is necessary to classify them into sub-families to derive conclusions about the new and uncharacterized lipases as it can shed light into their significance in comparison to the existing ones. Further it can help in designing experiments to enhance the stability of such lipases through protein engineering techniques. The industrial and pharmacological applicability of lipases have also been brought forward along with a need to understand and explore lipases. It is interesting to note that lipases are involved in the lipid metabolism of pathogens which makes them important candidates for drug development. This area still remains unexplored and requires further attention.

1.2. Introduction

Lipases are lipid-degrading enzymes found ubiquitously in nature (Sharma *et al*., 2001). They have been isolated from various sources: bacterial (45%), fungal (21%), animal (18%), plant (11%) and algal (3%) (Patil *et al*., 2011). They catalyze important biological reactions such as hydrolysis of acylglycerols and other complex esters (Gupta et al., 2004; Ali *et al*., 2012). They can be well understood based on the lipid substrate on which they act. A unique property of lipase is that it generally acts at the organic-aqueous interface, catalyzing the hydrolysis of ester bonds and releasing fatty acids (Pereira *et al*., 2003; Leal *et al*., 2002). In the presence of organic solvents, lipases are also capable of performing ester synthesis (Sharma *et al*., 2001). Such features make them an important group of biocatalysts with physiological and industrial potency as well as they represent one of the major products in the global industrial enzyme market with high growth potential (Verger, 1997; Haki and Rakshit, 2003). The various industrial applications of lipases include organic synthesis, modification of fats, flavor enhancement in food, and resolution of racemic mixtures (Sharma *et al*., 2001). Owing to the potential of lipases, there has always been a continuous effort towards enumerating the functional activity and stability of lipases. Thus research on lipases until now is focused on structural characterization, kinetics, cloning and sequencing of lipase genes (Alberghina *et al.*, 1991; Bornscheuer, 2000) but still there is a need for further exploration on the finer details about the biological and physicochemical features of lipases as this can develop future scopes for obtaining and engineering specific lipases with ease and for industrial benefits.

1.3. Biological and physiochemical features of lipases

The catalytic mechanism

A unique property of lipase is that it generally acts at the organic-aqueous interface, catalyzing the hydrolysis of ester bonds and releasing fatty acids (Pereira *et al*., 2003; Leal *et al*., 2002). Generally the amount of oil available at the interface determines the activity of the lipase (Verger *et al*., 1997). In the absence of water they carry out the reverse reaction of trans-esterification (TE). The reactions catalyzed by lipases have been presented in Figure 1.1.

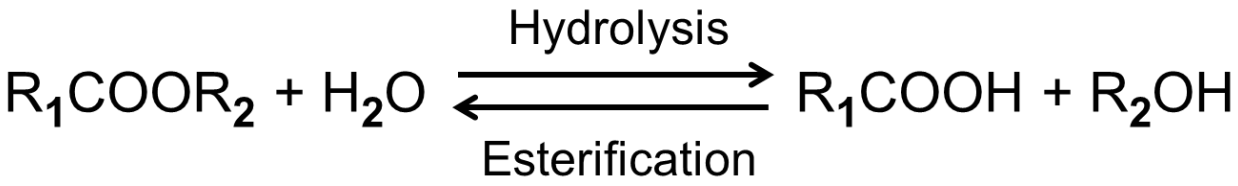


Figure 1.1: Catalysis mediated by lipases

The catalytic mechanism of lipases is often described by the ping-pong bi-bi model, releasing each product between each addition of substrate (Paiva *et al.*, 2000). Figure 1.2 illustrates the catalytic mechanism in detail. It involves four steps (Jaeger *et al.*, 1999), (i) lipid binding and the activation of a nucleophilic serine residue by a neighbouring active histidine, to which a proton from the serine hydroxyl group is transferred, (ii) attack by the oxygen atom of the serine hydroxyl group on the activated carbonyl carbon of the susceptible lipid ester bond and formation of oxyanion hole and a tetrahedral intermediate, (iii) the next stage is the deacylation step, in which an incoming water molecule hydrolyses the covalent intermediate (acyl enzyme) and the acid component of the substrate is esterified to the enzyme‘s serine residue and (iv) The histidine residue donates its additional proton to the oxygen atom of the active serine residue, which breaks the ester bond between serine and the acyl component, and releases the acyl product.

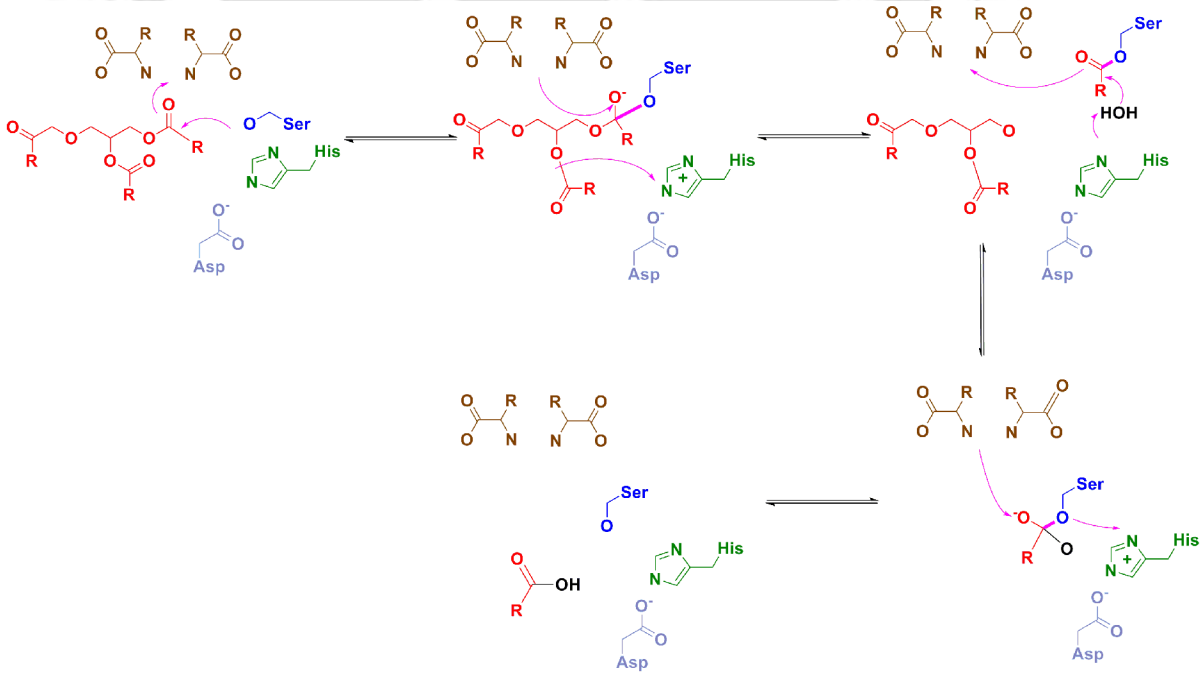


Figure 1.2: The ping pong model for lipase catalysis

The lid theory is the accepted model for the explanation of their catalytic mechanism. The lid is an amphiphilic region and it exposes the binding site to substrates in a water-lipid interface when the enzyme becomes active (Pleiss *et al.*, 1998; Balcáo *et al*., 1996). However, *Candida antarctica* lipase B, *Bacillus subtilis* lipase A are exceptions of the lid theory since it lacks the lid domain probably due to large hydrophobic surface surrounding the entrance channel of the active site (Uppenberg, *et al.*, 1994; Martinelle *et al.*, 1995).

Lipases show specificity from short-chain to long-chain esters. These substrate specificities vary significantly dependent on the structures of active sites. Some lipases like the *Candida antarctica* A show regioselectivity for ester bond at position 2 of the triglycerides while some, e.g. *Rhizopus oryzae* lipase have the specificity for 1 and 3 positions (Douchet *et al.*, 2003; Song *et al*., 2008; Du *et al.*, 2005). The versatility of lipases in substrate specificity and catalysis led to further research on the physicochemical aspects of lipases.

Sequence and structure

The number of database entries on lipase sequences in Uniprot as on July 2013 is 2165 and there are 120 entries of lipase structures in Protein Data Bank till date. The lipases have α/β hydrolase fold featuring parallel beta sheet with segregated alpha helices (Ollis *et al*., 1992; Lenfant *et al*., 2012) (Figure 1.3A). The active site consists of a nucleophilic serine, an acid (aspartate or glutamate), and histidine (Fojan *et al*., 2000). The conserved 'nucleophile elbow' formed by a γ-like turn, in which the catalytic serine resides in GXSXG pentapeptide motif (Jaeger and Eggert, 2002). The nucleophile elbow helps to stabilize the tetrahedral intermediate with the help of 'oxyanion hole' residues followed by the formation of acyl-enzyme complex. Then an attack of nucleophile forms tetrahedral intermediate followed by the release of product (Bornscheuer, 2002). Substrate diversity arises due to the variations in the C-terminal region which forms most of the substrate binding pocket. Experimentally determined structure of *Burkholderia cepacia* lipase in the open conformation with a substrate analog (PDB Id: 4LIP) shows that the active site has four binding pockets: oxyanion hole and three pockets which accommodate the substrate (Figure 1.3B) (Lang *et al*., 1998). The study also revealed that van der Waals interactions are the main forces that hold the substrate analog in binding pocket. Orientation of the acyl chain of substrate at their

respective pockets dictates selectivity of the *B. cepacia* lipase (Henke *et al*., 2002). Physiochemical nature of the substrate is the key property that differentiates the lipases from non-lipolytic esterases. Differentiating lipases from esterases is still an active area research (Ali *et al*., 2012).

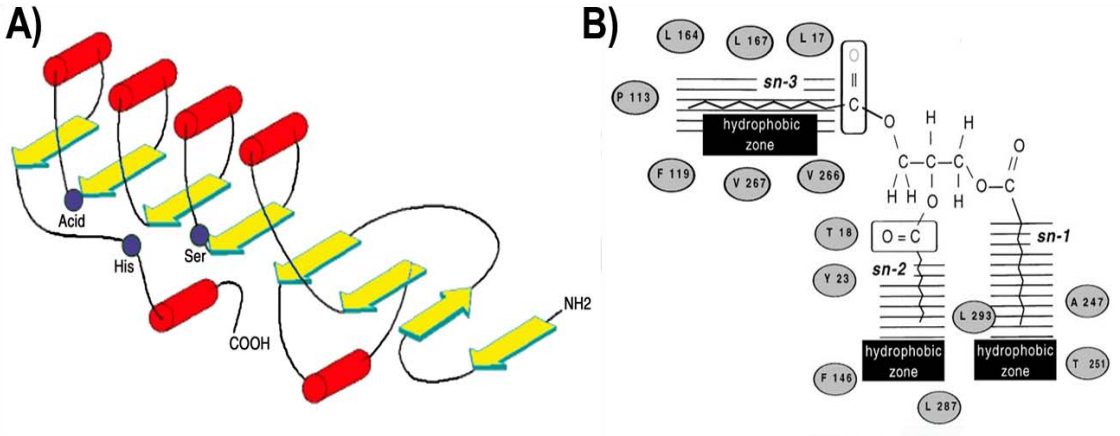


Figure 1.3: A) Canonical α/β hydrolase fold with catalytic residues depicted as dark circles. **B)** Substrate binding pocket of lipase (model system: *B. cepacia* lipase).

1.4. Databases on lipases

From the applications and versatility of lipases, it is clear that multitude of data are being generated on these enzymes. Therefore to understand and exploit the potential of lipases, a dedicated database with free access is a necessity to make more comprehensive and useful research. There are several databases in the public domain that provide information on lipases. They are the 'ESTerases and alpha/beta Hydrolase Enzymes and Relatives'(ESTHER) database, the Lipase Engineering Database' (LED), 'LIPAsedataBASE' (LIPABASE) and the 'Microbial Esterases and Lipases DataBase' (MELDB). The lipase entries in each of the database as on July 2013 are: MELDB: 455 entries; LIPABASE: 133 entries; and LED: 21090 lipase and lipase-like proteins. Among all these databases only Lipabase and MELDB are dedicated to triacylglycerol (TAG) acylhydrolases, a key lipase. None of these databases provide complete information and efficient clustering on TAG acylhydrolases. A dedicated database for TAG acylhydrolases is need for the time to explore its nature for applications in the field of pharmaceutics and other industrial applications.

1.5. Industrial Application of Lipases

The above mentioned catalytic advantages entitle lipases wide potential applications in many industries. Novozyme 435 is probably most well-known immobilized lipase and has been widely studied in academic research. The factors for the success of lipases as commercial enzymes are their versatility, stability, specificity, selectivity, no requirement of cofactor, availability of source and green chemistry (Salihu and Alam, 2012). Although lipases have many interesting applications in leather, biodiesel, dairy, textile, cosmetic and paper industries as well as in the synthesis of fine chemicals, agrochemicals and new polymeric materials; most significant industrial applications of lipases have been found in food, detergent, and pharmaceutical sectors (Ray, 2012) (Table 1.1).

Table 1.1: Industrial applications of lipases

Industry	Source	Catalytic Action	Product
Pharmaceutical	*Chromobacterium viscosum*	Hydrolysis & TE	Lipids & digestive aids Drug targets
Chemical	*Candida rugosa*	Enantio-selectivity	Biologically active analogues
Detergent	*Pseudomonas mendocina*	Synthesis & Hydrolysis	Synthesis and removal of cleaning agents
BioFuel	*Candida antarctica*	Hydrolysis & TE	BioFuels
Food & Beverages	*Geotrichum candidum*	Hydrolysis & TE	Quality & flavour Improvement of health food and bakery products
Cosmetics	*Candida cylindracea*	Synthesis	Emulsifiers and moisturizing agents
Dairy	*Rhizopus niveus*	Hydrolysis	Flavour agents, processing of cheese, butter, cocoa butter and margarine
Leather	*Thermomyces lanuginose*	Hydrolysis	Leather products
Paper	*Candida rugosa*	Hydrolysis	Paper products

The importance of organic solvent stable and thermostable lipases for different applications has also been growing rapidly (Haki, 2003). Their industrial uses still remain limited by their high production costs, commercialization in small amounts, and low performance of some lipase-mediated processes (Houde *et al*., 2004). Different methodologies have been

investigated from rational engineering to random mutagenesis towards increasing the stability, catalytic efficiency and organic solvent tolerance of lipases (Kourist *et al*., 2010). Thus there is a scope of further exploration of lipases.

1.6. Therapeutic importance of Lipases

Apart from industrial applications, lipases have been reported to have pharmacological importance. Russell *et al*. reported that the lipases of pathogens notably *Mycobacterium tuberculosis* have attracted considerable interest since they supply precursors for the lipid inclusion bodies and cell wall, key components for survival and virulence of pathogens (Russell *et al*., 2010). Lipases have also been reported to (i) elicit host immune-responsiveness and (ii) alter the host phago-lysosome resulting in favorable conditions for parasite growth, survival, and development (Shakarian *et al*., 2010). Current situation demand the need of new druggable targets and new drugs to reduce the therapy time and drug resistance of several infectious diseases (Nachega and Chaisson, 2003), which can be achieved by targeting both lipid metabolism and other key metabolisms simultaneously. Understanding the role of lipases in key biological process of pathogens will pave the way to combat tuberculosis, leishmaniasis and even HIV co-infection. Lipases were involved in the lipid metabolism especially lipid catabolism and also participates with other metabolisms of pathogens. It has been suggested that these lipids provide carbon source during the chronic infection phase, dormancy and also for the reactivation (Garton *et al.*, 2002; Daniel *et al.*, 2004). Also it is known that these lipids act as building blocks for virulent lipids (Russell *et al.*, 2010). Lipases, lipase-like enzymes and other lipid catabolizing enzymes likely play vital roles in the host lipid degradation process, virulence and host immune-responses in several pathogens (Singh *et al.*, 2010; Daleke *et al.*, 2011; Shakarian *et al.*, 2010). Thus such lipases can be targeted for drug development. Research in this area is still puerile.

This chapter brings out the importance of triacylglycerol lipases. The need for the creation of a database and a classification system of lipases has been brought out. Lipases were seen to be important both industrially and pharmacologically. Objectives were laid on the basis of the lacuna and the following chapters shall methodologically describe the further exploration of lipases in industry and therapeutics through *in silico* approaches.

1.7. Bioinformatics Techniques

1.7.1. Homology Modeling

Homology modeling is a computational technique to predict or model three dimensional structures of biological macromolecules (especially proteins) based on experimentally determined structures. Homology modeling builds a three dimensional (3D) model of the biological molecule of interest, while most of the available programs do not include its hydrogen atoms. As of July 2013, only 0.1% proteins have experimentally determined structures which reflect the lack of structural information of proteins (52163 protein structures in PDB while 41991850 protein sequences in UniProtKB). Homology modeling provides the structural information which could be utilized in several disciplines of biology, for instance, studying structural dynamics, mutational effects, even in drug discovery. The protein to be modeled is referred as 'target' and the protein with experimental determined structure is referred as 'template'. The key principle behind homology modeling is (i) 'Anfinsen's dogma' which states that the native structure of globular proteins is determined by its amino acid sequence and (ii) the protein structure is more conserved rather than a sequence in evolutionary process (divergent evolution). The three main approaches of homology modeling:

- Comparative modeling (to model homolog protein)
- Threading (to model homolog or analog protein)
- Ab initio modeling (when homolog structure is not available)

Most of the programs/web servers implement more than one approach to model the protein structure. The understanding of modeling programs/web servers and its strengths as well as weakness is prerequisite to develop reliable models with the knowledge towards structural biology of the target and template. As in Bioinformatics, homology modeling depends on the data availability and its quality, here of the experimentally determined structures. Homology modeling is a multi-step process which is self-explanatory. The key steps are grouped conceptually and are follows.

- Template recognition
- Initial alignment generation
- Backbone model building
- Loop modeling and side chain modeling
- Model optimization and validation

The first step is considered as important to predict reliable model which is dependent on the data availability and user's knowledge. The idea is to identify homolog protein of target with known structure followed by the determination of structurally conserved regions. The choice of template depends upon the purpose of modeling the target. It should be verified that the target and template were evolutionarily related with significant identity and similarity. The difference between target and template in the form of insertions or deletions (indel/gaps) in alignment reduces the reliability of modeled structure. The alignment should be refined to obtain optimal alignment with biological significance.

The backbone modeling is taken care by modeling programs which utilize the atomic coordinates of template. The reliability of backbone of target is inversely proportional to indel/gaps observed in the alignment. The other concern is existing challenges in loop and side chain modeling which is often worsened by the presence of many indel/gaps in alignment. The presence of indel of more than seven residues should be avoided due to the limitations of existing methodologies. Even though homology modeling is a well-developed method, the scope of research still exists in loop and side chain modeling to obtain biologically optimal 3D model of the target. Loop modeling is addressed by knowledge-based or energy-based methods while side chain modeling utilizes rotamer libraries. The artifacts introduced by the modeling algorithms should be optimized to yield best model. This is carried out to generate energetically favourable 3D conformation of using energy minimization and/or molecular dynamics simulations which takes care of bad contacts/steric clashes, side chains and loops. Several empirical and comparative methods are available to validate the predicted model with respect to distribution of fold, topology, bonded and non-bonded interactions. Critical assessment of techniques for protein structure prediction (CASP) community evaluate homology modeling methods and models which also shed light

evaluation of predicted model of target. To summarize, successful homology modeling depends upon the wise usage of existing knowledge of target and template as well as modeling and structure validation programs/web servers with emphasis on template identification and alignment generation.

1.7.2. Molecular Dynamics Simulation

Molecular dynamics (MD) simulations is one of the main tools in the theoretical study of biological molecules. It is based on the principle of statistical mechanics that calculates the time dependent behavior of the molecule of interest. MD simulations aid to understand and estimate the equilibrium and conformational changes of biomolecules notably proteins that complement experimental techniques. MD has several applications in diverse areas notably study of proteins such as stability, folding, molecular recognition and transport. Though MD is well accepted among researchers, it is still in the process of development to mimic real life solutions and computationally effective in terms of time. The fundamental idea behind MD simulation is to calculate the classical evolution of system in time through the integration of Newton's law of motion. MD explores the macroscopic properties of the system through microscopic simulations to study its thermodynamic and/or kinetic properties. The detailed information on MD is not in the context of the thesis and can be found elsewhere. The key concepts of MD simulation are outlined below:

***Ensemble*:** The biomolecule of interest is studied with the collection of different microscopic states but have an identical macroscopic state. Ensembles can be studied with different characteristics as: (i) micro-canonical [NVE], (ii) canonical [NVT], (iii) Isobaric-Isothermal Ensemble [NPT] and (iv) Grand canonical Ensemble [mVT] where the terms in square brackets kept were kept fixed. N, V, E, T and m represents number of atoms, volume, energy, pressure and chemical potential respectively.

Potential energy: The potential energy of system establishes the relative stabilities of different three dimensional structures. It is studies as a function of the atomic positions of all the atoms in the system. The equations of motion are solved numerically with integration algorithms such as Verlet, Leap-Frog, Velocity Verlet and Beeman's algorithm.

Solvent models: To understand the interaction and influence of the environment as well as to mimic to the real life, the biomolecules are studied in the presence of solvents generally water molecules. They are studied as implicit (use of dielectric constant) or explicit (inclusion of water molecules) treatment.

Several software/packages are available to carry out MD simulations such as AMBER, GROMACS, NAMD and the user has to choose them according to their need. The conceptual steps of MD simulation are:

i. Prepare initial coordinates and topology
ii. Solvate and neutralize the system in a periodic boundary conditions
iii. Energy minimization to avoid artifacts.
iv. Solvent equilibration followed by system equilibration
v. MD simulation

1.8. Scope of the Thesis

Though lipases are versatile and hold high industrial applicability they are still not fully explored. Triacylglycerol lipases have various applications and a database dedicated exclusively to Triacylglycerol lipases is a need. Moreover there is a need to classify such lipases into smaller clusters which can reveal not only their physicochemical characteristics but also shed light on their biological, industrial and therapeutic significance. This can help explore the benefits of still uncharacterized lipases and those extracted and purified from the meta-genome. There is also a need to improve thermo-stability and organic stability of industrially important enzymes however methods to accomplish the same still lack a guided approach. Thus developing a bioinformatics approach for the same can be of crucial importance. Lipases have been reported to be involved in pathogenicity of microorganisms. However this area of research still remains unexplored. Thus targeting such lipases for drug development by *in silico* approaches shall be beneficial to medical science.

1.9. Structure of the Thesis

The architecture of thesis comprises of five chapters. A short introduction of each chapter is given below:

Chapter 1. Introduction: Chapter 1 provides comprehensive information on the background of the applicability of lipases. It also presents a brief overview on the catalytic mechanism of lipases and its various applications. It highlights the need for a deeper understanding of lipases. This chapter sets the research question. Next the aim and objectives as developed from identified lacuna have been presented. It was observed that there was a need to develop a comprehensive database for triacylglycerol lipases and further explore their physicochemical and functional characteristics. Finally the scope of research has been outlined.

Chapter 2. Creation of triacylglycerol lipase database: This chapter presents the background information on types of biological databases and the requirement to develop a comprehensive triacylglycerol lipase database. It also proposes a systematic methodological framework for the development of the database along with the non-alignment methodology for sub-family classification of lipases. Additionally it addresses the construction of a web tool which can classify uncharacterized lipase sequences based on matching motifs into a subfamily which will further highlight its functional characteristics. The significance of each cluster/sub-family has been brought out. Each sub-family has a unique phylogeny, biochemical features and applications. This paves the path for exploiting each individual family for applications.

Chapter 3. Understanding industrially important yet unexplored lipases: From the created database in chapter 2, sub-family 5 and 22 whose representative members, staphylococcal and *Pseudomonas mendocina* lipases respectively were chosen for industrial application studies. They were chosen for enhancing their structural stability by optimizing their physiochemical properties through *in silico* mutagenesis. Characterization of the same was achieved through homology modeling, identification and validation of thermostable mutants and molecular dynamics simulation.

Chapter 4. Pharmaceutical applications to combat tuberculosis: From our database subfamilies comprising of lipases involved in lipid catabolism, virulence and host immune-responses in pathogens were identified. This chapter focuses on targeting lipases in tuberculosis which continues to be one of the deadliest diseases. Lipases from *Mycobacterium tuberculosis* from sub-family 10, 11 and 22 were targeted as these lipases have been reported to be involved in its virulence. The chapter involves drug target identification by virtual screening. Further potential multi-targeted inhibitors for tuberculosis were identified targeting mycolic acid synthesis pathway and host lipid catabolism simultaneously.

Chapter 5. Pharmaceutical applications to combat leishmaniasis: This chapter focuses on targeting lipases in leishmaniasis which is yet another therapeutic application of our database. Leishmaniasis was targeted as the currently available therapeutics is associated with high toxicity and low efficacy. The lipases involved in lipid metabolism of the pathogen plays a vital role in its pathogenesis. This can be exploited for therapeutic applications. The methodology in brief involves drug target identification by virtual screening followed by identification of potential drugs and its validation for leishmaniasis.

Chapter 6. Conclusions: After summarizing the whole research finding and the key findings, conclusions have been made. Achievement of objectives and precision of hypothesis have been addressed. Identification of the future research scope has also been brought out.

Chapter II

TRIACYLGLYCEROL LIPASE DATABASE

Chapter II

Triacylglycerol Lipase Database

2.1. Abstract

A relational database for triacylglycerol acylhydrolases has been created with integrated sequence, structure and bibliographical information. The collected sequences were classified for understanding important structural and functional features as well as to emulate the potential applications where they reflect phylogenetic and biochemical characteristics within the enzyme family. The database has a user-friendly graphical interface with browse and search capabilities. Non-alignment based sub-family classification of triacylglycerol acylhydrolases is achieved and implemented within database. It has unique workflow to classify new/uncharacterized true lipase into respective sub-family which would aid biologists to annotate and emulate their bio-physiochemical characteristics and functionalities.

2.2. Introduction

Databases have been used for decades to organize and integrate bulk of complex data (Xiong, 2006). Database management systems (DBMS) have been engaged to perform the fundamental database tasks of storing and organizing information, ensuring freedom from internal contradictions, enforcing specified data constraints, and returning consistent results to simultaneous queries from multiple users (Shekhar and Chawla, 2002). Thus such an approach increases the amount of meaningful information that can be derived from the database and leads to better communication with other databases.

Among the various databases correlating with their type of data, biological databases are of interest to us. There are innumerable biological databases. The database catalog DBcat contains 511 biological databases till June 2001 (DBC, 2001). High emphasis is on sequence and structural data of nucleotide and proteins related databases. Also, there are other types of databases pertaining to particular research, databases for organisms, metabolic pathways, enzymes, transcription factors and many others (SAL, 1998). It is interesting to conclude here, that bulk of biological databases were related to enzymes because of their importance.

Amongst all enzyme databases, there are several databases in public domain that provide information on industrially important and versatile enzymes as lipases. The 'Lipase Engineering Database' (LED) provides information on sequence and structure of lipases as well as proteins sharing alpha/beta (α/β) hydrolase fold to facilitate protein engineering (Fischer *et al.,* 2003). The 'ESTerases and alpha/beta Hydrolase Enzymes and Relatives' (ESTHER) database provide the analysis of protein and nucleotide sequences that are homologous to cholinesterases as well as belongs to α/β hydrolase superfamily (Hotelier *et al.,* 2004). Lipabase and MELDB are dedicated to triacylglycerol acylhydrolases. The 'Microbial Esterases and Lipases DataBase' (MELDB) is a database of carboxylesterases (EC 3.1.1.1) and triacylglycerol lipases (EC 3.1.1.3) of microbial origin with clustering of 455 lipase sequences through local pairwise alignment and a graph clustering algorithm (Kang *et al.,* 2006). The 'LIPAsedataBASE' (LIPABASE) is a compilation of triacylglycerol acylhydrolases that provides taxonomy, sequence, structure and statistical data on physiochemical properties (Messaoudi *et al.,* 2011). But none of these databases provide

complete information and clustering of triacylglycerol acylhydrolases. A dedicated database for triacylglycerol acylhydrolases is need for the time to understand and explore its nature for various applications.

Through this work we intend to present the development of a specialized triacylglycerol lipases database. Our relational model database comprises of sequence and structural data of triacylglycerol acylhydrolases. In addition, biological and bibliographic information about the protein is provided giving prime importance to their applicability.

2.3. Methodology

The methodology has three hierarchical tiers. The first tier is confined to data retrieval and integration of sequence, structure and bibliography into database. The second deals with classification of the sequences into sub-clusters/sub-families employing a robust model which can handle varying range of sequence data in length and diverse phylogeny. Finally, the subfamilies were further annotated to bring forward their biological and industrial importance through motif identification. The purpose was to retrieve the individuality of each sub-family through motif identification and analysis of newly characterized lipase sequences through a web tool based on such motifs which shall further aid in deduction of its biological importance and industrial applications.

2.3.1. Data Acquisition and database creation

Lipase sequences and structures were retrieved from public domain notably UniProtKB, Brenda and the Protein Data Bank (PDB). For the purpose of data curation, basic biological information such as gene name, organism name, protein existence and active site information were parsed with in-house scripts along with their bibliographical data. The data was further cleansed to remove redundancy with a cutoff of 100% sequence identity. It is noteworthy to mention that 113 proteins from different organisms were identical with their respective homologs and they were retained as separate entities, since they represented individual proteins. The final dataset comprised of 1072 protein sequence accession numbers and 131 structures (PDB ids) belonging to 43 lipases (Annexure 1 and 2). The data was then organized into multiple tables on the principles of “Entity-Relationship” model and the

database schema is presented in Annexure 3. The database was constructed with XAMPP architecture in quad-core Linux workstation where the user interface layer is driven by Apache/PHP and the data layer with MySQL. Tools like BLAST, ClustalW, JMol and JalView were also implemented to carry out sequence analysis, multiple sequence alignment and visualization of data respectively in the database. Knowledge generation on triacylglycerol lipases is the prime requisite that can be achieved through sub-family classification.

2.2.2. Sub-family classification

A classification should reflect the phylogeny and biochemical characteristics within the family. Further the classification should reflect strong basis of clustering. For this purpose, the available protein sequence classifiers SECATOR (Wicker *et al.*, 2001), PIRSF (Wu *et al.*, 2004), a hybrid method involving the Hidden Markov Model (Gaugh, 2006), SCI-PHY (Brown *et al.,* 2007) and CLUSS (Kelil *et al.,* 2007) were employed for the purpose of classifying the 1072 lipase sequences. After the analysis of aforementioned methods, we opted for an alignment-free algorithm as the data contained sequences of varying length ranging from 160 to 870 residues. Sub-family classification by CLUSS (Kelil *et al.,* 2008) was thus the method of choice as the other classification programs rely primarily on protein multiple sequence alignments. Moreover alignment-based classification still remains a challenge to yield biologically plausible results for hard-to-align sequences such as tandem repeats, multi-domain proteins and proteins of high difference in length. Further attraction was that CLUSS has also been well validated with large dataset and specific hard-to-align enzymes notably glycoside hydrolases. Sub-family classification of glycoside hydrolases clustered according to the substrate specificity and known biochemical activities (Kelil *et al.*, 2008).

2.2.3. Identification of motifs and development of web tool

We tried to intrude into the reason for sequence being clustered which led to motif identification that uniquely identifies each sub-family. Long motifs with high significant E-value as well as no gaps (indel) were identified with the help of Multiple EM for Motif Elicitation (MEME) (Timothy *et al.,* 1994). MEME represents motifs as the probability of

each possible letter at each position in the pattern. Such motifs were believed to characterize newly sequenced lipases into their correct sub-family and thus reveal their biological significance and industrial applicability. Motif alignment search tool (MAST) was further employed to carry out analysis on the new or uncharacterized lipase of interest submitted by the user. This was implemented to further exploit the identified motifs of each sub-family for creating a workflow for the sub-family identification of the uncharacterized lipase sequences submitted by the user.

2.3. Results and Discussion

2.3.1. Features of triacylglycerol lipase database (TLDB)

Triacylglycerol lipase database serve as a one pit stop for the comprehensive information on triacylglycerol lipases in terms of sequence, structure and bibliographic information. The database can be accessed through www.iitg.ernet.in/tldb and the home page of the database is presented in Figure 2.1. Sequence and structure belonging to lipases can be navigated by browse or by search interfaces. The browse interphase in the database implements organism, structure, and clusters search by UniProt accession, PDB id, protein name, gene name and reference. The search is text-based and case insensitive. Visualization of proteins is available in 2D and 3D representations. 2D representation is achieved by accessing images directly from EBI PDBsum. 3D visualization is implemented using Jmol program version 11.8.7. Crosslink information with other public databases such as UniProt, PDB, Pubmed are available which each protein facilitates the user to retrieve more information. BLAST searches can be performed against TLDB using the 'Search' interface. Inclusions of more features or working with structures are expected in future since TLDB is in the development stage. The schematic representation of TLDB is presented in Figure 2.2. Furthermore, TLDB led to the reliable classification and demonstrated the advantage of specialized as well as highly enriched database for triacylglycerol lipases.

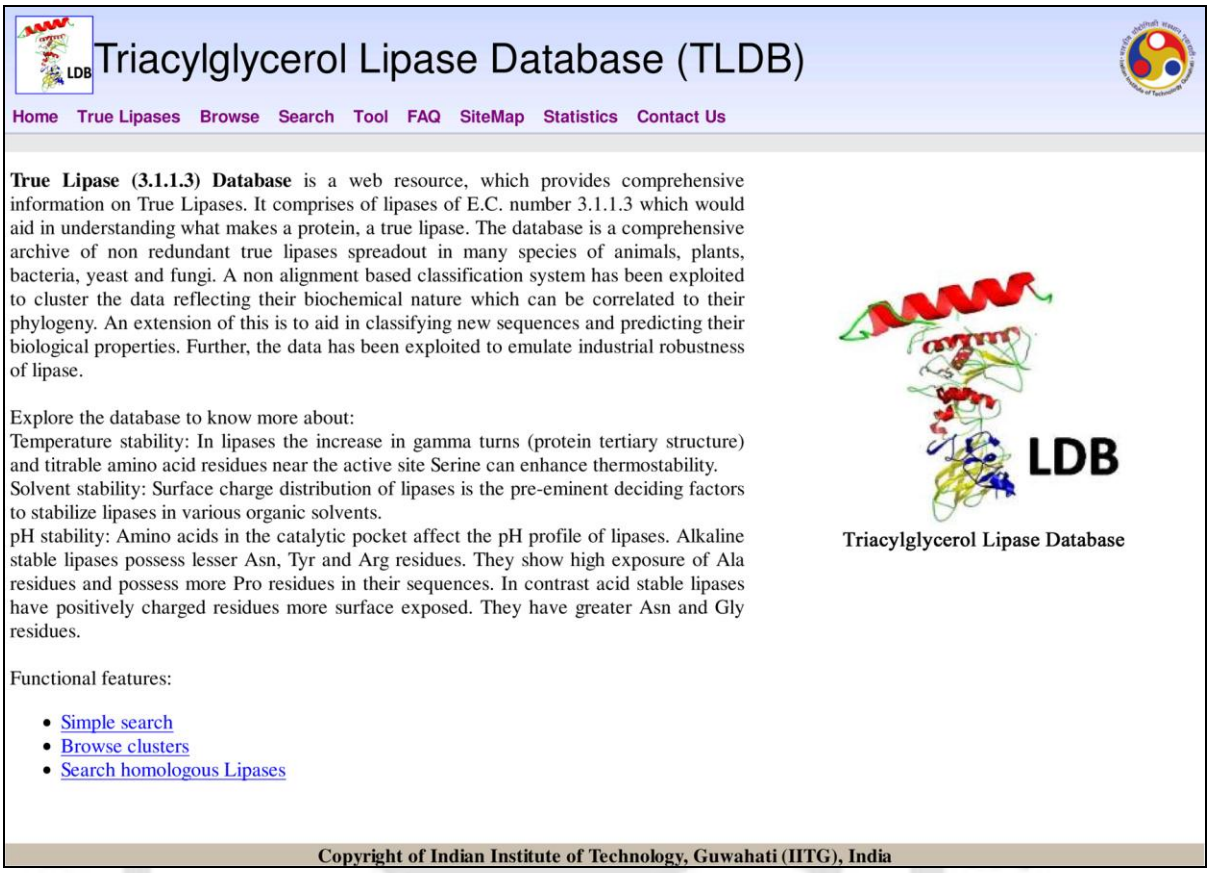


Figure 2.1: Homepage of the Triacylglycerol Lipase Database.

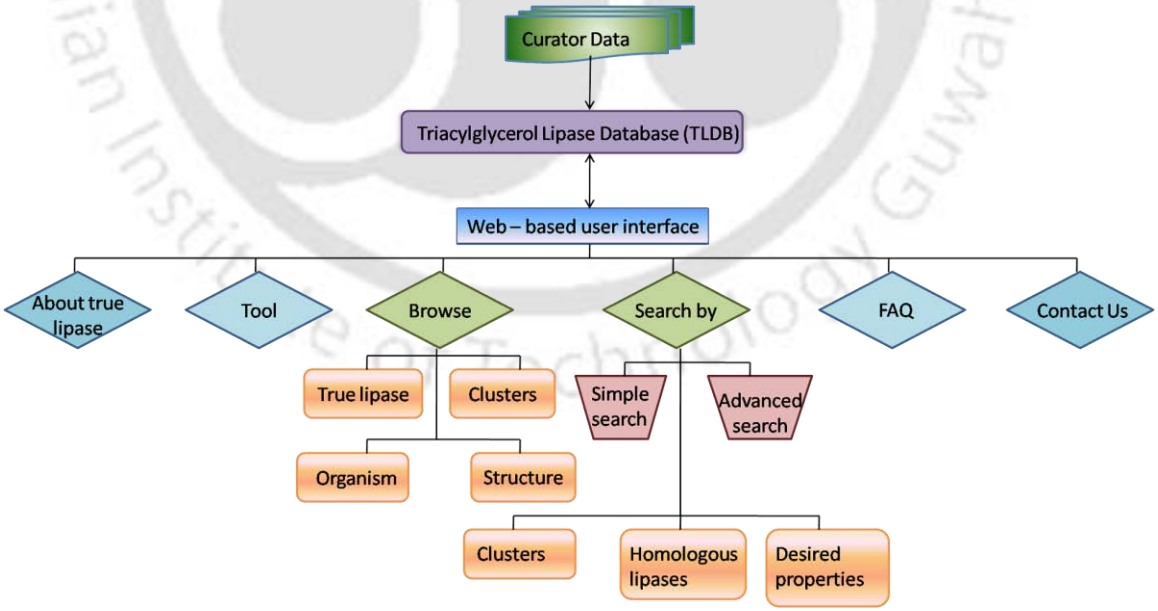


Figure 2.2: The schematic representation of Triacylglycerol Lipase Database.

2.3.2. The uniqueness of the database and its basis: sub-family classification

The non-alignment based classification of triacylglycerol lipases through CLUSS was used as a reliable framework for the systematic analysis of lipase homologous families. 74 sub-families were identified with 2 clusters of sub-families which are true representatives of triacylglycerol lipases (cluster 1: sub-family 4-7 and cluster 2: sub-family 8-11). Table 2.1 presents the distinct phylogeny in the sub-families and also validates the methodology employed in this work as proteins belonging to different phylogenetic group were clustered separately. The annotated information of sub-family classification and multiple sequence alignments of each sub-family can be downloaded on request from TLDB. Figure 2.3 illustrates the classification which reflects the phylogeny, biochemical and promiscuous characteristics of lipases.

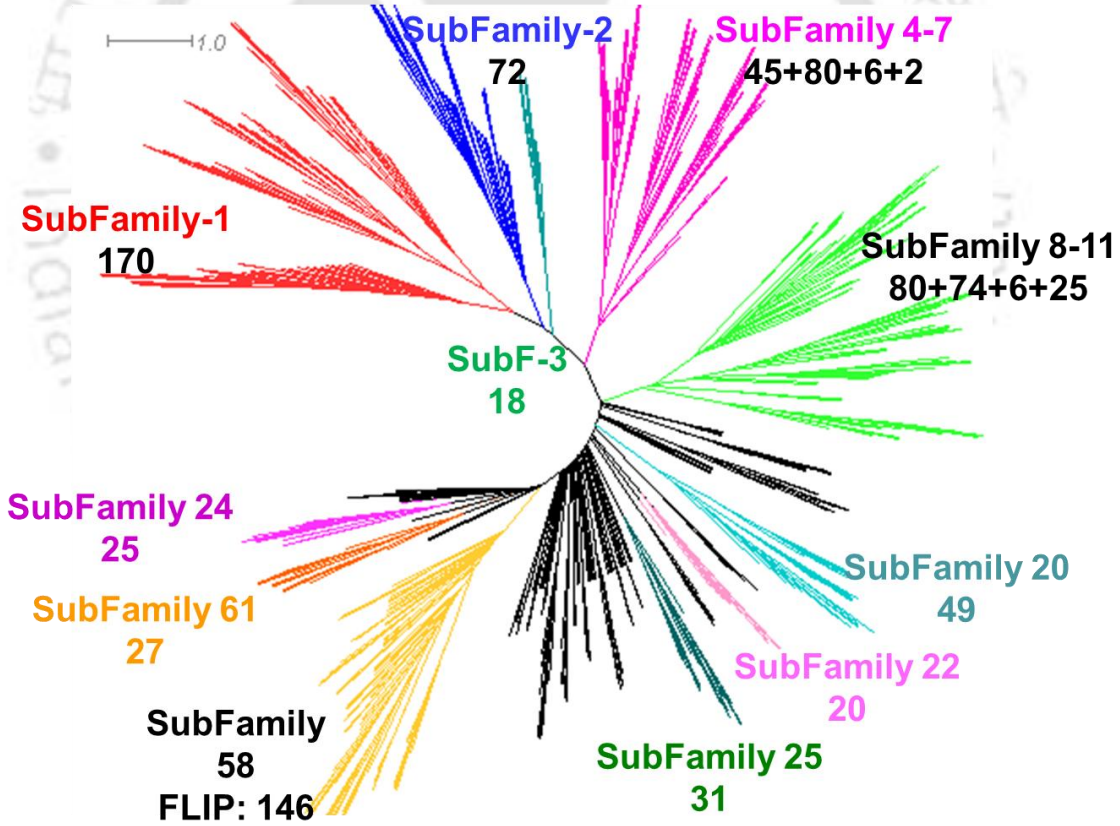


Figure 2.3: Sub-family classification of triacylglycerol lipases. Sub-families were obtained through CLUSS and revealed both biological and functional characteristics.

Table 2.1: Sub-family classification of triacylglycerol lipases showing distinct phylogeny.

Sub-family*	Description	Kingdom	Phylum
1 (170)	Promiscuous in nature	Archaea Bacteria Eukaryota Viruses	Crenarchaeota; Actinobacteria; Bacteroidetes; Cyanobacteria; Environmental Firmicutes; Planctomycetes; Proteobacteria; Alveolata; Euglenozoa; Fungi; Metazoa; Viridiplantae dsDNA viruses
2 (72)	PLRP-2	Bacteria Eukaryota	Firmicutes (1); Metazoa
3 (18)	contains cold lipases	Bacteria Eukaryota	Environmental; Firmicutes; Spirochaetes; Proteobacteria (14); Fungi
4 + 5 (45 + 80)	*Bacillus & Clostridium + Staphylococcus*	Bacteria	Firmicutes
8 + 9 (80 + 74)	*Pseudomonas*	Bacteria	Proteobacteria
11 + 20 (25 + 49)	therapeutically important subfamilies	Bacteria Eukaryota	Actinobacteria (23+23); Firmicutes (1+0) Fungi (1+26)
22 (20)		Bacteria	Actinobacteria; Proteobacteria (1)
25 (31)		Eukaryota	Fungi; Metazoa(1)
58 (136)	FLIP No Lipase motif	Bacteria	Actinobacteria; Aquificae; Bacteroidetes; Chloroflexi; Firmicutes; Planctomycetes; Proteobacteria; Synergistetes; Thermotogae; Verrucomicrobia
61 (27)	therapeutically important subfamilies	Bacteria	Proteobacteria; Environmental (1)
64 (25)		Bacteria	Proteobacteria

***** indicates the number of sequences in subfamily is denoted in parenthesis
+ indicates the sub-families have similar biological characteristics

The sub-family classification revealed that CLUSS not only classified triacylglycerol lipases in accordance with their phylogeny but also showed differential classification based on their functional importance. The groupings based on functionality and biological importance has been described here.

I. Industrially important sub-families: Lipases from *Bacillus* and *Staphylococcus* have been reported to be important for detergent, paper and oleochemical industries especially when they are thermo and organic solvent stable (Hasan *et al.,* 2006). Such lipases were grouped together in Sub-family 4 which contains lipases from *Bacillus* and *Clostridium* class which are thermo- and organic solvent- tolerable lipases. Sub-family 5 have triacylglycerol lipases from *Staphylococcal* source which are mainly extracellular lipases also known to be

thermostable and important in detergent industries. Sub-family 15 contains minimum α/β hydrolases from *Bacillus subtilis*. Such lipases do not possess any lid and thus do not involve interfacial catalysis (Dartoris *et al.,* 1992) Thus they are better catalysts for industrial applications. Sub-family 53 contains lipases with photo-transduction activity and has been also characterized to accept diacyglycerol as substrates (EC 3.1.1.34). Sub-family 36 have *Lysinibacillus sphaericus* non-polar organic solvent-tolerant lipase with preference towards long-chain substrates and lipases stable in non-polar solvents are important for industries dedicated towards chemical synthesis (Sharma *et al.,* 2001).

II. Biologically important sub-families: These are subfamilies possessing triacylglycerol lipases with enzymatic function additional to lipases such as kinases and transferases. Sub-family 1 is highly promiscuous in nature in terms of activity, biochemical characteristics, length and source of organism. For instance, lipase from *Drosophila melanogaster* (UniProt Id: Q9VKT1) is of length 1073 residues while lipase from *Melanogrammus aeglefinus* (UniProt Id: Q802C5) is of length 234 residues. Sub-family 22 has 20 lipases from *Actinobacteria* containing signal domain for the twin-arginine translocation (Tat) pathway which serves the role of transporting folded proteins across energy-transducing membranes. Interestingly, sub-family 58 comprises of 136 lipases which participate in flagellar biosynthesis and motility from bacterial source. Sub-family 37 contains 12 lipases with patatin-like phospholipase domain-containing protein. This domain is structurally and functionally related to the animal cytosolic phospholipase A2. Sub-family 73 has transferases (EC 2.7.11.1) with significant lipase activity from *Toxoplasma gondii* and sub-family 57 contains 4 lipases that belong to Type-B carboxylesterase/lipase family (EC 3.1.1.13). Sub-family 71 contains a LIM domain with zinc finger-containing lipase which is again predicted to be heterogeneous serine/threonine kinase (EC 2.7.-.-).

III. Therapeutically important subfamilies: Sub-families dominated by lipases with pharmacological importance can lay the platform of future drug development and bring forward their uniqueness to be clustered in separate subfamilies. Sub-family 2 contains pancreatic lipase and pancreatic lipase related proteins (PLRP). Such lipases hydrolyze dietary fat molecules in the human digestive system. Under the conditions of pancreatitis, the

pancreas autolyse and release pancreatic lipases into serum. Through measurement of serum concentration of pancreatic lipase, acute pancreatitis can be diagnosed (Koop *et al.*, 1984). Sub-family 20 was observed to be dominated by *Corynebacterium*, *Mycobacterium* and *Candida* lipases. Particularly interesting is *Mycobacterium* which is the causative agent of the pandemic disease, tuberculosis. Sub-family 34 contains two lipases from *Trypanasoma* which is similar towards leishmanial lipases. Sub-family 65 contains bacterial lipases with 'Lipase_bact_N_dom' domain. This domain usually occurs in the N-terminal region of bacterial virulence factor lipases. Sub-family 40 has lipases which has been predicted as Tubulin-specific chaperone E from human parasite, *Pediculus humanus* which belong to Type-B carboxylesterase/lipase family (EC 3.1.1.13). Sub-family 70 contains 5 sequences among which 4 belong to PE-PGRS protein family from mycobacterial source and sub-family 1 is a lipase with CAP-GLY domain from *Toxoplasma gondii*. Sub-family 28 contains 4 alkaline lipases from *Penicillium* and *Neosartorya* source. Such lipases are usually involved in the pathogenesis of the microorganisms. Hence they can qualify as good drug targets.

Comparison of sub-family classification with other lipase databases using representative lipases showed that the classification implemented in our databases is scientifically agreeable in the context of phylogeny and function. Also, our classification reflects the substrate specificity of each sub-family which could discriminate the short- and long-chain acting lipases as well as promiscuous lipases. Table 2.2 presents comparison of sub-family classification implemented in our database with other lipase databases. Our database is better in function as it could classify lipases both according to phylogeny and function as well as physicochemical characteristics.

Table 2.2: Comparison of sub-family classification with other lipase databases

Representative lipases from CLUSS (sub-family number)	**LED sub-family (superfamily)**	**LED class**[$]
Lipa (15)	Bacillus lipases (Bacillus lipases)	GX
Lactonizing lipase (8)	*Burkholderia cepacia* lipase like (Burkholderia lipases)	GX
Lipase (9)		
Lipase b (50)	*Candida antarctica* lipase B like (*Candida antarctica* lipase like)	GX
Lipase (1)*****		GGGX
Lipase (1)*****	Gastric lipases (Gastric lipases)	GX
Hepatic lipase (2)	Hepatic lipases (Lipoprotein lipases)	GX
Endothelial lipase (2)	Lipoprotein lipases (Lipoprotein lipases)	GX
Bile-salt-activated lipase (57)	Mammalian bile salt activated lipase like (Carboxylesterases)	GGGX
Lipase 1 (22)	Moraxella lipase 1 like (Moraxella lipase 1 like)	GX
Arylacetamidedeacetylase (1)*****	Moraxella lipase 2 like (Moraxella lipase 2 like)	GGGX
Esterase (1)*****		
Lipase 1 (3)[#]	Moraxella lipase 3 like (Moraxella lipase 3 like)	GX
Pancreatic lipase/ PLRP-2 (2)	Pancreatic lipases (Lipoprotein lipases)	GX
Lipa (61)	Pseudomonas lipases (Pseudomonas lipases)	GX
Lipase (1)*****	*Rhizomucor miehei* lipase like (Filamentous fungi lipases)	GX
Lipase 2 (6)	*Saccharomyces cerevisiae* lipase 2 like (Burkholderia lipases)	GX
Lipase (5)	*Staphylococcus aureus* lipase like (Burkholderia lipases)	GX
Carboxylesterase (35)	*Yarrowia lipolytica* lipase like (*Yarrowia lipolytica* lipase like)	GGGX

***** contains promiscuous lipases
[#] contains cold lipases
[$] LED class reflects the substrate specificity of lipases and lipase related proteins while GGGX has preference towards carboxylesterases as well as short-chain acting lipases and GX has preference towards medium as well as long-chain acting lipases.

2.3.3. Identification of motifs and the web tool

MEME identified motifs which helped to differentiate each sub-family from the other. Table 2.3 presents the motifs identified for five sub-families. The details of other motifs have been tabulated and presented in Appendix I of the thesis. These motifs were utilized for a tool development with the aim to characterize new lipases sequences into sub-families which can be indicative of their biological and industrial importance.

Table 2.3: Non-alignment based consensus sequence of the long motifs

SubF	Description	Motifs*
1	Promiscuous in nature	IALFG[GT]DPYK[VLI]x[VYI][VT]GHS[AQL]G[GAT] NGPEQSL[AG][FY][IL]LADAG[YF]DVW[LM]GNSRG[NT][RT]YS [KY][HY][GN]YPVE[ET][HY][ET][VI][TV]TEDGYILT[LI][HFQ]RIPR[GS][RK]N
2	PLRP	GRITGLDPA[EG]P[LCY]F[QE]GTP[EP]EVRL[DS]P[SDT]DAxFVDVIHT G[LY]SP[SE]NVH[LV]IG[HF]SLGA[HQ]VAG[FYE]AGRRLKGx PG[LG]G[FL]GMSQP[VI]GH[VL]DF[YF]PNGG
3	Contains cold lipases	H[VL]IIPDL[PL]G[FS]G[ED][ST]S[KV]PMD[AF]DY[SR][VS][PE][AN][QL]A[ET][RK]L[RH][ER][FL]V[EQ] H[IV][AG]G[HN]S[ML]GG[AS]I[AS][VL][LA]YA[AG][KQ]YP[KF][ED][VT]KSL[WF]LVDSAG[IFV][WF][SR][AS][AG]N[TPS]K[YS] [IY][DN][AI]PTL[IV][LV]WG[DK][KQ]D[KQR]II[NK][VP]E[VT]A[NE][EL][LI]KR[LI][LI][KP][NQ]A
4	Bacillus & Clostridium	G[YH][RTE][TV]YT[AL][AT]VGP[VL]SSNWDRACE[LA]YA[QY][IL]VGGTVDYG[AE]AH[AS][KA]K[HY]GH[ANE]R[YF]GRTY [SQ][IV]YD[FL]KLDQWGL[KR][RK][QK]PGES[FQ][DLR][HD]YF[EN]R[VL][KL]RS[PN][IV]W[TKN]ST[DK][DT][AI][RS][YV] [NG][KR][IV]H[LI]I[GA]HS[MQ]GGQT[IA]R[MT]L[VT][QS]LL[ES][NE]GS[QE]EE[RI][ENS]Y
5	Staphylococcus	AFGSNYDRAVELYYYIKGGRVDYGAAHAAKYGHERYGKTY[EK]GI[YM][KP][DN]W[KE]PGK[VI]HLVGHSMGGQTIR[QL][LM]EE[LF]LRNGN[KRPQ]EE[IV][EA]Y[QH]K[AQK]HGGEISP LF[KT]G[GN][HQ][DN]NM [YV]TN[AD][TE][DP][AK][TI][QR][KR]G[IV]WQV[TK]P[TIV][KI][HQ][DG]WDHVDF[VI]G[QV]D[SF][SLT]D[TF][KV]R[TK][RG][EA]EL[QA][DQN]F[WY][HT][GH][LI][AI]

* catalytic motif is underlined

The work flow has been illustrated in Figure 2.1. Once the user submits a sequence of interest in the 'Search Homologous Lipases', the workflow/tool is invoked (Figure 2.4). First, the locally installed BLAST tool is utilized to search for the user sequence in our database. When the sequence identity is 100% between the user sequence (query) and lipase from our databases (subject), the length of both the sequence is compared to conclude whether the user sequence exists in our database or not. Once the existence of user sequence is verified, MAST tool is utilized to search for the motifs for each sub-family in a sequential manner followed by ranking the results according to the E-value implemented in MAST. Then the top ranked sub-family is assigned as the sub-family for the user sequence. Finally, the user sequence is aligned with the sequences of sub-family revealing the conservation of amino acids between them. This alignment provides insight on the conservation of conserved and non-conserved regions that could be correlated with the structurally and functionally important amino acids of lipases from the TLDB database. This workflow aids the user to

understand the sequence and structural features of the lipase of interest that could aid to engineer them for improved/modified biochemical characteristics. Additionally it can be also utilized for the engineering of protein mutants with improved biochemical properties. Moreover if the new sequence is from a metagenome it can give insight about its source organism. Figure 2.5 illustrates the outcome of a submitted sequence to the web tool.

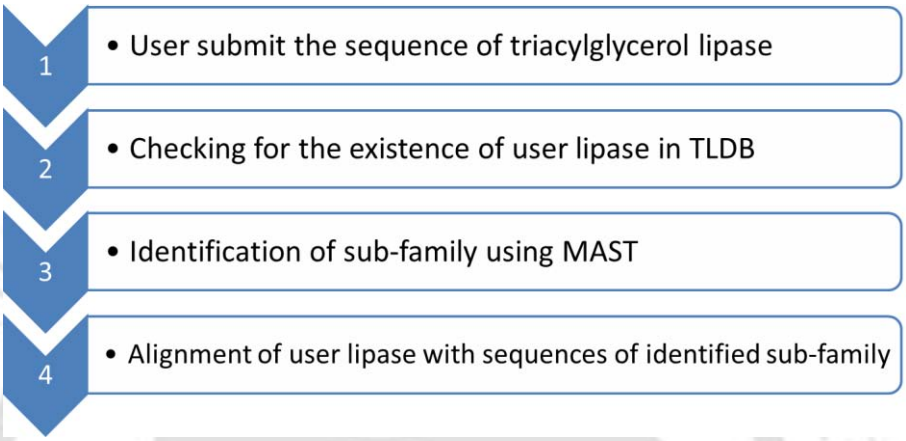


Figure 2.4: Work flow of the web tool implemented in Triacylglycerol Lipase Database.

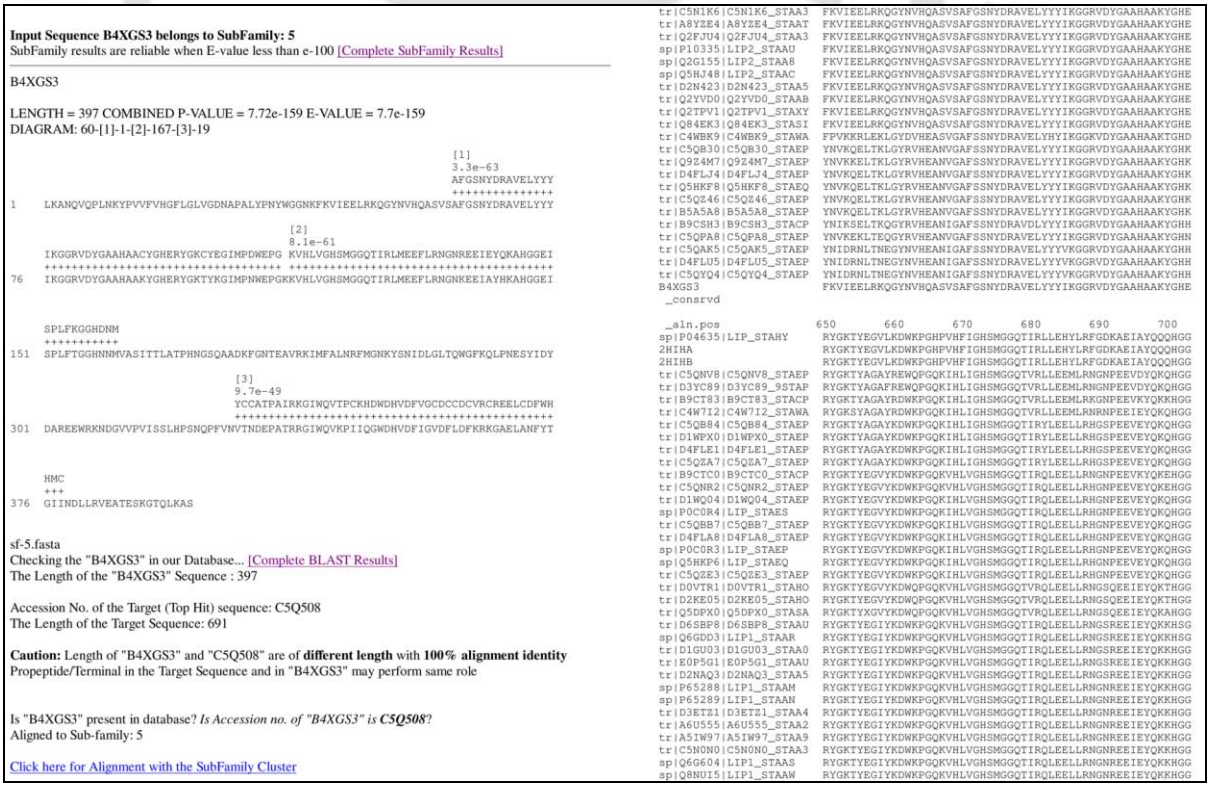


Figure 2.5: Sub-family classification and alignment of user-submitted sequence predicted by the workflow implemented in Triacylglycerol Lipase Database.

2.4. Conclusions

Triacylglycerol Lipase Database has been created with simple graphical user interface which comprises information of 1072 sequences, 132 structures and 272 bibliographies. Our database has been implemented with non-alignment based classification methodology making it unique. Workflow was implemented in our database to identify sub-family classification and conserved patterns of the user-submitted sequence of lipase which would aid in understand and engineer new/novel triacylglycerol lipases. The established framework of our database can be extended to other lipases/enzymes and would be instrumental in several fields ranging from bioinformatics, protein engineering and drug-discovery. Key sub-families depicted in the database bring out the significant applicability of lipases. They are sub-family 5 which have triacylglycerol lipases from *Staphylococcal* source, industrially important sub-family 15 containing minimum α/β hydrolases from *Bacillus subtilis* and sub-family 20 was observed to be dominated by *Corynebacterium*, *Mycobacterium* and *Candida* lipases involved in pathogenesis, hence pharmacologically important. The upcoming chapters of the thesis will utilize the lipases from various clusters obtained through classification for their applicability both in industries and therapeutics.

Chapter III

UNDERSTANDING AND EXPLOITING LIPASES

Chapter III

Understanding and Exploiting Lipases

3.1. Abstract

This chapter brings forward the industrial importance of lipases and the need for more versatile lipases in the terms of structural stability. The need of alkaline thermostable lipases has been growing rapidly as they are highly attractive for the production of detergents, biodiesel, pharmaceuticals agents and various other applications. From our sub-family classification it was clear that certain lipases though potent for various industrial applications are yet under-explored. Such lipases from *Pseudomonas* and *Staphylococcus* belonging to sub-family 5 and 22 were chosen from our TLDB. They were subjected to *in silico* mutagenesis for enhancing their structural stability and the effect of the mutations have been analyzed through molecular dynamics simulations. The various parameters analyzed through simulation suggested that the *in silico* mutants stabilized the local conformation and improved the structural stability. In summary, this study prefigures well as a prototype for mutation and protein engineering studies of the candidate lipases chosen from our TLDB database.

3.2. Introduction

Lipases are attractive biocatalysts for industries as they show thermostability, organic solvent stability, substrate specificity, chemo-, regio- and enantio-specific behavior (Malcata *et al.*, 1990). Thus, they have found use in food, organic synthesis, pharmaceuticals, oleochemical, paper, leather and detergent industries. In food and dairy industries they are useful in cheese ripening, flavor development, and enzyme modified cheese (EMC) technology. In pharmaceutical industry, they have been employed in the racemic resolution of drugs such as naproxen and ibuprofen. Irrespective of the aforesaid the industrial uses still remain limited by their high production costs, commercialization in small amounts, and low performance of some lipase-mediated processes. Thus there is a need for emulating/engineering lipases. Different methodologies have been researched towards increasing the stability, catalytic efficiency and organic solvent tolerance of the lipases (Dartoris *et al.*, 1992). Protein engineering techniques such as site-directed mutagenesis, directed evolution may be instrumental in making the lipases more acceptable at large scale production in industries. Rational engineering of lipases was carried out to increase interactions contributing to stability: disulphide bonds, hydrogen bonds, electrostatic interactions, hydrophobic interactions and metal ion binding and also by increasing protein rigidity. Introduction of disulphide bond which holds N- and C-terminus of *Penicillium camembertii* lipase increased its optimum temperature by 10ºC and optimum pH was inclined acidically by 0.7 (Yamaguchi *et al.*, 1996). In *Fusarium heterosporum* lipase, C-terminal extension was proven to show increased temperature stability by 10ºC (Nagao and Shimada, 1998). In *Candida antartica* B lipase, T103G mutation in the consensus active site signature GXSXT increased the thermal stability by 4ºC (Patkar and Vind, 1998). Computer simulations indicated the importance of the interaction of charges in the opening of the lid of lipases. In *Humicola lanuginosa* lipase, mutation in lid (Glu87) into either neutral residue or lysine changes nature of the electrostatic interactions from being repulsive to attractive towards contact zone making interfacial activation more favourable (Peters and Toxvaerd, 1997).

As guided rules for enhancing the stability of lipases are still elusive, Bioinformatics can inspire rational engineering of lipase stability. For example, the selectivity for triacylglycerol substrates in *Rhizopus oryzea* lipase was studied through modeling studies. Mutants L258A and L258S favoured the sn-1 hydrolysis where mutants L258F and L258F/L254F favoured the sn-3 hydrolysis (Scheib *et al.,* 1998). Other than the importance of substrate selectivity for chemical industries, the need for alkaline thermostable lipases has been growing rapidly as they are highly attractive for the production of detergents, biodiesel and also instrumental for the production of pharmaceutical agents (Illanes 1999; Haki 2003). Detailed understanding of the structure of lipases would aid in engineering them to function in the desired industrial environments (Benjamin and Pandey, 1998; Jaeger and Eggert, 2002). In order to check all possible mutations with improved structural stability, the experimental approaches are time- and resource-intense process. Therefore, *in silico* rational approach is useful to select optimistic mutations that could be validated experimentally (Looger *et al.*, 2003).

It was observed from the work of TLDB and sub-family classification that lipases were clustered into functionally and industrially important groups. Most important were sub-family 22 and 5 whose representative members, *Pseudomonas mendocina* lipase (PML) and staphylococcal lipases respectively show potential in the context of industrial applications like fat stain remover in detergent industries. From sub-family 5, lipase from *Staphylococcus aureus* (SAL3) was reported to be active at temperatures between 40-60°C and having a pH range of 8-10 with optimum pH of 9.5 (Horchani *et al.*, 2009). The lipase from *S. xylosus,* SXL2 was reported to remain active at temperatures between 40-60°C, with an optimum of 55°C and have a pH optimum of 8.5 (Horchani *et al.*, 2011). Therefore both the lipases are thermostable and alkaline-stable which makes them potent for detergent industry. However, 3D structure of SAL3 and SXL2 were not available. They can be treated as subjects for enhancing their structural stability by optimizing their physiochemical properties by prediction of their tertiary structure and analyzing the hot spots for *in silico* mutagenesis which will be instrumental in enhancing industrially desirable proteins. This study may make them potent commercially viable lipases.

3.3. Material and Methods

This chapter attempts to understand and exploit the industrial potential using the 3D structure of the candidate lipases PML, SAL3 and SXL2 chosen from our TLDB database. The crystal structure of PML (PDB Id: 2FX5) was utilized. Since the experimental structures of staphylococcal lipases SAL3 and SXL2 are unknown, experimental mutagenesis at all possible sites with possible mutation is time and resource consuming. Rational approach has been implemented which begins with prediction of 3D structure followed by *in silico* characterization to identify functionally important amino acid residues. The identification of residues that are structurally/strategically important for structural stability leads to prediction of potential *in silico* mutants and thereby mutagenesis hotspots. The *in silico* mutants were studied and validated with molecular dynamics simulations.

3.3.1. Homology Modeling of Staphylococcal Lipases

Multiple sequence analysis of SAL3 (UniProt Id: B4XGS3) and SXL2 (UniProt Id: E7CRL0) with its staphylococcal and thermostable homologs was performed to identify the conserved amino acid residues. Multiple sequence analysis was performed using ClustalX (Thompson *et al.,* 1994). The lipase from *Staphylococcus hyicus* (SHL) was identified as template to predict the structure of SAL3 and SXL2 using NCBI-BLASTP analysis against Protein Data Bank (PDB). MODELLER9v6 has been implemented to generate the three dimensional model for SXL2 (Marti-Renom *et al.,* 2000) with the help of MOLPDF and DOPE scores. Quality of the 3-D structure homology models were assessed in NIH SAVES server using Ramachandran (Laskowski *et al.,* 1993) and ERRAT plots (Colovos *et al.,* 1993).

To establish the reliability of the predicted structures, a 20ns molecular dynamics (MD) simulation was performed in an explicit water environment with the periodic boundary conditions applied in three dimensions. The net charge of system was neutralized by the addition of sodium ions by replacing water molecules that are at least 3.50 Å from the protein surface. The solvent was equilibrated for 100ps by restraining the solute atoms

through a harmonic force constant of 1000 kJ mol^{-1} nm^{-2} in NVT ensemble followed by NPT ensemble. NPT production run was carried out for 10ns with no restraints using 2fs of integration time. All MD simulations were carried out with temperature of 300K with velocity rescaling thermostat in which protein and non-protein atoms were coupled to separate temperature coupling baths. The pressure was controlled at 1 atm using Parrinello-Rahman barostat. The Particle Mesh Ewald (PME) summation method was used for calculating the long-range electrostatic interactions with cut-off of 12 angstroms (Å). The linear constraint solver (LINCS) algorithm was used to constrain the bonds involving hydrogen atoms. 2000 conformations from production run were analyzed to check whether the SXL2 model is energetically stable.

3.3.2. Prediction of in silico mutants

The candidate lipases, SAL3, SXL2 and PML were investigated for the stability prediction upon point mutation with ProSA (prosa2003 program) (Sippl, 1993). The predicted model of SAL3 and SXL2 as well as experimental structure of PML (PDB Id: 2FX5) was utilized for the prediction of their *in silico* mutants. The ProSA generates potentials and z-scores reflecting the quality of protein structures on carrying out the predicted mutations of all possible sites using knowledge-based potentials. The predicted mutations were classified as stabilizing and de-stabilizing mutations which aids us to identify the mutagenesis hotspots of candidate lipases SAL3, SXL2 and PML. Also, MD simulations of the wild type of candidate lipases identified respective 'dynamic residues' for mutant studies. The dynamic residues were based on root-mean square fluctuations (RMSF) that can be correlated to temperature factors which reflects the local structural flexibility. The mutants were selected based on the presence of 'thermolabile' residues.

3.3.3. Validation of identified *in silico* mutants through MD simulations

The identified *in silico* mutants were studied in detail to provide insight at the atomic level with the help of 20ns MD simulations with the optimized parameters of their corresponding wild type of candidate lipases SAL3, SXL2 and PML. The comparative MD analysis between the mutants and wild type identify potential mutants with better

structural stability. MD trajectories were analyzed with built-in Gromacs tools. The root-mean square deviation (RMSD) and RMSF of the backbone Carbon atoms were studied with g_rmsd and g_rmsf respectively to understand the degree of conformational changes in the respective 3D structures of the candidate lipases and its mutants. The secondary structure analysis was carried out using the Kabsch and Sander algorithm incorporated in their Dictionary of Secondary Structure for Proteins (DSSP) program (Kabsch and Sander, 1983) which was installed into Gromacs to analyze the variation of protein secondary structure changes. Program g_gyrate was used to measure the radius of gyration which reflects the compactness of the structure. Program g_sas was used to compute interaction surface areas between solvent molecules and complexes. Sigmaplot was utilized to prepare the plots of trajectory analysis from MD simulations and the figures were produced using Adobe PhotoShop.

3.4. Results and Discussion

3.4.1. Homology Modeling of Staphylococcal Lipases

Mature SAL3 and SXL2 comprise of 397 amino acid residues and markedly differ in substrate chain length preference and stability with its homologs. Surprisingly, SXL2 has similar functional parameters as SAL3. From the conservation of amino acid residues (Figure 3.1), the catalytic triad of SAL3 and SXL2 (red star) were identified as Ser119, Asp310 and His352 as well as the calcium binding at C-terminus (green triangle) comprised of residues Gly286, Asp351, Asp354, Asp359 and Asp362. Zinc ion coordination having the residues Asp64, His84, His90 and Asp236 at the N-terminus of SAL3 and SXL2 that overlaps the nucleotide binding motif ‘p-loop’ containing the residues from Gly89 to Thr96 differentiates them from commercial *Bacillus* lipases. The other important functional regions of SAL3 and SXL2 as deduced from the multiple sequence alignment (Figure 3.1) are (i) the oxyanion hole having the residues Phe17 and Met117 implicated in tetrahedral intermediate formation with substrate (ii) the lid comprising of a helix followed by loop containing the residues from Asn180 to Leu206, and (iii) the residues believed to be involved in substrate specificity are Leu18 and Val355.

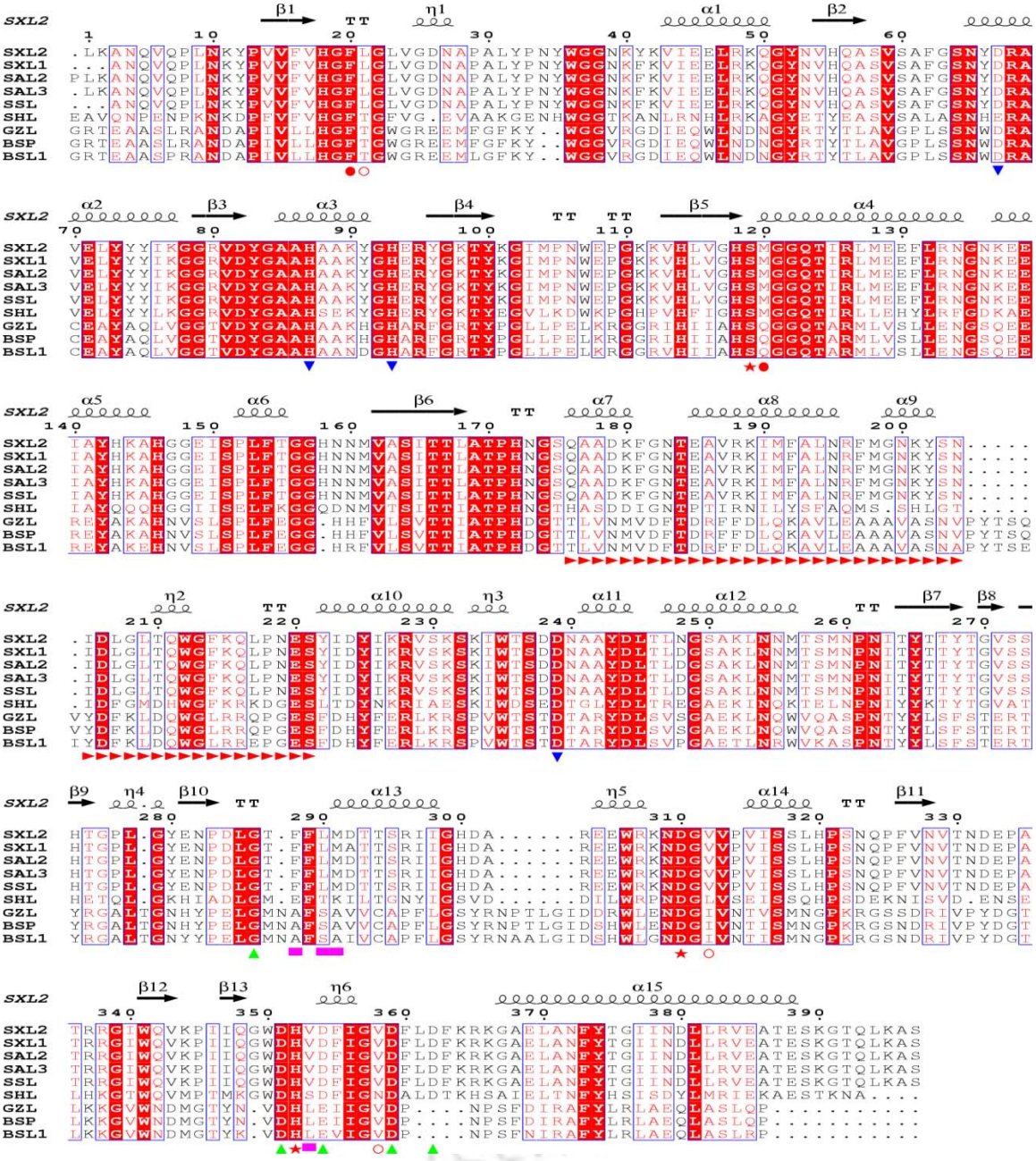


Figure 3.1: Sequence alignment of the mature peptides of SAL3 and SXL2 with other staphylococcal and homologous thermostable lipases using ClustalX. Red star: catalytic residues, Red sphere: oxyanion hole, Red circle: chain selectivity, Red right-faced triangle: potential lid region, Blue inverted triangle: zinc binding residues, Green triangle: calcium binding residues, Pink rectangle: phospholipase binding residues

Comparison with the residues that bind with phospholipid substrate (Lys295, Ser356, Glu292 and Thr294) of SHL reveal that the corresponding residues in SAL3 and SXL2 were found to be Met288, Val350, Phe285 and Leu287. The substitution of the apolar

residues in SAL3 and SXL2 could be a possible reason for the loss of phospholipase activity and may facilitate the binding with hydrophobic substrates. Staphylococcal lipases (SAL3 and SXL2) and SHL, a homologous lipase with significant lipase activity and also having high A1 and minor A2 phospholipase activities) share 52% and 53% sequence identity respectively yet markedly differ in substrate chain length preference and thermostability (Figure 3.1) (Tiesinga *et al.,* 2007). Ion coordination of SAL3 and SXL2 differentiates them from commercial *Bacillus* lipases.

Homology Modeling: The tertiary model of SAL3 and SXL2 in open conformation has been predicted with the aid of SHL as template (PDB id: 2HIH, Resolution: 2.86Å), the only staphylococcal lipase with experimental structure (53% sequence identity and 67% sequence similarity with the SHL in the modeled 900 residues) (van Oort *et al.,* 1989; Simons *et al.,* 1998). SAL3 and SXL2 have the α/β hydrolase fold with βαβ structural motif. The fold consists of seven-stranded central parallel β-sheet covered by helices in SHL (Figure 3.2). Comparison with the commercial bacterial lipases reveals significant difference at the N-terminus: (i) the first two strands of the central β-sheet are absent and (ii) the topology also contains a β1-α3-β2 insertion between helix α4 and strand β5 as observed in SHL (Brzozowski *et al.,* 1991). The GXSXG pentapeptide motif (GHSMG, residues 117–121) found in γ turn that positioned the catalytic Ser119 optimally between His352 and the binding pocket (Noble *et al.,* 1993; Derewenda and Sharp 1993; Heikinheimo *et al.,* 1999). Structure validation reveals that the predicted structure of SAL3 and SXL2 is satisfactory. The reasonable statistics was reflected by the distribution of the backbone dihedral angles in the Ramachandran plot with 94.2% of the residues in the allowed region, 5.2% in additionally allowed region, and 0.3% in the generously allowed region. MD indicates that the validated model of SAL3 and SXL2 is energetically stable and reasonably accurate for further analysis. The tertiary structure of SAL3 and SXL2 can be accessed with the accession number PM0076011 and PM0079083 respectively in Protein Model DataBase**.**

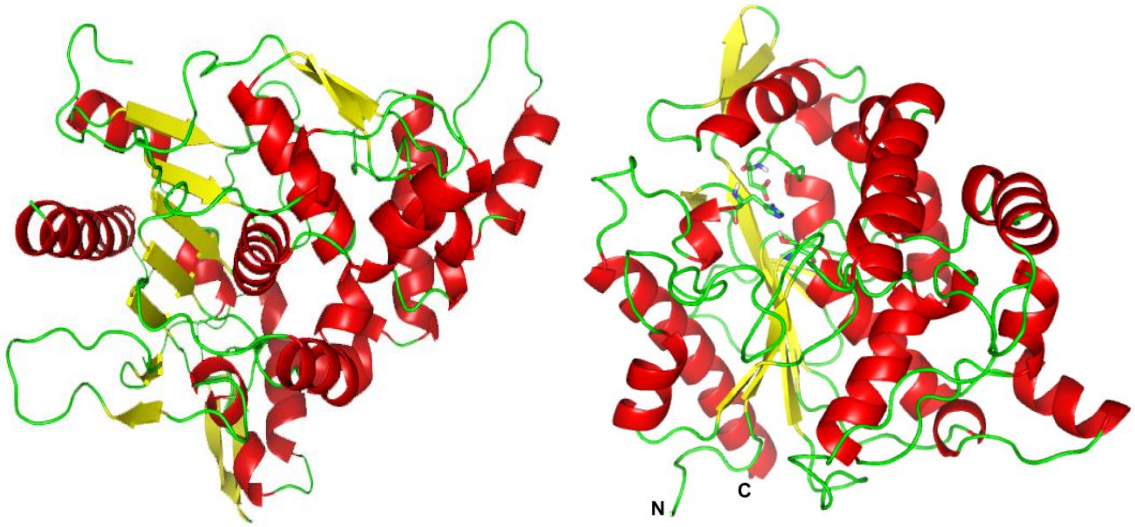


Figure 3.2: Tertiary structure of SAL3 and SXL2 in cartoon representation predicted with Modeller. Catalytic residues were depicted as sticks while N- and C- terminus were denoted as N and C respectively. Alpha helices, beta strands and loops are shown in red, yellow and green respectively.

Comparison of the candidate lipases with homologous thermostable and mesostable lipases: The structural analysis of SAL3 and SXL2 with its homologs showed that the secondary structural elements and functionally important amino acid residues are conserved. Also, the overall fold of SAL3 and SXL2 with commercial lipases used in the study is similar. The core beta sheet with the nucleophilic elbow is well conserved, as observed in known lipases (Heikinheimo *et al.*, 1999). The insertion in the lid region of thermostable bacillus lipases may involve in its movement, and may also potentially be associated with the thermostability and specificity. In addition, the zinc coordination which overlaps with the P-loop at the N-terminus is conserved, while the calcium-binding residues are not similar between SWL2 and to the known structures of thermostable lipases from *Bacillus spp*. SXL2 structure is highly similar with thermostable BSP; hence the factors that contribute to thermostability could be similar with zinc-binding domain. SAL3 and SHL have an unusually high proportion of aromatic residues. Thermostability of an enzyme is conferred by a combination of numerous contributing factors and a prominent factor of SAL3 and SXL2 could be its zinc-binding domain as seen in other thermostable lipases (Watanabe *et al.*, 2005)

3.3.2. In silico Mutagenesis of the Candidate Lipases

Our observations from ProSA identified four classes of residues where mutation may lead to destabilization in SAL3 and SXL2. In the first class, positions were predicted at which substitution of any one of these positions by the rest of nineteen amino acids will lead to destabilization of the protein (Class 1). The second class comprises of positions which can be mutated by one corresponding amino acid without affecting the stability (Class 2). Similarly, the third and fourth class comprises of positions which can be mutated by two and three corresponding amino acid respectively without affecting the stability (Class 3 and 4). The details of the point mutation specific to each class and their corresponding secondary structure of SAL3 and SXL2 have been presented in Table 3.1 and 3.2 respectively. Based on this framework, we could identify 127 and 118 amino acid residues as structurally/strategically important for stability of SAL3 and SXL2 respectively. The distributions of residues of SXL2 belonging to the four classes at the conserved region reveal that the functionally important residues could contribute to the structural stability involved notably (i) G19 in HG dipeptide, (ii) G92, G97, T99 in P-loop, (iii) G286, D362 in calcium binding, and (iv) E185, L194, P218, S221 in lid region. Further five residues Glu94, Glu139, Phe289, Lys308 and Leu361 of SAL3 and thirteen residues Ala3, Glu139, Met161, Val164, His172, Ile205, Trp212, Ala302, Ala305, Pro321, Pro325, Pro334, and Ala335 of SXL2 were identified in which substitution by any amino acid does not affect protein stability while the local conformation is considered as mutagenesis hotspots (Figure 3.3). Hence mutation can be carried out at these mutagenesis hotspots for emulating the structural stability of SAL3 and SXL2.

Table 3.1: Summary of predicted mutation sites leading to destabilization of SAL3.

Secondary Structure	Predicted Residues	Stability Class			
		I	II	III	IV
	127	47	33	21	26
Coil	40	13	15	5	7
α-Helix + 3_{10}	58	24	8	11	15
β-Strand	17	5	7	2	3
Turn	12	5	3	3	1

Table 3.2: Summary of predicted mutation sites leading to destabilization of SXL2.

Secondary Structure	Predicted Residues	Stability Class			
		I	II	III	IV
	118	41	26	32	19
Coil	30	15	6	6	3
α-Helix + 3_{10}	43	11	10	14	8
β-Strand	18	8	3	5	2
Turn	27	7	7	7	6

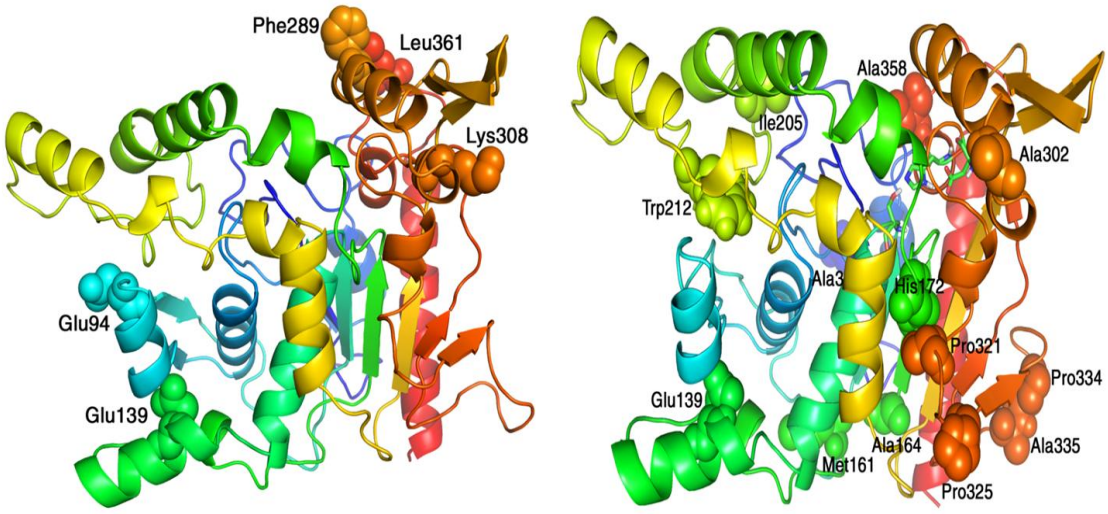


Figure 3.3: Mutagenesis hotspots of SAL3 and SXL2 predicted based on the thirteen residues whose substitution by any residue does not affect protein stability.

In silico mutagenesis of Pseudomonas mendocina lipase: Point mutation was carried out for each residue of PML to check the change in overall stability using ProSA (Sippl, 1993). Our observations identified four set of residues based on *in silico* mutagenesis which lead to destabilization in PML. In the first set, 24 positions were predicted at which substitution of any one of these 24 positions by the rest of nineteen amino acid residues will lead to destabilization of the protein (Class 1). The second set comprises of 15 positions which can be mutated by one corresponding amino acid residue without affecting the protein stability (Class 2). Similarly, set 3 and 4 comprises of 26 and 32 positions respectively which can be mutated by two and three corresponding residue without affecting the protein stability. Based on this framework, we identified 97 amino

acid residues as structurally/strategically important for PML which are distributed throughout the protein. Furthermore five residues Pro7, Pro10, Glu83, Pro187 and Pro219 were identified as stabilizing mutants in which the substitution by any other amino acid residues does not affect the protein stability and the local regions were identified as mutagenesis hotspots for PML (Figure 3.4).

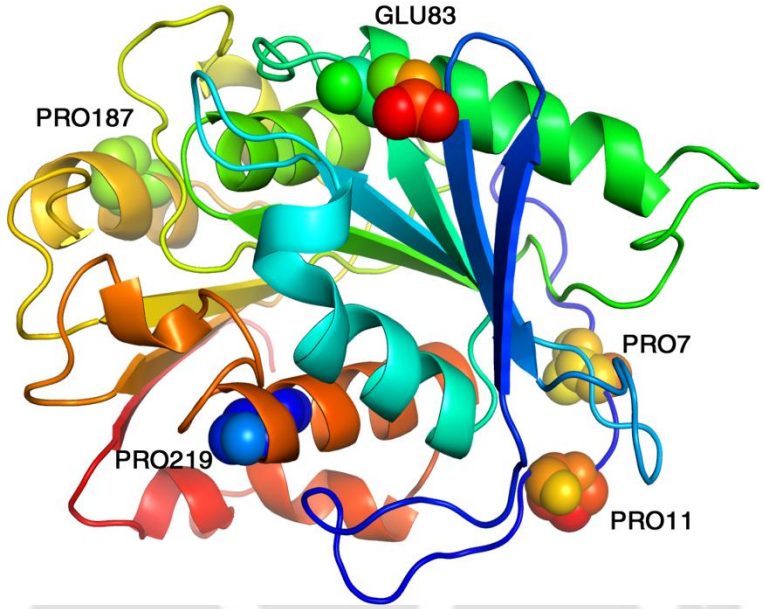


Figure 3.4: Mutagenesis hotspots with stabilizing mutants of PML.

Also, MD simulations of the candidate lipases, SAL3, SXL2 and PML identified respective 'dynamic residues' for mutant studies. The 'thermolabile' dynamic residues were selected based on RMSF that can be correlated to temperature factors which reflects the local structural flexibility (Table 3.3).

Table 3.3: Thermolabile 'dynamic' residues based on RMSF of the candidate lipases.

Candidate Lipase	Residues	Secondary Structure*
SAL3	Asn204	loop
	Asn331	loop
SXL2	Asn219	turn
	Glu220	loop
PML	Arg91	α-helix
	Arg163	loop
	Gln242	loop

* the secondary structure where thermolabile 'dynamic' amino acid residues resides.

3.4.3. Validation of in silico mutagenesis with molecular dynamics

The respective selected residues in the candidate lipases were mutated with a residue that could improvise the structural stability. The *in silico* mutants were investigated and validated with MD utilizing built-in Gromacs tools.

3.4.3.1. SAL3 MD analysis: The mutants identified from ProSA and RMSF study were studied with the wild type of SAL3 with the utility tools from Gromacs.

Root Mean Square Deviation (RMSD):

The degree of conformational changes of wild type and identified mutants of SAL3 from ProSA and RMSF study is monitored by RMSD of alpha Carbon atom during the course of MD simulation. The RMSD of conformations from production run relative to its initial structure have been presented in Figure 3.5 and 3.6 respectively. The MD study reflects that the E94A mutant is relatively unstable than the wild type which is evident from the Figure 3.5B. The time-dependent RMSD analysis reflects that the wild type and its mutants of PML have less deviation in RMSD throughout the simulation and implies that the 3D structures are relatively stable except the E94A mutant. The mutants from ProSA study of SAL3 were comparatively similar while N204P mutant has comparatively better RMSD values. Among mutants identified from RMSF studies of wild type SAL3, N204P mutant was found comparatively stable than N331P mutant, perhaps due to stabilization

of the loop where N204P resides. The average and standard deviation of RMSD values of wild type and mutants of SAL3 have been summarized in Table 3.4 which reflects the order of stability from the average values of RMSD as (i) F289P < wild type from ProSA study and (II) N204P < N331P < SAL3 from RMSF study of SAL3. The RMSD analysis reflects that the structural integrity of mutants is affected to certain extent while the N204P mutant found to have better RMSD values throughout the simulation. The N204P mutant might lead to the stabilization of the loop and its local conformation. The F289P mutant was also found to be more stable among the identified mutants while the other mutants have high fluctuations of RMSD values throughout the simulation which is evident from the Figure 3.5 and 3.6.

Radius of Gyration (Rg):

The radius of gyration has been analyzed in a time-dependent manner to investigate the compactness of wild type and its mutants of SAL3 from ProSA and RMSF study and has been presented in Figure 3.7 and 3.8 respectively. The radius of gyration reflects the packing of residues throughout the simulation thereby stability and folding rate of SAL3. The wild type and identified mutants have similar profile of Rg throughout the simulation. The order of Rg values from Table 3.4 reflects that the mutants E94A and F289P from ProSA study have better Rg profile than wild type SAL3 as well as the mutants N204P and N331P. The N331P has comparatively more Rg values while N204P and F289P reaches steady equilibrium.

Root Mean Square Fluctuation (RMSF):

The local deformability of the protein has been analyzed through root mean square fluctuations of alpha Carbon atoms from MD simulations. The RMSF plotted against the residues of wild type and its mutants of SAL3 from ProSA and RMSF study has been presented in Figure 3.9 and 3.10 respectively. Appreciable difference of RMSF between wild type and mutants of SAL3 is observed at the mutation site and their respective local regions. The average values of RMSF indicate that the mutants are considerably similar to that of wild type SAL3. The mutants E94A and N204P stabilize the fluctuations of the

N- and C-terminus while wild type has high flexibility at N- terminus. SAL3 might have an inherent flexible terminus which could have been stabilized by its quaternary form.

Solvent Accessible Surface Area (SASA):

The trend of SASA analysis from Figure 3.11 and 3.12 reflects that the N331P mutant from RMSF study have more hydrophobic area than any other mutants and wild type SAL3. The hydrophobic area of N331P mutant is also consistent throughout the simulation. The average values of SASA indicate that E94A mutant from ProSA study has considerably less hydrophobic area than wild type SAL3. The order of average values of SASA is as (i) E94A < SAL3 < F289P from ProSA study and (ii) SAL3 < N204P < N331P from RMSF study of SAL3. The mutant N331P reaches steady equilibrium in the initial stage of simulation reflecting that the mutation might lead to the stabilization of the local conformation where it resides.

End to End chain distance between first and last alpha Carbon atoms:

The distance between first and last alpha Carbon atoms of wild type and mutants of SAL3 from ProSA and RMSF study were analyzed throughout the simulation which would aid to understand the structural integrity of wild type and its mutants of SAL3 and have been presented in Figure 3.13 and 3.14 respectively. Overall, a similar trend of end to end chain distance between the mutants has been observed while it is found to be consistent for wild type SAL3 throughout the simulation. The order of stability from the average values of 'End to End chain' distance is (i) SAL3 < F289P < E94A from ProSA study and (ii) SAL3 < N204P < N331P from RMSF study of SAL3.

Analysis of Secondary Structure:

The secondary structure analysis was carried out using the Kabsch and Sander algorithm incorporated in their DSSP program (Kabsch and Sander, 1983). The plots show the structural variation of each residue of wild type and mutants of SAL3 from ProSA and RMSF study during the time course of simulation and have been presented in Figure 3.15 and 3.16 respectively. The stabilization of the corresponding secondary structure of

mutants was observed (Table 3.5). The mutant F289P further stabilizes the consecutive bend while the consecutive turn was destabilized or distorted.

Table 3.4: Summarized parameters of MD simulations for wild type and *in silico* mutants of SAL3.

Method*	Variant	End to End Distance (nm)	Radius of gyration (nm)	RMSD (nm)	RMSF (nm)	SASA (nm^2)
	SAL3	3.16±0.28	1.99±0.02	0.26±0.05	0.14±0.08	101.85±2.60
PROSA	E94A	4.79±1.19	1.97±0.02	0.24±0.04	0.14±0.07	100.86±3.29
	F289P	4.60±1.56	1.97±0.02	0.25±0.04	0.14±0.07	102.20±2.97
RMSF	N204P	3.76±1.20	1.99±0.01	0.22±0.04	0.14±0.07	102.10±2.19
	N331P	4.87±0.98	2.01±0.01	0.25±0.04	0.14±0.07	104.82±1.81

* methodology implemented for the identification of the mutants

Table 3.5: Stabilization of the secondary **s**tructure by the mutants of SAL3.

STUDY	SAL3	Secondary Structure*
PROSA	E94A	bend
	F289P	loop
RMSF	N204P	loop
	N331P	loop

* the secondary structure where the mutants resides

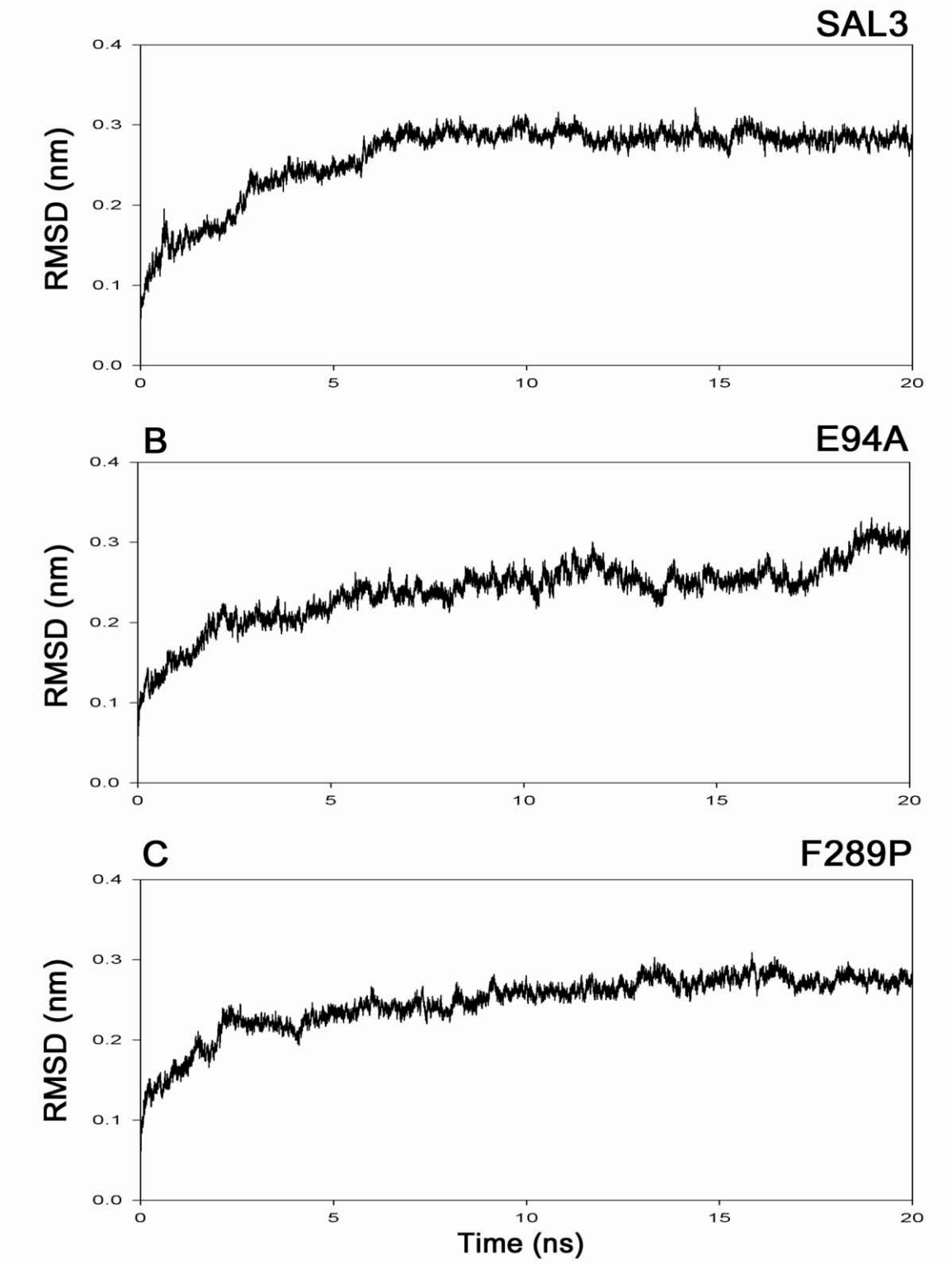


Figure 3.5: The root-mean-square deviation (RMSD) to the starting structure as a function of time of wild type and mutants of SAL3 from ProSA study during the time course of simulation is shown. (A) SAL3 (Wild type). (B) Mutant E94A. (C) Mutant F289P.

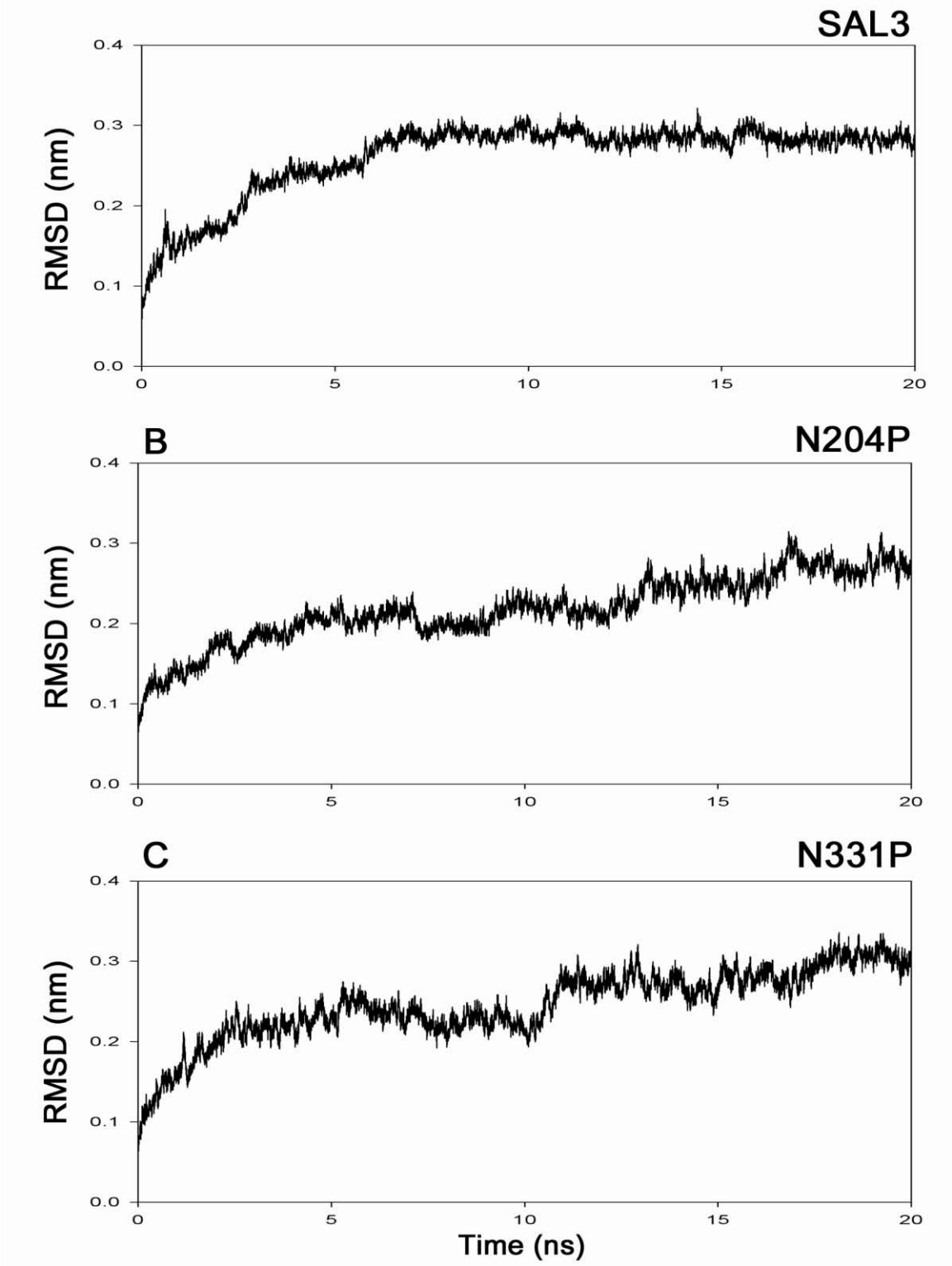


Figure 3.6: The root-mean-square deviation (RMSD) to the starting structure as a function of time of wild type and mutants of SAL3 from RMSF study during the time course of simulation is shown. (A) SAL3 (Wild type). (B) Mutant N204P. (C) Mutant N331P.

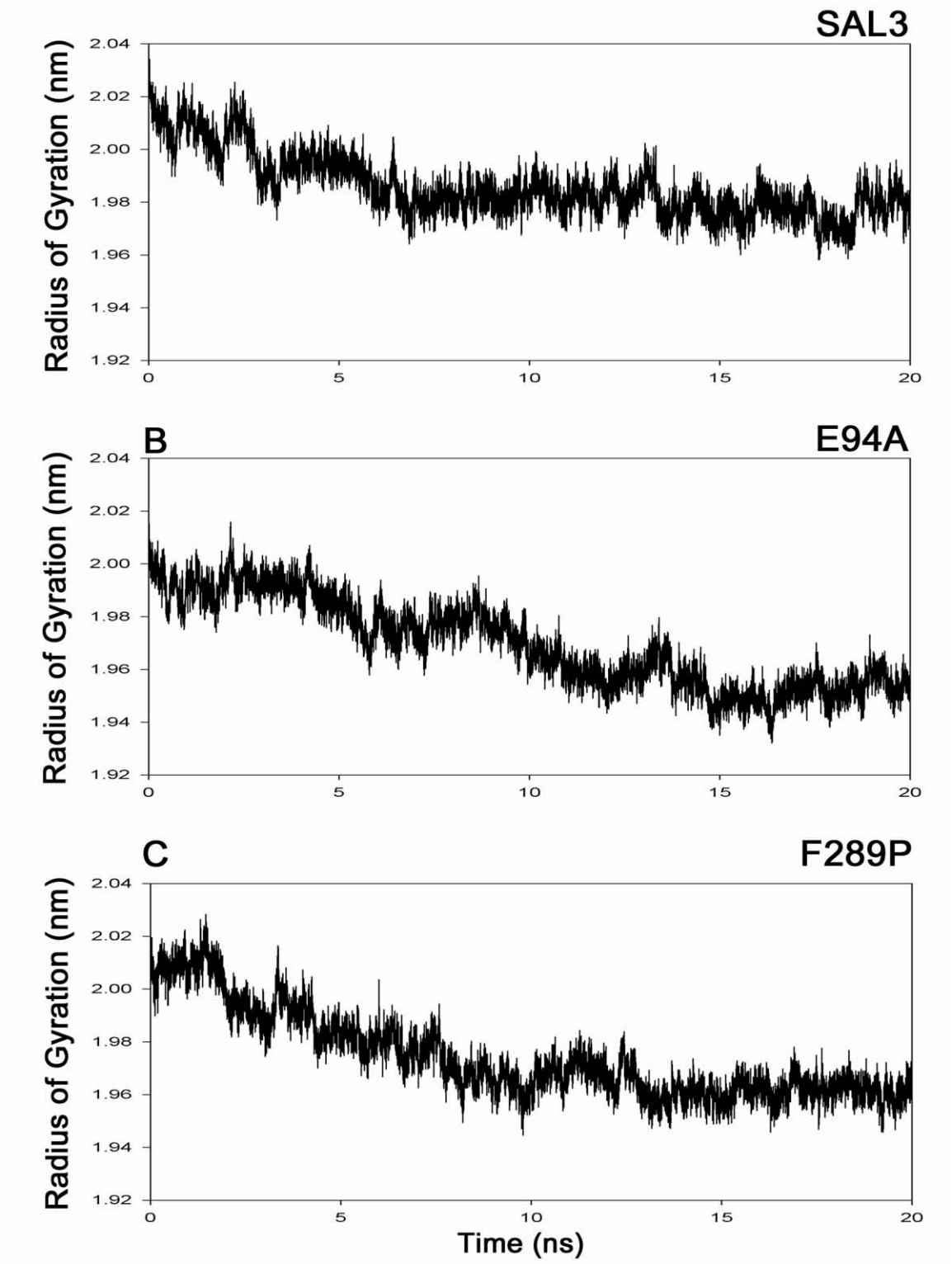


Figure 3.7: Radius of gyration of α-carbon atoms as a function of time of wild type and mutants of SAL3 from ProSA study during the time course of simulation is shown. (A) SAL3 (Wild type). (B) Mutant E94A. (C) Mutant F289P.

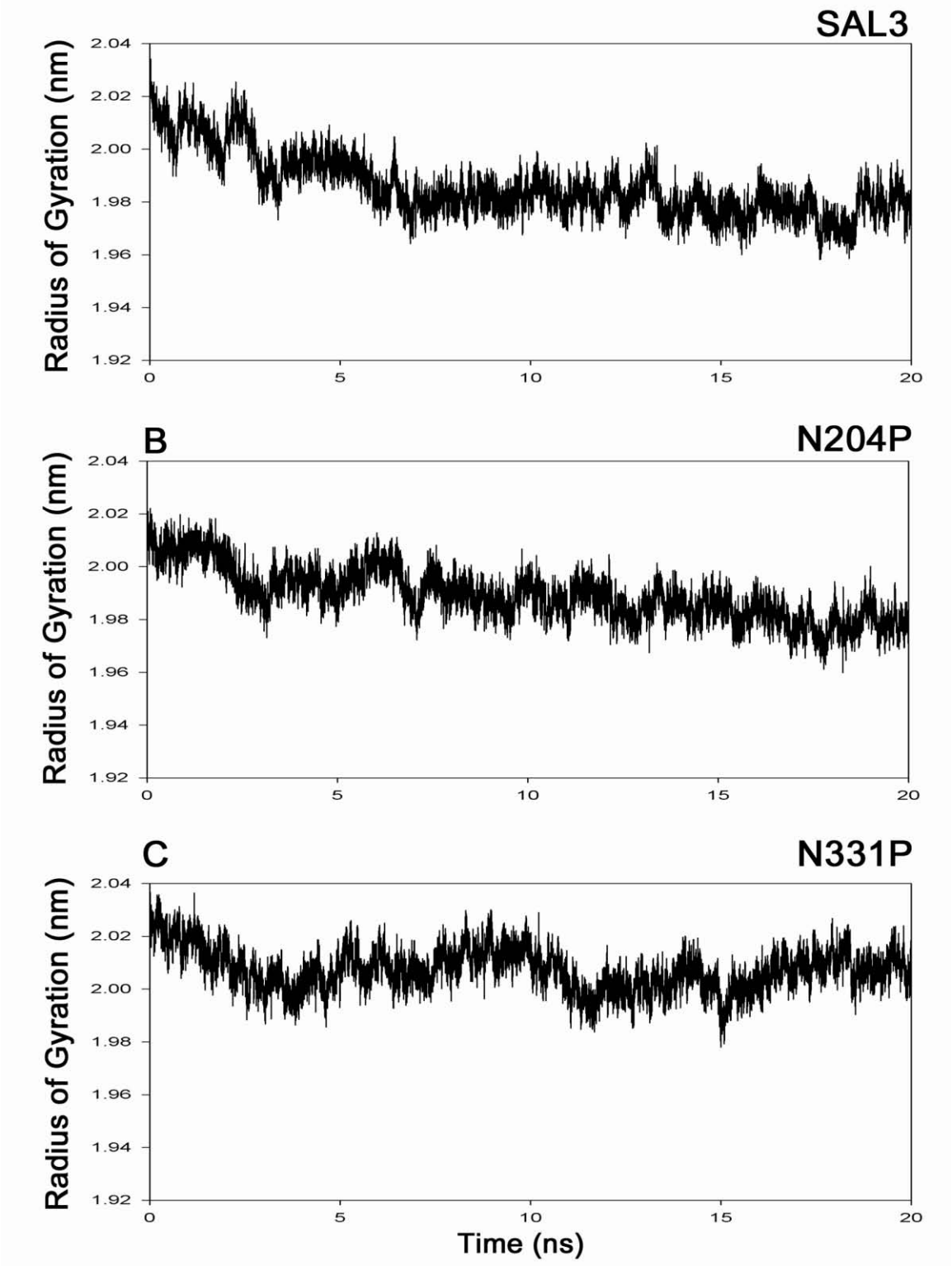


Figure 3.8: Radius of gyration of α-carbon atoms as a function of time of wild type and mutants of SAL3 from RMSF study during the time course of simulation is shown. (A) SAL3 (Wild type). (B) Mutant N204P. (C) Mutant N331P.

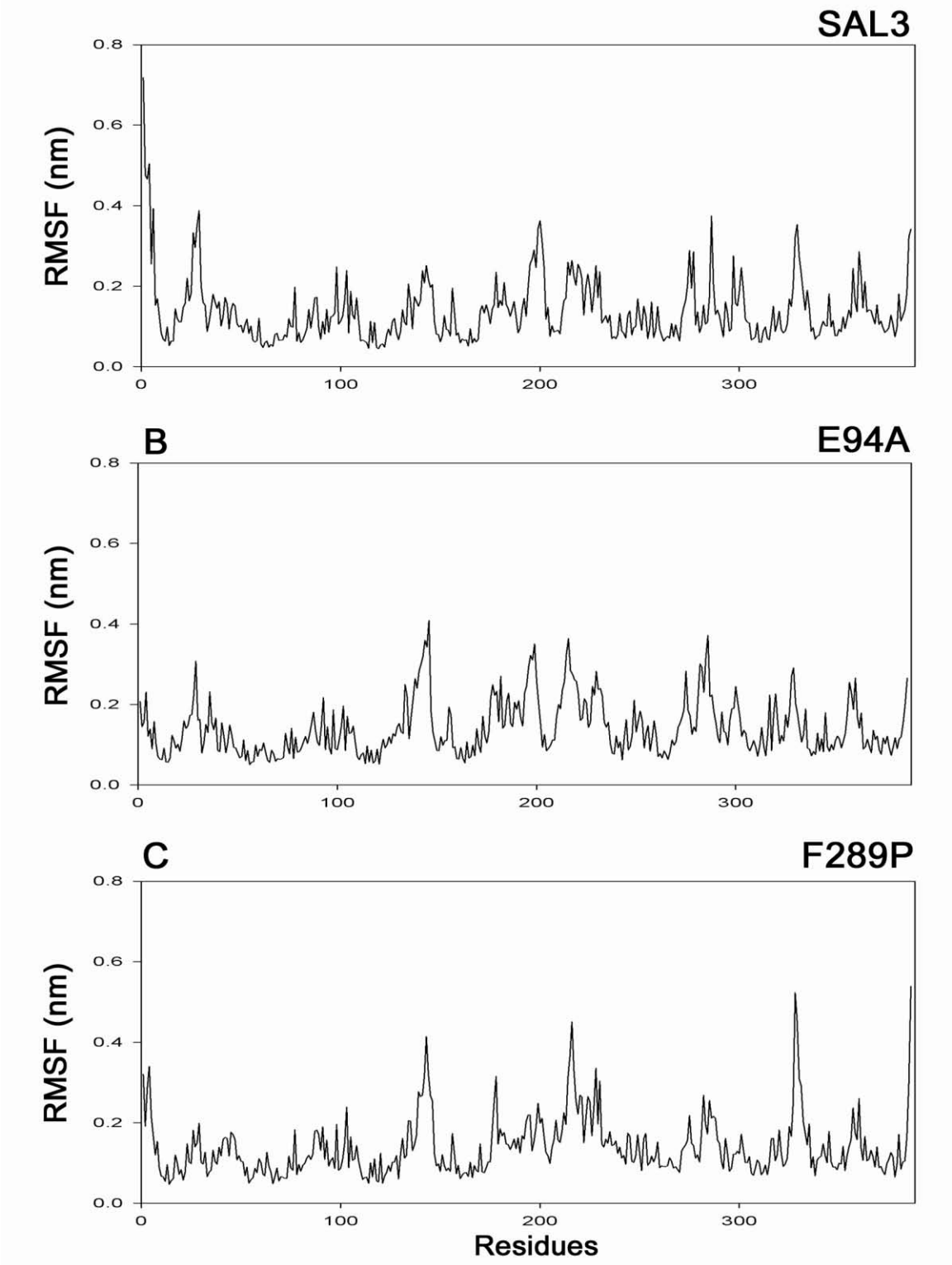


Figure 3.9: Root mean square fluctuation of α-carbon atoms as a function of residue of wild type and mutants of SAL3 from ProSA study during the time course of simulation is shown. (A) SAL3 (Wild type). (B) Mutant E94A. (C) Mutant F289P.

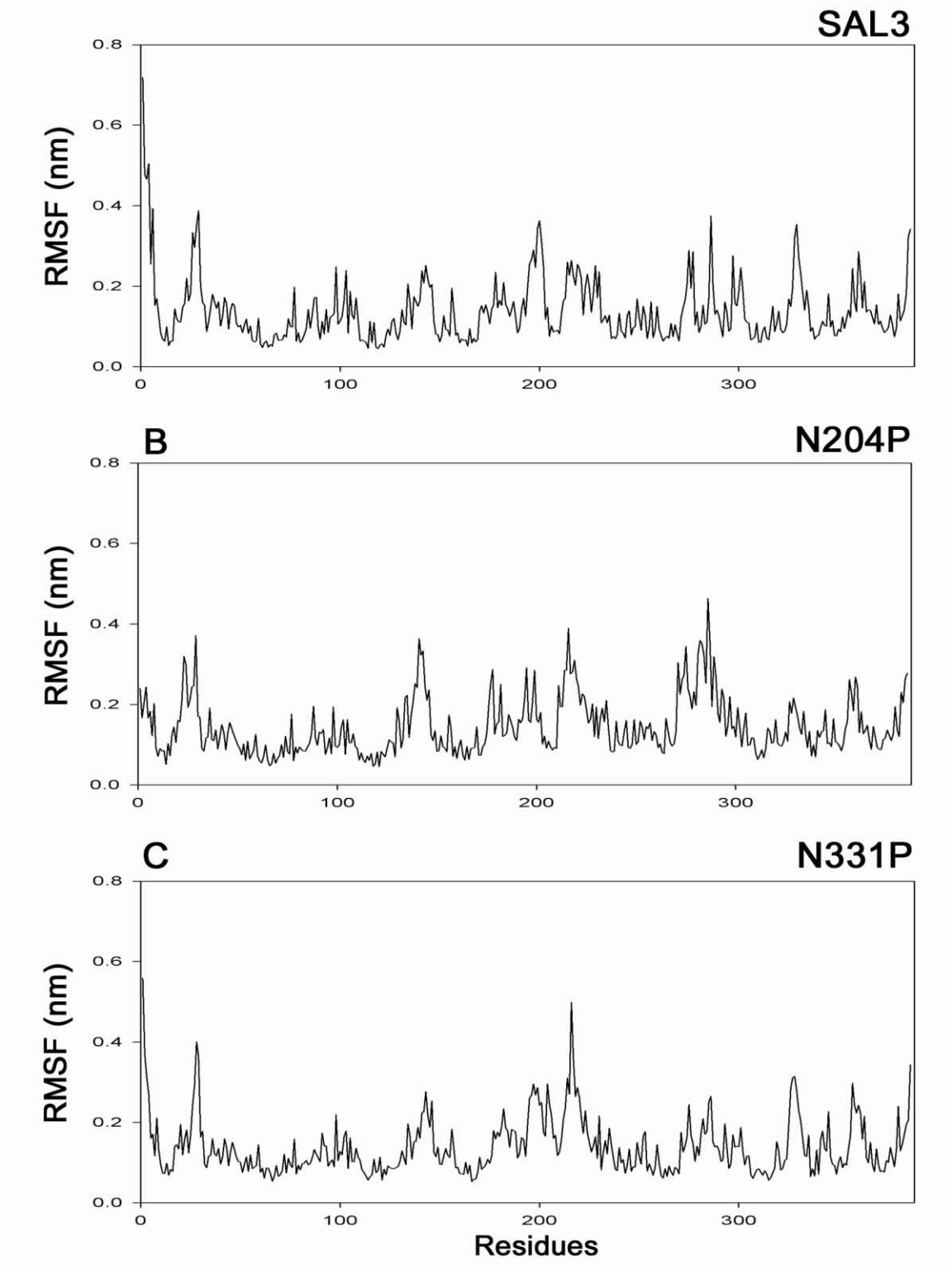


Figure 3.10: Root mean square fluctuation of α-carbon atoms as a function of residue of wild type and mutants from RMSF study during the time course of simulation is shown. (A) SAL3 (Wild type). (B) Mutant N204P. (C) Mutant N331P.

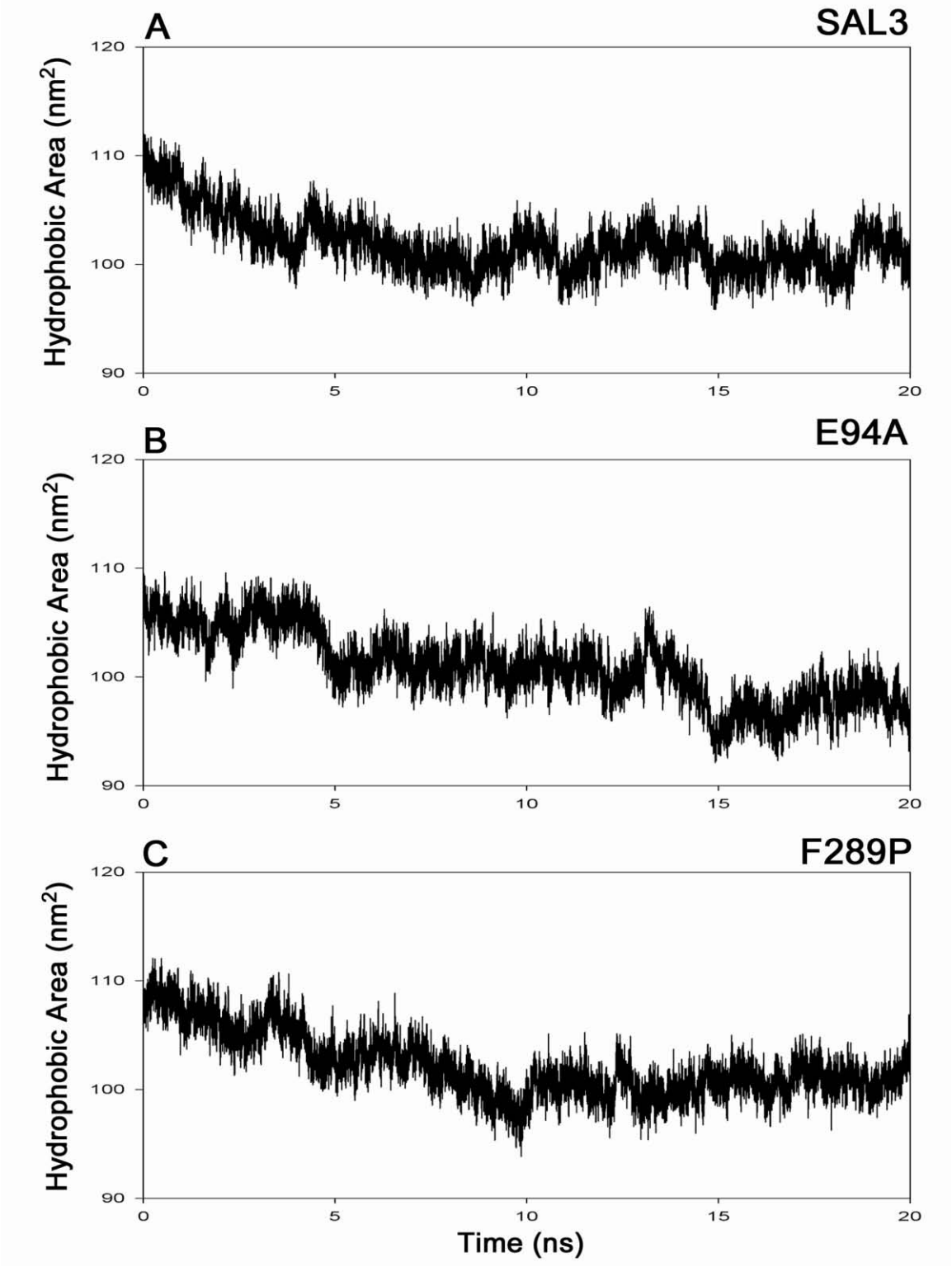


Figure 3.11: The solvent accessible surface area (SASA) of wild type and mutants of SAL3 from ProSA study during the time course of simulation is shown. (A) SAL3 (Wild type). (B) Mutant E94A. (C) Mutant F289P.

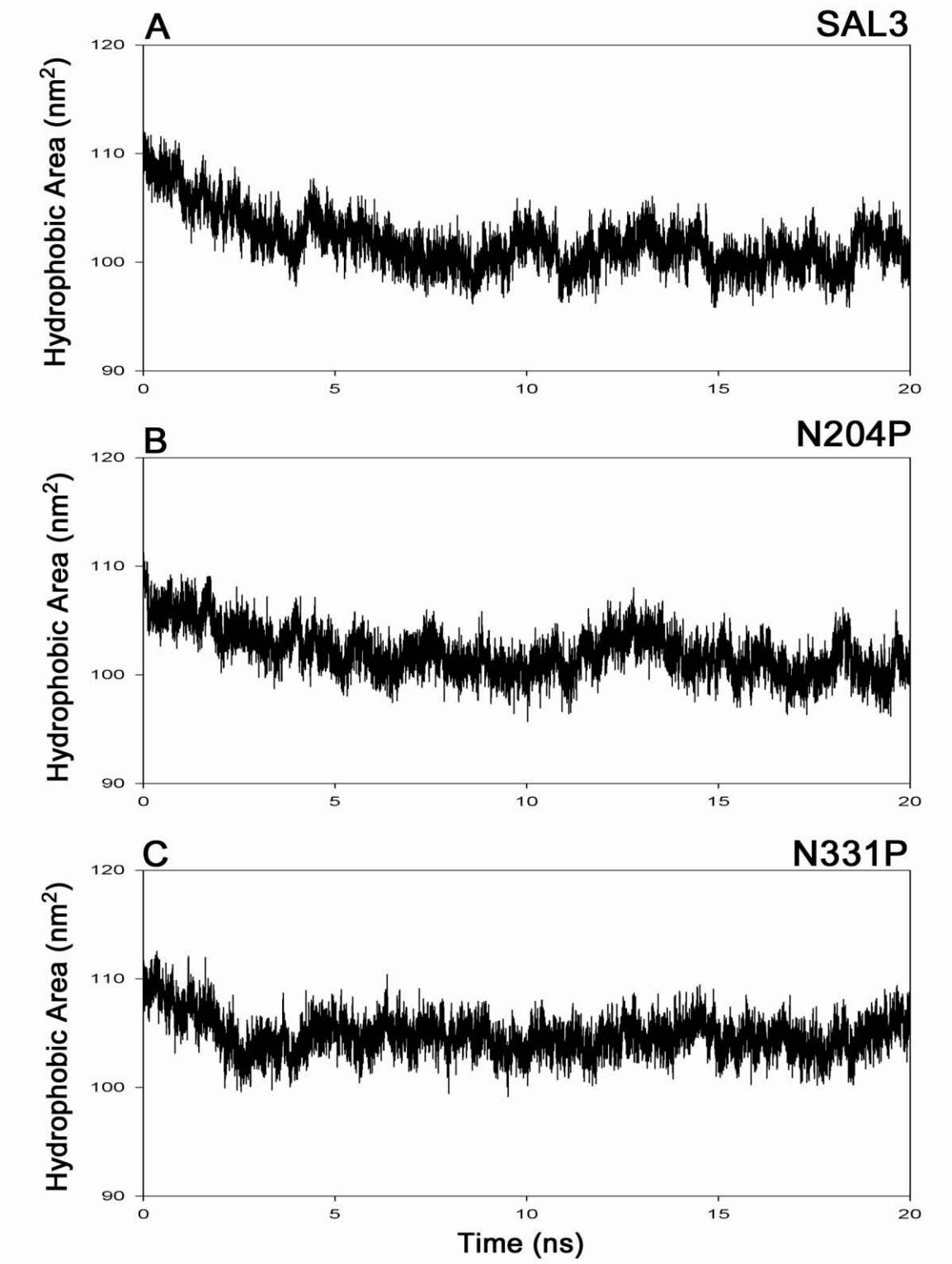


Figure 3.12: The solvent accessible surface area (SASA) of wild type and mutants of SAL3 from RMSF study during the time course of simulation is shown. (A) SAL3 (Wild type). (B) Mutant N204P. (C) Mutant N331P.

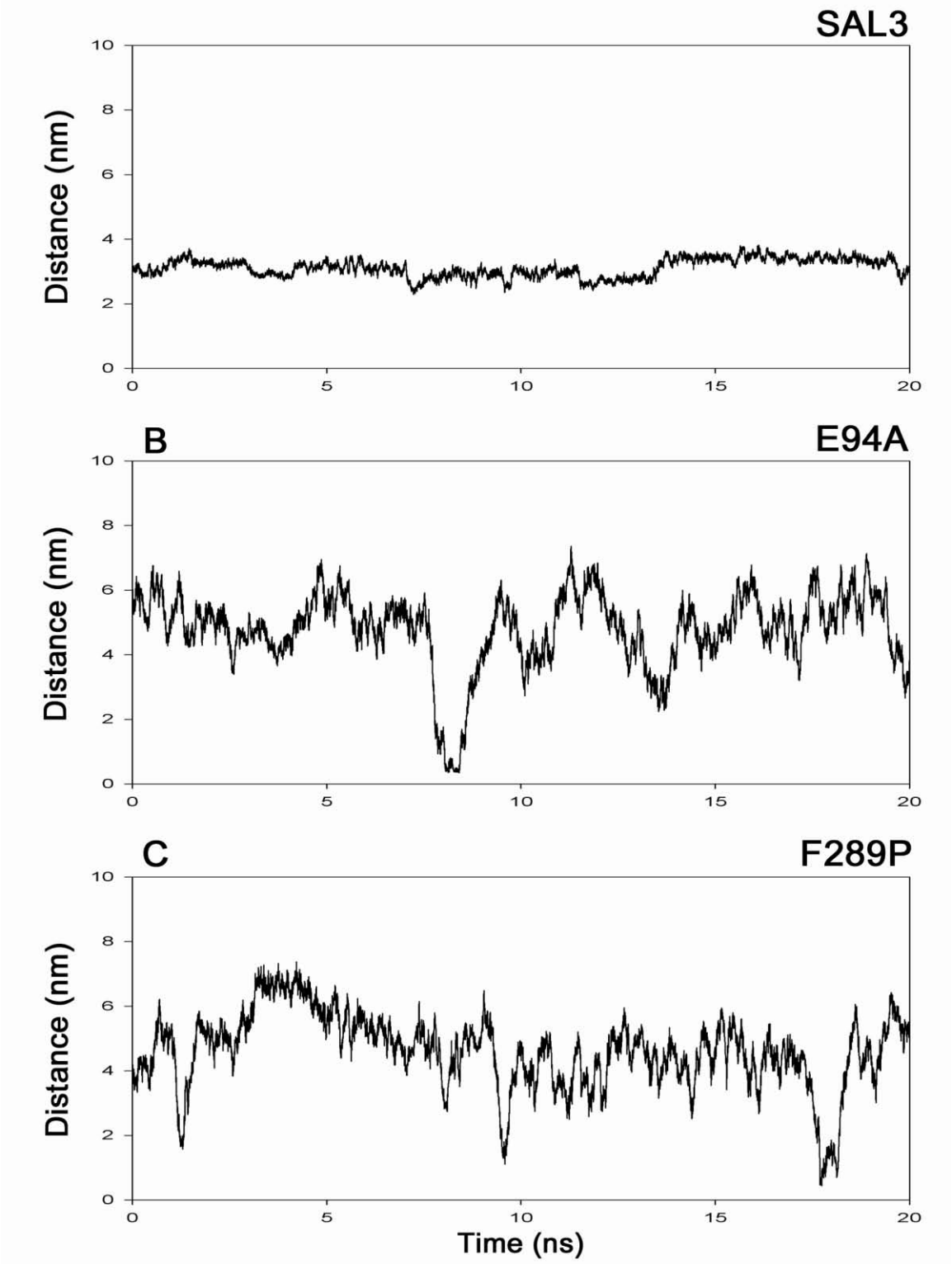


Figure 3.13: The distance between first and last alpha Carbon atoms of wild type and mutants of SAL3 from ProSA study during the time course of simulation is shown. (A) SAL3 (Wild type). (B) Mutant E94A. (C) Mutant F289P.

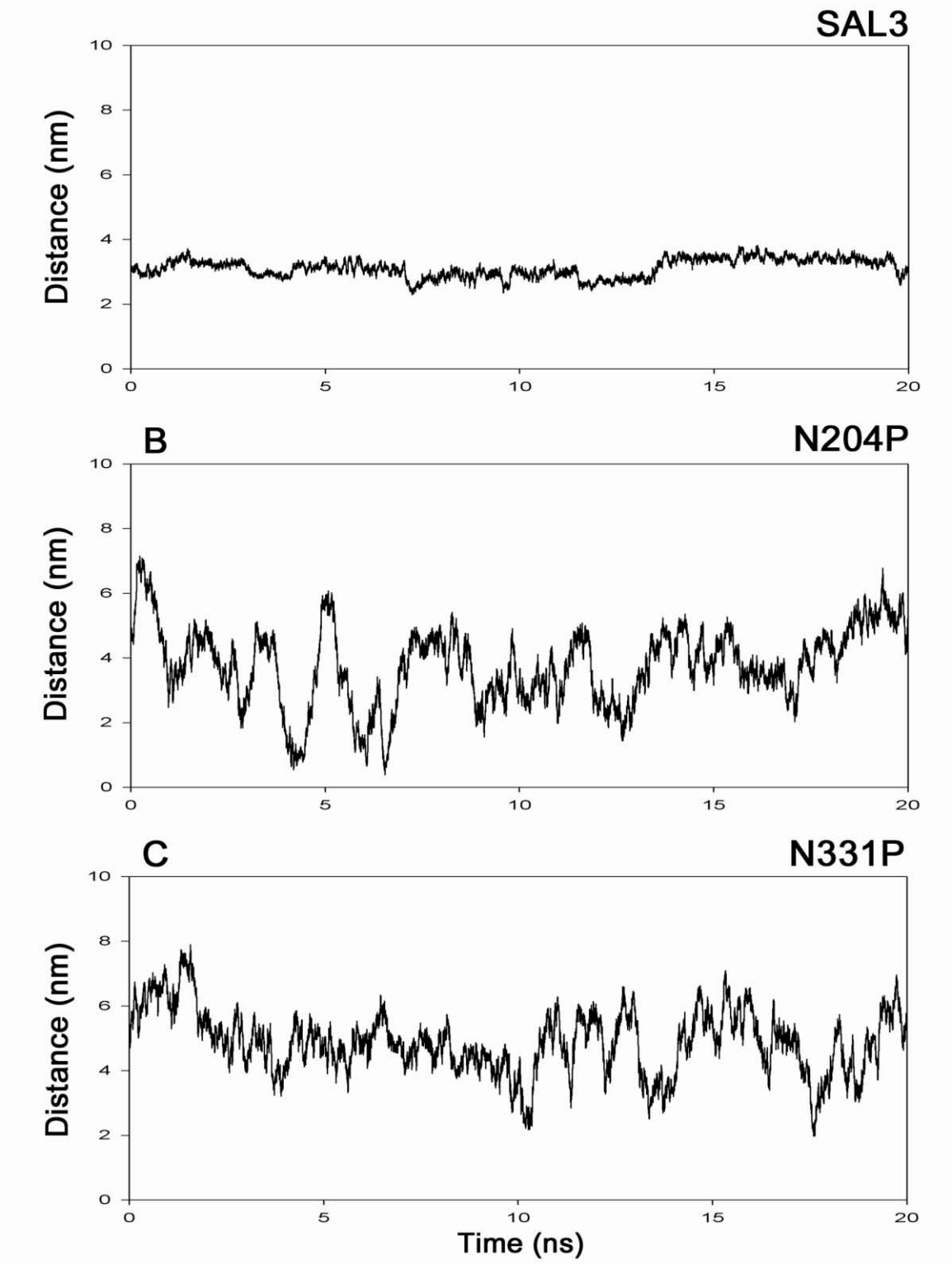


Figure 3.14: The distance between first and last alpha Carbon atoms of wild type and mutants of SAL3 from RMSF study during the time course of simulation is shown. (A) SAL3 (Wild type). (B) Mutant N204P. (C) Mutant N331P.

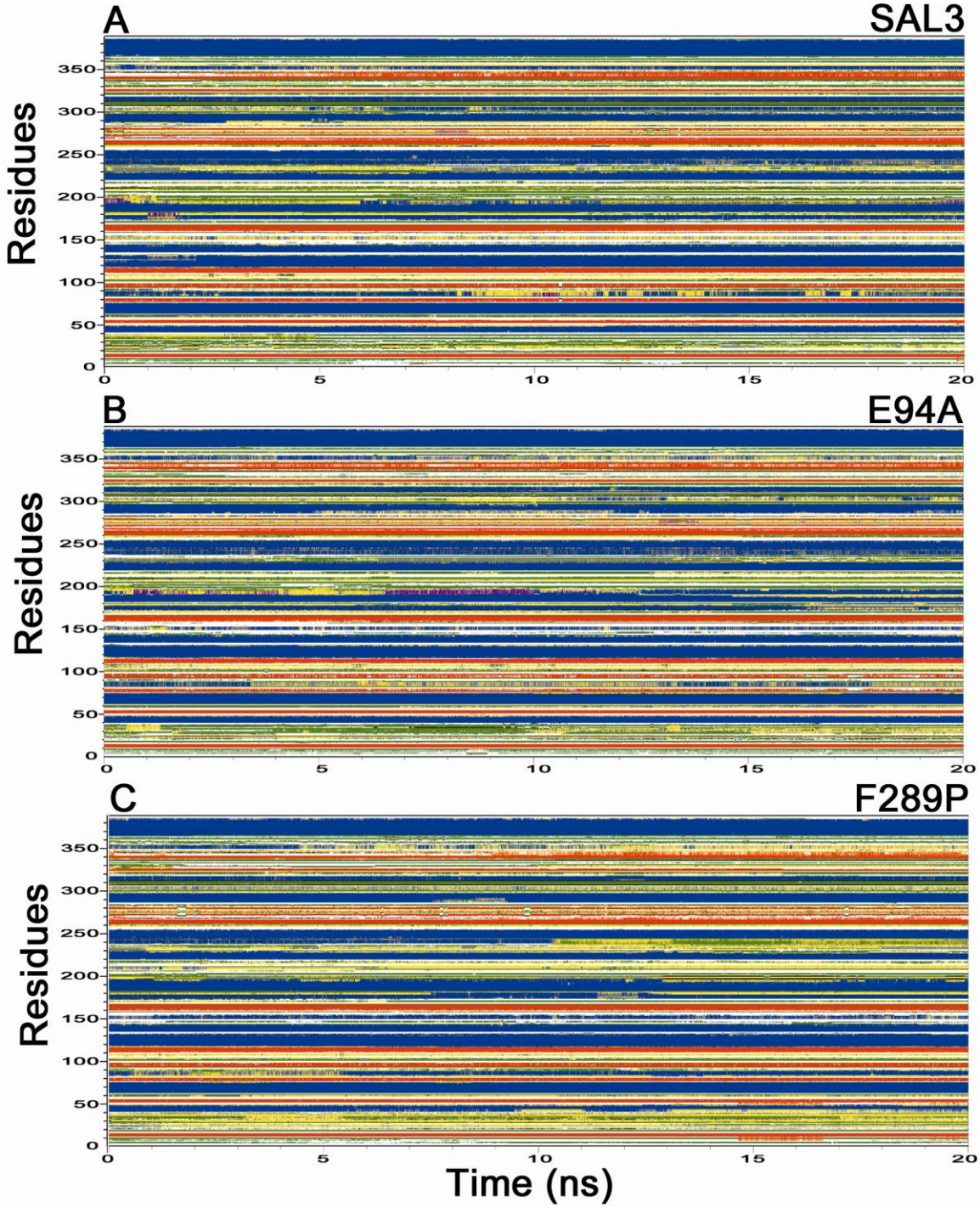


Figure 3.15: The secondary structural analysis of wild type and mutants of SAL3 from ProSA study during the time course of simulation are shown. (A) SAL3 (Wild type). (B) Mutant E94A. (C) Mutant F289P.

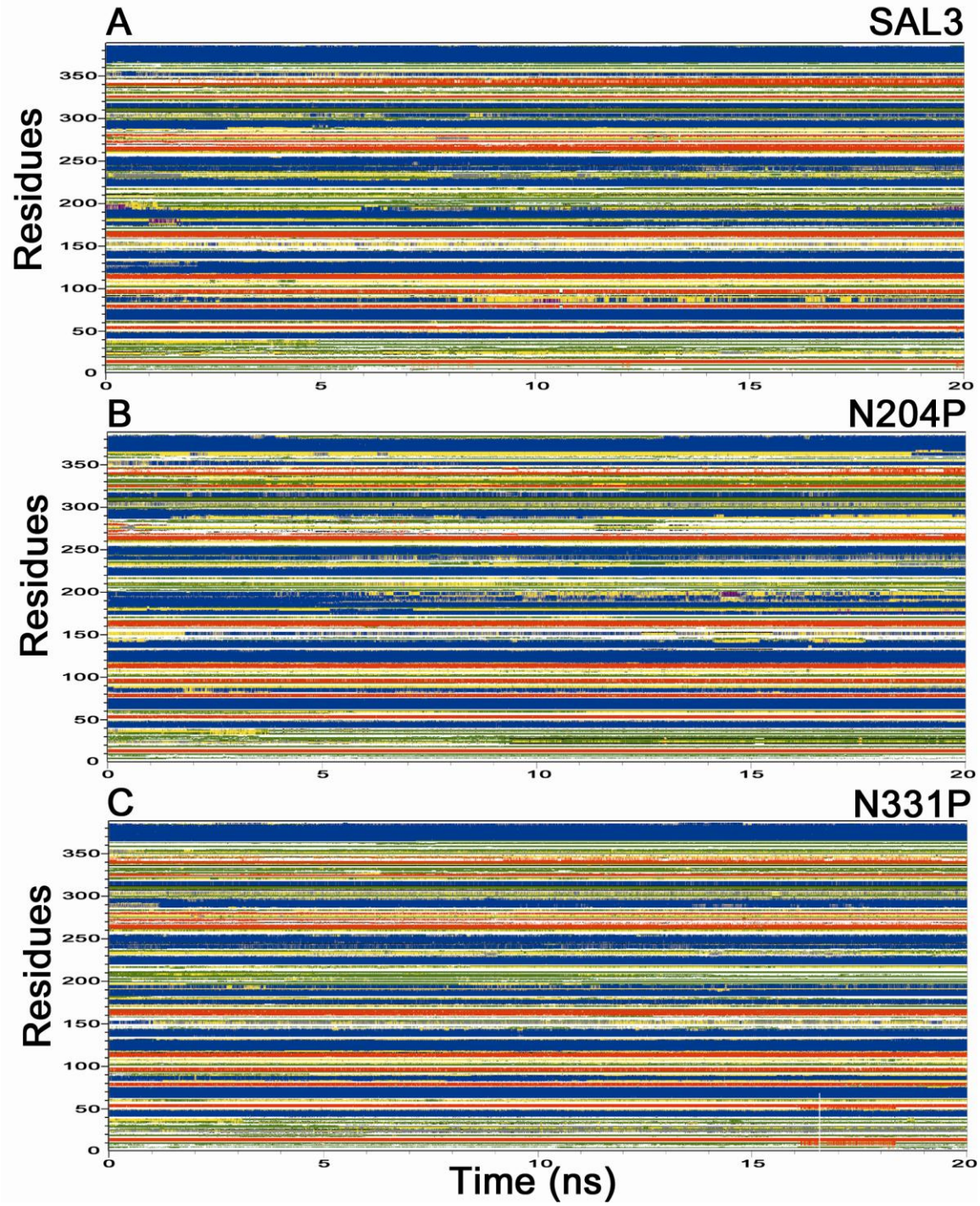


Figure 3.16: The secondary structural analysis of wild type and its mutants of SAL3 from RMSF study during the time course of simulation is shown. (A) SAL3 (Wild type). (B) Mutant N204P. (C) Mutant N331P.

3.4.3.2. SXL2 MD analysis: The mutants identified from ProSA and RMSF study were studied with the wild type of SXL2.

Root Mean Square Deviation (RMSD):

The degree of conformational changes of wild type SXL2 and its identified mutants is monitored by root mean square deviation of alpha Carbon atom during the course of molecular dynamics simulation. The backbone RMSD of conformations from production run relative to its initial structure has been studied for wild type SXL2 with mutants identified from ProSA and RMSF study were represented in Figure 3.17 and 3.18 respectively. The time-dependent RMSD analysis reflects that the wild type and its mutants of SXL2 have less deviation in RMSD throughout the simulation and implies that the 3D structures are relatively stable except the N219G mutant. Among the identified mutants of SXL2, N219G mutant has comparatively greater RMSD values. Among the identified mutants, E220G mutant from RMSF study was found comparatively stable than the other mutants as well as wild type SXL2, perhaps due to the stabilization of the loop where E220G resides as well as the local conformation. The average and standard deviation of RMSD values of wild type and its mutants of SXL2 have been summarized in Table 3.6 which reflects the order of stability from the average values of RMSD as (i) M161F < SXL2 from ProSA study and (ii) E220G < SXL2 < N219G from RMSF study of SXL2. The RMSD analysis reflects that the structural integrity of the mutants is affected to certain extent than wild type SXL2. The E220G mutant had steady equilibrium and had better RMSD values throughout the simulation. The M161F mutant also had steady equilibrium with less fluctuation than the N219G mutant.

Radius of Gyration (Rg):

The radius of gyration has been analyzed in a time-dependent manner to investigate the compactness of wild type and its mutants of SXL2 from ProSA and RMSF study and has been presented in Figure 3.19 and 3.20 respectively. The radius of gyration reflects the packing of amino acid residues throughout the simulation thereby stability and folding

rate of SXL2. The wild type and identified mutants have similar profile of Rg throughout the simulation. The order of Rg values from Table 3.6 reflects that the mutants M161F from ProSA study have better Rg profile than wild type SXL2 and other mutants from the RMSF study. The order of Rg from the average values of 'End to End chain' distance is (i) M161F < SXL2 from ProSA study and (ii) SAL3 < N219G < E220G from RMSF study of SAL3. The M161F and E220G mutants had steady Rg profile while E220G had comparatively more Rg values.

Root Mean Square Fluctuation (RMSF):

The local deformability of the protein has been analyzed through root mean square fluctuations of alpha Carbon atoms from MD simulations. The RMSF plotted against the residues of wild type and its mutants of SXL2 from ProSA and RMSF study has been presented in Figure 3.21 and 3.22 respectively. Appreciable difference of RMSF between wild type and mutants of SXL2 is observed at the mutation site and their respective local regions. The average values of RMSF indicate that the mutants are considerably similar to that of wild type SXL2. The order of average values of RMSF is (i) E220G < M161F < SXL2 < N219G. The mutants stabilize the fluctuations at the N-terminus comparatively with wild type while the M161F and E220G have better RMSF profile that the N219G mutant which is evident from the RMSF plots. Similar to SAL3, SXL2 might have an inherent flexible terminus which could have been stabilized by its quaternary form

Solvent Accessible Surface Area (SASA):

The trend of SASA analysis from Figure 3.23 and 3.24 reflects that the N219G and E220G mutants from RMSF study have more hydrophobic area than M161F mutant and wild type SXL2. The hydrophobic area of E220G mutant is also consistent throughout the simulation. The average values of SASA indicate that M161F mutant from ProSA study has considerably less hydrophobic area than wild type SXL2. The order of average values of RMSF is (i) M161F < SXL2 < N219G < E220G. It is evident from the SASA plots that the M161F and E220G mutants had comparatively steady SASA profile that the wild type and N219G mutant which had high fluctuations. These mutants might stabilize the

hydrophobic contacts and favoured the structural stabilization of their respective local conformation.

End to End chain distance between first and last alpha Carbon atoms:

The distance between first and last alpha Carbon atoms of wild type and mutants of SXL2 from ProSA and RMSF study were analyzed throughout the simulation which would aid to understand the structural integrity of wild type and its mutants of SXL2 and have been presented in Figure 3.25 and 3.26 respectively. The order of stability from the average values of 'End to End chain' distance is SXL2 < M161F < E220G < N219G. Overall, a similar trend of end to end chain distance between the mutants from RMSF study has been observed while it is found that M161F mutant from ProSA study is consistent throughout the simulation and better than wild type SXL2. Considerable changes have been observed in end to end chain distance in mutants N219G and E220G that during the time course of simulation which reflects the inherent flexible terminus of SXL2 and is consistent with RMSF analysis.

Analysis of Secondary Structure:

The secondary structure analysis shows the structural variation of each residue of wild type and mutants of SXL2. The plots of simulation from ProSA and RMSF study have been presented in Figure 3.27 and 3.28 respectively. The structural stabilization of the corresponding secondary structure where the identified mutants M161F, N219G and E220G resides were observed (Table 3.7).

Table 3.6: Summarized parameters of MD simulations for wild type and *in silico* mutants of SXL2.

Method*	Variant	End to End Distance (nm)	Radius of gyration (nm)	RMSD (nm)	RMSF (nm)	SASA (nm^2)
	SXL2	2.89±0.28	1.97±0.02	0.24±0.05	0.14±0.07	101.23±3.85
PROSA	M161F	3.04±0.11	1.95±0.01	0.24±0.03	0.13±0.07	100.00±3.17
RMSF	N219G	3.77±0.29	1.99±0.01	0.29±0.06	0.16±0.08	104.05±4.31
	E220G	3.14±0.34	1.99±0.01	0.20±0.03	0.12±0.07	104.22±1.84

* methodology implemented for the identification of the mutants

Table 3.7: Stabilization of the secondary structure by the mutants of SXL2.

STUDY	SXL2	Secondary Structure*
PROSA	M161F	bend
RMSF	N219G	turn
	E220G	loop

* the secondary structure where the mutants resides

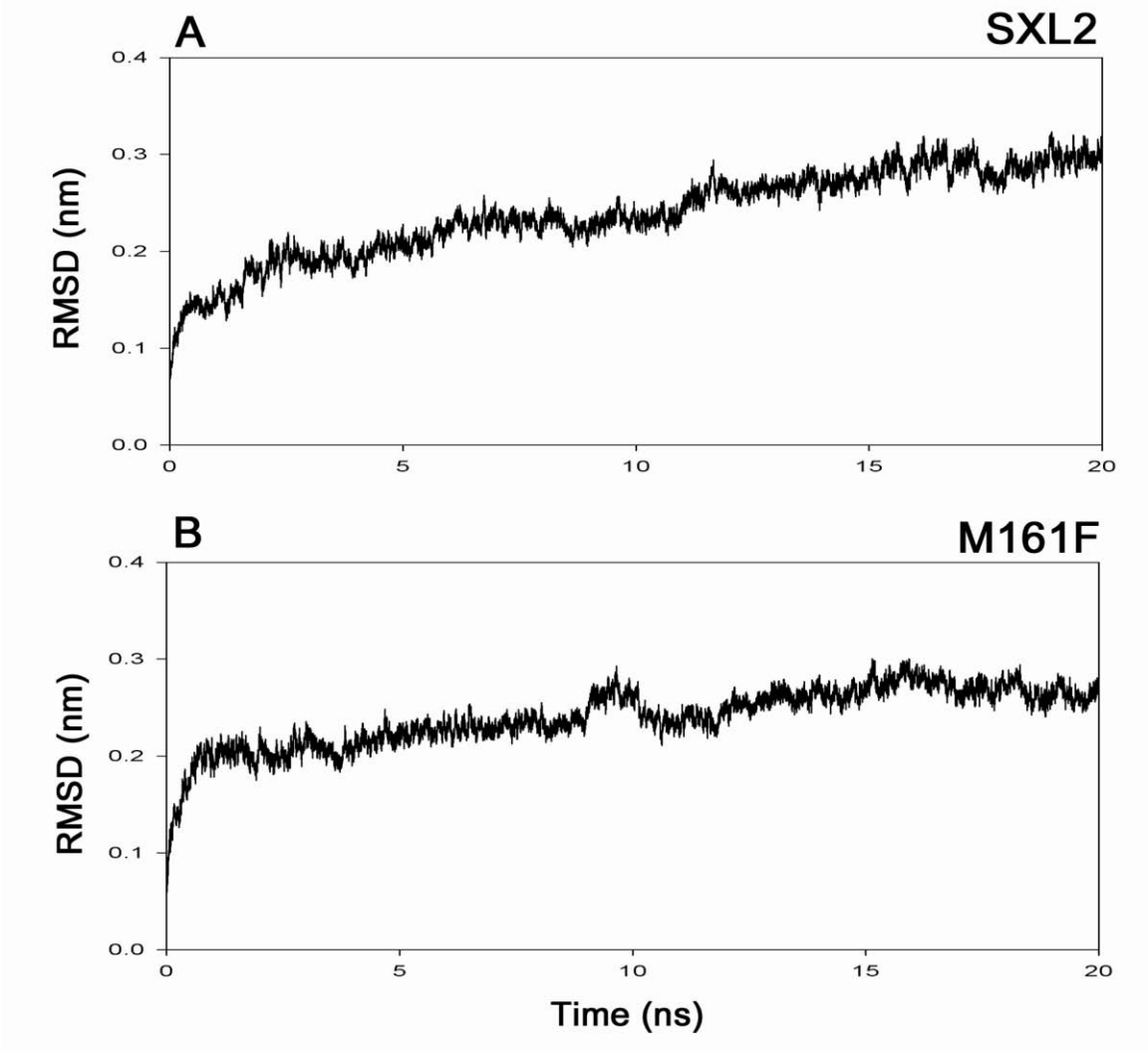


Figure 3.17: The root-mean-square deviation (RMSD) to the starting structure as a function of time of wild type and mutants of SXL2 from ProSA study during the time course of simulation is shown. (A) SXL2 (Wild type). (B) Mutant M161F.

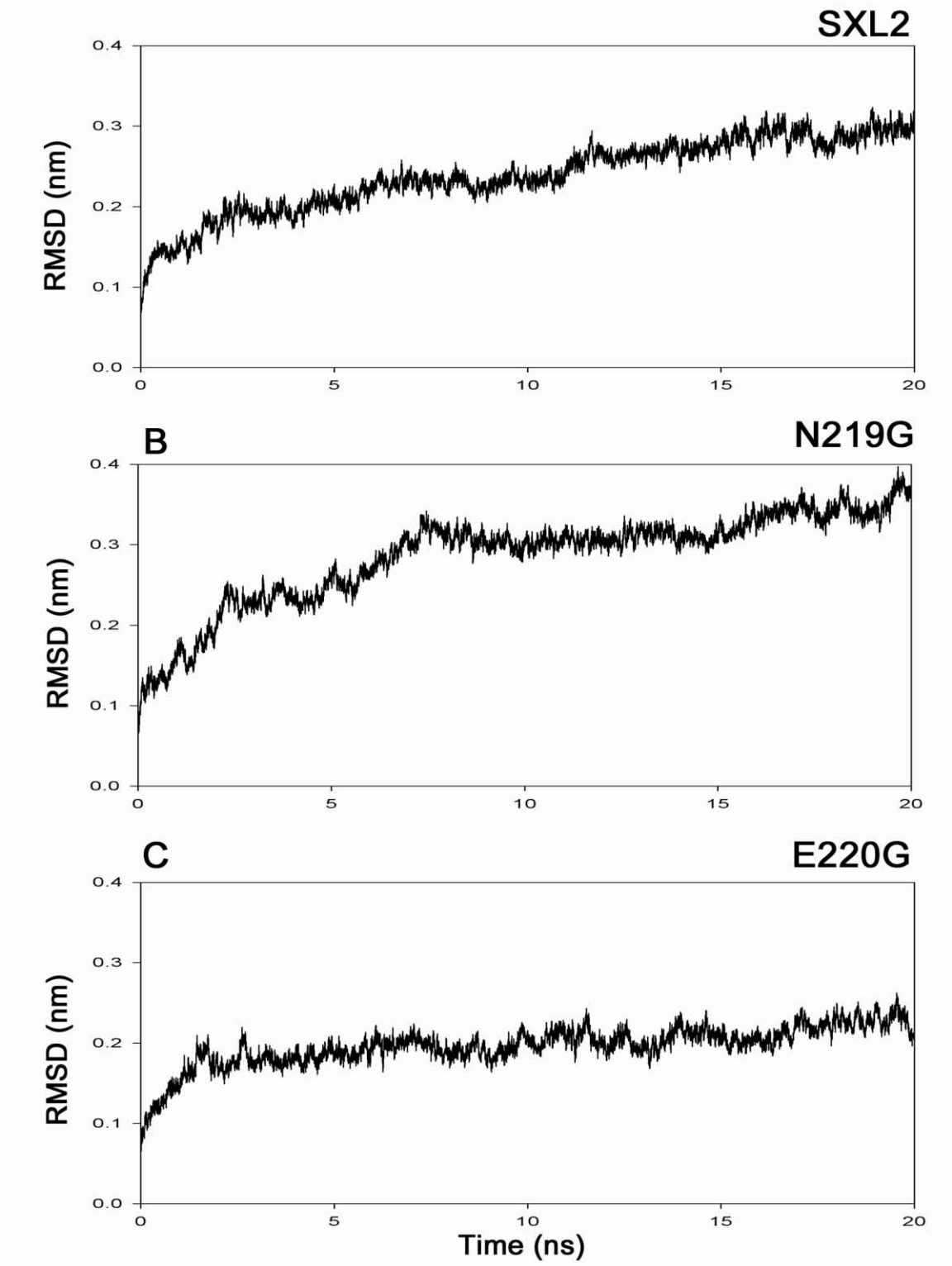


Figure 3.18: The root-mean-square deviation (RMSD) to the starting structure as a function of time of wild type and mutants of SXL2 from RMSF study during the time course of simulation is shown. (A) SXL2 (Wild type). (B) Mutant N219G. (C) Mutant E220G.

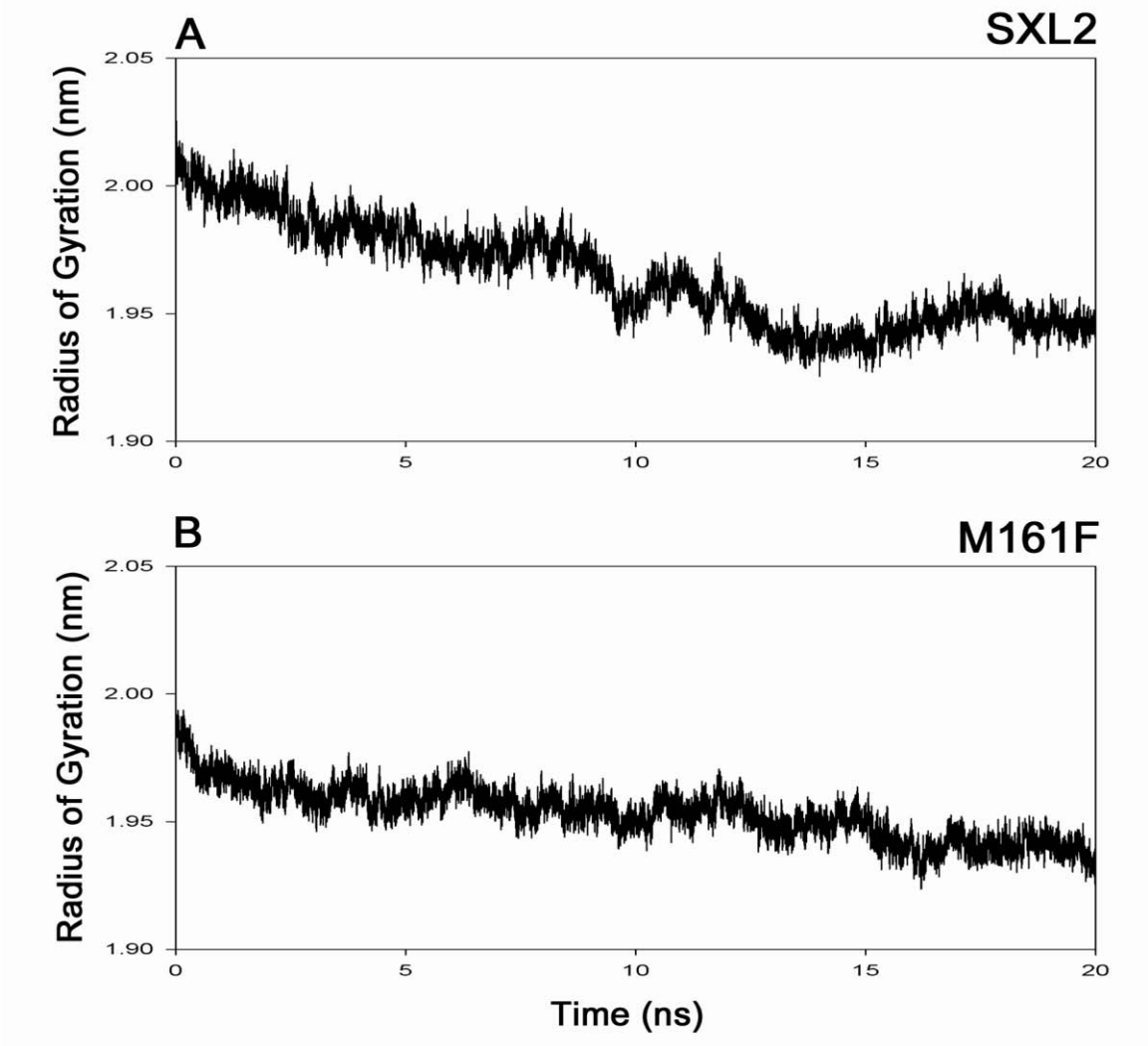


Figure 3.19: Radius of gyration of α-carbon atoms as a function of time of wild type and mutants of SXL2 from ProSA study during the time course of simulation is shown. (A) SXL2 (Wild type). (B) Mutant M161F.

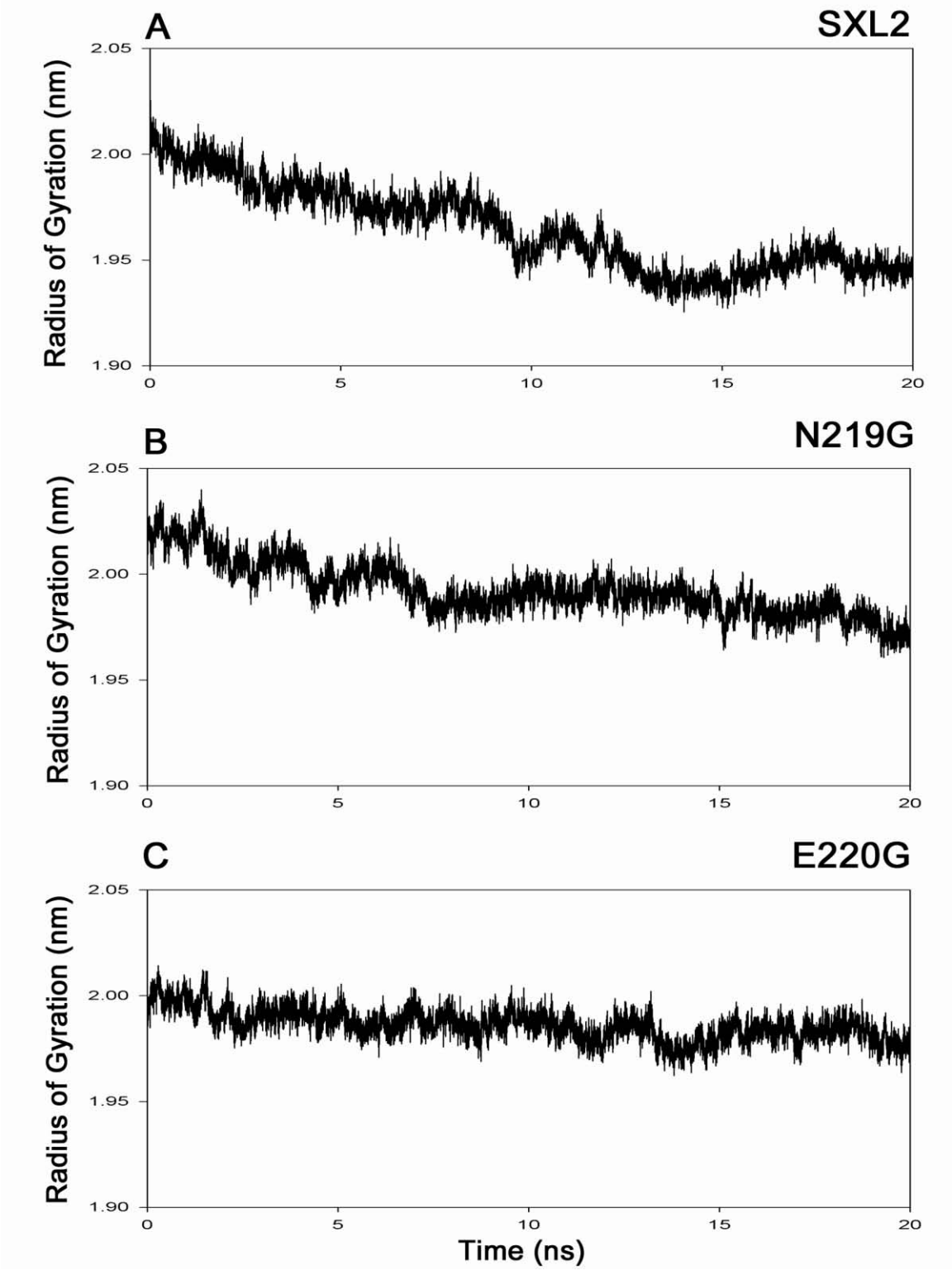


Figure 3.20: Radius of gyration of α-carbon atoms as a function of time of wild type and mutants of SXL2 from RMSF study during the time course of simulation is shown. (A) SXL2 (Wild type). (B) Mutant N219G. (C) Mutant E220G.

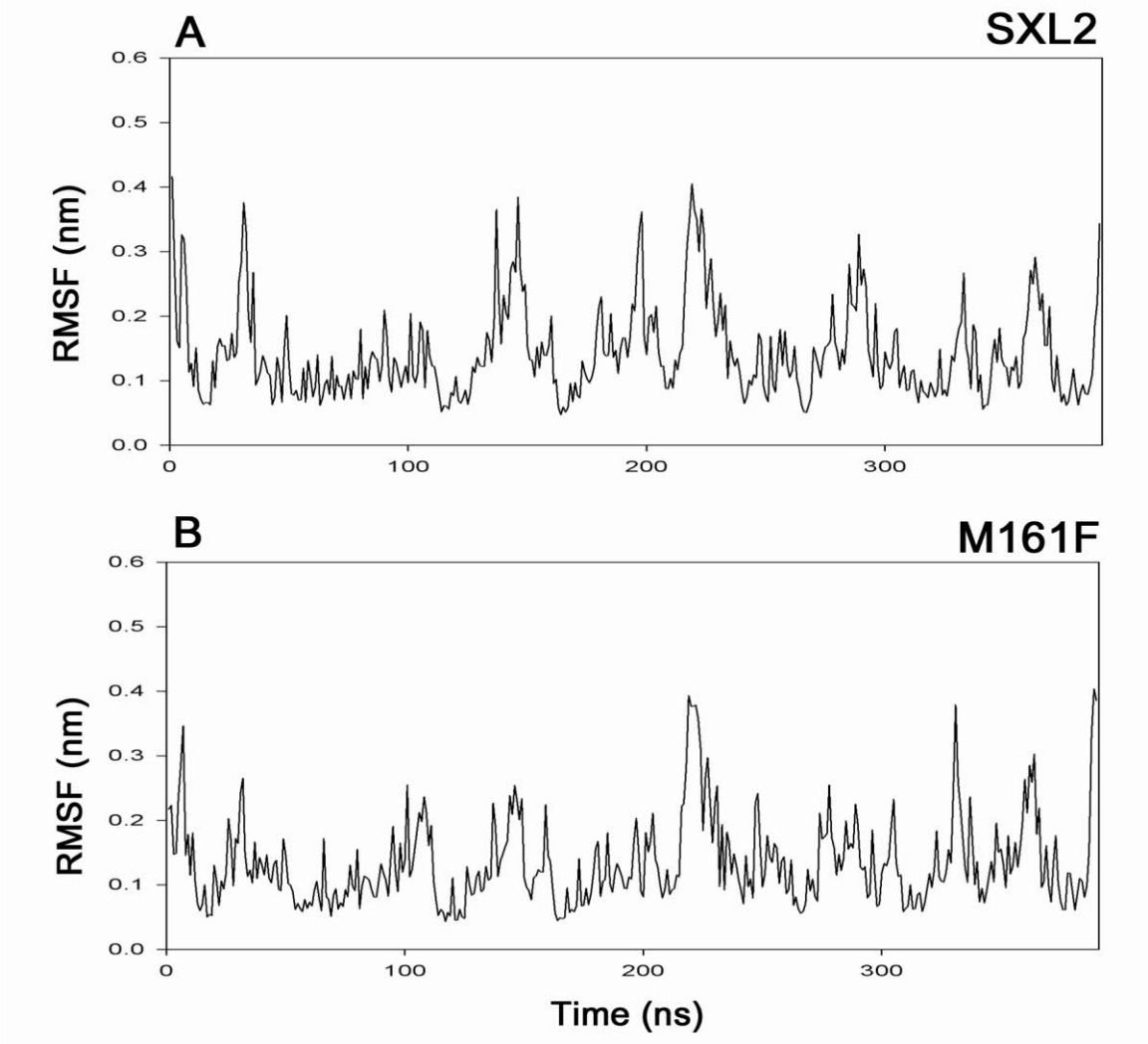


Figure 3.21: Root mean square fluctuation of α-carbon atoms as a function of residue of wild type and mutants of SXL2 from ProSA study during the time course of simulation is shown. (A) SXL2 (Wild type). (B) Mutant M161F.

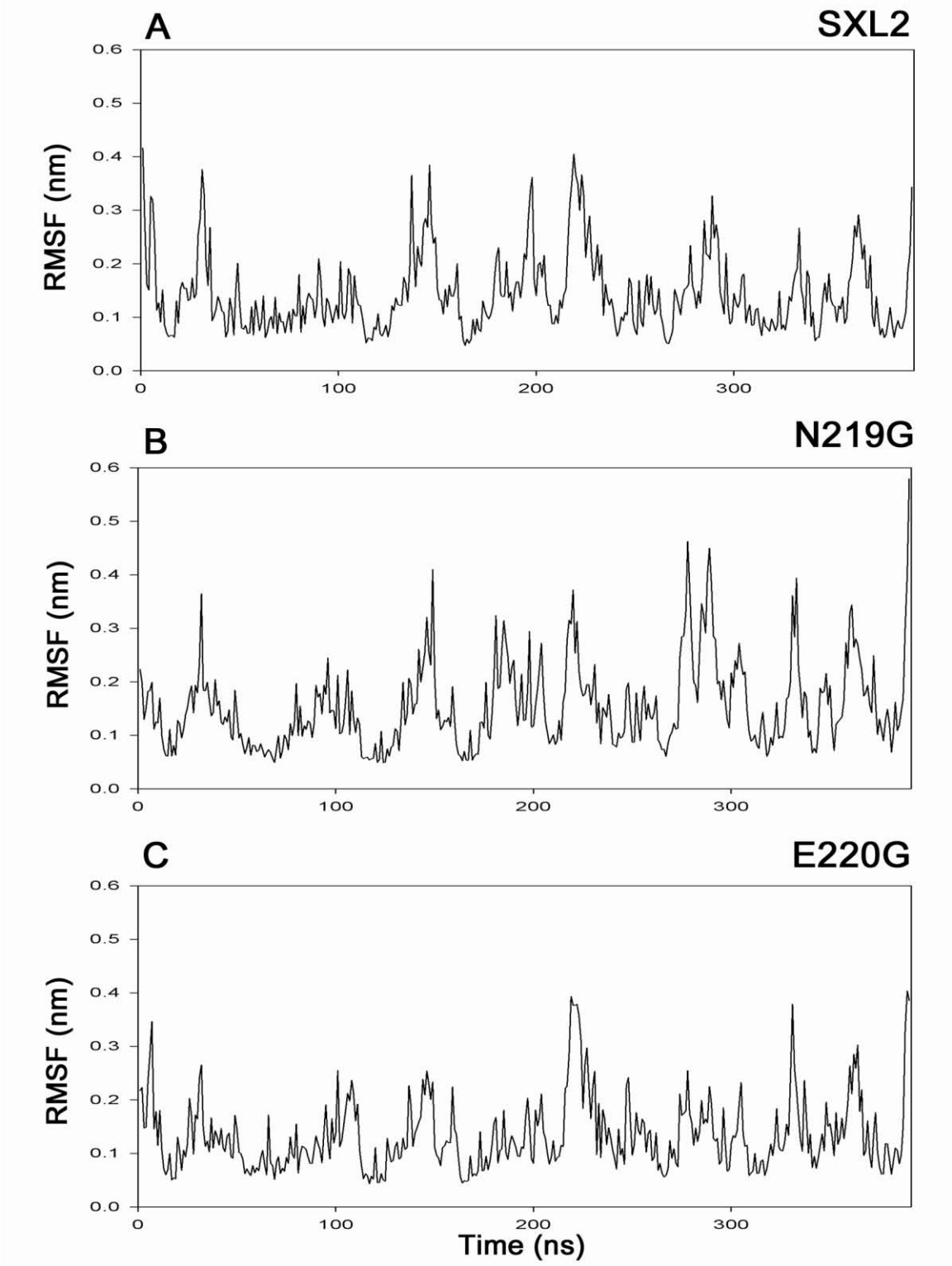


Figure 3.22: Root mean square fluctuation of α-carbon atoms as a function of residue of wild type and mutants of SXL2 from RMSF study during the time course of simulation is shown. (A) SXL2 (Wild type). (B) Mutant N219G. (C) Mutant E220G.

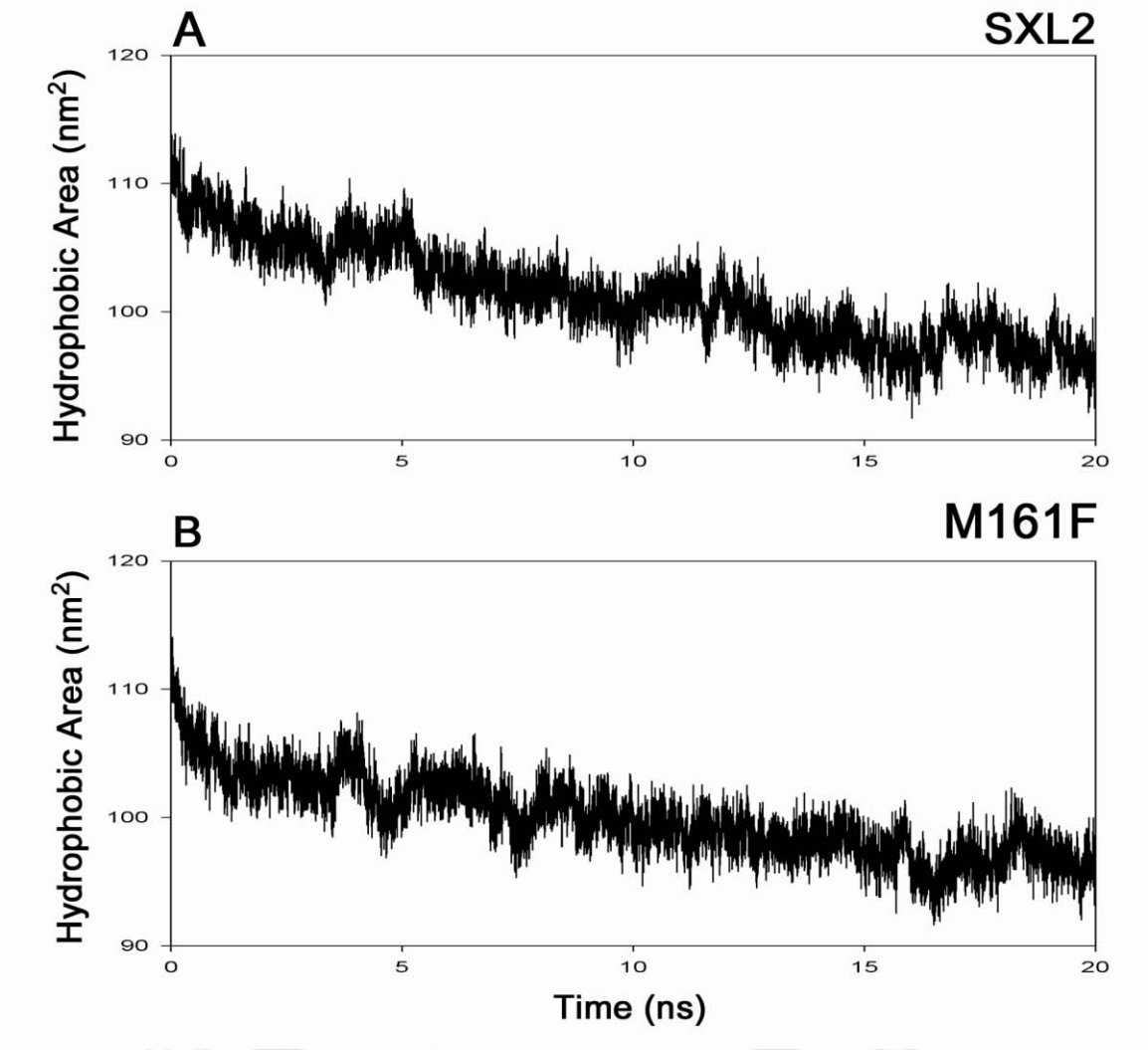


Figure 3.23: The solvent accessible surface area (SASA) of wild type and mutants of SXL2 from ProSA study during the time course of simulation is shown. (A) SXL2 (Wild type). (B) Mutant M161F.

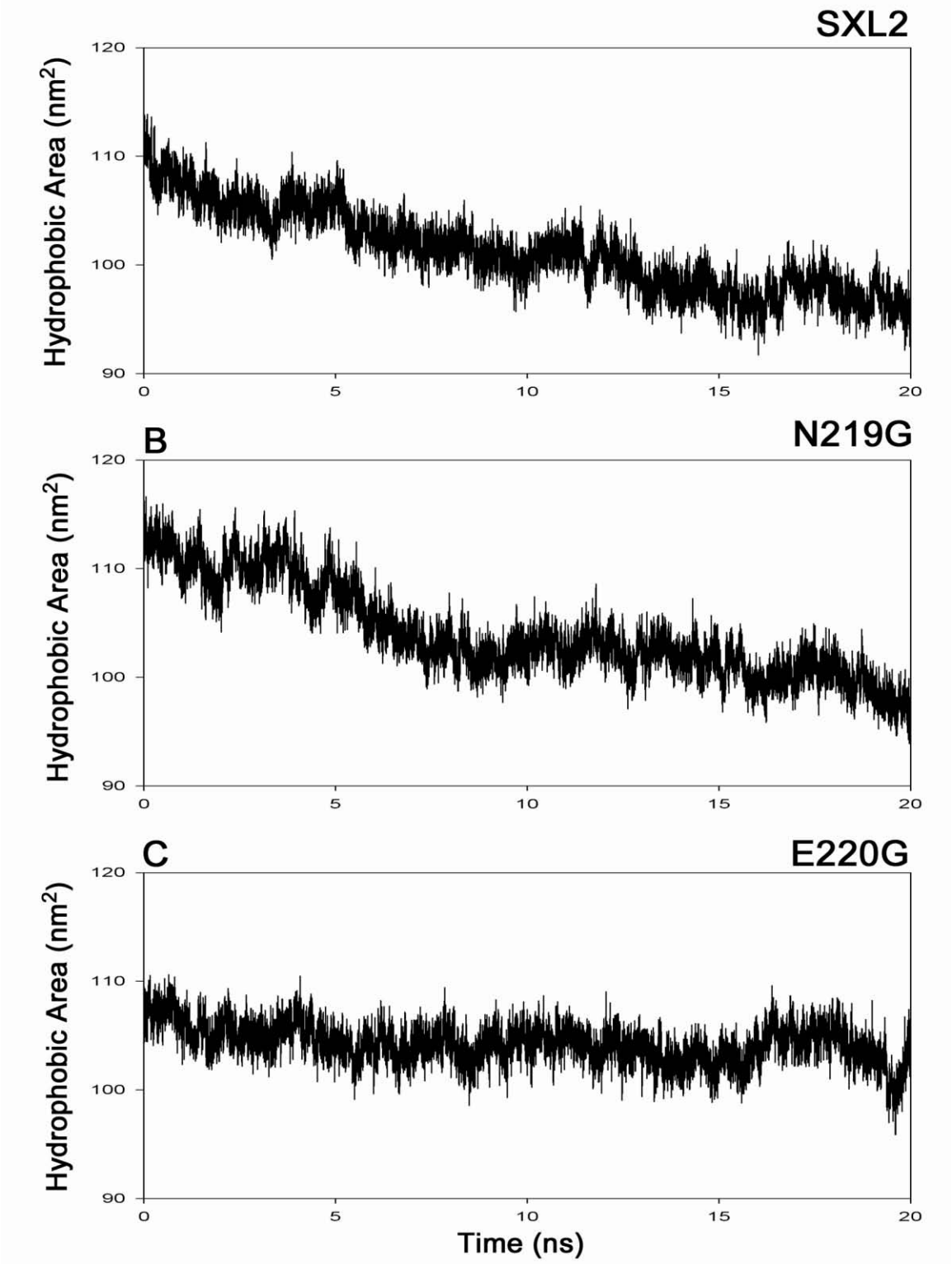


Figure 3.24: The solvent accessible surface area (SASA) of wild type and mutants of SXL2 from RMSF study during the time course of simulation is shown. (A) SXL2 (Wild type). (B) Mutant N219G. (C) Mutant E220G.

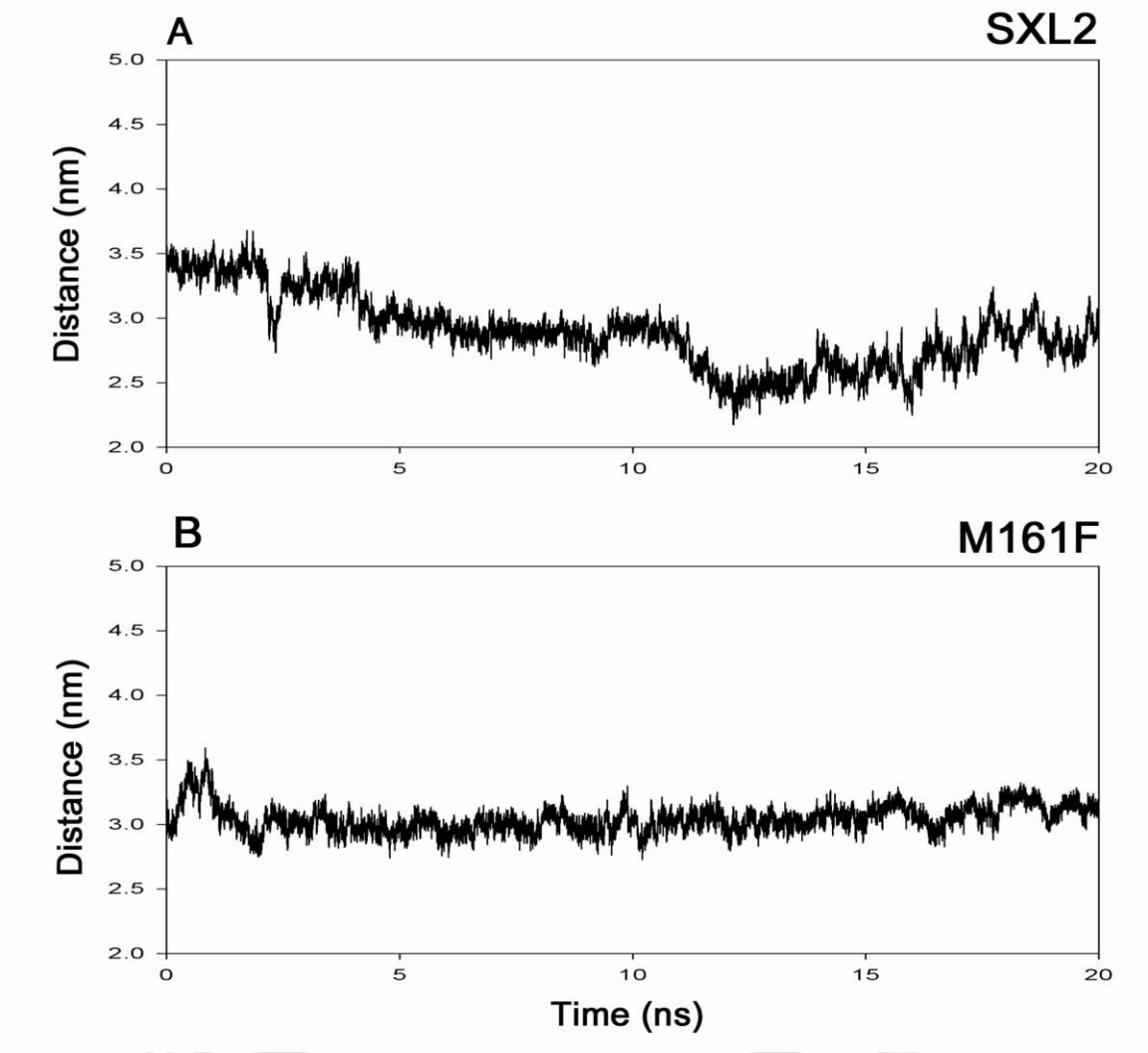


Figure 3.25: The distance between first and last alpha Carbon atoms of wild type and mutants of SXL2 from ProSA study during the time course of simulation is shown. (A) SXL2 (Wild type). (B) Mutant M161F.

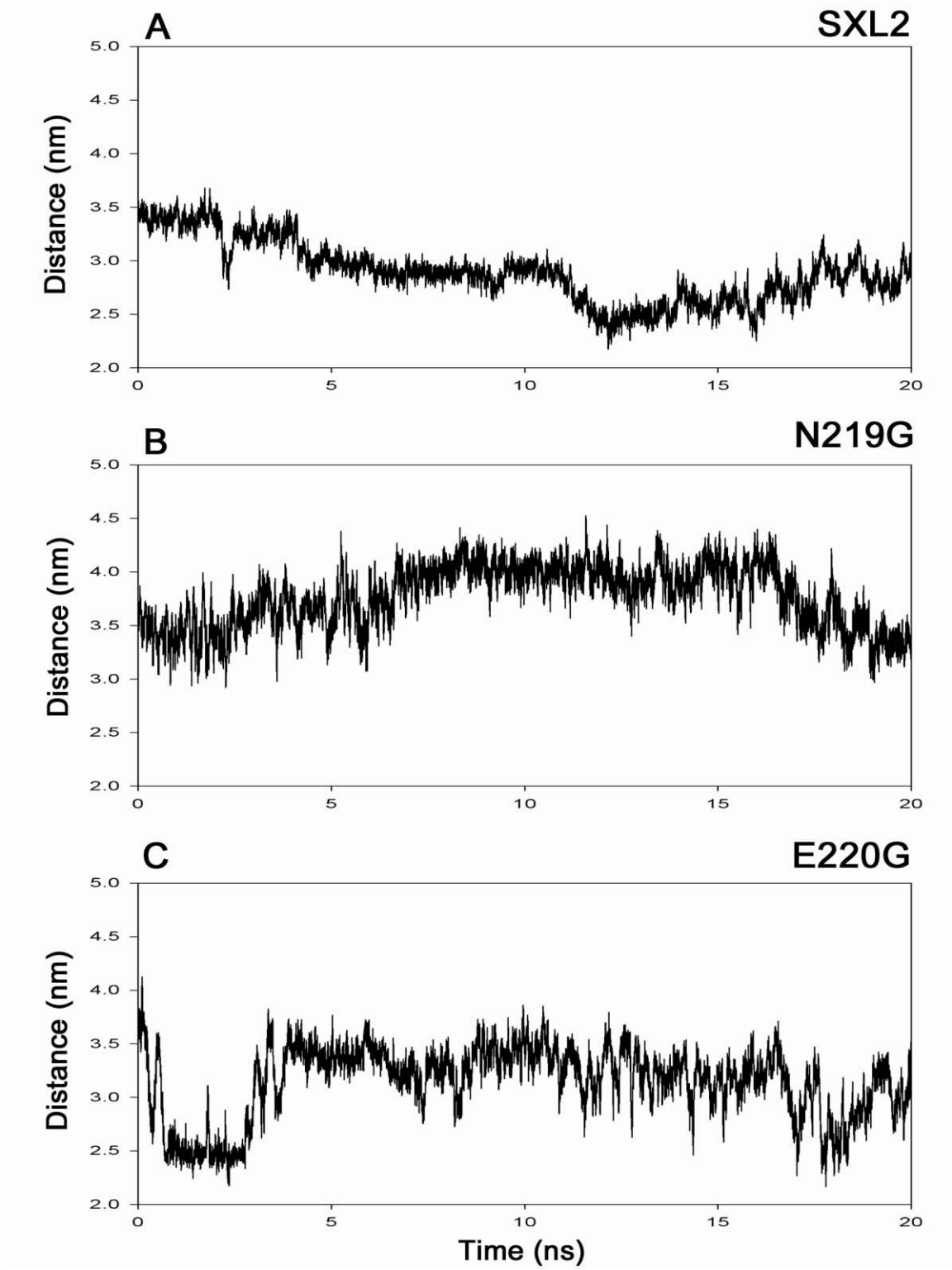


Figure 3.26: The distance between first and last alpha Carbon atoms of wild type and mutants of SXL2 from RMSF study during the time course of simulation is shown. (A) SXL2 (Wild type). (B) Mutant N219G. (C) Mutant E220G.

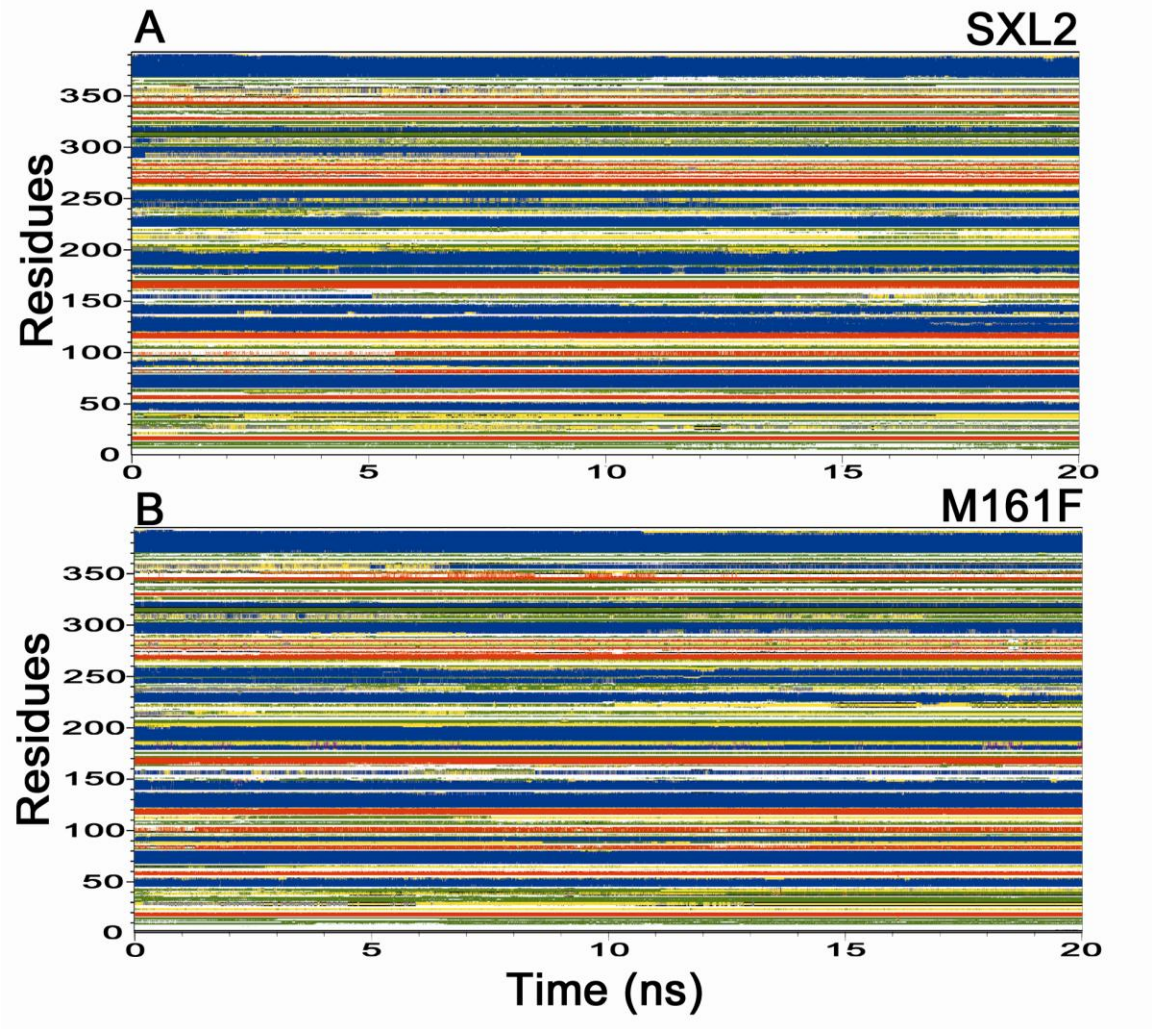


Figure 3.27: The secondary structural analysis of wild type and mutants of SXL2 from ProSA study during the time course of simulation is shown. (A) SXL2 (Wild type). (B) Mutant M161F.

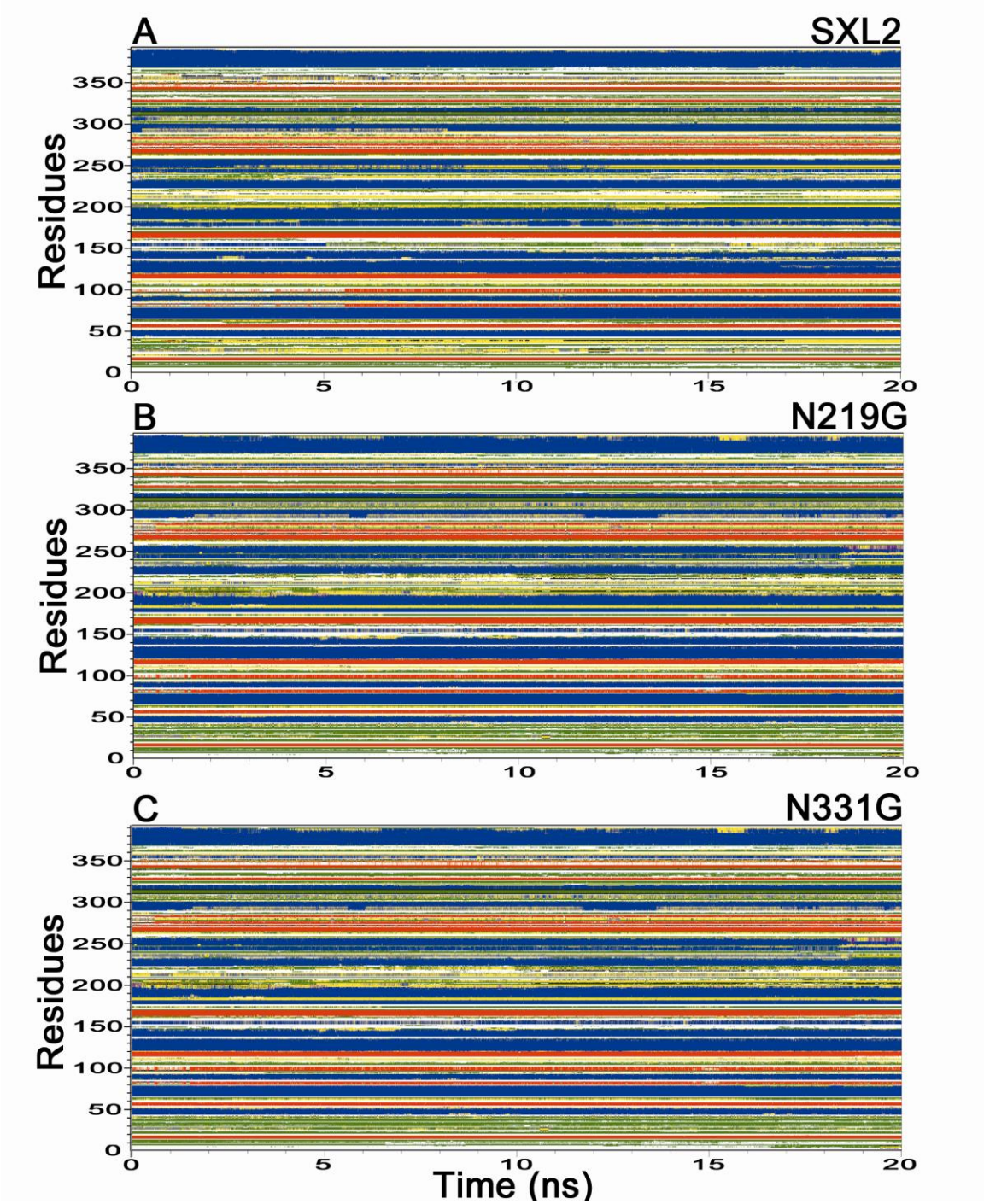


Figure 3.28: The secondary structural analysis of wild type and mutants of SXL2 from RMSF study during the time course of simulation is shown. (A) SXL2 (Wild type). (B) Mutant N219G. (C) Mutant E220G.

3.4.3.3. MD Analysis of PML: The mutants identified from ProSA and RMSF study were studied with the wild type of PML.

Root Mean Square Deviation (RMSD):

The degree of conformational changes of wild type wild type and its mutants of PML identified is monitored by root mean square deviation of alpha Carbon atom during the course of molecular dynamics simulation. The backbone RMSD of conformations from production run relative to its initial structure has been studied for wild type PML with mutants identified from ProSA and RMSF study were represented in Figure 3.29 and 3.30 respectively. The time-dependent RMSD analysis reflects that the wild type and its mutants of PML have less deviation in RMSD throughout the simulation and implies that the 3D structures are relatively stable. Among ProSA mutants of PML, P219A mutant has comparatively greater RMSD values. This happens in P219A perhaps due to the stabilization of nearby secondary structures as well as the local conformation of the α-helix and especially the beta hairpin in which Pro219 resides. Among mutants identified from RMSF studies of wild type PML, Q242P mutant was found comparatively stable than R91A and R163P mutant, perhaps due to the stabilization of the loop where Glu242 resides and also the stabilization of the consecutive turn. The average and standard deviation of RMSD values of wild type and its mutants of PML have been summarized in Table 3.8 which reflects the order of stability from the average values of RMSD as (i) PML < P219A < P10A <E83R <P7A < P187A from ProSA study and (ii) PML <R91A < R163P < Q242P from RMSF study of PML. The RMSD analysis reflects that the structural integrity of P219A, R91A and R163P mutants is affected to certain extent than other mutants while the P187A mutant found to have better RMSD values throughout the simulation. The mutation of Pro187 into Alanine might lead to the stabilization of the helix and its local conformation. The P7A and Q242P mutants were also found to be more stable among the identified mutants.

Radius of Gyration (Rg):

The radius of gyration has been analyzed in a time-dependent manner to investigate the compactness of wild type and its mutants of PML from ProSA and RMSF study and has been presented in Figure 3.31 and 3.32 respectively. The radius of gyration reflects the packing of amino acid residues throughout the simulation thereby stability and folding rate of PML. The wild type and identified mutants have similar profile of Rg throughout the simulation. The order of Rg values from Table 3.8 reflects that the mutants P10A and P219A from ProSA study as well as the mutants R91A and Q242P have better Rg profile than wild type PML. The Q242P mutant could be better mutant in the context of Rg and RMSD analysis.

Root Mean Square Fluctuation (RMSF):

The local deformability of the protein has been analyzed through root mean square fluctuations of alpha Carbon atoms from MD simulations. The RMSF plotted against the residues of wild type and its mutants of PML from ProSA and RMSF study has been presented in Figure 3.33 and 3.34 respectively. Appreciable difference of RMSF between wild type and mutants of PML is observed at the N terminus specifically the first seventeen residues. Also, appreciable difference at the mutation site is also observed at their respective regions. The P187A and R91A mutant show global reduction in the highly flexible regions while the average values of RMSF indicates the E83R mutant have considerable reduce in the RMSF values globally and the R163P mutant have considerable increase in the RMSF values globally.

Solvent Accessible Surface Area (SASA):

The trend of SASA analysis from Figure 3.35 and 3.36 reflects that the mutants and wild type of PML are similar and consistent with Rg analysis. The average values of SASA indicate that the mutants has considerably less hydrophobic area than wild type SXL2 (106.16 ± 1.92 nm^2) notably P219A and Q242P mutants. It is evident from the SASA plots that the mutants had comparatively steady SASA profile than the wild type.

End to End chain distance between first and last alpha Carbon atoms:

The distance between first and last alpha Carbon atoms of wild type and mutants of PML from ProSA and RMSF study were analyzed throughout the simulation which would aid to understand the structural integrity of wild type and its mutants of PML and have been presented in Figure 3.37 and 3.38 respectively. Overall, a similar trend of end to end chain distance has been observed except the mutants Q242P and R163P. Drastic changes have been observed in end to end chain distance in mutants Q242P and R163P during the time course of simulation.

Analysis of Secondary Structure:

The secondary structure analysis shows the structural variation of each residue of wild type and mutants of PML from ProSA and RMSF study during the time course of simulation and has been presented in Figure 3.39 and 3.40 respectively. The structural stabilization of the corresponding secondary structure where the identified mutants R91A, P187A, P219A, R163P and Q242P were observed while noticeable change in structural stabilization was not observed in case of the mutants P7A, P10A and E83R. Reduced stabilization of the consecutive bend was observed in case of the mutant R91A and R163P while the mutant R163P further destabilizes a nearby turn (Table 3.9). Apart from structural stabilization of the α-helix where the mutants P187A and P219A resides, they also stabilize the nearby secondary structure, turn and bend respectively.

Table 3.8: Summarized parameters of MD simulations for wild type and *in silico* mutants of PML.

Method*	Variant	End to End Distance (nm)	Radius of gyration (nm)	RMSD (nm)	RMSF (nm)	SASA (nm^2)
	PML	2.50±0.23	1.63±0.01	0.18±0.03	0.11±0.07	106.16±1.92
PROSA	P7A	2.08±0.33	1.63±0.01	0.14±0.02	0.11±0.06	104.66±2.03
	P10A	2.14±0.32	1.62±0.01	0.15±0.02	0.11±0.06	104.57±2.53
	E83R	1.78±0.22	1.63±0.01	0.15±0.02	0.10±0.05	105.12±1.97
	P187A	2.40±0.19	1.63±0.01	0.13±0.02	0.11±0.05	104.38±1.61
	P219A	1.90±0.29	1.62±0.01	0.18±0.02	0.11±0.06	103.41±1.91
RMSF	Q242P	2.23±0.31	1.62±0.01	0.14±0.02	0.11±0.05	103.60±2.22
	R163P	3.53±1.20	1.63±0.01	0.16±0.03	0.12±0.06	105.50±1.71
	R91A	4.06±1.01	1.62±0.01	0.17±0.02	0.11±0.07	104.13±2.35

* methodology implemented for the identification of the mutants

Table 3.9: Stabilization of the secondary structure by the mutants of PML.

STUDY	PML	Secondary Structure*
PROSA	P187A	α-helix
	P219A	α-helix
RMSF	Q242P	loop
	R163P	loop
	R91A	α-helix

* the secondary structure where the mutants resides

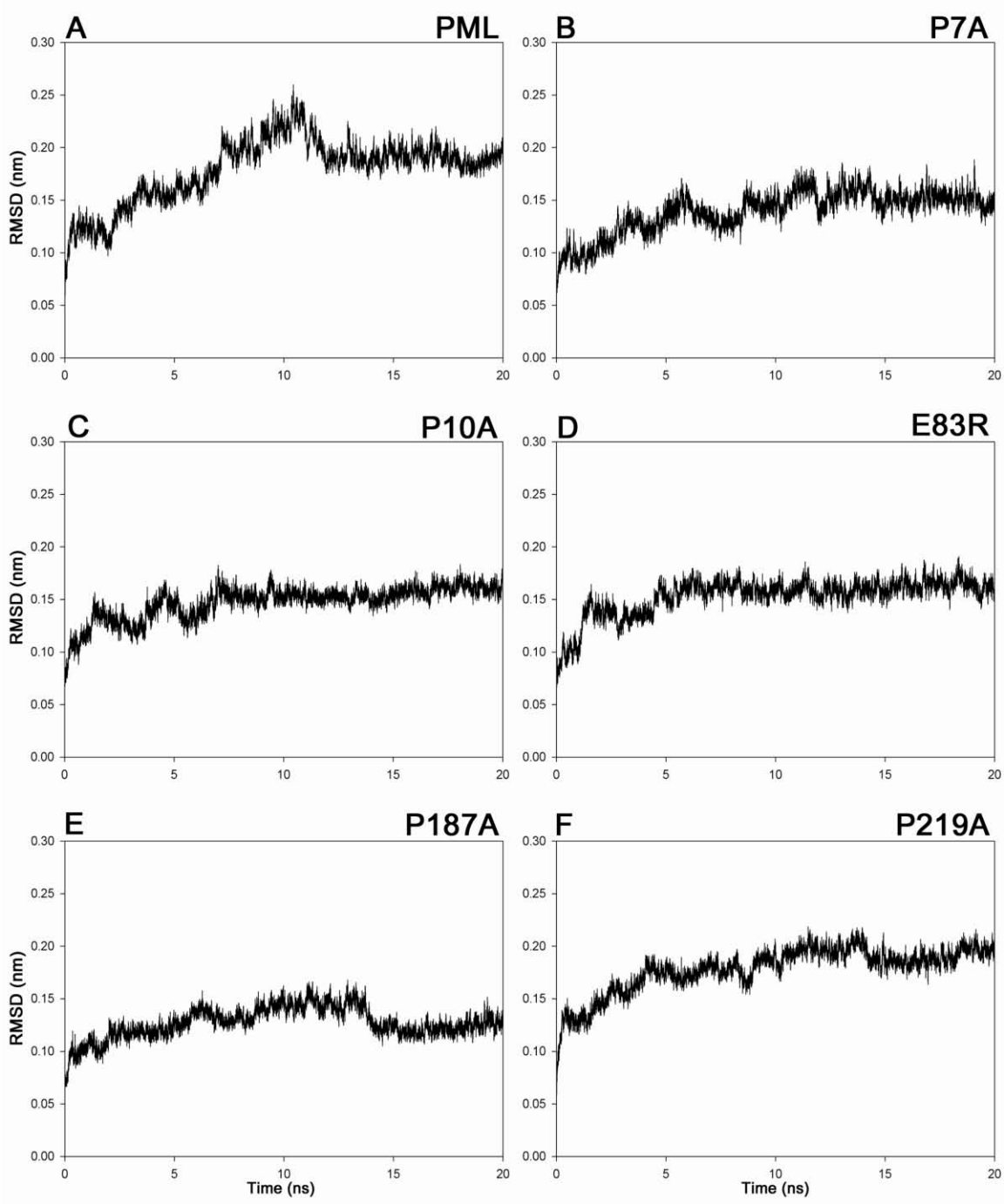


Figure 3.29: The root-mean-square deviation (RMSD) to the starting structure as a function of time of wild type and mutants of PML from ProSA study during the time course of simulation is shown. (A) PML (Wild type). (B) Mutant P7A. (C) Mutant P10A. (D) Mutant E83R. (E) Mutant P187A. (F) Mutant P219A.

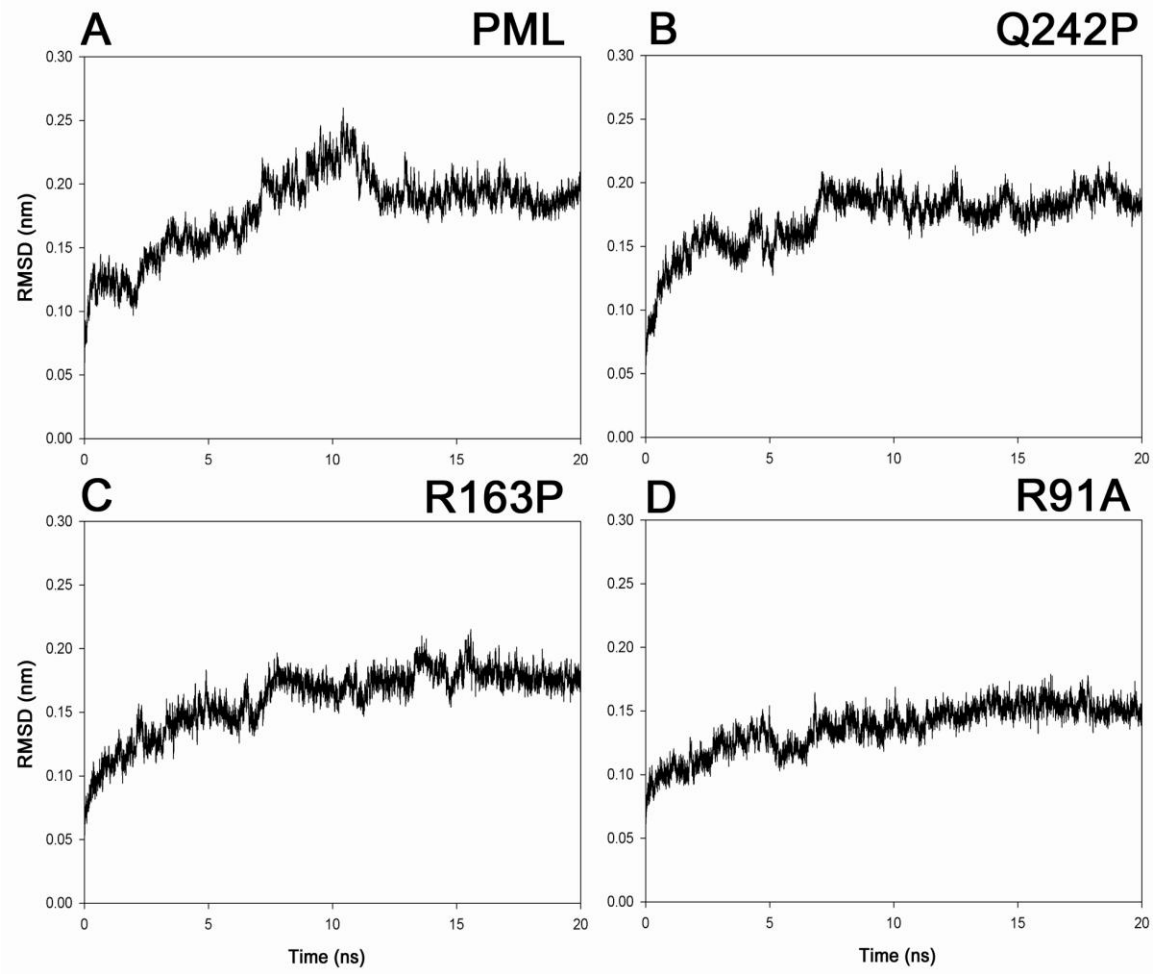


Figure 3.30: The root-mean-square deviation (RMSD) to the starting structure as a function of time of wild type and mutants of PML from RMSF study during the time course of simulation is shown. (A) PML (Wild type). (B) Mutant Q242P. (C) Mutant R163P. (D) Mutant R91A.

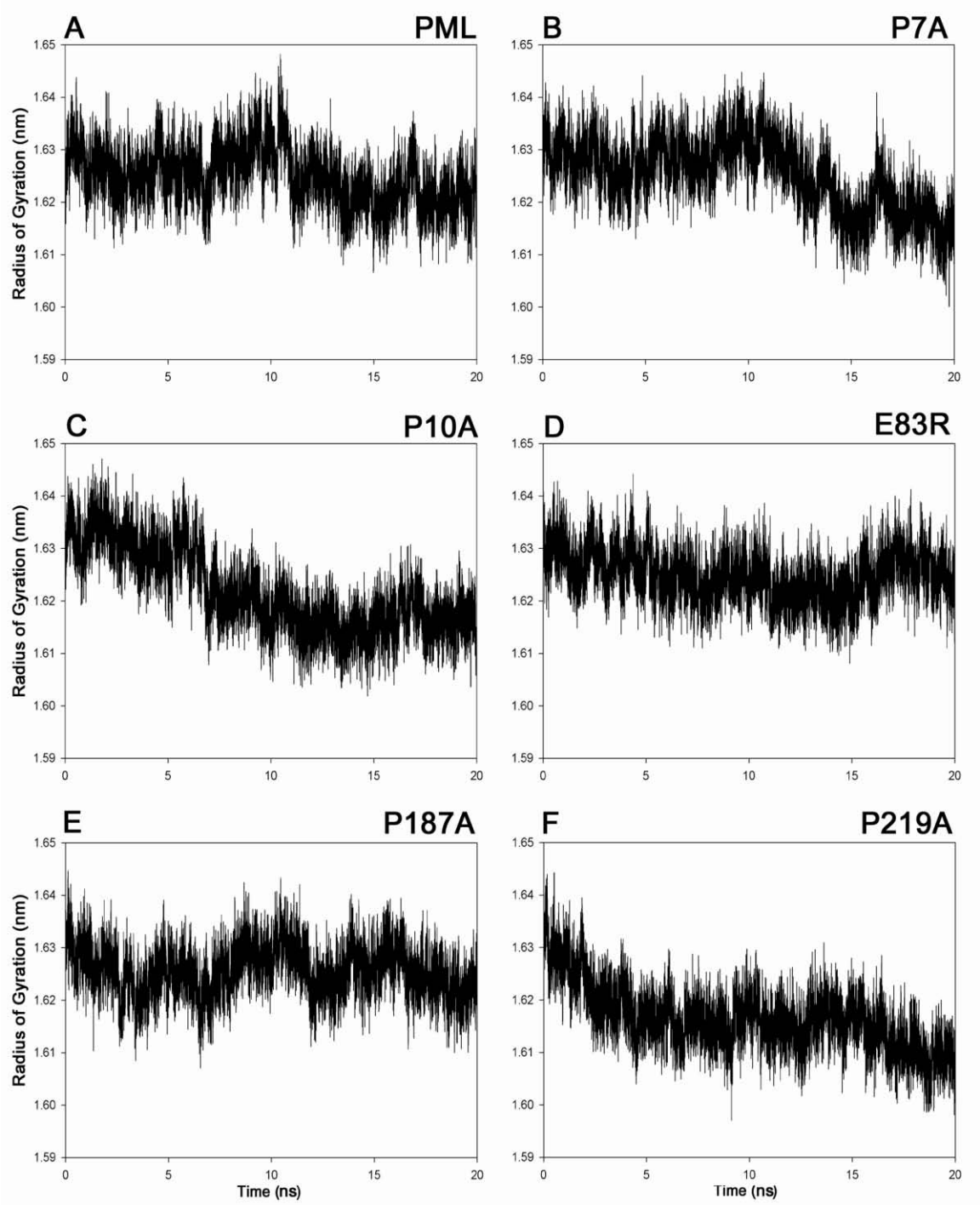


Figure 3.31: Radius of gyration of α-carbon atoms as a function of time of wild type and mutants of PML from ProSA study during the time course of simulation is shown. (A) PML (Wild type). (B) Mutant P7A. (C) Mutant P10A. (D) Mutant E83R. (E) Mutant P187A. (F) Mutant P219A.

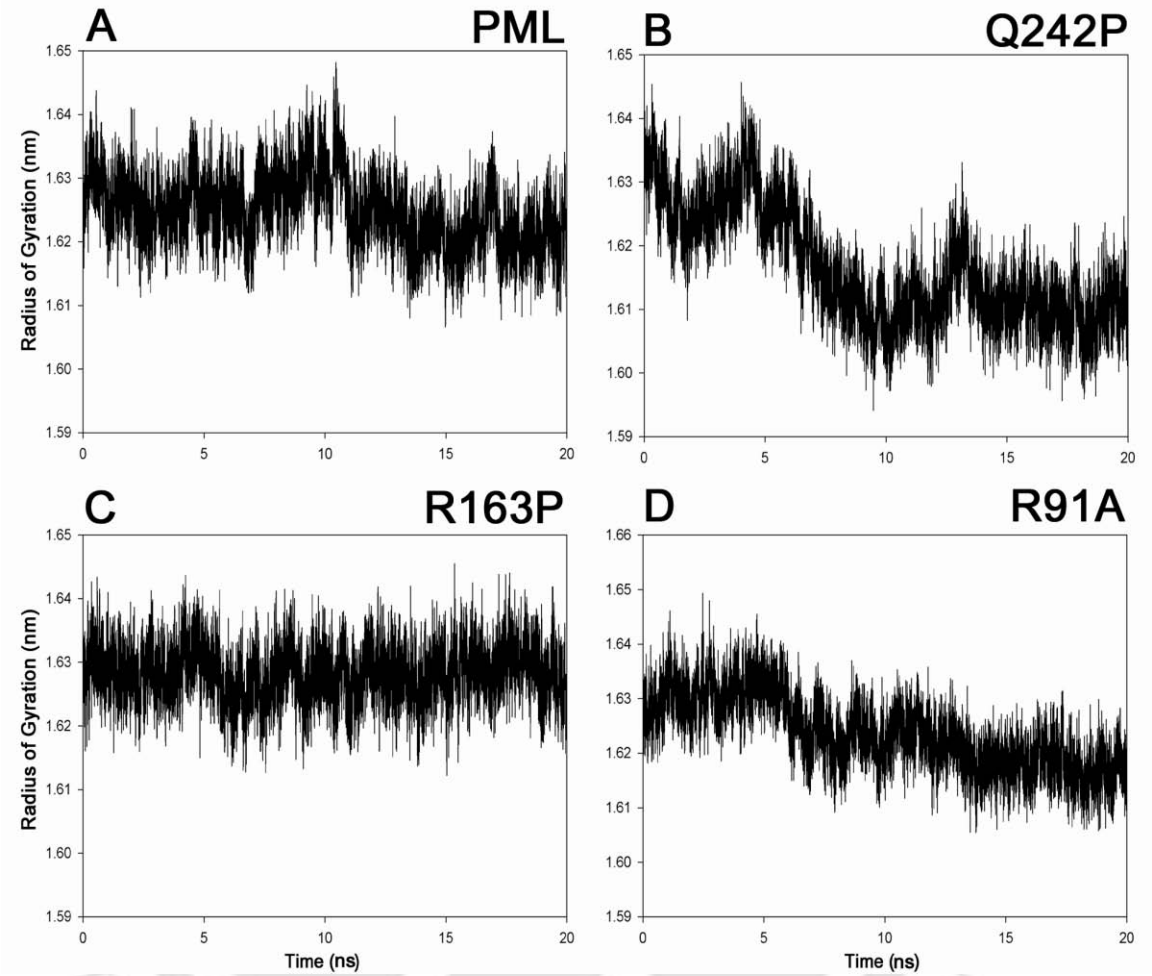


Figure 3.32: Radius of gyration of α-carbon atoms as a function of time of wild type and mutants of PML from RMSF study during the time course of simulation is shown. (A) PML (Wild type). (B) Mutant Q242P. (C) Mutant R163P. (D) Mutant R91A.

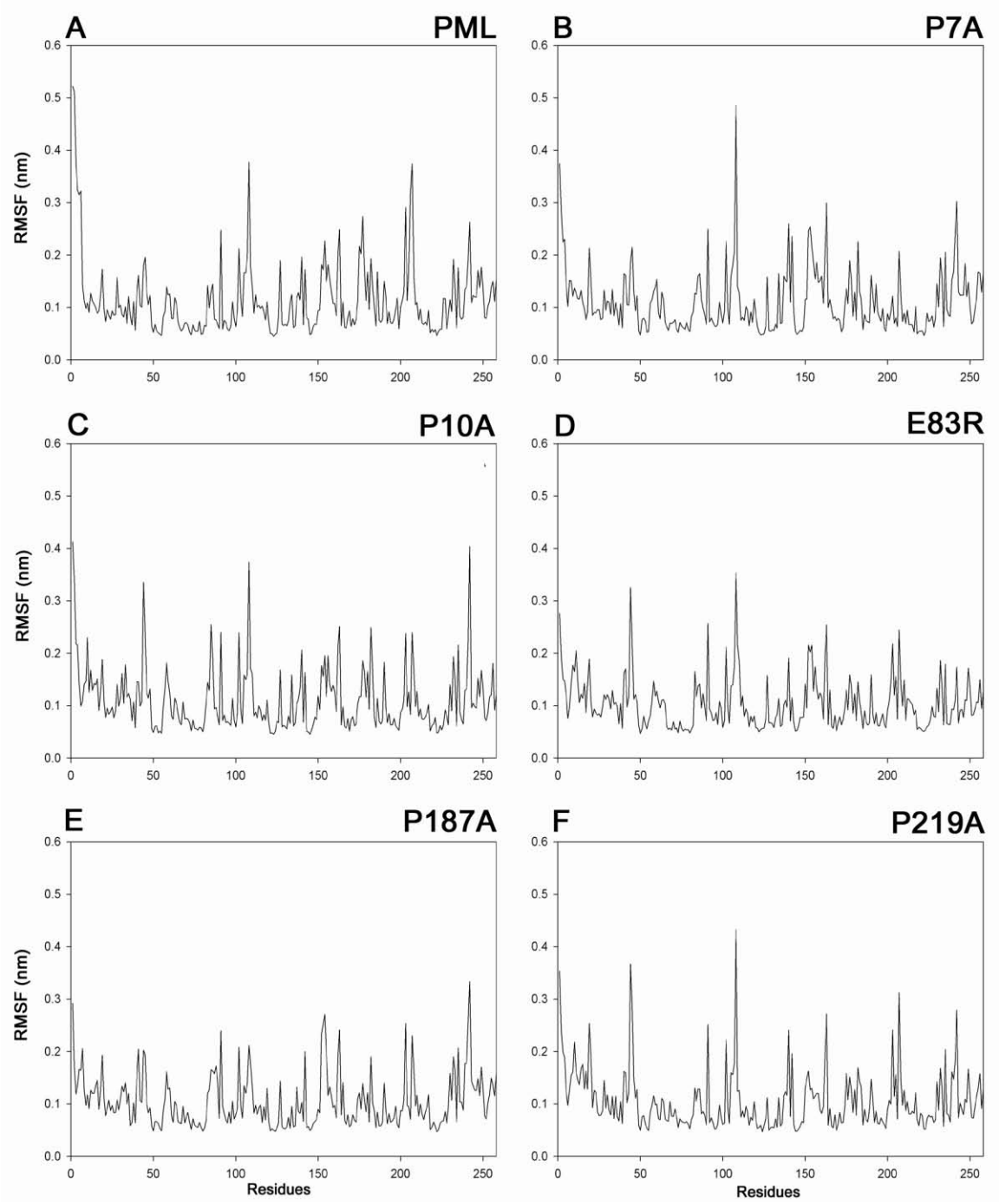


Figure 3.33: Root mean square fluctuation of α-carbon atoms as a function of residue of wild type and mutants of PML from ProSA study during the time course of simulation is shown. (A) PML (Wild type). (B) Mutant P7A. (C) Mutant P10A. (D) Mutant E83R. (E) Mutant P187A. (F) Mutant P219A.

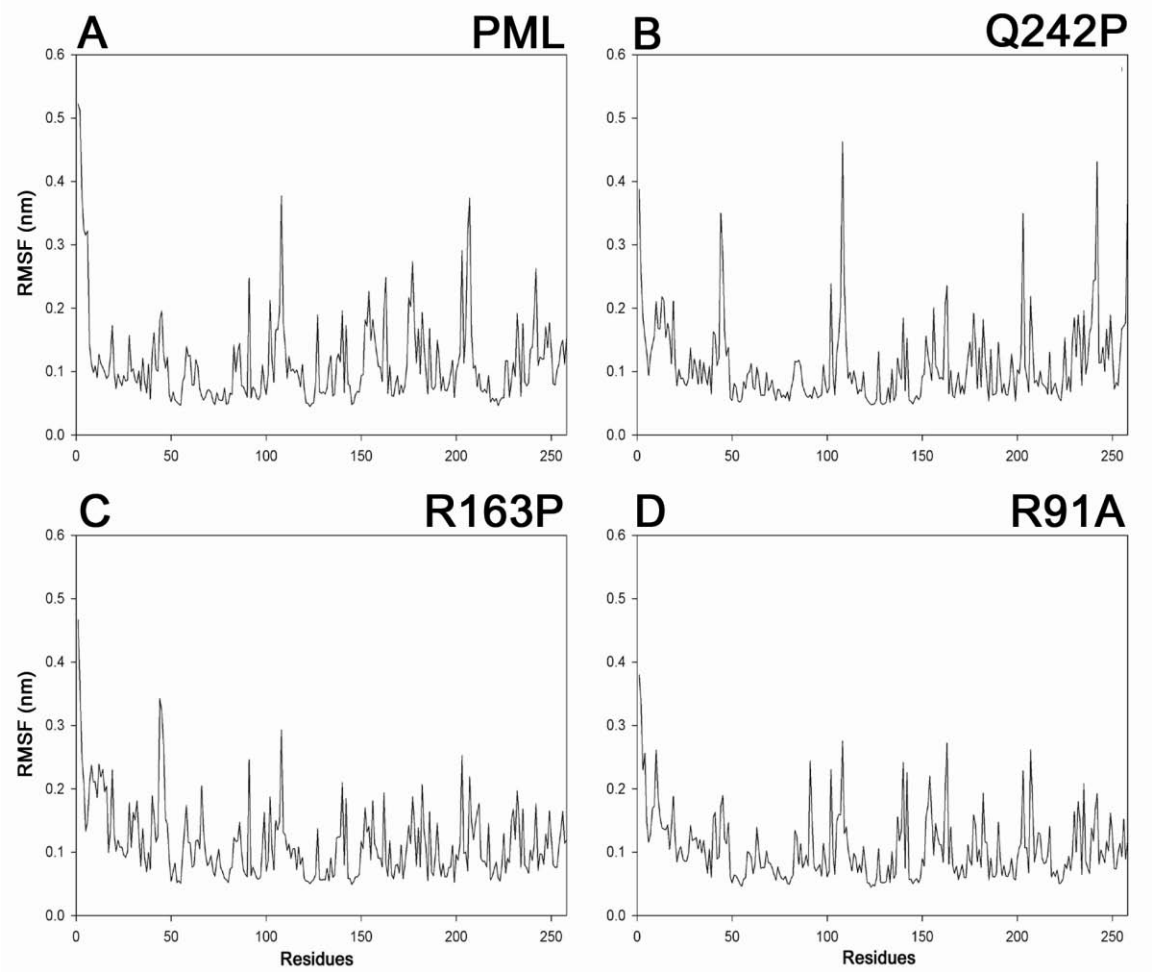


Figure 3.34: Root mean square fluctuation of α-carbon atoms as a function of residue of wild type and mutants of PML from RMSF study during the time course of simulation is shown. (A) PML (Wild type). (B) Mutant Q242P. (C) Mutant R163P. (D) Mutant R91A.

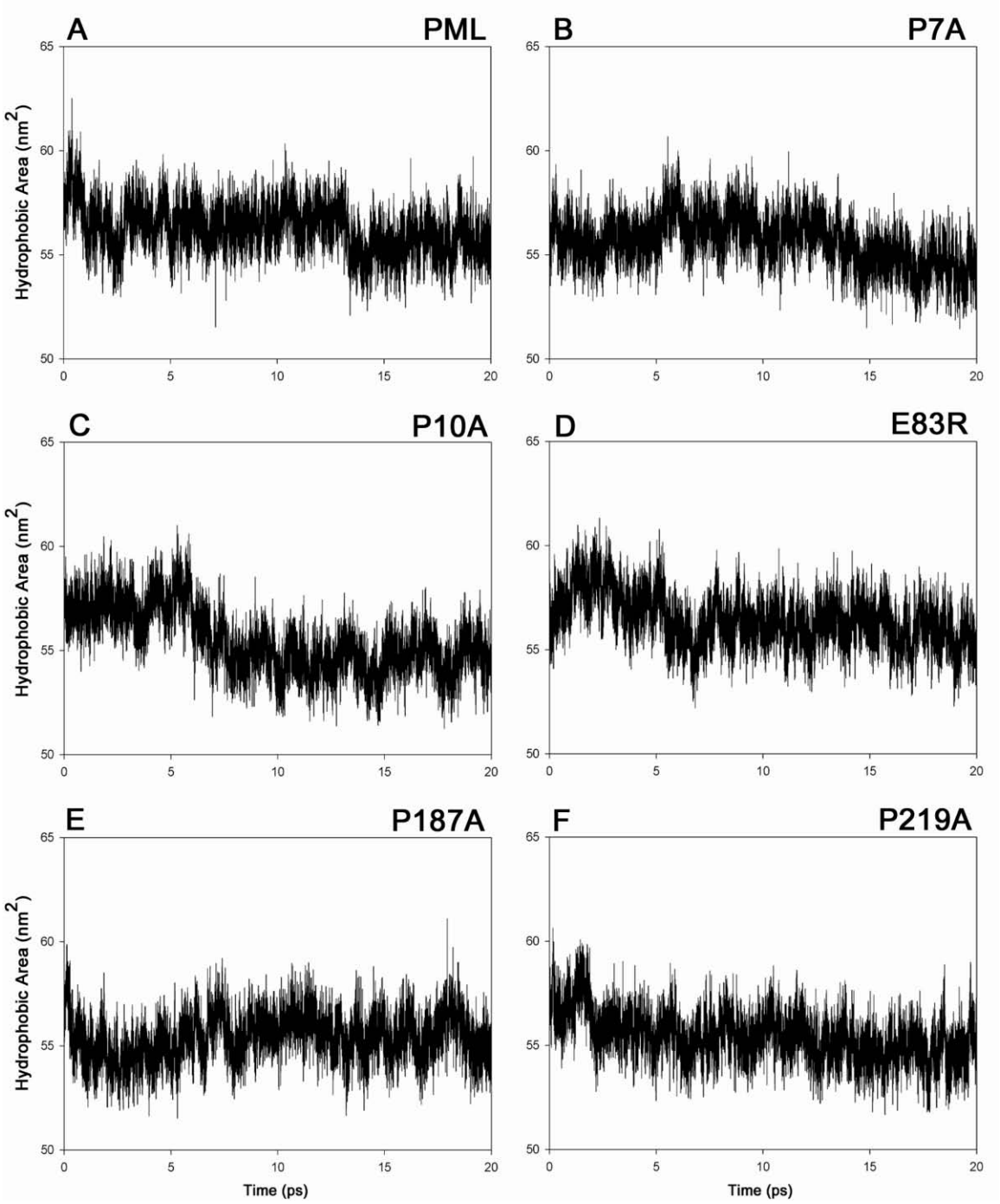


Figure 3.35: The solvent accessible surface area (SASA) of wild type and mutants of PML from ProSA study during the time course of simulation is shown. (A) PML (Wild type). (B) Mutant P7A. (C) Mutant P10A. (D) Mutant E83R. (E) Mutant P187A. (F) Mutant P219A.

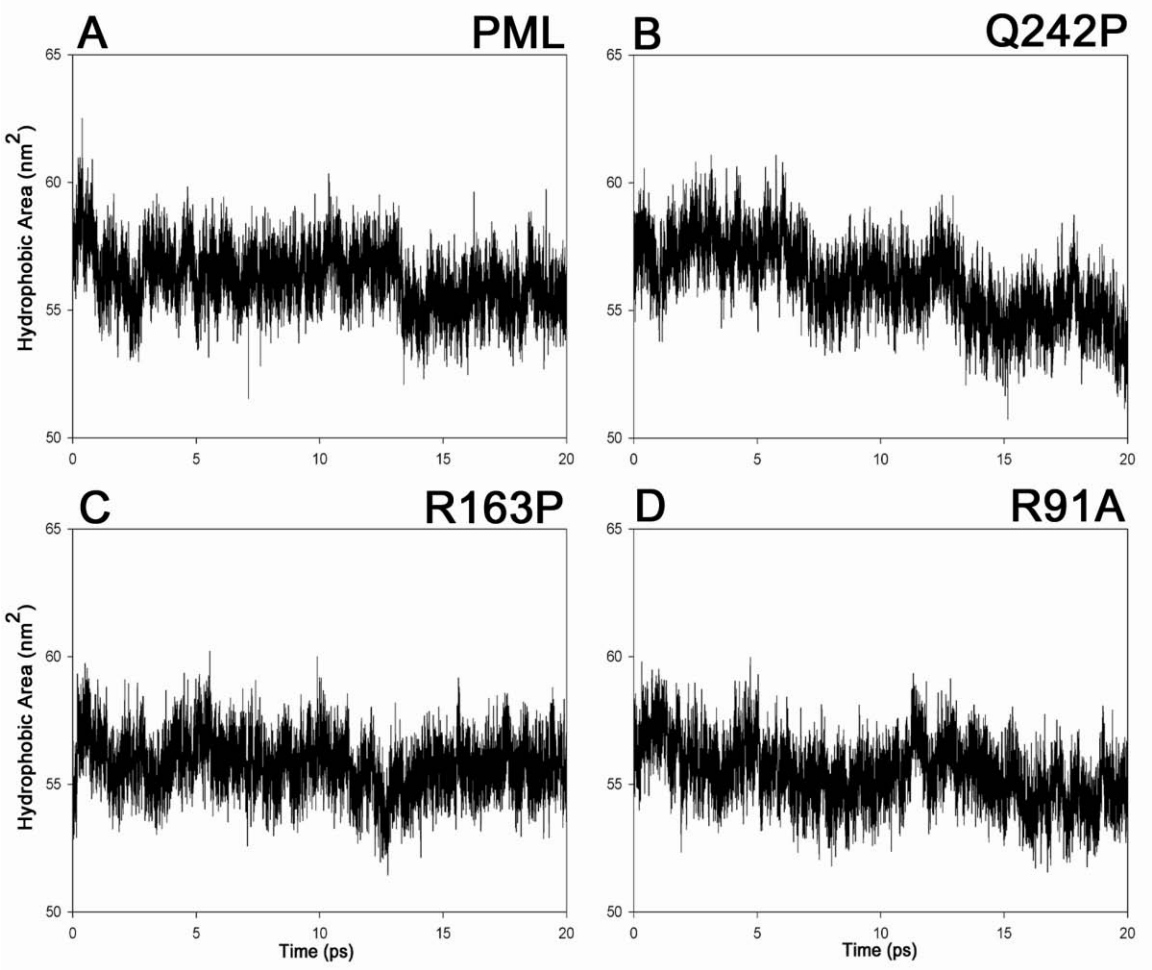


Figure 3.36: The solvent accessible surface area (SASA) of wild type and mutants of PML from RMSF study during the time course of simulation is shown. (A) PML (Wild type). (B) Mutant Q242P. (C) Mutant R163P. (D) Mutant R91A.

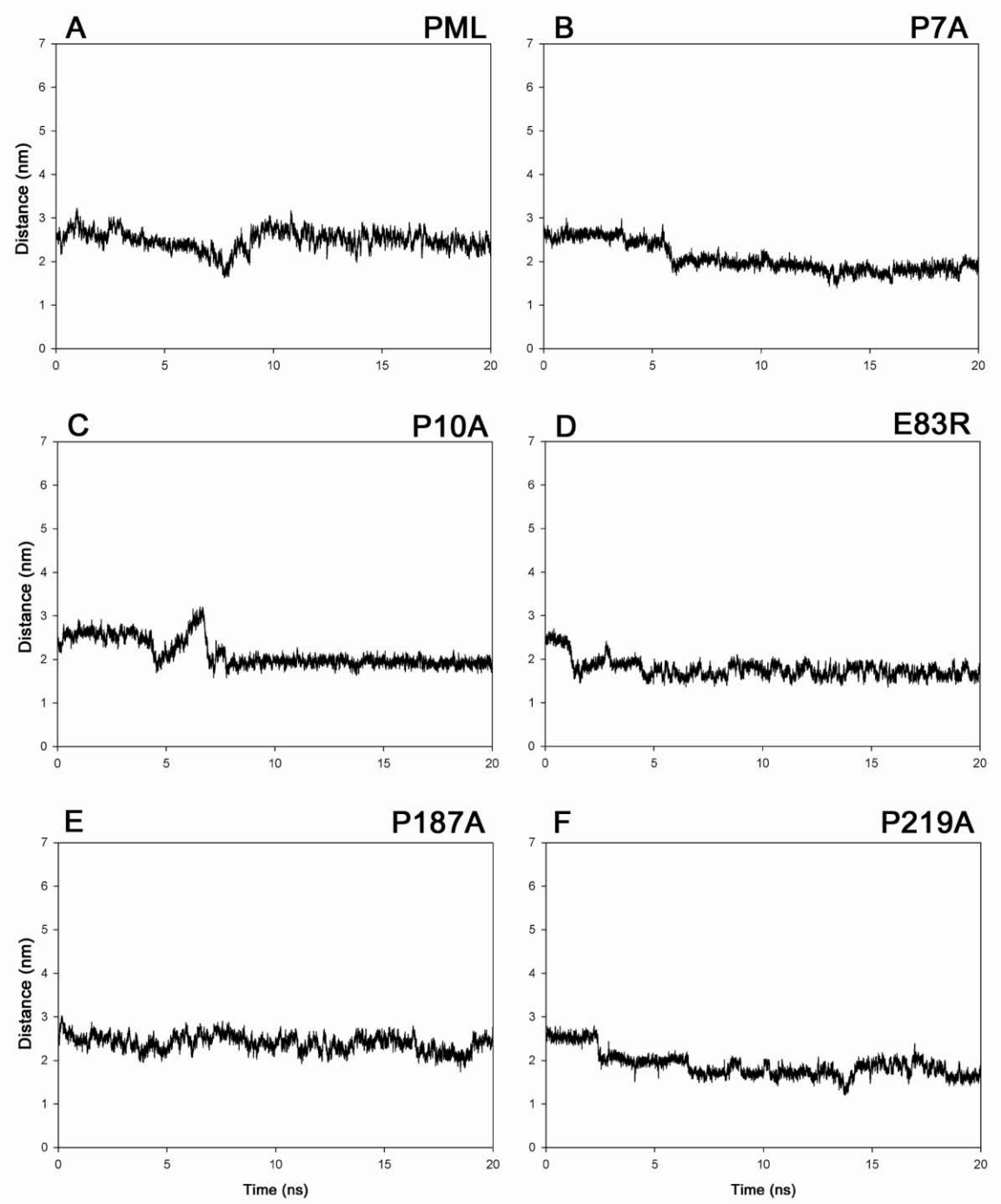


Figure 3.37: The distance between first and last alpha Carbon atoms of wild type and mutants of PML from ProSA study during the time course of simulation is shown. (A) PML (Wild type). (B) Mutant P7A. (C) Mutant P10A. (D) Mutant E83R. (E) Mutant P187A. (F) Mutant P219A.

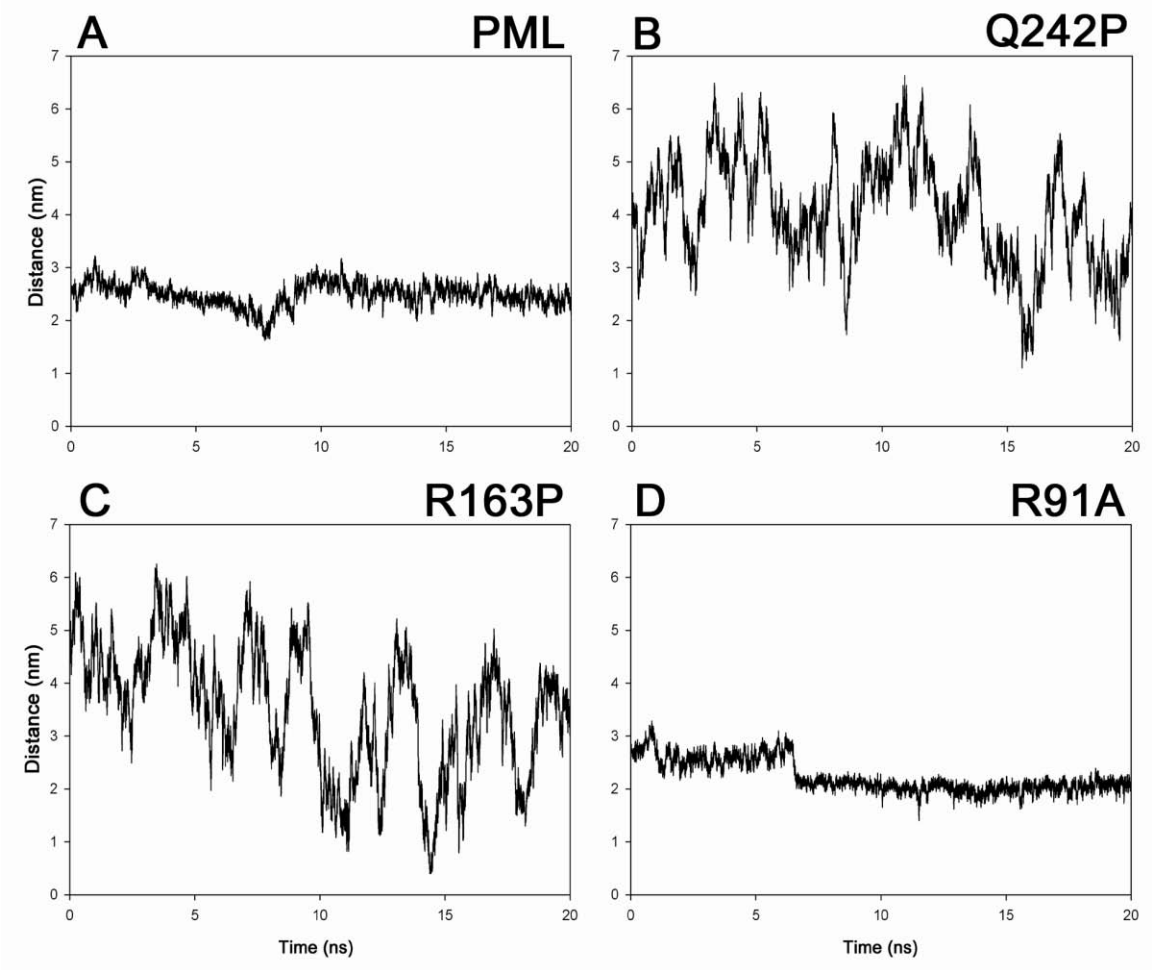


Figure 3.38: The distance between first and last alpha Carbon atoms of wild type and mutants of PML from RMSF study during the time course of simulation is shown. (A) PML (Wild type). (B) Mutant Q242P. (C) Mutant R163P. (D) MutantR91A.

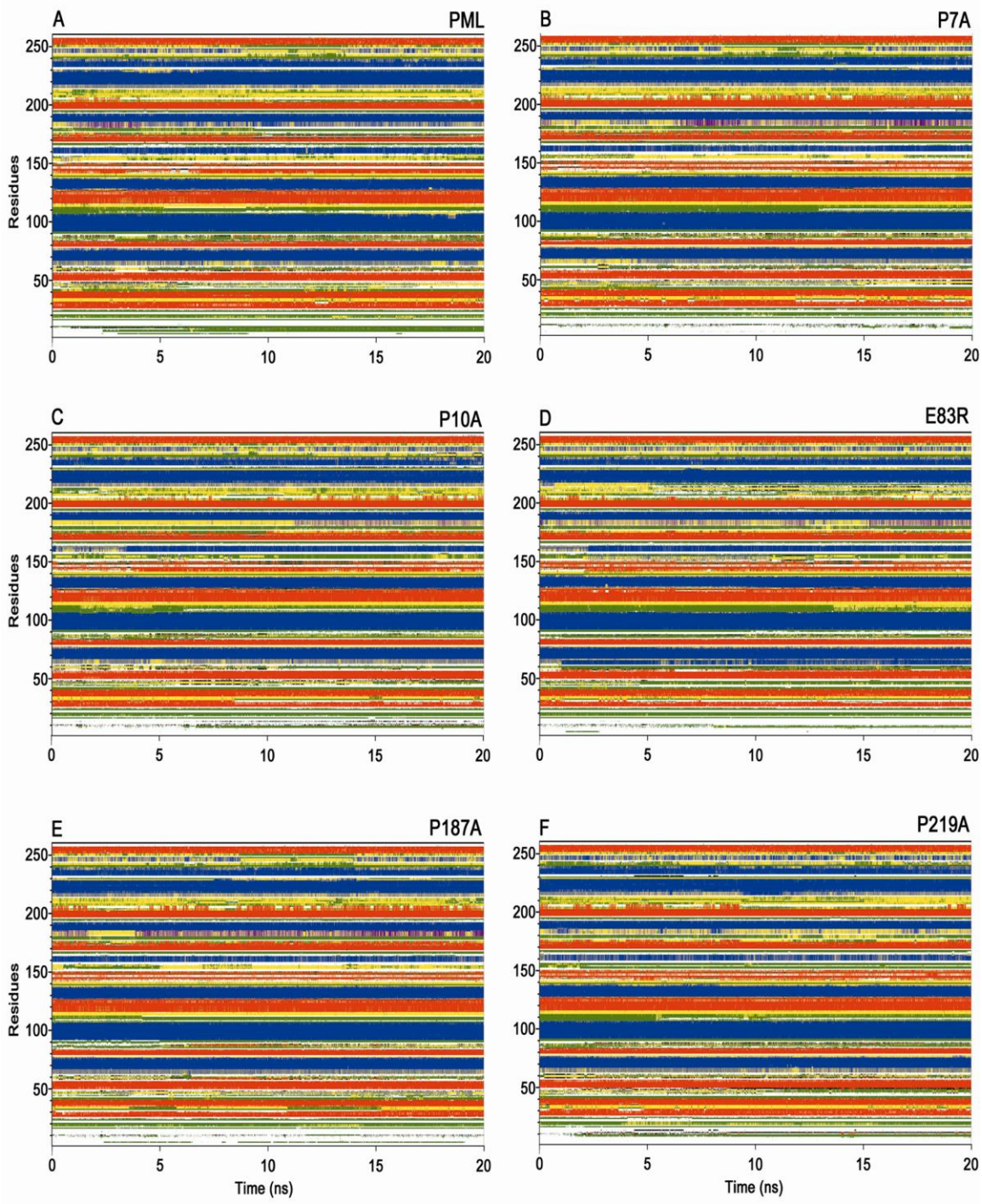


Figure 3.39: The secondary structural analysis of wild type and mutants of PML from ProSA study during the time course of simulation is shown. (A) PML (Wild type). (B) Mutant P7A. (C) Mutant P10A. (D) Mutant E83R. (E) Mutant P187A. (F) Mutant P219A.

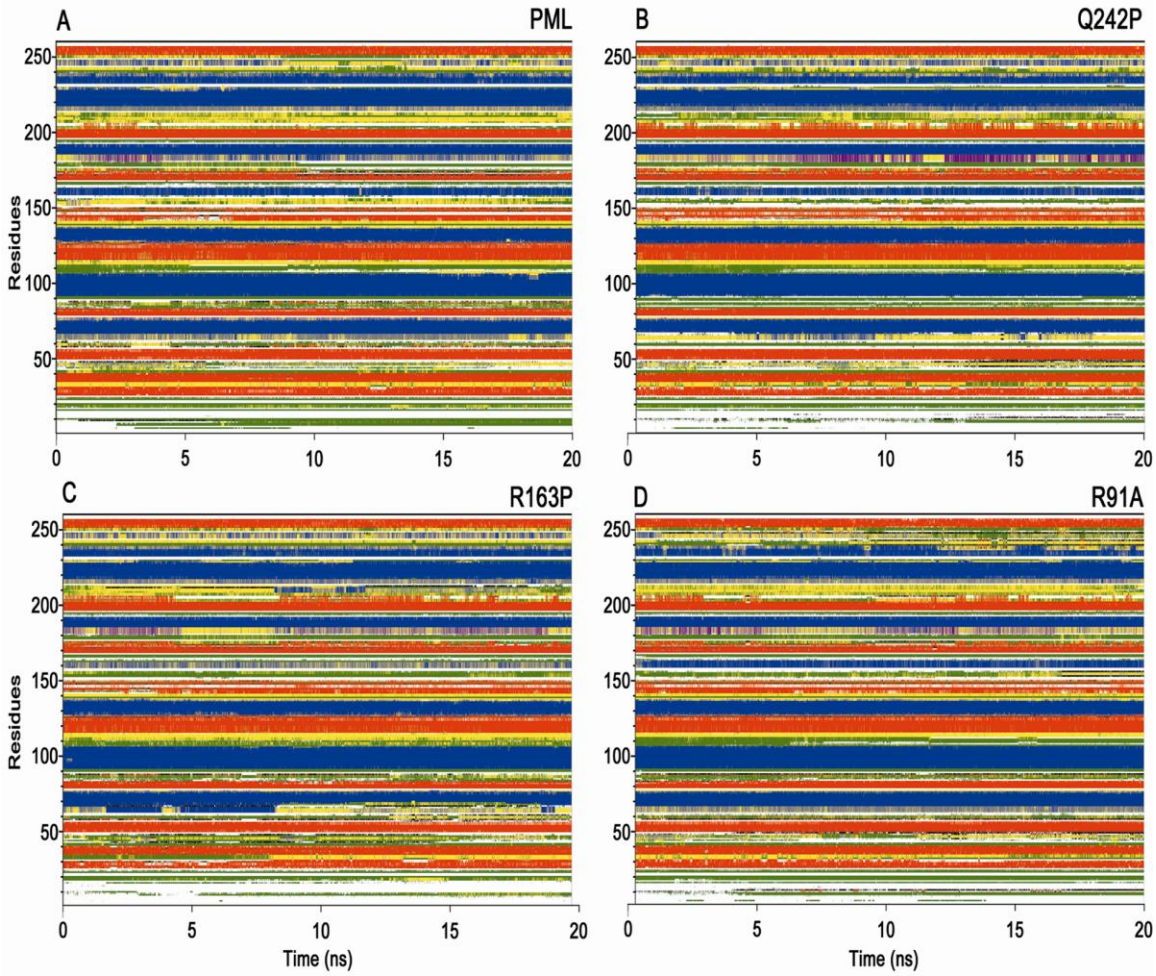


Figure 3.40: The secondary structural analysis of wild type and mutants of PML from RMSF study during the time course of simulation is shown. (A) PML (Wild type). (B) Mutant Q242P. (C) Mutant R163P. (D) Mutant R91A.

The MD simulations of wild type and its respective mutants of the candidate lipases reflects that (i) the mutants N204P of SAL3; E220G of SXL2, and E83R, P219A of PML have better RMSD values (ii) all the identified mutants have similar trend of Rg as their respective wild type while the mutants R91A, R163P of PML; M161F, E220G of SXL2 and F289P, N204P of SAL3 have steady equilibrium than their respective wild type proteins, (iii) RMSF analysis indicates that the mutants of candidate lipases have comparative values with their respective wild type proteins while the, mutants N204P of SAL3; E220G of SXL2, and E83R, P219A of PML have better RMSF values (iv) the mutants N204P of SAL3; M161F of SXL2, and P219A, Q242P of PML have lesser

SASA values than its wild type which reveals decrease in solvent accessible hydrophobic area thereby the possibility of increase in compactness of hydrophobic core, (v) the mutants N204P of SAL3 and M161F of SXL2 have similar values with their respective wild type proteins while E83R and P219A of PML have better 'end-to-end chain distance' values than its wild type and (vi) the secondary structure variation analysis reveals that the mutants increased the stability of secondary structure where they resides notably the mutants F289P (loop) of SAL3; M161F (bend), E220G (loop) of SXL2 and R91A (α-helix), P187A (α-helix), P219A (α-helix), R163P (loop) of SAL3. The discussed mutants of candidate lipases display increase in structural stability, reduced flexibility and lesser solvent exposed hydrophobic area which validates the *in silico* mutants. We attempted rational engineering of the candidate lipases utilizing structural bioinformatics approach which provide insights that could help to understand the structure of triacylglycerol lipases and also help to devise better strategies to design mutants to emulate their industrial applications.

3.5. Conclusions

From the Triacylglycerol lipase database, PML from subfamily 22 as well as SAL3 and SXL2 from subfamily 5 were selected to exploit for industrial applications based on their uniqueness and homology towards commercial lipases. *In silico* characterization and mutational studies on the candidate lipases identified structurally and functionally important residues. Mutagenesis hotspots and *in silico* mutants were identified which were validated through molecular dynamics. The parameters taken into account were z-scores which reflected the quality of protein structures on carrying out the predicted mutations of all possible sites using knowledge-based potentials and root-mean square fluctuations (RMSF) that can be correlated to temperature factors which reflects the local structural flexibility. The final mutants were selected based on the presence of thermolabile residues. Notably, the mutants PML-P187A, PML-P219A, SAL3-F289P, SXL2-M161F and mutants from RMSF study were found with improved structural stability and suitable for their use in industrial applications. These predicted *in silico* mutants can be taken further towards *in vitro* validations.

Chapter IV

TARGETING LIPOLYTIC ENZYMES TO COMBAT TUBERCULOSIS

Chapter IV

Targeting Lipolytic Enzymes to Combat Tuberculosis

4.1. Abstract

Many lipases in microorganisms have significant role in their existence and pathogenicity. These lipases are significant in the metabolism of the microorganism playing crucial role in lipid catabolism. This fact has also been brought out in TLDB. One such pathogen is *Mycobacterium tuberculosis* which causes tuberculosis, one of the leading infectious diseases. Extensive use of the lipid metabolism for survival highlights the importance of lipid catabolizing lipases. Recent studies validated the lipases Rv0183 and Rv3802c as attractive drug targets. The chapter reports potential inhibitors for drug targets followed by dual potential inhibitors using integrated computational approach. Potential inhibitors were identified through virtual screening. Additionally, dual inhibitors targeting mycolic acid synthesis pathway and host lipid catabolism simultaneously were also discovered. Experimental validation by the TB research community should validate the therapeutic utility of the identified potential inhibitors against Rv0183 and Rv3802c bringing out the significance of pathogenic lipases.

4.2. Introduction

It has been 90 years since BCG was first administered in humans as the vaccine against tuberculosis still it continues to be one of the deadliest diseases (Ritz and Curtis, 2009). Despite long and widespread usage of the vaccine and the drugs for tuberculosis across the globe, two billion people equating one-third of world's population is still infected (Corbett *et al.,* 2003). Tuberculosis caused 1.45 million deaths and 8.8 million fresh cases were reported in 2010 (World Health Organization, 2011). Tuberculosis is caused by *Mycobacterium tuberculosis*, a highly complex and remarkably successful pathogen that resides in macrophages of the host. The pathogen usually establishes infection in the mammalian respiratory system (Pulmonary tuberculosis), but extra-pulmonary tuberculosis also occurs (McKinney *et al.,* 2000). Tuberculosis is also associated with multiple drug resistance and HIV co-infection. The effective control of the disease is a matter of concern (Corbett *et al.,* 2003; Nachega and Chaisson, 2003). In this regard, understanding the biology of persistence will help in designing new anti-tubercular drugs (Zhang, 2006) which will kill both active and persistent *M. tuberculosis.* Current situation urges the need of new anti-tubercular drugs and TB drug targets which can reduce the treatment time and overcome the drug resistance problem (Wilson, 2009).

The parasite switches between different physiological states in the latent, active, dormant and reactivated infection stages within the host. The adaptation of pathogen to the host environmental stress by utilizing its lipid metabolism is critical for survival of the microorganism (Cotes *et al.,* 2008; Ducati *et al.,* 2006). Enzymes involved in the lipid metabolism of pathogen have attracted considerable interest since they supply fatty acids for the formation of intracellular lipid inclusion bodies (LIB) and cell membrane which are the key components for its survival and pathogenicity. It has been suggested that these lipids may provide a carbon source during the chronic infection phase, dormancy and also for the reactivation (Garton *et al.,* 2002; Daniel *et al.,* 2004). Also it is known that these lipids act as building blocks for virulent lipids (Russell *et al.,* 2010). Lipases, lipase-like enzymes and other lipid catabolizing enzymes likely play vital roles in the host lipid degradation process, virulence and host immune-responses in *M. tuberculosis* as well as in several other pathogens (Singh *et al.,* 2010; Daleke *et al.,* 2011; Shakarian *et al.,* 2010). Several lipase gene families

in *M. tuberculosis* highlight their physiological importance (Table 4.1). Lipases Rv0183 and Rv3802c have been chosen as the drug targets for the current study based on their physiological importance in the pathogen. Biological relevance for drug discovery and targeting the chosen drug targets have been investigated and listed in Table 4.2.

Table 4.1: List of lipase gene families found in *M. tuberculosis* with its representative genes.

Gene Family	Active site consensus sequence	Representing Genes	α/β hydrolase fold	References
Lip	GXSXG	LipC to LipZ	Yes	Deb *et al.*, 2006
Cutinase	G-[YF]-S-[QL]-G	Culp1 to Culp7	Yes	Longhi *et al.*, 1999
β-Lactamase	S-X(T)-X(S)-K	LipE, LipL, LipP	Yes	Hugonnet *et al.*, 2009
PE, PPE	PE/PPE	LipY, LipX	No	Bottai and Brosch, 2009
Phospholipases	GXSXG	plcA, plcB, plcC	Yes	Srinivas *et al.*, 2008

Table 4.2: The drug targets chosen for the present study and its biological relevance for drug discovery to combat tuberculosis.

Rv3802c	Rv0183
• Phospholipase A and Thioesterase • Essential cell wall lipase • Homolog in *M. smegmatis* proven to have role in integrity and internal structure of cell wall	• Extracellular Monoglyceride Lipase • Essential for host lipid catabolism • Homolog in *M. smegmatis* proven to have role in cell physiology, colony morphology and cell-cell interaction
• Conserved throughout mycobacterial species even in *M. leprae* • Significantly divergent from mammalian lipases	

4.2.1. Drug Target I: Rv0183, a monoglyceride lipase

M. tuberculosis Rv0183 is a lipid degrading enzyme with preference towards monoacylglycerol followed by other acylglycerols (Côtes *et al.*, 2007). It may contribute to the growth and pathogenicity of *M. tuberculosis* by participating in host lipid hydrolysis and cell wall remodeling process. The Rv0183 ortholog in *M. smegmatis* has a proven role in cell

physiology, colony morphology and cell interaction as well as showed similar biochemical behavior and cell localization patterns of Rv0183 (Dhouib *et al.*, 2010). Rv0183 is also found in *M. leprae*, which contains minimal genes needed for survival within host cell. By minimal genome concept and physiological role, Rv0183 is considered to be essential for survival and pathogenicity.

4.2.2. Drug Target II: Rv3802c, an essential cell wall lipase

M. tuberculosis Rv3802c is found as an integral part of the operon which is dedicated to the synthesis of mycobacterial cell wall and could participate in cell wall metabolism, suggesting its distinct role in mycobacteriology (Marmiesse *et al.*, 2004; Song *et al.*, 2008; Parker *et al.*, 2009). The highly complex and unique cell wall is critical for survival and infection of mycobacterium within the host cells (Barry, 2001). It is reinforced by the fact that 15.5% of the total coding region of its genome is associated with cell wall and related proteins (Cole *et al.*, 1998; Cole, 1999). The cell wall core - mycolyl arabinogalactan peptidoglycan (mAGP) complex found beyond the plasma membrane is insoluble when cell wall is disrupted (Brennan, 2003). Mycolic acids in mAGP complex are long-chain α-alkyl β-hydroxy fatty acids, the signature lipids of the hydrophobic mycobacterial cell wall (Todar, 2004). Mycolic acids are essential for the survival as well as contribute to the pathogenesis of *M. tuberculosis* through a variety of mechanisms (Guenin-Macé *et al.*, 2009; Pieters, 2001). They are also the determining factor for permitting molecules through the cell wall (Liu *et al.*, 1996). Mycolic acids are a key characteristic of the entire Corynebacteria-Mycobacteria-Nocardia group and alterations in these molecules have huge effects on stability and permeability of the cell envelope (Boshoff and Barry, 2006). The insoluble core is essential for the viability of the cell and thus targeted in the context of discovery of new drugs (Brennan and Crick, 2007). As the pathogen depends on its lipid metabolism for many of its needs, secreted lipases at the key stages of *M. tuberculosis* infection like Rv3802c need to be studied in detail to understand the mechanism of pathogenesis (Armstrong and Hart, 1971; Beatty and Russell, 2000; Russell, 2001; Canaan *et al.*, 2004). Conditional disruptions of Rv3802c homolog MSMEG_6394 in *M. smegmatis* lead to the loss of cell wall integrity and internal structure of the cell (Crellin *et al.*, 2010). The tetrahydrolipstatin (THL) inhibits both MSMEG_6394 and

Rv3802c (Crellin *et al.*, 2010; West *et al.*, 2011). All these evidences point to the essentiality of Rv3802c that motivates us to search for effective potential inhibitors.

Considering the recent increase in incidences of multi drug resistance (MDR) and extreme drug resistance (XDR) tuberculosis (Shah *et al.*, 2007; Berry and Kon, 2009), finding a new drug target and drugs to cripple pathogen's survival and infection is of high importance which could be achieved by targeting both cell wall and lipid metabolism with the help of Rv0183 and Rv3802c.

4.3. Materials and Methods

Structure-based drug discovery (SBDD) is the most preferred technique to identify novel potential inhibitors against the protein of interest and the opted approach for present study has been presented (Fig. 4.1). The approach explores representative molecular scaffolds followed by similarity based virtual screening to identify potentially new molecules with anti-tubercular activity which is also cost effective in terms of computational resources and time.

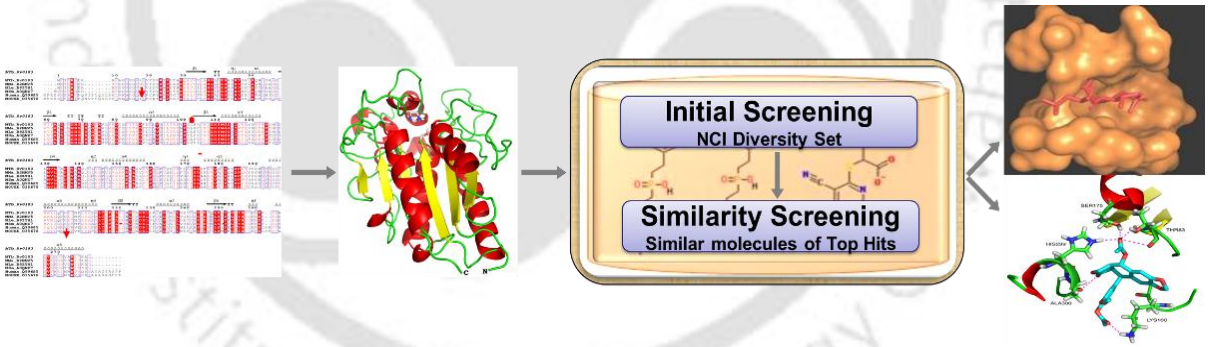


Figure 4.1: Structure-based drug discovery approach implemented in the present study.

4.3.1. Homology Modeling

Respective sequence of the drug targets along with mycobacterial and mammalian orthologs were retrieved from UniprotKB database. Accession numbers, source organisms and identity towards Rv0183 and Rv3802c of the collected orthologs have been listed in Table 4.3 and Table 4.7 respectively. It is noteworthy to mention that Rv3802c have no sequence and structural homologs in the human proteome. Multiple sequence analysis of the respective drug target and its orthologs were carried out with ClustalW to identify the

functionally/structurally important residues, motifs and domains (Thompson *et al.*, 1994). Template for homology modeling was identified using 'BLASTP' program against Protein Data Bank (PDB) database to identify the closest structural homologs. Sequence alignment between the drug target and its respective template was used as input for model building process in MODELLER9v7 independently (Marti-Renom *et al.*, 2000). The quality of hundred generated three dimensional (3D) models was ranked by probability density function (MOLPDF) and discrete optimized protein energy (DOPE) scores (Shen and Sali, 2006). In order to identify the best model of respective drug targets, the ten top ranked models with low MOLPDF and DOPE score were assessed using PROCHECK validation package (Laskowski *et al.*, 1993) and ERRAT plot (Colovos and Yeates, 1993) in NIH Structural Analysis and Validation server. The best model of the respective drug target was energy minimized using the steepest descent algorithm and subjected to molecular dynamics (MD) simulations with the GROMACS3.2 package using GROMOS96 43a1 force field and simple point charge (SPC) water model (Berendsen *et al.*, 1995; Lindahl *et al.*, 2001; van der Spoel *et al.*, 2005). Native protein fold evaluation on the identified best model was carried out with knowledge-based energy profile of ProSA program (Sippl, 1993). Using MD, the energy minimized model of the respective drug targets was analyzed for its structural stability with 3D periodic boundary conditions. The net charge of system was neutralized with sodium as counter-ions (Rv0183-seven; Rv3802c-eleven) by replacing water molecules that are at least 3.50 angstroms (Å) from the protein surface. The solvent was equilibrated for 100ps by restraining the solute atoms through a harmonic force constant of 1000 kJ mol^{-1} nm^{-2} in NVT ensemble followed by NPT ensemble. NPT production run was carried out for 10ns with no restraints using 2fs of integration time. All MD simulations were carried out with temperature of 300 K with velocity rescaling thermostat where protein and non-protein atoms were coupled to separate temperature coupling baths. The pressure was controlled at 1 atm using Parrinello-Rahman barostat. Particle mesh Ewald (PME) summation method was used for calculating the long-range electrostatic interactions with cut-off of 12 Å. The linear constraint solver (LINCS) algorithm was used to constrain the bonds involving hydrogen atoms (Hess *et al.*, 1997; Hess, 2008). 2000 conformations were generated and analyzed with root-mean-square deviation (RMSD) plot to check whether the modelled structure of respective drug targets is energetically stable in the solvent system.

4.3.2. Structure-Based Virtual Screening

Structure-based virtual screening is a viable and successful method for the identification of hits for drug discovery. Virtual screening is also considered as an essential part that enriches lead molecules in pharmaceutical industry for translational research (Lyne, 2002). Step-wise virtual screening methodology was opted in the present study to identify molecular scaffolds against the drug targets of interest followed by the identification of drug-target specific potential inhibitors (Fig. 4.2).

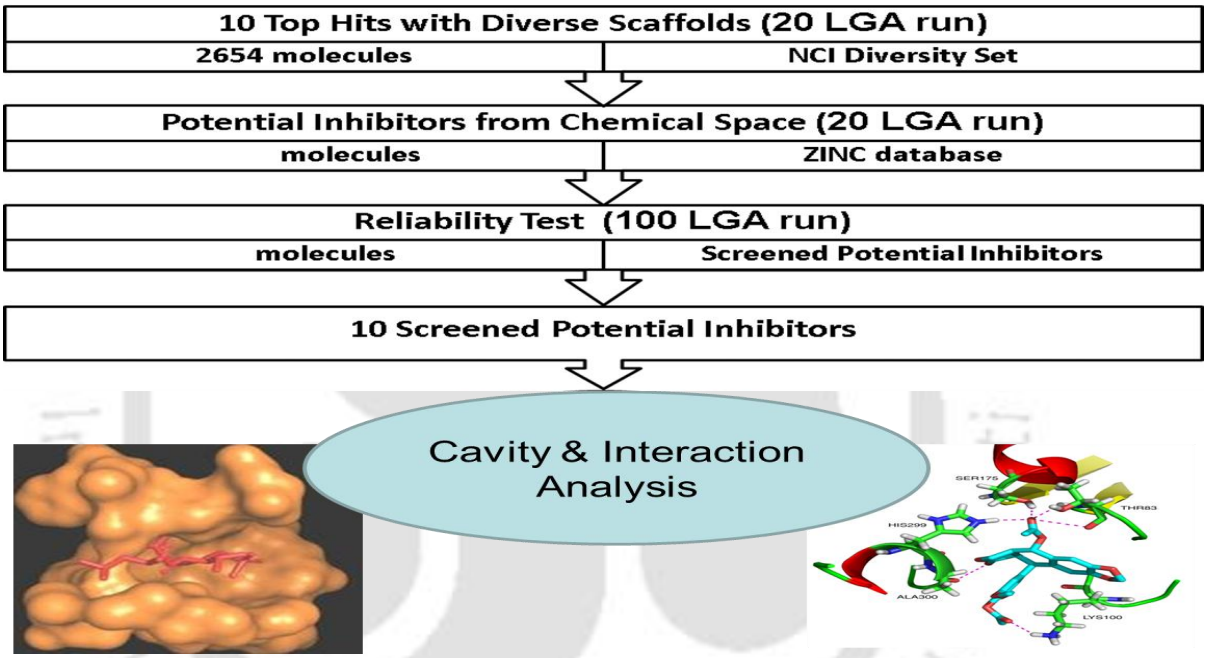


Figure 4.2: Schematic representation of virtual screening methodology opted in the present study.

Virtual screening was performed against the energy minimized model of the respective drug targets with AutoDock4.2 using lamarckian genetic algorithm (LGA) which is considered as the best for small molecule docking studies (Morris *et al.*, 1998; Cosconati *et al.*, 2010; Morris *et al.*, 2009; Park *et al.*, 2009). The small molecule dataset used in the study was NCI diversity set II containing 1880 structurally diverse chemical compounds (2654 molecular structures including isomers). The choice of NCI diversity set greatly reduces the computational time as it contains only the representative molecules as well as identifies the molecular scaffolds that could bind with lipases. Hydrogen was added to all the residues of respective drug targets followed by addition of Gasteiger-Marsili charges. Then non-polar

hydrogen atoms were merged onto their respective heavy atoms and atom types were fixed using AutoDockTools (Morris *et al.,* 2009). All ligands were similarly pre-processed using 'prepare_ligand4.py'script available with AutoDockTools distribution. Grid maps were generated for all the atom types present in the respective drug targets and its small molecules along with electrostatic and desolvation maps using AutoGrid utility with grid points to cover the entire region occupying the active site pocket of respective drug targets (Rv0183: 62 x 58 x 60; Rv3802c: 90 x 90 x 90) keeping catalytic Serine at the center of grid box with grid spacing of 0.375 Å. All docking simulations were performed with an initial population size of 300 and 20 independent LGA runs. In each run the best individual from each generation was propagated to the next generation and remaining docking parameters were set to default. The docking results were ranked on the basis of their predicted free energy of binding (ΔG). The whole docking procedure was then repeated with the closest human structural ortholog - monoglyceride lipase (MGL) (Bertrand *et al.,* 2010) to find the potential inhibitors (top hits) based upon their difference in ΔG that could be specific towards their respective drug targets. The molecules were visually inspected for hydrogen bond interactions and van der Waals interactions with the drug targets. The specific interactions of top hits with active site residues of their respective drug targets were further analyzed (Saravanan *et al.,* 2010). Ten top hits were chosen based upon their difference in ΔG (potential mycobacterial specific inhibitors) as the initial dataset to carry out the next step of virtual screening. Virtual screening process was repeated with similar molecules of the initial dataset (Tanimoto coefficient > 0.6) from ZINC database (Irwin and Shoichet, 2005) on both the respective drug targets and human MGL to identify effective and specific inhibitors. Docking protocol was validated using human MGL (PDB id: 3PE6) (Schalk-Hihi *et al.,* 2011), the structural homolog from the human proteome which is co-crystallized with reversible inhibitor (PDB ligand id: ZYH). Rigid docking calculations could mimic the pose of reversible inhibitor (ZYH) reasonably well while docking with the side chain of catalytic Serine flexible was not satisfactory. The rigid docking protocol used in our virtual screening studies would be effective and preferable to screen large number of ligands.

4.3.3. Identification of Potential Dual Inhibitors

Multi-targeted inhibitors are more effective and make pathogen less vulnerable to adaptive resistance due to the multiple actions exerted simultaneously. The approach is to target various biological processes notably cell wall and host lipid catabolism of *M. tuberculosis* simultaneously. The identified drugs targets Rv0183 and Rv3802c were considered for the study with their respective prioritized top hits. Docking of Rv0183 top hits were carried out against Rv3802c and *vice versa* with high stringency.

4.4. Results and Discussion

4.4.1. Drug Target I: Rv0183

M. tuberculosis Rv0183 is a monoacylglycerol lipase comprised of 279 amino acids with a molecular mass of 30 kDa (Côtes *et al.,* 2007). Since its homolog in *M. smegmatis* has roles in cell physiology, colony morphology, and cell-cell interaction, Rv0183 could play an important role in the survival of *M. tuberculosis* as well (Dhouib *et al.,* 2010). From sequence analysis (Table 4.3 and Fig. 4.3), the catalytic residues Ser110, Asp226 and His256 were identified along with conserved HG dipeptide motif which contributes to substrate binding. Rv0183 has Gly-His-Ser-Met-Gly (Gly108-Gly112) as penta-peptide catalytic motif which is strictly conserved in its homologous MGLs used in this study. Rv0183 does not possess the lid domain and triglyceride lipase motif ([LIV]-{KG}-[LIVFY]-[LIVMST]-G-[HYWV]-S-{YAG}-G-[GSTAC]) according to PROSITE database yet it possesses significant diacylglycerol and triacylglycerol lipase activity (Côtes *et al.,* 2007). Enzyme classification number for Rv0183 in UniProtKB is EC 3.1.-.-, unclassified hydrolase and it could be updated as EC 3.1.1.23, monoacylglycerol lipase. Sequence analysis revealed that Rv0183 is highly conserved in mycobacterial species notably in *M. bovis*, *M. leprae* and *M. smegmatis with* 100%, 79% and 68% identity respectively. Rv0183 also has ~34% identity with known mammalian MGLs which are known to be involved in fat mobilization process in adipose tissues (Karlsson *et al.,* 1997; Karlsson *et al.,* 2000). Even though good similarity exists between mycobacterial and mammalian MGLs, sequence analysis reveals significant differences such as Valine in place of Glutamate following the catalytic Histidine and Serine in place of Alanine before HG dipeptide in mammalian species.

Table 4.3: Rv0183 and its orthologs used for sequence analysis in the current study.

UniProt Database Accession	Short Name	Organism	Identity (%)
O07427	Rv0183	*M. tuberculosis*	
B2HMV5	Mma_B2HMV5	*M. marinum*	89
B8ZTH1	Mle_B8ZTH1	*M. leprae*	79
A0QNZ7	Msm_A0QNZ7	*M. smegmatis*	68
Q99685	Human_Q99685	*Homo sapiens*	34
O35678	MOUSE_O35678	*Mus musculus*	36

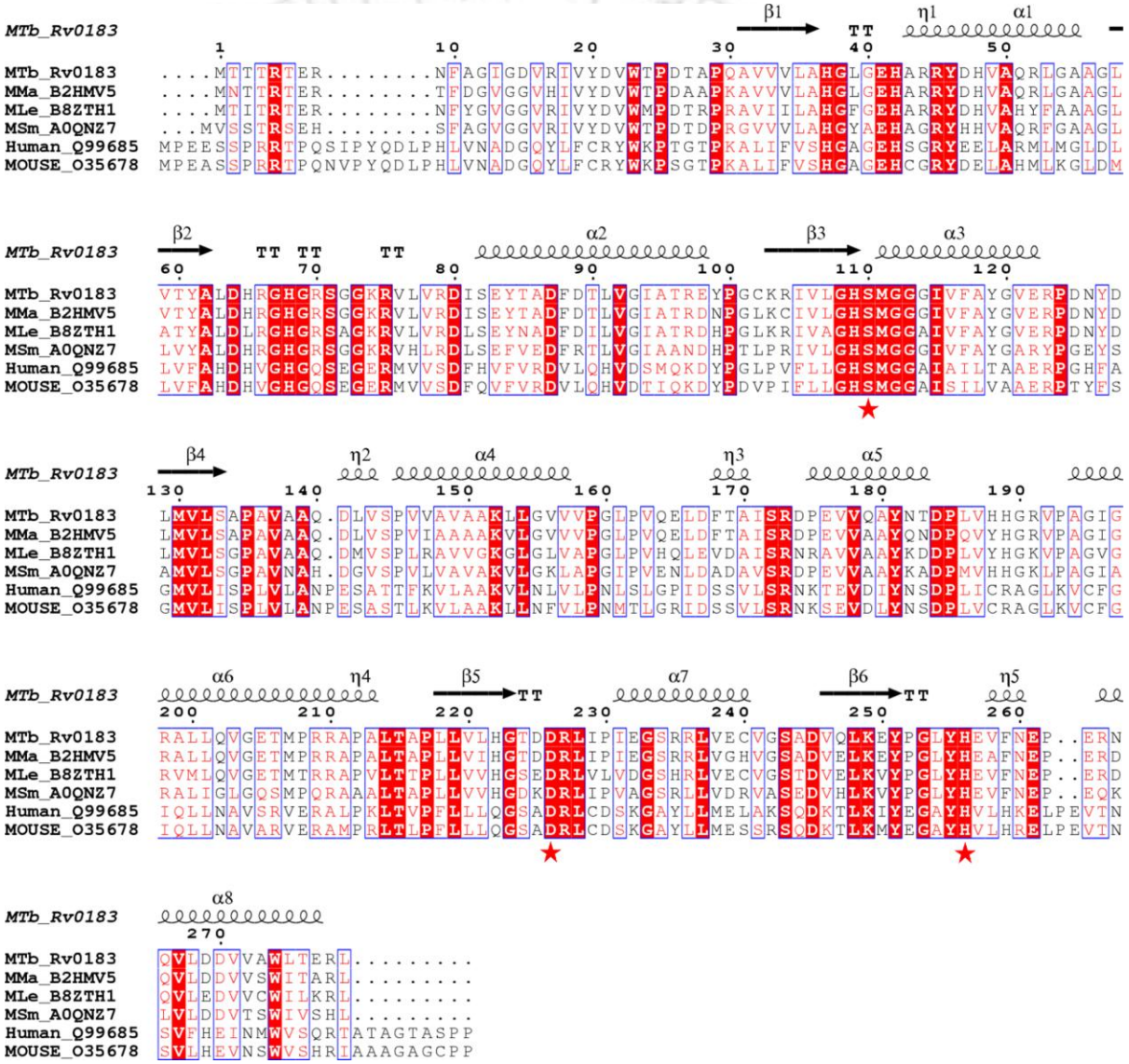


Figure 4.3: Multiple sequence alignment of monoacylglycerol lipases between mycobacterial and mammalian species. Numbering of amino acids is denoted with respect to Rv0183. This figure was produced using ESPript.

Homology Modeling: In view of the limited structure availability of mycobacterial MGLs, molecular modeling approaches have assisted to predict its 3D structure. Selection of human MGL (2.20 Å in its native form with PDB id: 3HJU) as template for structural modeling was due to its functional homology and global sequence similarity (37% identity and 56% similarity in the aligned region) (Labar *et al.*, 2010). In Fig. 4.3, the sequence alignment between Rv0183 and human MGL shows that the conservation of amino acids is throughout the protein.

The energy minimized homology model of Rv0183 takes the α/β hydrolase fold commonly found in lipases with a central parallel β-sheet covered by N- and C- terminus on one side and α-helices on the other as shown in Fig. 4.4. Rv0183 also has a typical conformation of the "nucleophilic elbow" including tightly constrained γ-turn positioning the catalytic Serine near the binding pocket which lies in the disallowed region in the Ramachandran plot (Noble *et al.*, 1993; Heikinheimo *et al.*, 1999). Ramachandran plot also shows that 89.2% of the residues lie in the most favored regions and 9.9% in the additional allowed regions, which indicates that the homology model does not display significant geometrical distortions from expected values. Homology model of Rv0183 also showed good ERRAT plot with the overall quality factor of 95.565% as well as comparable ProSA energy profile with the template human MGL (Fig. 4.5 and Fig. 4.6).

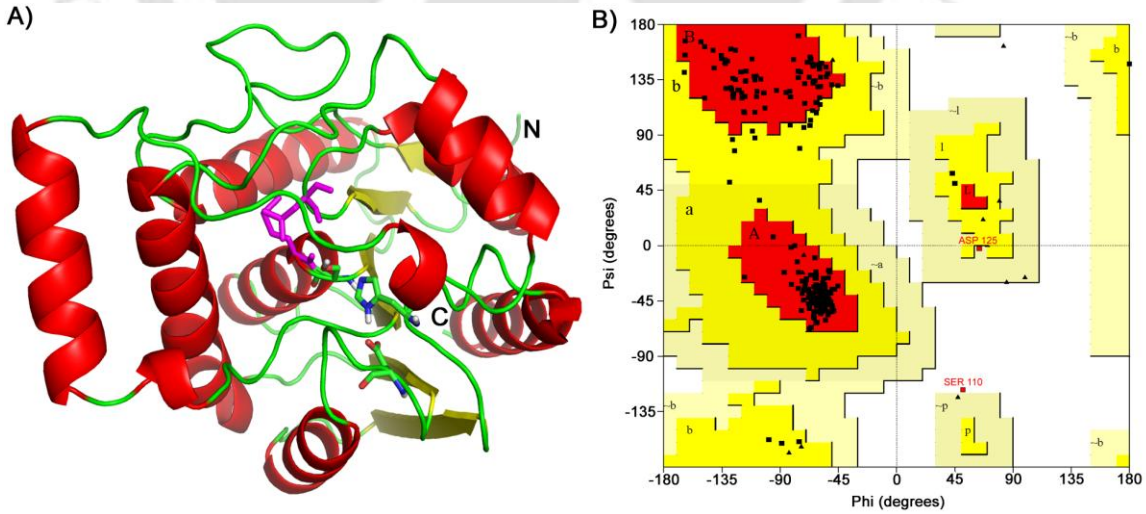


Figure 4.4: A) Rv0183 modelled structure shown in cartoon representation. Catalytic residues and oxyanion hole (magenta) residues are shown in sticks. This figure was produced using PyMOL. B) Ramachandran plot of Rv0183 model revealing the acceptable statistics of dihedral angle distribution of amino acids.

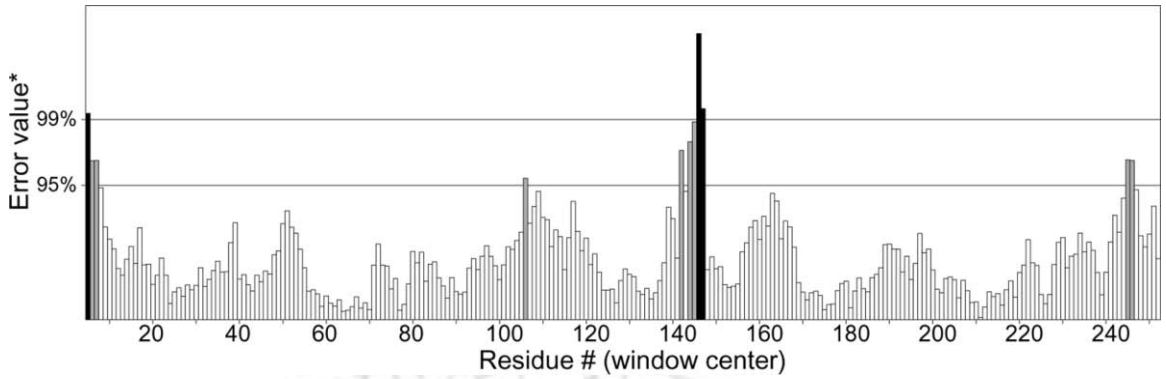


Figure 4.5: Errat plot of Rv0183 model signifying that the structure has considerably low steric hindrance.

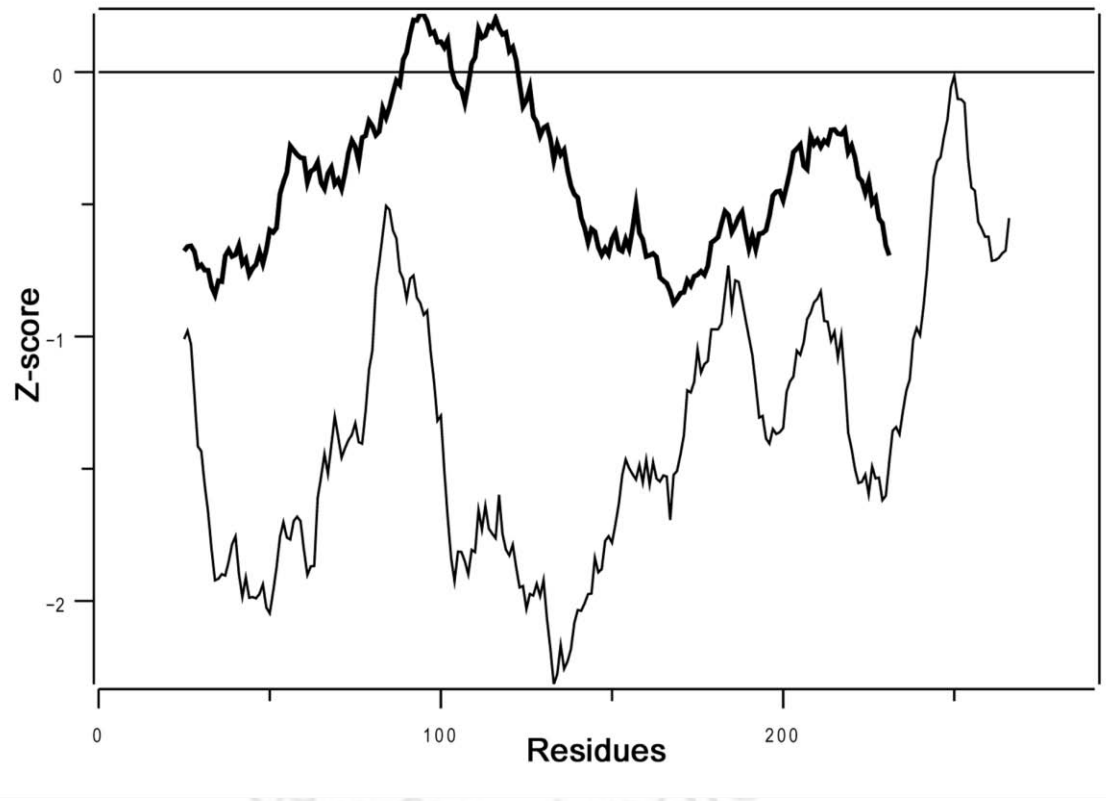


Figure 4.6: Comparative ProSA energy profile of Rv0183 and its template, human MGL in thick and thin line respectively reflects that both have similar fold.

During MD simulation, the RMSD value of Rv0183 backbone atoms from production MD is plotted as time-dependent function (Fig. 4.7). The RMSD values of the backbone atoms in the system tend to be convergent after 0.5ns with fluctuations less than 1 Å; this demonstrates the conformational stability of the Rv0183 model. The RMSD between energy minimized structure and final structure from MD calculated by PyMOL was considerably

low (0.387 Å). Structure validations suggest that the energy minimized Rv0183 model is sufficiently accurate for virtual screening process.

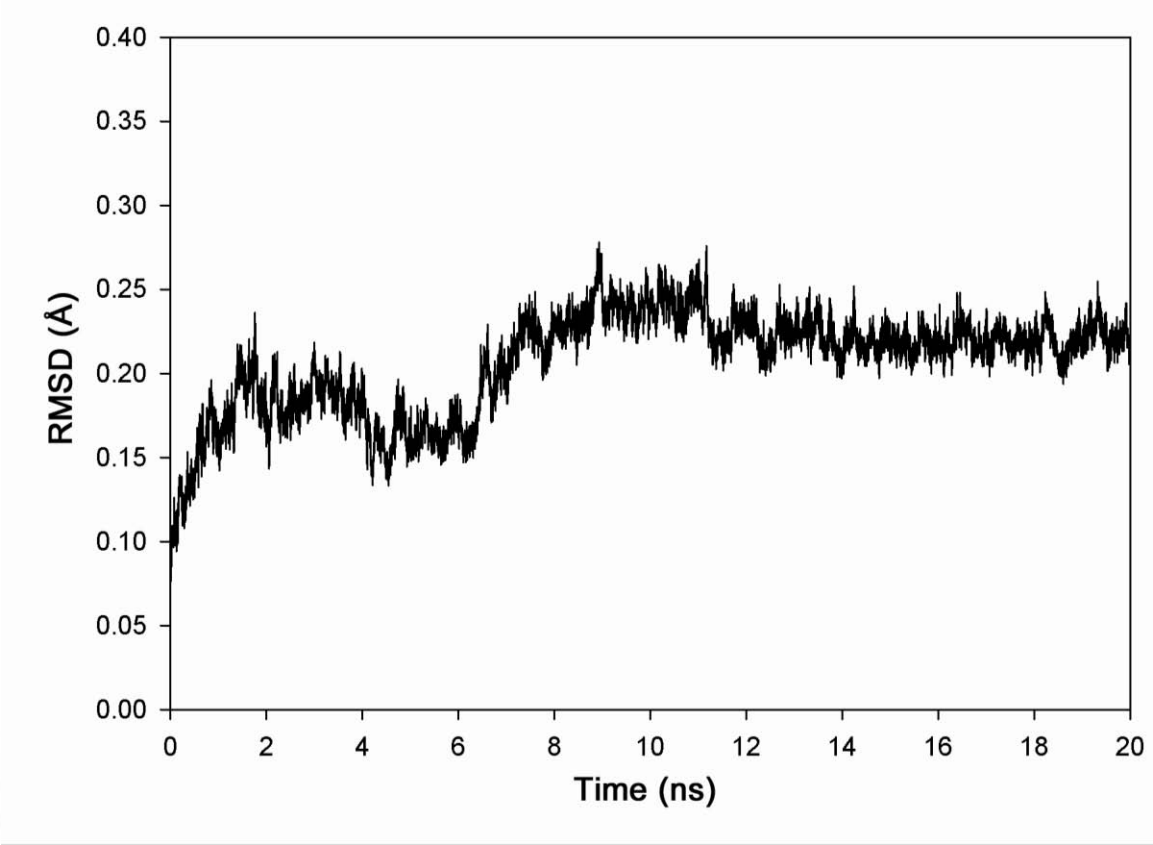


Figure 4.7: Time-dependent RMSD plot of Rv0183 backbone atoms from production MD suggests that the energy minimized structure of Rv0183 is stable.

Discovery of the potential inhibitors against Rv0183 by virtual screening: Validation of the docking methodology using co-crystallized structure of human MGL (PDB id: 3PE6) reproduced the binding mode of reversible inhibitor (PDB ligand id: ZYH) very well. Poses of superimposition of the inhibitors between crystal structure of human MGL and docking calculations are shown in Fig. 4.8. Flexible docking has been carried out for reversible inhibitor again by keeping the side chain of catalytic Serine of flexible. The flexible docking calculations were not satisfactory since they could not mimic the pose of inhibitor of the crystallized structure. Consequently, rigid docking (protein was kept rigid and the ligand was flexible) was carried out in the virtual screening studies.

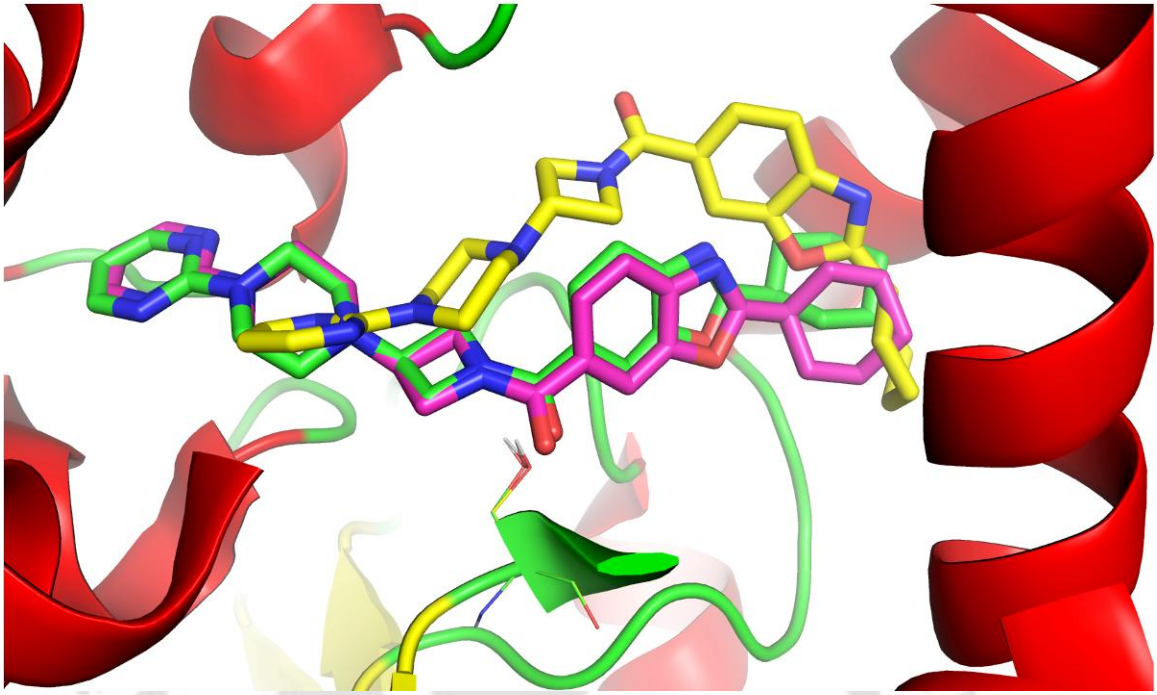
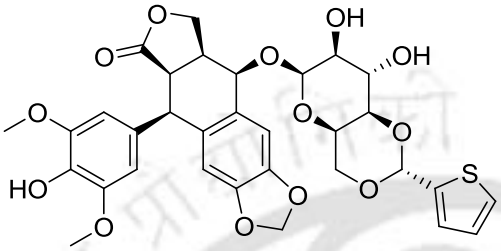
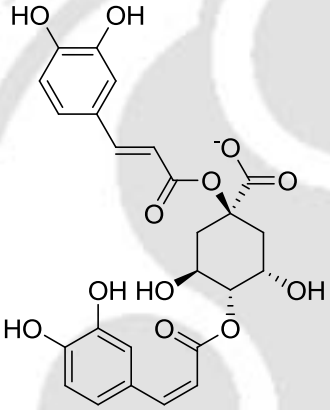
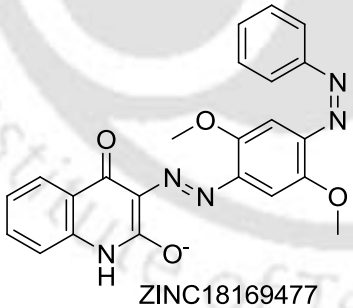
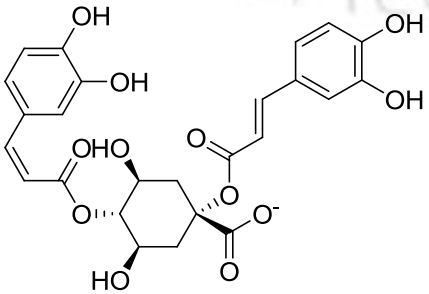
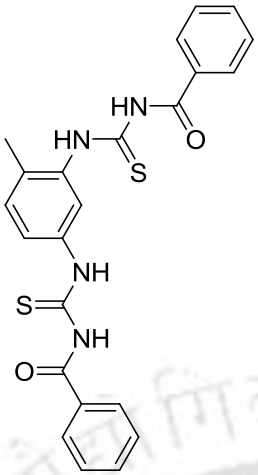
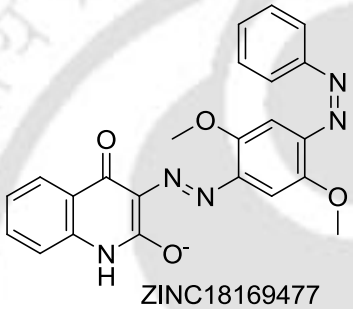
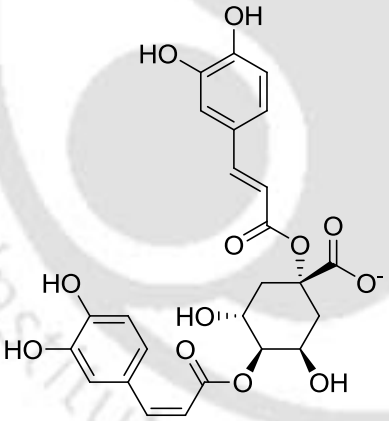
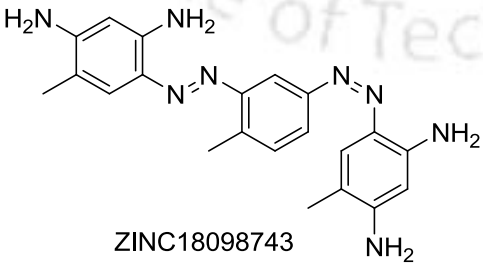


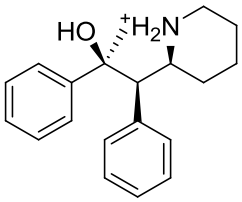
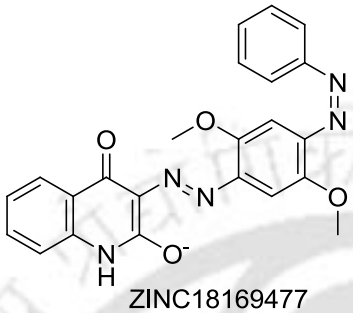
Figure 4.8: Superimposition of the reversible inhibitor poses between crystal structures [PDB id: 3PE6] and docking calculations on human MGL reflect that the rigid docking is favourable to carry out virtual screening. Human MGL is depicted in cartoon representation. Catalytic Serine is depicted as lines. Inhibitors are depicted in stick representation. Inhibitor poses from crystal structure, rigid docking and flexible docking are represented in green, magenta and yellow respectively.

Initial virtual screening was performed with NCI diversity dataset II from ZINC database against Rv0183 model and human MGL to discover potential hits. The difference between the computed binding affinities towards Rv0183 and human MGL was set as the criterion to identify potential inhibitors which are specific towards Rv0183. The top 10 hits of initial screening are listed in Table 4.4.

Table 4.4: Ten top hits selected from the initial virtual screening on NCI diversity set based on the difference in free energy of binding.

S. No.	Molecule	Free Energy of Binding (kcal/mol)		
		Rv0183	Human MGL	Difference
1	 ZINC11616855	-11.89	-8.36	-3.53
2	 ZINC04918963	-10.63	-7.2	-3.43
3	 ZINC18169477	-9.94	-7.6	-2.34
4	 ZINC04918956	-9.95	-7.65	-2.3

5	 ZINC01687247	-12.14	-10.11	-2.03
6	 ZINC18169477	-9.91	-8.03	-1.88
7	 ZINC04918969	-9.39	-7.54	-1.85
8	 ZINC18098743	-9.26	-7.41	-1.85

9	 ZINC04783480	-9.42	-7.6	-1.82
10	 ZINC18169477	-9.79	-7.97	-1.82

For the final screening, similarity search of ten top hits has been performed on ZINC database which resulted in 9033 molecular structures (alternative isomers and protonation states) of 5350 molecules. Virtual screening was carried out again on the similar molecules to find top hits which are specific towards Rv0183. The top hits with free energy of binding more than 8kcal/mol were screened and re-docked with 100 independent LGA runs to increase the reliability of the docking predictions (reliability test). The best conformation of the largest cluster from the reliability test was considered as optimal conformation of the screened hits with LdLip3. Top hits of computational studies and their interaction were analyzed for the presence of hydrogen bonds, van der Waals and possible hydrophobic interactions with Rv0183 using PyMOL (DeLano, 2002).

Contact footprinting through clustering analysis of docked poses from screened hits: Bouvier *et al.* have demonstrated that contact based clustering using docking poses can provide insights on the binding mode of small molecules with the protein of interest (Bouvier *et al.*, 2010). The twenty screened hits were clustered into 5 clusters which have been investigated at the level of sub-clusters to discriminate the contacts of each sub-cluster from each other (Fig. 4.9). Clustering revealed that the top hits were clustered based on the possible interactions as well as the contacts with Rv0183 according to the orientation of the top hits at the active site of Rv0183. The sub-clusters observed within the clusters showed

minor variation in the mode of interactions which are predominant with charge and hydrogen bond interactions (Table 4.5).

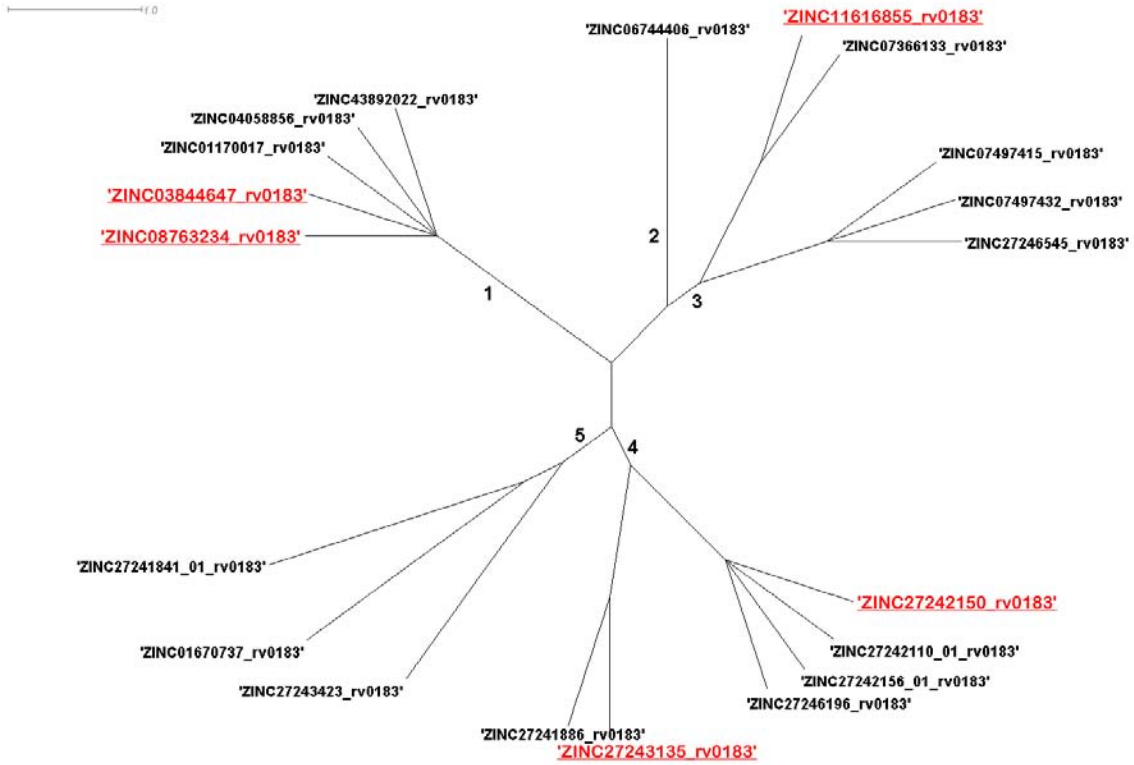


Figure 4.9: Tree representation of contact footprints clustering for twenty top hits with Rv0183. Numbers on branches represent each of the clusters while ZINC database accession of molecule is presented. The five top hits were considered as positive hits for SOM training and represented in bold.

Table 4.5: Conserved contacts of screened hits at the ligand binding pocket of Rv0183 within clusters and various sub-clusters in each cluster and screened hits belonging to the sub-cluster.

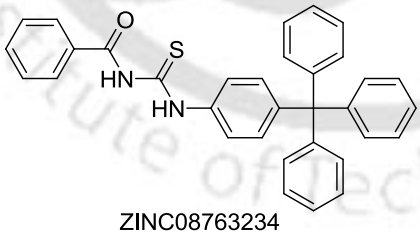
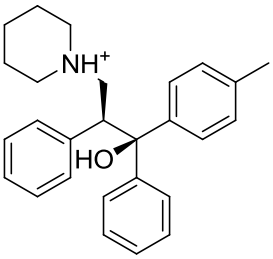
Cluster	Contact residues Conserved within cluster	Sub-cluster [Molecule ZINC DB IDs]	Contact residues conserved within subcluster	Hydrogen bond interactions within subcluster
1	Leu39, Ser110, Leu228	1. [ZINC08763234, ZINC03844647, ZINC01170017, ZINC04058856, ZINC43892022]	- Leu166[H] Leu166[H] - Leu166[H], Val192[H]	

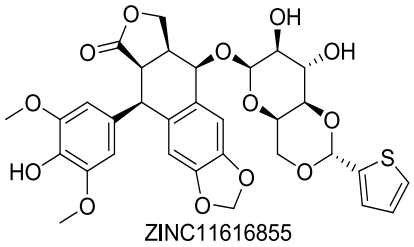
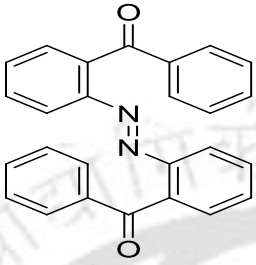
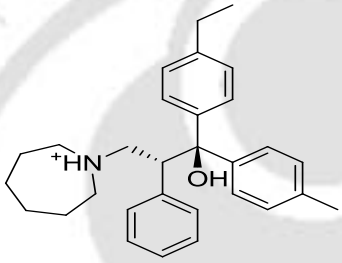
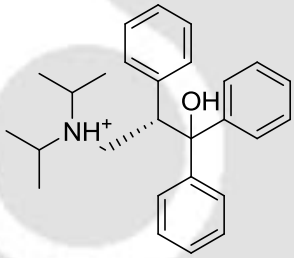
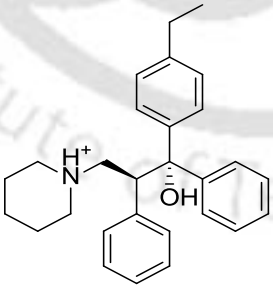
2	Gly38, Ser110, Met111, Ala136, Val137, Ala138, Ala139, Gln140, Leu142, Gly197, Leu201, Val203, Gly204, Leu228, Ile229, Pro230, His256	1. [ZINC06744406]		
3	Ala138, Ala139, Gln140, Leu142, Gly197, Leu200, Leu201, Ile229	1. [ZINC11616855, ZINC07366133] 2. [ZINC07497415, ZINC07497432, ZINC27246545]	Gly38 Leu39, Met111, Ala136, Val203, Ile229	Gln140[H] Gln140[H] Gln140[H] Gln140[H] -
4	Leu39, Ser110, Ala136, Val137, Ala138, Ala139, Gln140, Gly197, Leu200, Leu201, Gly204, Leu228, Ile229, Pro230, His256	1. [ZINC27242110, ZINC27242150, ZINC27242156, ZINC27246196] 2. [ZINC27243135, ZINC27241886]	 Leu142, Val203, Arg227	Leu206[O] - Leu206[O] - - -
5	Leu39, Ser110, Ala136, Val137, Ala138, Ala139, Gln140, Gly197, Leu200, Leu201, Gly204, Leu228, Ile229, Pro230, His256	1. [ZINC01670737, ZINC27241841] 2. ZINC27243423	Met111 Leu142, Met207, Pro208, Arg227,	- - -

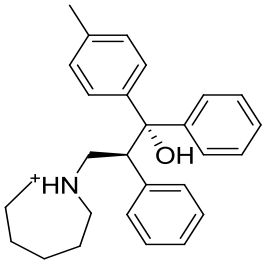
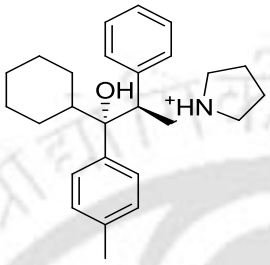
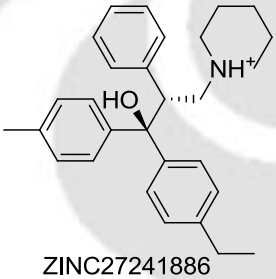
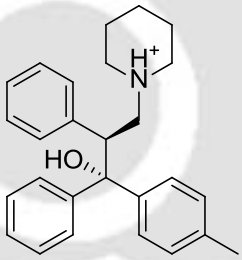
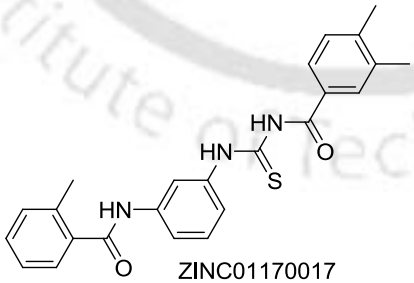
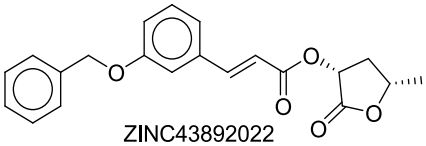
The twenty top hits were screened as potential inhibitors which were listed in Table 4.6. It is noteworthy to mention that these differences in predicted binding energies were larger than the standard error in the AutoDock4.2 scoring function which is ~2.5 kcal/mol (Morris *et al.*, 2009). Most of the known lipase inhibitors are irreversible inhibitors which form covalent bond with catalytic residue Serine, for instance SAR269 (Bertrand *et al.*, 2010) and THL (West *et al.*, 2011). Our current study concentrates on finding potential inhibitors which may lack adversities of the irreversible inhibitors. From virtual screening studies, the interactions of top hits with the Rv0183 model were visually inspected and the interactions suggest that the top hits interact with the key residues of active site pocket of Rv0183. The top hits were bound at the active site of Rv0183 and also share most of the contact residues at 4 Å, which are Leu39, Ser110, Met111, Ala136, Ala138, Ala139, Gln140, Leu142, Leu200, Leu201, Val202, Gly204, Leu228, Ile229 and His256. It is noteworthy to mention that top hits bound

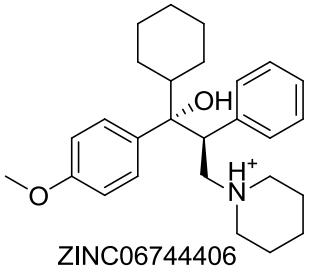
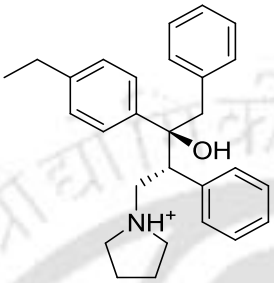
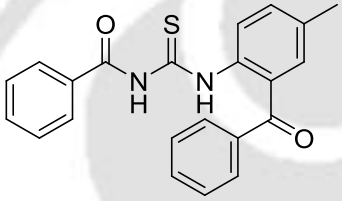
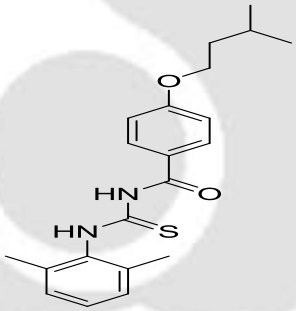
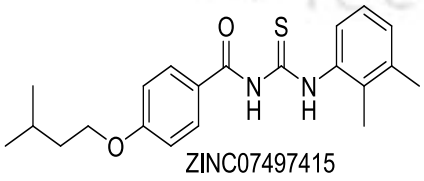
at the vicinity of catalytic residues Ser110 and His256. The binding pocket of Rv0183 with human MGL have been investigated through structural comparison studies as well as contact footprinting which highlights that the ligand binding pocket with respect to the screened hits of Rv0183 is comparatively hydrophobic in nature (Fig. 4.10). Significant differences between the contact residues of some of the top hits with Rv0183 and human MGL were observed, as Leu142, Val147, Gln164, Glu165, Phe168, His188, Arg191, Arg198, Pro208, Ile279, Pro230, Glu257 and Phe259 vs. Ser, Lys, Gly, Pro, Ser, Cys, Gly, Ile, Glu, Cys, Asp, Val and His, correspondingly. Apart from this, considerable differences at the semi-conserved contact residues were observed, as Leu39, Val76, Gly114, Ala136, Ala138, Gln140, Ala148, Val163, Leu166, Ala170, Ile171, Val187, His189, Gly190, Val192, Val203, Gly204 and Met207 in Rv0183 whose counterpart in human MGL is Ala, Met, Ala, Leu, Leu, Asn, Val, Leu, Ile, Val, Leu, Ile, Arg, Ala, Leu, Ala, Val and Val respectively. These differences in amino acids reflect that the active site of Rv0183 was considerably polar in nature when compared with human MGL.

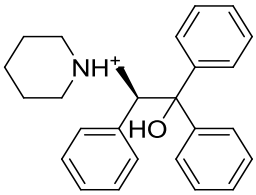
Table 4.6: Screened potential inhibitors selected from virtual screening based on the difference in predicted free energy of binding and number of conformations in the largest cluster is given in parenthesis. Hydrogen bond interactions were predicted with AutoDockTools. Atoms of the residues involved in hydrogen bonds have been indicated in the square brackets.

S. No.	Molecule ZINC DB ID	Free Energy of Binding (kcal/mol)			Residues forming hydrogen bond
		Rv0183	HuMGL	Difference	
1	 ZINC08763234	-12.50 (78)	-7.82 (27)	-4.68	Ser110[HG], His256[HE2]
2	 ZINC27242150	-11.36 (65)	-7.24 (21)	-4.12	-

3	 ZINC11616855	-10.16 (60)	-6.10 (18)	-4.06	Gln140[H]
4	 ZINC03844647	-12.73 (41)	-9.18 (26)	-3.55	Leu166[H]
5	 ZINC27243135	-11.50 (63)	-8.05 (27)	-3.45	Gln140[H]
6	 ZINC01670737	-9.21 (44)	-5.78 (22)	-3.43	-
7	 ZINC27241841	-11.05 (60)	-7.66 (21)	-3.39	-

8	 ZINC27246545	-11.17 (52)	-7.84 (51)	-3.33	-
9	 ZINC27246196	-10.82 (53)	-7.51 (17)	-3.31	-
10	 ZINC27241886	-11.08 (56)	-7.83 (39)	-3.25	-
11	 ZINC27242156	-11.33 (43)	-8.09 (23)	-3.24	Leu228[O]
12	 ZINC01170017	-12.04 (59)	-8.85 (17)	-3.19	Leu166[H]
13	 ZINC43892022	-11.16 (71)	-7.98 (23)	-3.18	Ser110[HG], Leu166[H], Val192[H]

14	 ZINC06744406	-10.45 (59)	-7.36 (17)	-3.09	-
15	 ZINC27243423	-10.70 (56)	-7.66 (17)	-3.04	-
16	 ZINC04058856	-11.36 (87)	-8.34 (23)	-3.02	Ser110[HG]
17	 ZINC07497432	-10.76 (80)	-7.79 (31)	-2.97	Gln140[H]
18	 ZINC07366133	-10.49 (92)	-7.56 (33)	-2.93	Gln140[H]
19	 ZINC07497415	-10.81 (51)	-7.89 (33)	-2.92	Gln140[H]

20	 ZINC27242110	-10.78 (62)	-7.95 (25)	-2.83	Leu228[O]
Substrate	Glycerol monodecanoate	-7.26 (28)	-7.49 (12)	0.23	-
Inhibitor	Orlistat	-8.46 (12)	-7.64 (2)	-0.82	Gln140[H]
Inhibitor	ZYH	-9.40 (89)	-11.14 (46)	1.74	Leu39[H], Leu200[H]

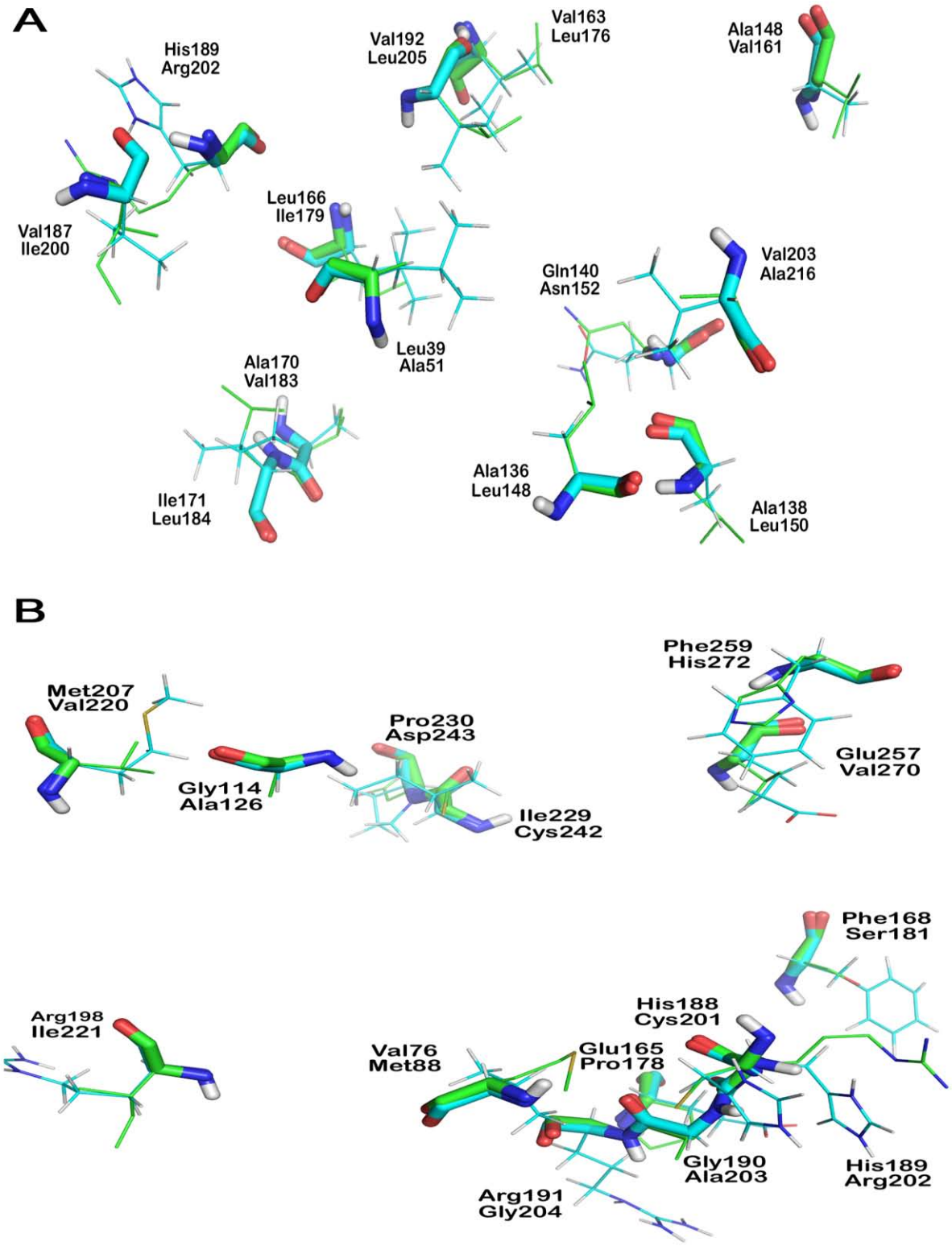


Figure 4.10: Comparison study of the hits binding pocket of Rv0183 with human MGL reveal that the hits binding pocket of Rv0183 is comparatively polar in nature.

The best hit, ZINC08763234 have strong van der Waals interaction (-18.9607 kcal/mol) with Rv0183 and does not have hydrogen bond interactions (-0.0021 kcal/mol) which is reflected in Fig. 4.11. The best hit may also form hydrophobic interactions with Rv0183. Also, some of the top hits form hydrogen bonds with Gln140, Leu166 and Val192 which provides specificity since their counterpart residues in human MGL is Asparagine, Isoleucine and Leucine respectively. The interactions were also stabilized by the hydrophobic residues of the inner cavity. The inhibition mode of screened hits was expected to be competitive. With drugs targeting Rv0183, it could be harder for *M. tuberculosis* to evolve drug resistance against the top hit, which addresses the significant medical problem of multi-drug-resistant and extremely-drug-resistant strains of tuberculosis.

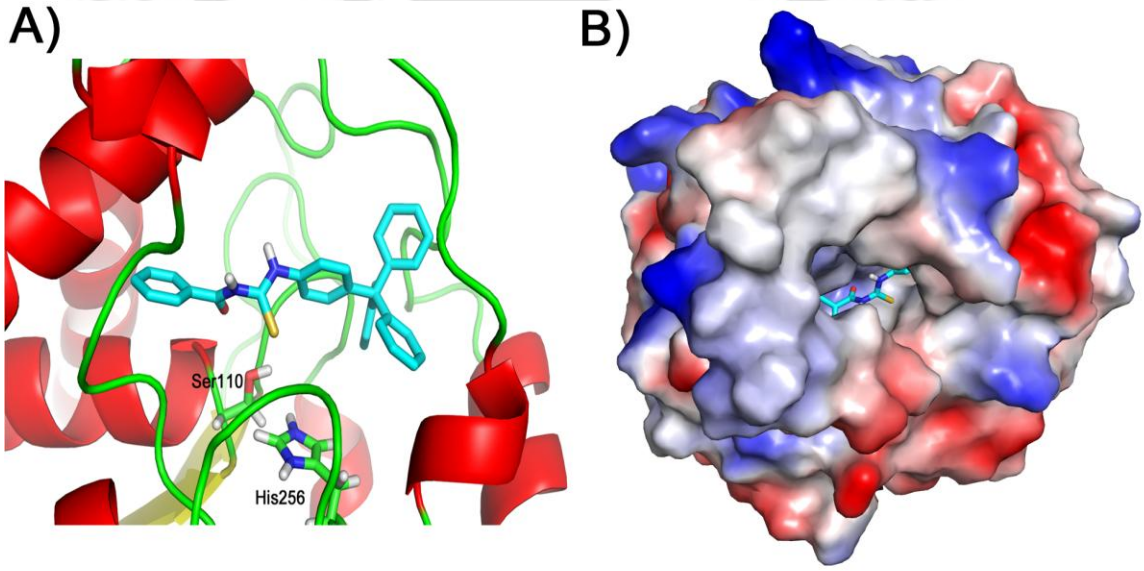


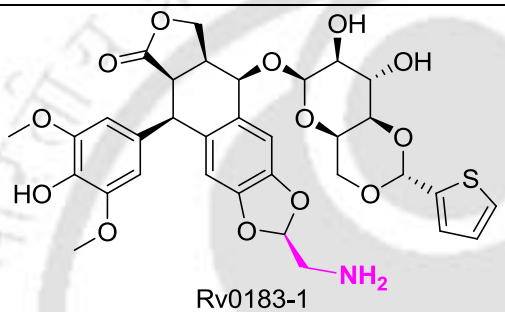
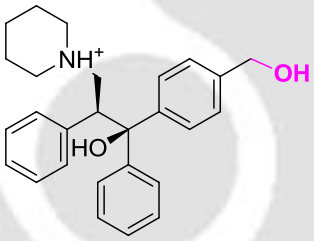
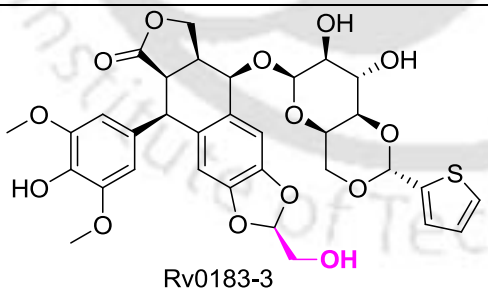
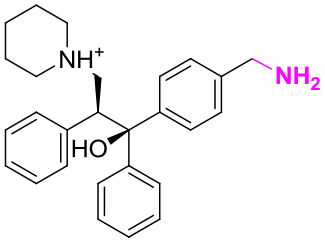
Figure 4.11: A) The interaction mode of best hit ZINC08763234 in the vicinity of catalytic residues of Rv0183 as suggested by AutoDock. B) The best hit ZINC08763234 bound well in the active site of Rv0183. This figure was produced using PyMOL.

Design of potential inhibitors against Rv0183:

SBDD provided insights on the binding mode of interaction between the top hits with Rv0183 which aid to design new molecules with improved predicted free energy of binding. The visual inspection of binding mode of the best hits concurrent with contact footprinting analysis revealed that the modification of the best hits through inclusion of functional groups that may favor the ionic interactions between the best hits and Rv0183. The detailed information of new molecules is summarized in Table 4.7 and Figure 4.12. The NH_2, NO_2 OH and CH_2OH have been incorporated in

the methyl group of phenyl ring. As expected, the incorporation of functional groups improvised the free energy of binding and slight change in conformation of binding mode of the new molecules. The study provides insight on the design and development of lead molecules against tuberculosis.

Table 4.7: The new molecules with better free energy of binding designed based on SBDD studies against Rv0183. Number of conformations in the largest cluster is given in parenthesis. Hydrogen bond interactions were predicted with AutoDockTools. Atoms of the residues involved in hydrogen bonds have been indicated in square brackets. The inclusion/replacement of functional group on the top hits is depicted with magenta color.

S. No.	Molecule*	Free Energy of Binding (kcal/mol)			Residues forming hydrogen bond
		Rv0183	Human MGL	Difference	
1.	 Rv0183-1	-18.08 (69)	-11.60 (24)	-6.48	Leu39[H], Ser110[HG], His256[HE2]
2.	 Rv0183-2	-12.49 (59)	-8.02 (18)	-4.47	-
3.	 Rv0183-3	-13.22 (51)	-9.78 (18)	-3.44	-
4.	 Rv0183-4	-12.27 (67)	-9.39 (18)	-2.88	-

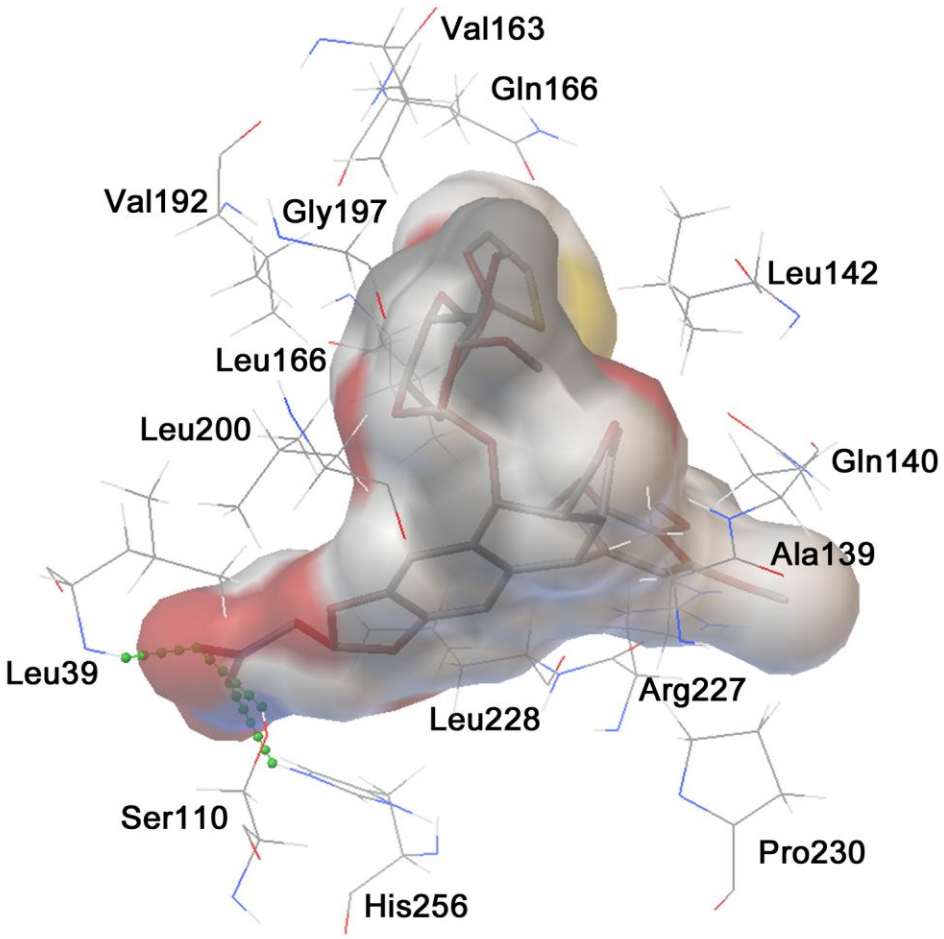


Figure 4.12: Binding mode of new designed molecule Rv0183-1 in the active site pocket of Rv0183.

4.4.2. Drug Target II: Rv3802c

M. tuberculosis Rv3802c is a lipase with Phospholipase A and Thioesterase activity located in genetic cluster Rv3799c to Rv3804c (Parker *et al.,* 2009). This genetic cluster is dedicated to the synthesis of core structures of mycobacterial cell wall as well as it is found to be conserved in actinomycetes including *Corynebacterium glutamicum* (Takayama *et al.,* 2005; Malen *et al.,* 2010). Rv3802c ortholog in *C. glutamicum* is Ncgl2775 which plays a key role in the regulation of outer lipid composition under heat stress and their orthologs were proposed to be envelope lipids regulation factor (ElrF) family (Meniche *et al.,* 2005). Rv3802c gained importance because of its genetic location dedicated to mycolic acid biosynthesis (Sassetti *et al.,* 2003). Rv3802c is also conserved in *M. leprae*, which is considered a minimal genome (Vissa and Brennan, 2001). Of seven cutinases in *M. tuberculosis*, Rv3802c is the only one reported as essential and translocated to cell wall

(West *et al.,* 2008; West *et al.,* 2009). Based on the experimental studies on its essentiality and role in the organism's cellular integrity from *M. smegmatis*, Rv3802c is considered as good drug target for treating tuberculosis (Crellin *et al.,* 2010; West *et al.,* 2011).

M. tuberculosis Rv3802c comprises of 336 amino acid residues and found active towards all fatty acid chain lengths of nitrophenyl esters (Parker *et al.,* 2009). Sequence analysis reveals that Rv3802c is conserved within most of the mycobacterial species including *M. leprae* and no sequence homolog was found in human (Table 4.8). Sequence analysis identified Ser175, Asp268 and His299 as the catalytic residues of Rv3802c and the conserved "nucleophilic elbow" (Gly-Phe-Ser-Gln-Gly; Gly173-Gly177) (Fig. 4.13). As reported in MSMEG_6394, Rv3802c also has a consensus of Gly-(Phe/Tyr)-Ser-Gln-Gly pentapeptide, a remarkable feature found in cutinase family. The domain studies using sequence search in PFAM (Finn *et al.,* 2010) also suggest the presence of cutinase domain extends from Ile110 to Ala273 in Rv3802c but no cutinase activity was reported. PE-PPE domain was identified in Rv3802c by NCBI Conserved Domain Database ranging from Val124 to Pro305 (Adindla and Guruprasad, 2003; Marchler-Bauer *et al.,* 2011). Interestingly, MSMEG_6394 has been used by Sultana *et al.* (Sultana *et al.,* 2011) to predict the structure of serine hydrolases with PE/PPE domain. MSMEG_6394 and Rv3802c might be serine hydrolases integrated with the PE/PPE domain which might be essential for the survival of the pathogen. The N-terminal region of Rv3802c was predicted to have transmembrane spanning region from 12 to 34 residues using TMHMM and a eukaryotic signal peptide from 1 to 38 residues using SIGNALP in InterProScan sequence search at the European Bioinformatics Institute (McWilliam *et al.,* 2009) which might have role in translocating Rv3802c to mycobacterial cell wall.

Table 4.8: Rv3802c and its representative mycobacterial homologs used in sequence analysis.

S. No.	UniProt Database Accession	Short Name	Organism	Identity/Similarity (%)
1.	O53581	Rv3802c	*M. tuberculosis*	
2.	Q7TVM8	Mb3832c	*M. bovis*	100/100
3.	Q9CDB3	Ml0099	*M. leprae*	75/84
4.	Q745L1	MAP_0218	*M. paratuberculosis*	75/85
5.	A0R619	Msmeg_6394	*M. smegmatis*	69/78

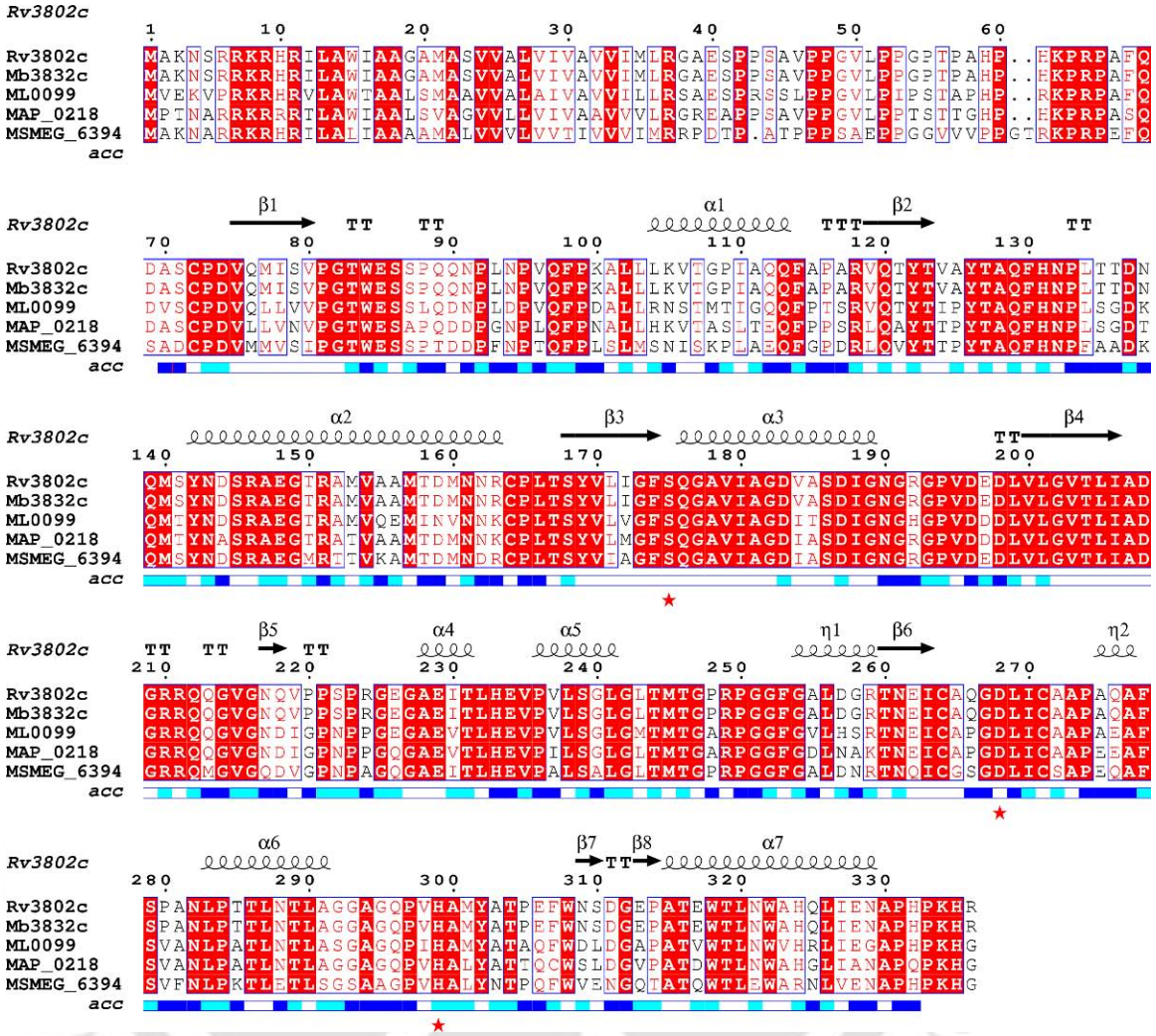


Figure 4.13: Multiple sequence alignment of *M. tuberculosis* Rv3802c with its homologs reflect that the amino acids are conserved throughout the protein. Numbering of amino acids has been denoted with respect to Rv3802c and its secondary structure has also been shown. The catalytic residues are shown with star symbol. Surface accessibility has been depicted in the modeled region with white, cyan and blue colour for buried, intermediate and accessible residues respectively.

Homology Modeling: From sequence analysis and NCBI search against PDB database, the functional homolog of Rv3802c in *M. smegmatis*, MSMEG_6394 with a resolved crystal structure of 2.9 Å was identified as the template for homology modeling studies. Another contributing factor for its selection was its inhibition by THL. The modelled structure predicted that Rv3802c is a member of cutinase family of α/β hydrolases (Fig. 4.14). The first 69 residues are not modelled due to the lack of structural information and believed not to have any functional role in enzymatic activity of Rv3802c. The modelled region (73% identity and 82% similarity towards MSMEG_6394), has been represented with "acc" legend which reflects surface accessibility of the amino acids, has been depicted in Fig. 4.13. The

putative function of the non-modelled region could be anchoring Rv3802c in the cell wall so that it could participate in cell wall remodeling. As observed in MSMEG_6394, Rv3802c has α/β hydrolase domain with six-stranded parallel beta sheets covered by helices and "lid" sits atop of active site. The active site was formed by optimally positioning the local conformation of the catalytic residues shown as sticks in Fig. 4.14. The nucleophilic Serine (Ser175) was found at the bend of the tight turn in the "nucleophilic elbow" and other catalytic residues (Asp268 and His299) are positioned adjacent to each other within the active site. Taken together, structural features of Rv3802c are similar to that of MSMEG_6394 and other serine esterases. Also in the energy minimized Rv3802c model, the nucleophilic Serine was observed in disallowed region of the Ramachandran plot, a typical feature found in most of the energy minimized serine lipases/esterases. Apart from conserved catalytic residues, strict conservation was observed at the active site cavity namely Trp84, Glu85, Ser86, Thr127, Ala128, Gln129, Met140, Phe174 and Ala300 which might be involved in the substrate binding and recognition. The oxyanion hole which is involved in substrate binding and in stabilizing the tetrahedral intermediate could not be identified from sequence analysis because of the lack of HG dipeptide. The comparison of Rv3802c model with human MGL and cutinases [PDB id: 1CEX] (Longhi *et al.*, 1997) identifies Thr83 and Gln176 as the putative oxyanion hole residues (Sultana *et al.*, 2011).

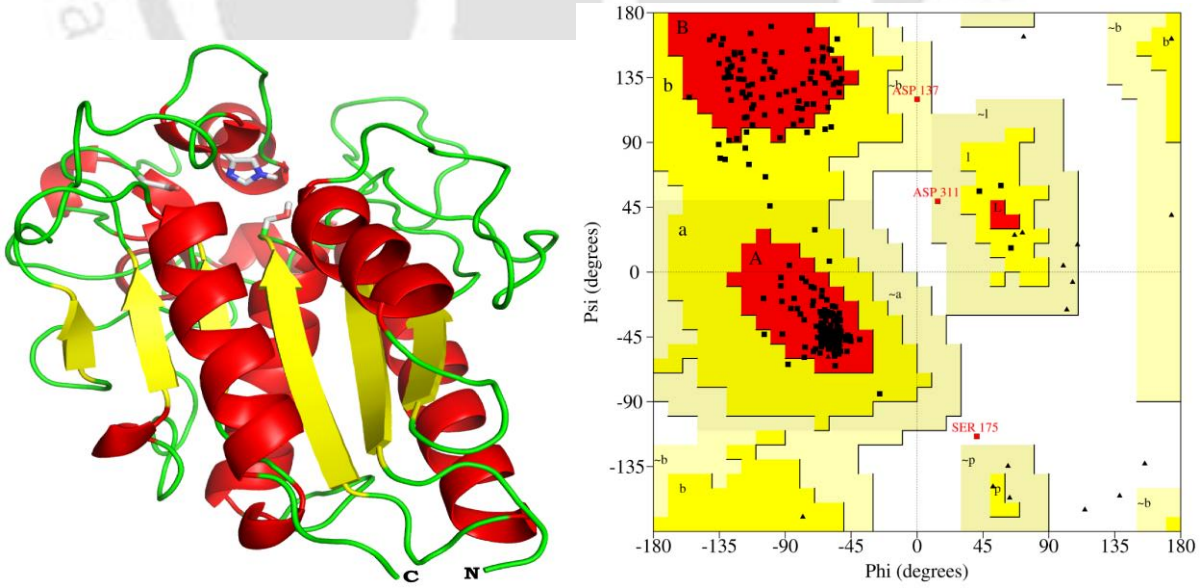


Figure 4.14: A) Rv3802c modelled structure shown in cartoon representation. Catalytic residues are shown in stick representation. This figure was produced using PyMOL. B) Ramachandran plot of Rv3802c model revealing the acceptable statistics of dihedral angle distribution of amino acids.

The validated modelled structure of Rv3802c has acceptable statistics of backbone dihedral angle distribution of amino acids in Ramachandran plot with 91% in core, 7.6% in additionally allowed and 0.9% in generously allowed region as presented in Fig. 4.14. G-factor of Rv3802c was -0.07 which again indicates acceptable statistics of distribution of phi and psi along with chi1, chi2 and chi3 angles (G-factor less than -0.5 is considered to be unusual). Errat plot of the final model of Rv3802c showed an overall score for structural quality of 78.906 with minor 'structure error' that reflects the steric hindrance between few amino acids (Fig. 4.15). Verify-3D also revealed that 98.87 % of the amino acids in the current structure of Rv3802c have compatible 1D-3D score greater than 0.2. The Rv3802c model has Z-score value of -6.71 in the range of native conformations of crystal structures and also comparable with MSMEG_6394 (Z-score: -6.47) with ProSA-web (Wiederstein and Sippl, 2007). The fold quality of Rv3802c from ProSA was comparable with experimentally determined MSMEG_6394 indicating the acceptability of the modelled Rv3802c structure (Fig. 4.16).

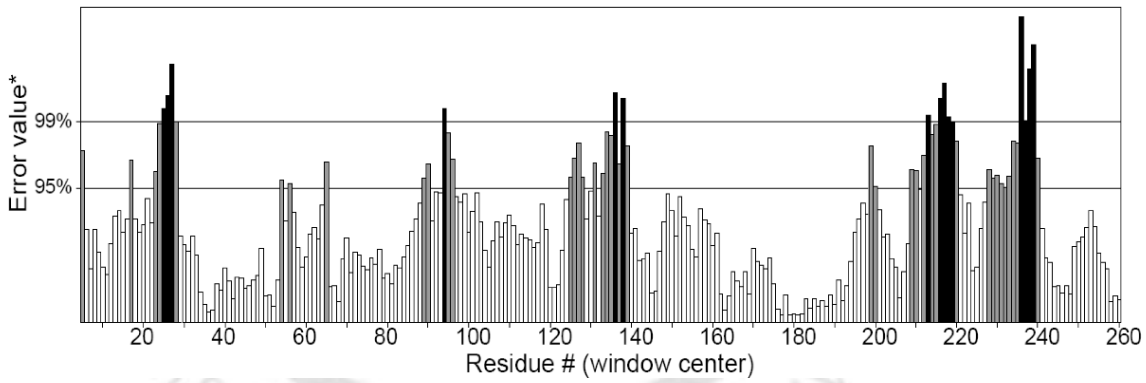


Figure 4.15: The Errat Plot of modeled Rv3802c showing that the structure is considerably of low steric hindrance reflected by very few bars of 99% error value.

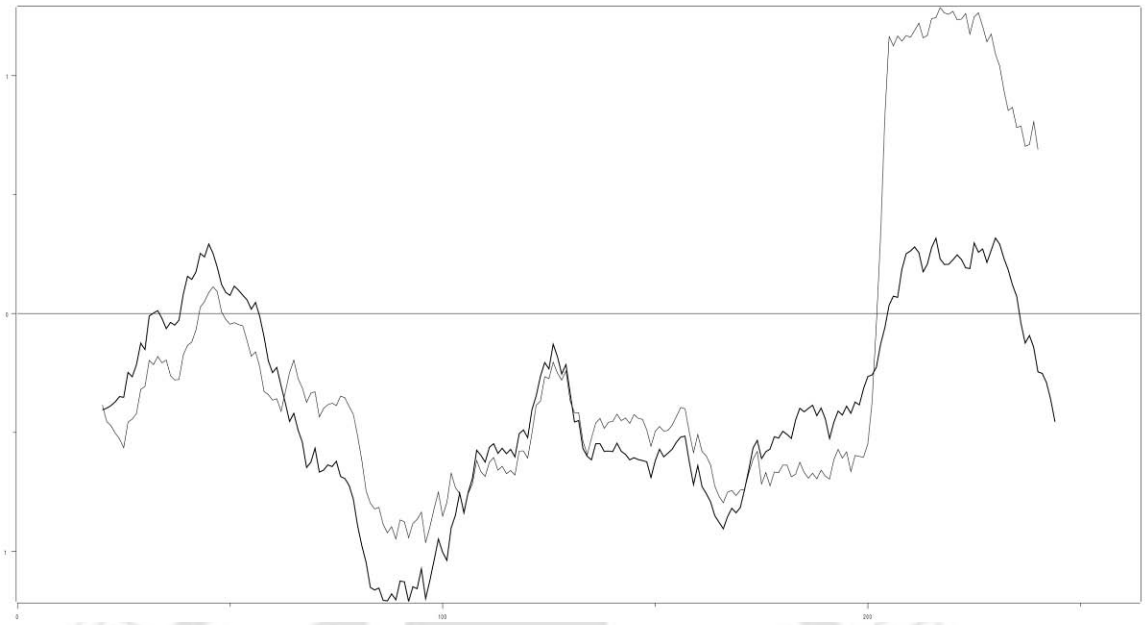


Figure 4.16: Comparative ProSA energy profile of Rv3802c and its homolog counterpart in *M. smegmatis* in thick and thin line respectively reflecting both have similar fold.

Structural comparison of Rv3802c model with MSMEG_6394 on Cα-backbone atoms shows overall RMSD of 0.189. RMSD of 32 out of 45 Cα-backbone atoms in the active site region (4 Å around catalytic residues for simplicity) was 0.158 and residues Tyr126, Gly173, Asp208, Gly292, Pro297 and Ala300 were not superimposed and Tyr142, Ala181, Arg221, Leu290, Gly293 and Ala294 in Rv3802c was observed in the active site region under study. This might could be the reason that THL inhibits Rv3802c effectively when compared with MSMEG_6394 and further experimental studies will provide insights in specific inhibition potential of THL. In order to check the stability of Rv3802c model, RMSD of backbone atoms from MD production run was plotted as time-dependent function in Fig. 4.17. The graph clearly indicates that there is significant change in RMSD for the initial 5ns and the system converged with fluctuations less than 0.5 Å. The RMSD between energy minimized model of Rv3802c and final structure from MD was low (2.522 Å). Furthermore, structural comparison of energy minimized structure with structures generated throughout the MD production run indicates that the energy minimized Rv3802c model represents a stable conformation. Structure validation results suggest that the energy minimized Rv3802c model is sufficiently accurate for virtual screening process. Modelled structure of Rv3802c was submitted to Protein Model Database with the identification number PM0077735.

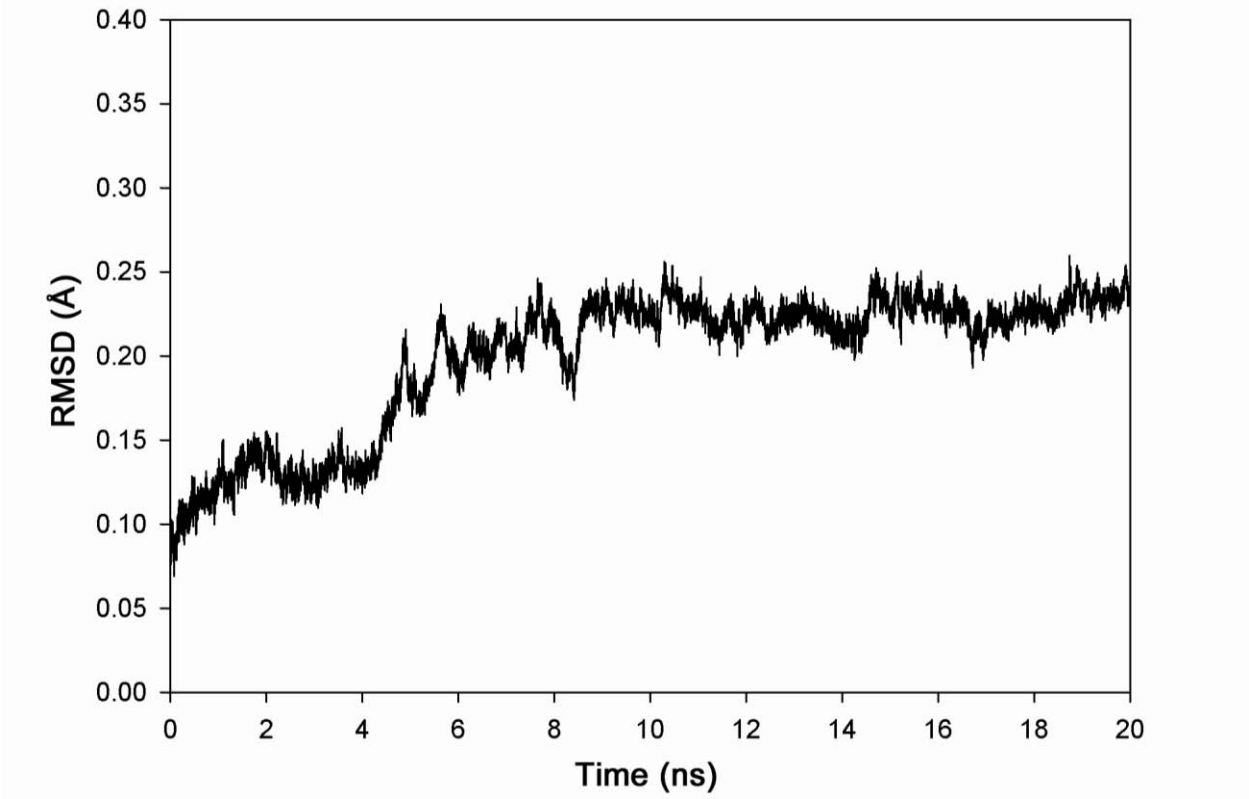
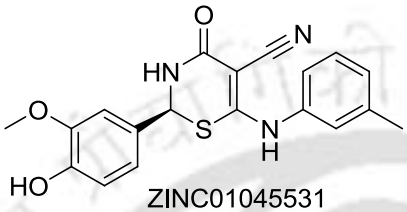
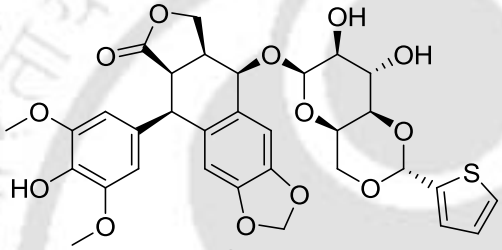
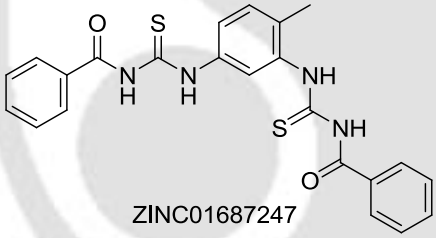
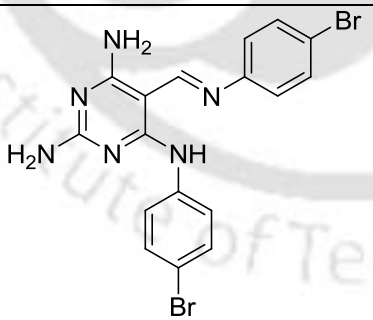
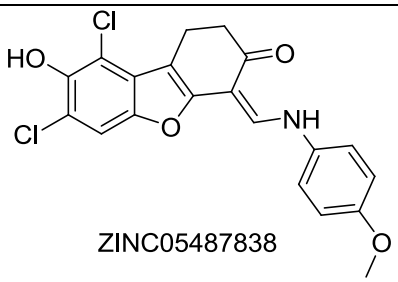


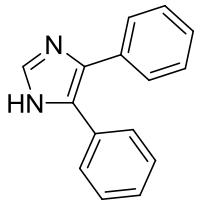
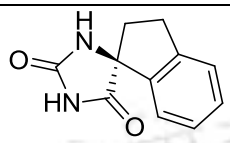
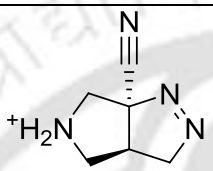
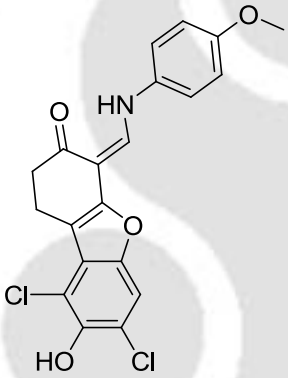
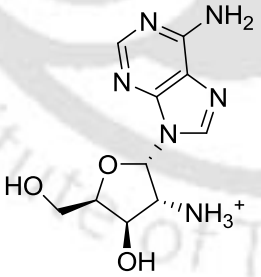
Figure 4.17: MD of modeled Rv3802c shows that the energy minimized structure is energetically stable.

Discovery of the potential inhibitors against Rv3802c by virtual screening: Rv3802c structure was modelled for the first time, since little is known about its inhibitors or the structural features required for better inhibition. To serve as possible starting point for translational research, structure based virtual screening was used to identify potential inhibitors targeting Rv3802c active site. To summarize, structurally diverse compounds were virtually screened to identify potential inhibitors for drug therapy against tuberculosis. Most of the known lipase inhibitors are irreversible inhibitors which form covalent bond with catalytic Serine, for instance SAR269 (Bertrand *et al.*, 2010), JZL184 (Long *et al.*, 2009[a]; Long *et al.*, 2009[b]) and THL (Beatty and Russell, 2000). Our current study implemented two-tier screening approach to find potential inhibitors effective towards Rv3802c targeted against its substrate binding pocket. In the first approach, initial virtual screening was carried out using NCI diversity dataset II against the active site of Rv3802c model to discover all possible structurally diverse potential hits. Similarly, screening was also carried out for human MGL. The molecules with higher binding affinity towards Rv3802c than human MGL (based on the difference in predicted free energy of binding) was set as the criterion to

identify potential inhibitors against Rv3802c. The top hits of initial screening have been listed in Table 4.9.

Table 4.9: Top 10 hits selected on Initial Screening using NCI Diversity Set based on the difference in predicted free energy of binding according to AutoDock.

S. No	**Molecule** **ZINC database ID**	**Free Energy of Binding (kcal/mol)**		
		Rv3802c	Human MGL	Difference
1.	 ZINC01045531	-9.91	-6.64	-3.27
2.	 ZINC11616855	-9.27	-6.35	-2.92
3.	 ZINC01687247	-10.4	-7.73	-2.7
4.	 ZINC13152247	-9.6	-7.11	-2.49
5.	 ZINC05487838	-11.2	-8.72	-2.46

6.	 ZINC00065175	-8.12	-5.72	-2.4
7.	 ZINC03258337	-8.2	-5.9	-2.3
8.	 ZINC05541927	-7.61	-5.33	-2.28
9.	 ZINC05487838	-10.9	-8.64	-2.27
10.	 ZINC07996353	-7.2	-4.93	-2.27

The recently identified leads with inhibitory action against Rv3802c (four potent molecules than known inhibitor, tetrahyrolipstatin) was also considered as the initial dataset (West *et al.*, 2011). Similarity search (60% similarity using tanimotto coefficient) of 10 top hits from initial screening along with the potent compounds reported by West *et al.* was performed on ZINC database resulted in 23479 and 8023 molecules including isomers respectively. The

virtual screening with above protocol was performed again with 20 GA runs on the similar molecules to find potential inhibitors towards Rv3802c (secondary screening).

Contact footprinting of screened hits of Rv3802c: Contact based clustering using docking poses of Rv0183 demonstrated that can provide insights on the binding mode of small molecules. The twenty screened hits of Rv3802c were clustered into 5 clusters which have been investigated at the level of sub-clusters to discriminate the contacts of each sub-cluster (Fig. 4.9). As expected, clustering revealed that the top hits were clustered based on the possible interactions as well as the contacts with Rv3802c. The sub-clusters observed within clusters showed minor variation in the interaction mode which are predominant with charge and hydrogen bond interactions (Table 4.5).

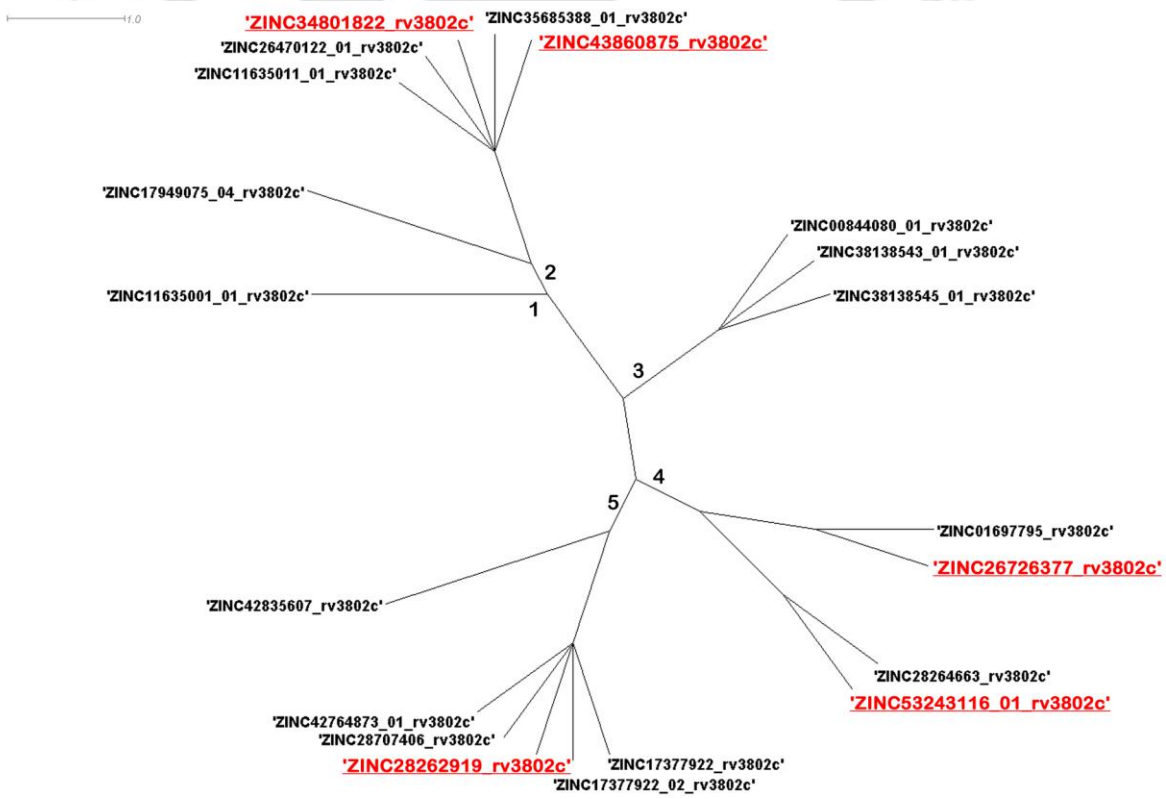


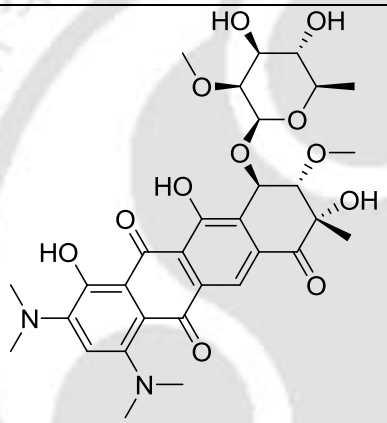
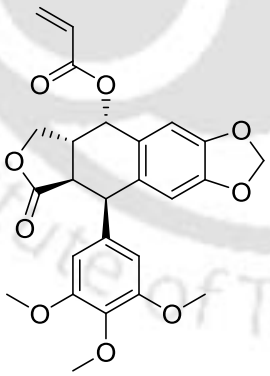
Figure 4.17: Tree representation of contact footprints clustering for 20 top hits with Rv3802c. Numbers on branches represent each of the five clusters while ZINC database accession of each molecule is presented.

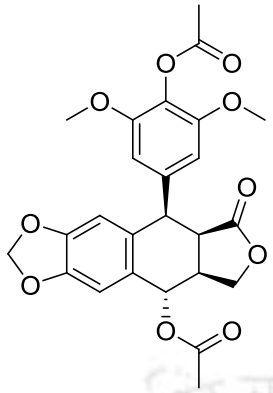
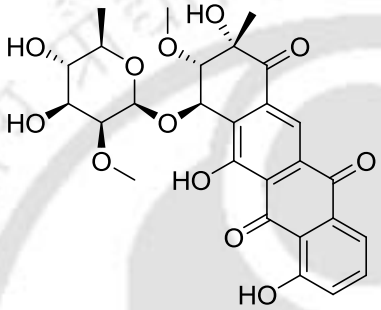
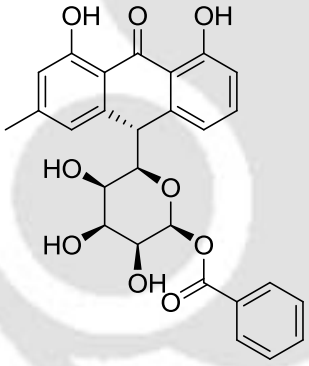
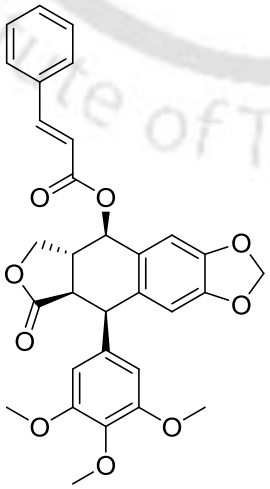
Table 4.10: Conserved contacts of screened hits at the ligand binding pocket of Rv3802c within clusters and various sub-clusters in each cluster and screened hits belonging to the sub-cluster.

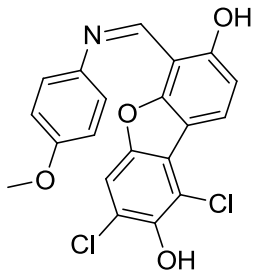
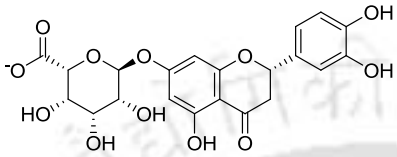
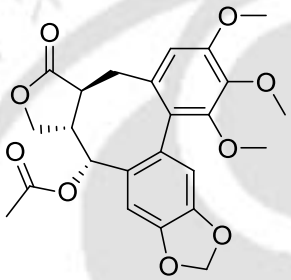
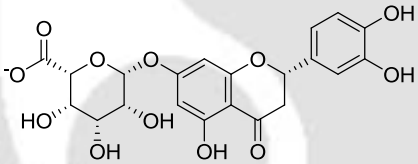
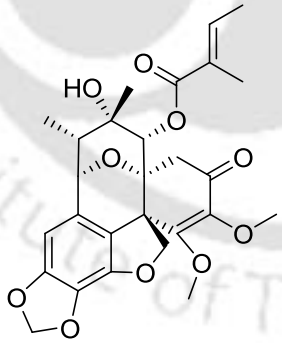
Cluster	Contact residues Conserved within cluster	Sub-cluster [Molecule ZINC DB IDs]	Contact residues conserved within sub-cluster	H-bond interactions
1	Thr83, Trp84, Glu85, Phe98, Lys100, Ala101, Leu102, Leu104, Lys105, Phe174, His299, Ala300, Met301, Tyr302, Ala303	1. [ZINC11635001_01]		Ala303[NH]
2	Thr83, Trp84, Glu85, Phe98, Lys100, Ala101, Leu102, Phe174, His299, Tyr302	2. [ZINC11635011_01, ZINC26470122_01, ZINC34801822, ZINC35685388_01, ZINC43860875]		Ala303[NH] Glu85[OE2], His299[HE2], Gly293[O] Lys100[O], Lys100[NH], Gly293[O] Ala303[NH] Lys100[O], Lys100[NH], Gly293[NH]
		3. [ZINC17949075_04]	Lys100[O], Tyr302[NH]	Lys100[O], Tyr302[NH]
3	Thr83, Trp84, Glu85, Phe98, Lys100, Ala101, Leu102, Phe174, His299, Ala300, Met301, Tyr302, Ala303	1. [ZINC00844080_01, ZINC38138543_01, ZINC38138545_01]		- Trp84[O], Lys100[HZ2], Tyr302[NH] Trp84[O], Lys100[HZ2], Tyr302[NH]
4	Trp84, Glu85, Lys100, Leu102, Phe174, G292, Gly293, His299, Ala300	2. [ZINC01697795, ZINC26726377]	Gly82, Thr83, Phe98, Ser175, Tyr302	Ala300[O], Ser175[HG], His299[NE2], Thr83[OG1], His299[HE2] Trp84[O], Lys100[HZ3], Ser175[HG], His299[O], His299[HE2]
		1. [ZINC28264663, ZINC53243116_01]	Ala101	Trp84[O], His299[O] Thr83[O], Thr83[OG1], Ser175[HG],
5	Thr83, Trp84, Glu85, Lys100, Ala101, Leu102, Phe174, G292, Gly293	1. [ZINC17377922, ZINC17377922_02, ZINC28262919, ZINC28707406, ZINC42764873_01]	Phe98, Ser175	Lys100[O] - Trp84[O], Ser175[HG], His299[O], His299[HE2] Trp84[O], Gly293[O] Thr83[O], Lys100[O]
		2. [ZINC42835607]	Pro99, Leu104, Lys105, His299, Ala300, Met301, Ala303 T316	-

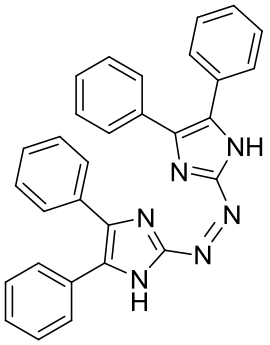
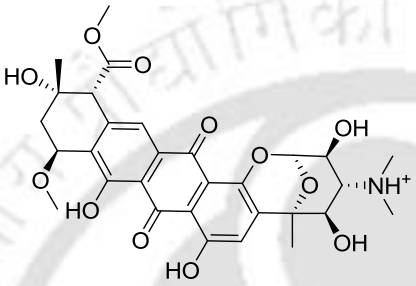
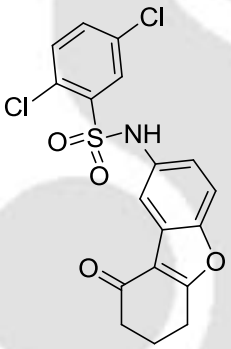
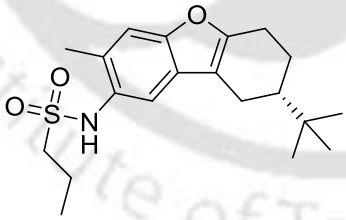
The top hits of secondary screening are listed in Table 4.11. Top hits of computational studies were visually inspected and analyzed for all possible hydrogen bond and hydrophobic interactions with Rv3802c.

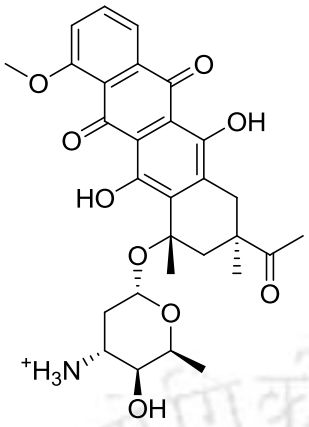
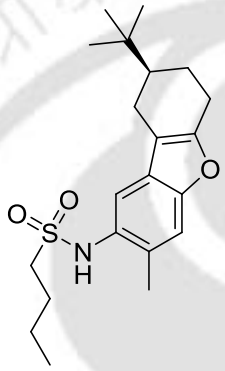
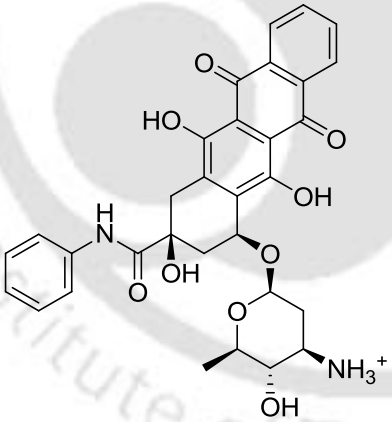
Table 4.11: Twenty potential inhibitors screened from secondary similarity screening using ZINC database considering initial dataset from NCI diversity set and *in vitro* inhibitors derived from THL based on the difference in predicted free energy of binding according to AutoDock. Number of conformations in cluster with best energy top hits is given in square brackets.

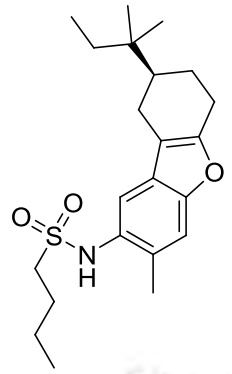
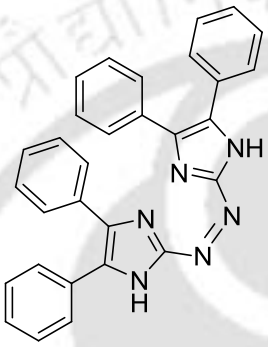
S. No.	Molecule ZINC DB ID	Free Energy of Binding (kcal/mol)			Residues forming hydrogen bond
		Rv3802c	HuMGL	Difference	
1	 ZINC43860875	-10.97 (40)	-7.19 (17)	-3.78	Lys100[O], Lys100[NH], Gly293[NH]
2	 ZINC28262919	-9.43 (56)	-6.83 (31)	-2.60	Trp84[O], Ser175[HG], His299[O], His299[HE2]

3	 ZINC26726377	-10.52 (48)	-8.00 (14)	-2.52	Trp84[O], Lys100[HZ3], Ser175[HG], His299[O], His299[HE2]
4	 ZINC34801822	-11.07 (40)	-8.59 (22)	-2.48	Lys100[O], Lys100[NH], Gly293[O]
5	 ZINC53243116_01	-8.4 (77)	-6.13 (31)	-2.27	Thr83[O], Thr83[OG1], Ser175[HG],
6	 ZINC28264663	-9.68 (62)	-7.45 (22)	-2.23	Trp84[O], His299[O]

7	 ZINC17949075_04	-10.54 (83)	-8.33 (35)	-2.21	Lys100[O], Tyr302[NH]
8	 ZINC38138545_01	-11.55 (71)	-9.38 (10)	-2.17	Tyr302[NH], Trp84[O], Lys100[HZ2]
9	 ZINC01697795	-9.23 (86)	-7.28 (47)	-1.95	Ala300[O], Ser175[HG], His299[NE2], Thr83[OG1], His299[HE2]
10	 ZINC38138543_01	-11.60 (48)	-9.68 (9)	-1.92	Tyr302[NH], Trp84[O], Lys100[HZ2]
11	 ZINC28707406	-9.08 (80)	-7.32 (17)	-1.76	Trp84[O], Gly293[O]

12	 ZINC17377922_02	-9.89 (94)	-8.16 (52)	-1.73	-
13	 ZINC42835607	-9.53 (78)	-7.96 (8)	-1.57	-
14	 ZINC00844080_01	-9.69 (41)	-8.24 (32)	-1.45	-
15	 ZINC35685388_01	-9.26 (85)	-7.91 (40)	-1.35	Ala303[NH]

16	 ZINC42764873_01	-8.98 (41)	-7.70 (17)	-1.28	Thr83[O], Lys100[O]
17	 ZINC11635001_01	-9.36 (96)	-8.11 (42)	-1.25	Ala303[NH]
18	 ZINC26470122_01	-12.00 (43)	-10.76 (10)	-1.24	Glu85[OE2], His299[HE2], Gly293[O]

19	 ZINC11635011_01	-9.5 (88)	-8.27 (35)	-1.23	Ala303[NH]
20	 ZINC17377922	-9.74 (91)	-8.6 (45)	-1.14	Lys100[O]
Substrate	p-nitrophenyl butyrate	-6.39 (99)	-7.45 (19)	1.06	Lys100[O], Tyr302[NH]
Inhibitor	Orlistat	-8.13 (6)	-7.64 (2)	-0.49	Trp84[O], Ser175[HG], His299[O], His299[HE2]
Inhibitor	ZYH	-9.41 (54)	-11.14 (46)	1.73	Lys100[O], Ala300[O]

[a] Molecule at different protonation or tautomeric state.

The hydrogen bond interaction mode of the best hit is shown in Fig. 4.18 which is energetically favourable and may provide effective inhibition. Best hit ZINC43860875 forms hydrogen bond interactions with Lys100 and Gly293. Lys100 is also forming a charge-π interaction with the aromatic ring of the ZINC43860875. The best hit was also stabilized by the hydrophobic residues of the cavity mainly by Lys100 and Ala300 through aromatic-aliphatic hydrophobic interaction.

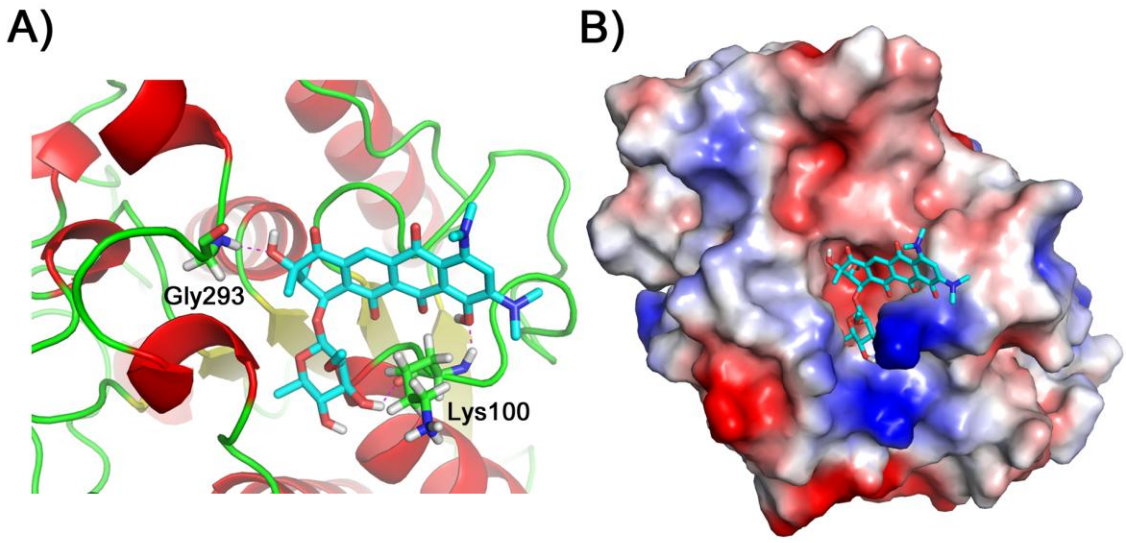


Figure 4.18: A) The interaction mode of best hit ZINC43860875 with Rv3802c as suggested by AutoDock. B) The best hit ZINC43860875 bound well in the active site of Rv3802c. This figure was produced using PyMOL.

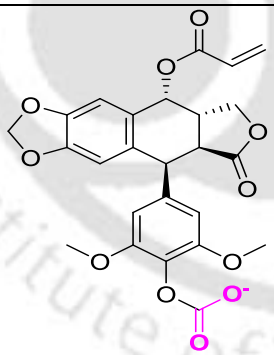
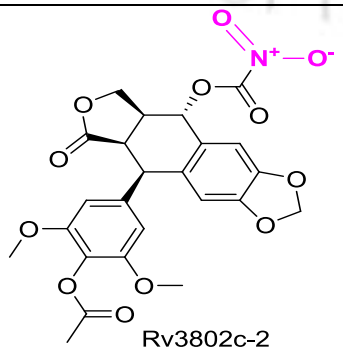
Also, the top hits were bound at the active site of Rv3802c and share most of the conserved residues at contact region of 4 Å, which are Thr83, Trp84, Glu85, Phe98, Lys100, Ala101, Leu102, Lys105, Phe174, Ser175, Gly292, Gly293, H299, Ala300, Met301, Tyr302 and Ala303. It is noteworthy to mention that the top hits were bound at the vicinity of catalytic residues Ser175 and His299. Key differences in contact residues were observed in Lys105 and Gly293 with exceptions in *M. smegmatis* MSMEG_6394 which has Asparagine and Serine respectively. Significant differences in the contact residues were observed, notably Thr83, Trp84 and Lys105 whose counterparts in human MGL are Alanine, Glycine, and Glutamate respectively in the structurally aligned region. The binding mode of top hits was visually inspected and it suggests that the interactions of top hits with the key residues of active site pocket were clustered around the substrate binding pocket. The mode of inhibition is expected to be competitive.

Drugs targeting both cell wall and lipid metabolism can be used with DOTS (directly observed treatment, short-course) therapy (Frieden and Munsiff, 2005) which may prevent the pathogen enter dormancy, may reduce the therapy time and may be active on natural variant strains of *M. tuberculosis*.

Design of potential inhibitors against Rv3802c:

SBDD provided insights on the binding mode of interaction between the top hits with Rv0183 which aid to design new molecules with improved predicted free energy of binding. The visual inspection of binding mode of the best hits concurrent with contact footprinting analysis revealed that the modification of the best hits through inclusion of functional groups that may favor the ionic interactions between the best hits and Rv3802c. The detailed information of new molecules is summarized in Table 4.12 and Figure 4.19. The NH_2, NO_2 OH and CH_2OH have been incorporated in the methyl group of phenyl ring. As expected, the incorporation of functional groups improvised the free energy of binding and slight change in conformation of binding mode of the new molecules. The study provides insight on the design and development of lead molecules against tuberculosis.

Table 4.12: The new molecules with better free energy of binding designed based on SBDD studies against Rv3802c. Number of conformations in the largest cluster is given in parenthesis. Hydrogen bond interactions were predicted with AutoDockTools. Atoms of the residues involved in hydrogen bonds have been indicated in square brackets. The inclusion/replacement of functional group on the top hits is depicted with magenta color.

S. No.	Molecule*	Free Energy of Binding (kcal/mol)			Residues forming hydrogen bond
		Rv3802c	Human MGL	Difference	
1.	 Rv3802c-1	-14.63 (48)	-6.86 (13)	-7.77	Trp84[O], Lys100[HZ3], Ser175[HG], His299[O]
2.	 Rv3802c-2	-10.73 (54)	-7.60 (15)	-3.13	Glu85[OE2]

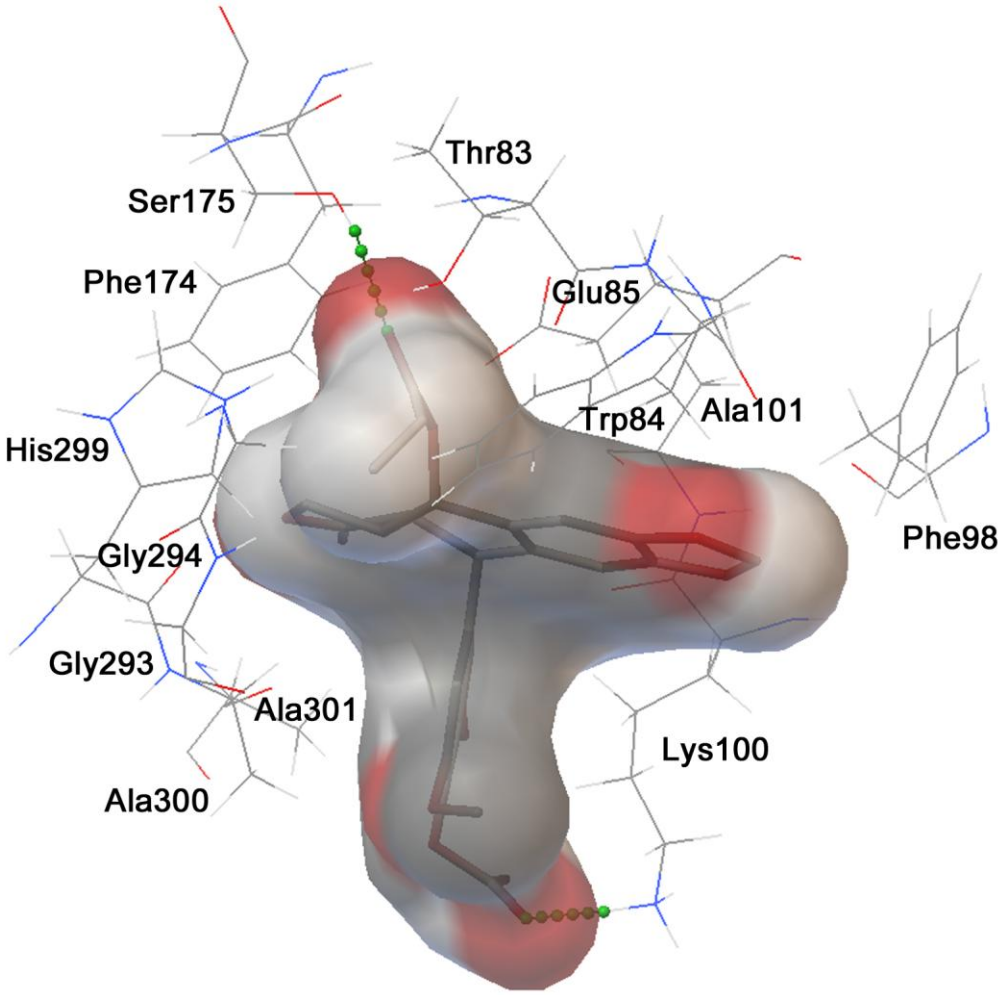
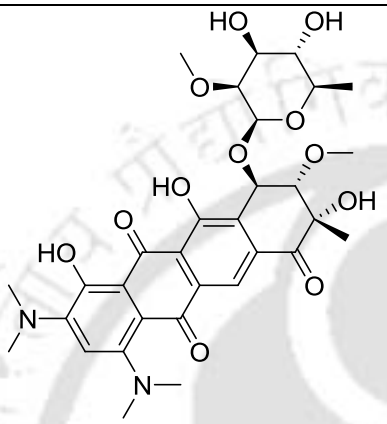
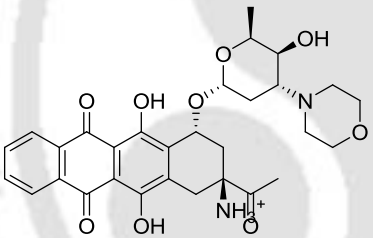
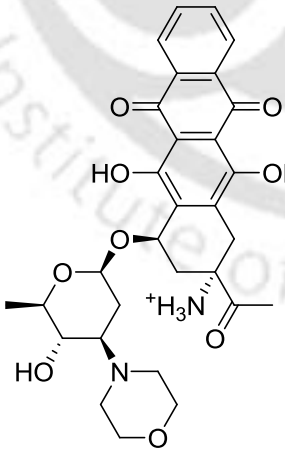


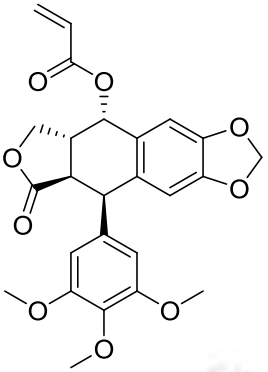
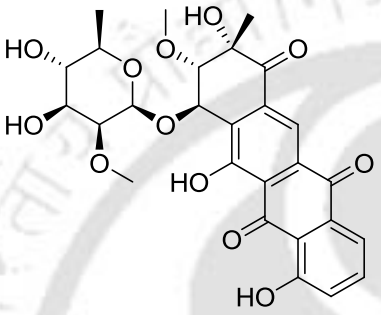
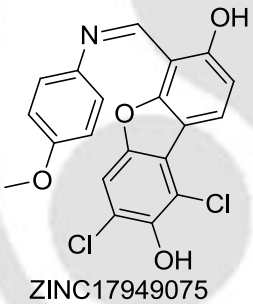
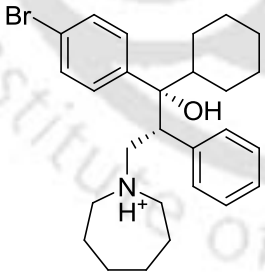
Figure 4.19: Binding mode of new designed molecule Rv3802c-1 in the active site pocket of Rv3802c.

4.4.3. Identification of Potential Dual Inhibitors

Multi-targeting approach is to target various biological processes notably cell wall and host lipid catabolism of *M. tuberculosis* simultaneously against the drugs targets Rv0183 and Rv3802c with their respective prioritized top hits (Rv0183: 117; Rv3802c: 172). Stringent docking of Rv0183 top hits were carried out against Rv3802c and vice versa. The dual inhibitors were screened based on the difference in free energy of binding of both the drug targets against human MGL (Table 4.13). The best hit ZINC43860875 is identified with difference of ΔG of -3.61 and -3.85 kcal/mol for Rv0183 and Rv3802c, respectively, and predicted ΔG of -7.19 kcal/mol for human MGL (Fig. 4.20).

Table 4.13: Potential dual inhibitors identified from multi-targeting strategy based on the difference in predicted free energy of binding (ΔG) according to AutoDock. Number of conformations in the largest cluster of top hits is given in parenthesis.

S. No.	ZINC ID	Free Energy of Binding (kcal/mol)			Rv0183-HuMGL	Rv3802c-HuMGL
		Human MGL	Rv0183	Rv3802c		
1	 ZINC43860875	-7.19 (17)	-10.80 (44)	-10.97 (40)	-3.61	-3.78
2	 ZINC59340279	-6.78 (17)	-11.31 (46)	-9.82 (48)	-4.53	-3.04
3	 ZINC59340278	-6.99 (18)	-10.12 (42)	-9.79 (42)	-3.13	-2.80

4	 ZINC28262919	-6.83 (31)	-8.79 (53)	-9.43 (56)	-2.00	-2.60
5	 ZINC34801822	-8.59 (22)	-11.20 (40)	-11.10 (40)	-2.65	-2.48
6	 ZINC17949075	-8.33 (35)	-11.62 (86)	-10.54 (83)	-3.29	-2.21
7	 ZINC27244815	-8.13 (13)	-10.54 (55)	-10.17 (62)	-2.41	-2.04

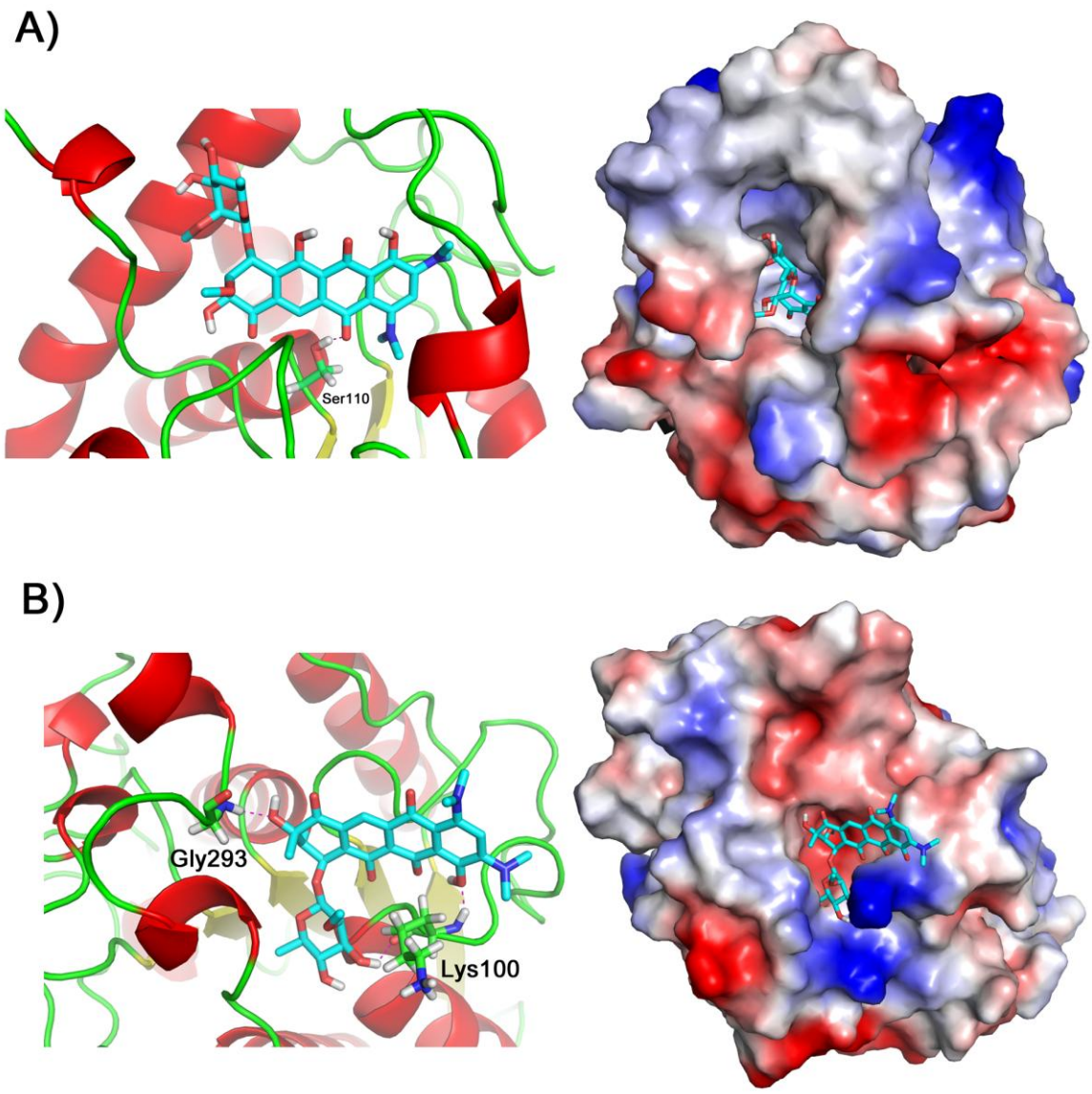


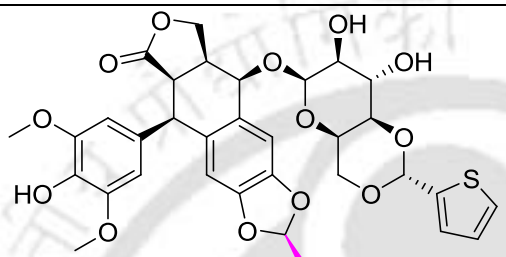
Figure 4.20: The best hit ZINC43860875 bound well in both the respective active sites of the identified drug targets A) Rv0183 and B) Rv3802c.

Design of new dual potential inhibitors to combat tuberculosis

The new molecules designed for drug targets Rv0183 and Rv3802c were docked against each other as well as the modification of functional groups of identified dual inhibitors were also carried out to identify new dual inhibitors with improved free energy of binding. The molecule Rv0183-1 was identified as new dual inhibitor against both Rv0183 and Rv3802c with free energy of binding -18.08 and -16.14 kcal/mol respectively which is summarized in Table 4.14 and Figure 4.21. The

incorporation of NH_2 group improvised the binding affinity of the new molecules. The study provides insight on the design and development of lead molecules against tuberculosis.

Table 4.14: The new molecules with better free energy of binding designed based on SBDD studies. Number of conformations in the largest cluster is given in parenthesis.

S. No.	ZINC ID	Free Energy of Binding (kcal/mol)			Rv0183-HuMGL	Rv3802c-HuMGL
		Human MGL	Rv0183	Rv3802c		
1	 Rv0183-1	-11.60 (24)	-18.08 (69)	-16.14 (63)	-6.48	-4.54

* the inclusion/replacement of functional group on the top hits are depicted with magenta color.

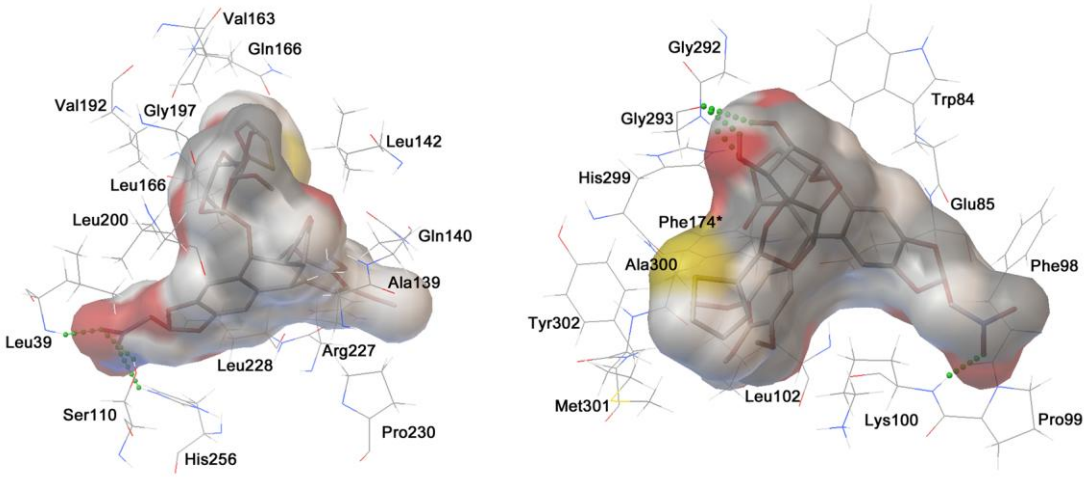


Figure 4.21: Binding mode of designed molecule Rv0183-1 as dual inhibitor in the active site pocket of Rv0183 and Rv3802c. * represents Phe174 of Rv3802c is behind the dual inhibitor.

4.5. Conclusions

The current study targeted cell wall and lipid metabolism of *M. tuberculosis* as starting arsenal to combat tuberculosis. Knowledge about the cell wall and lipid metabolism of *M. tuberculosis* can lead to resilience against tuberculosis. Rv0183 and Rv3802c were identified as potential drug targets. ZINC08763234 and ZINC28262919 were discovered as potential inhibitor which might be specific towards mycobacterial MGL (Rv0183) and essential cell

wall lipase (Rv3802c) respectively. ZINC43860875 was discovered as potential dual inhibitor targeting mycolic acid synthesis pathway and host lipid catabolism simultaneously. Contact footprinting analysis aid us to design new molecules with better predicted free energy of binding against the drug targets that might facilitate the design of new anti-tubercular leads. Experimental validation of the identified potential inhibitors by the TB research community should validate the therapeutic utility of identified targets against Rv0183 and Rv3802c.

Chapter V

LIPASES AS DRUG TARGETS TO TREAT LEISHMANIASIS

Chapter V

Lipases as Drug Targets to Treat Leishmaniasis

5.1. Abstract

Leishmaniasis is a neglected tropical disease caused by several species of *Leishmania*, which is one of the successful opportunistic pathogen. *Leishmania* being lipid scavenging pathogen, the rational approach is to target its lipid metabolism especially lipid catabolizing lipases. The LdLip3 lipase is considered as drug target based on its physiological importance for the pathogen. Implementing computational and experimental approaches, the chapter reports four new anti-leishmanial agents. The screened molecules had anti-leishmanial activity as well as negligible toxicity towards mouse macrophages. New molecules was designed based on the identified anti-leishmanial agents with better predicted free energy of binding which could be validated by the research community to validate further the therapeutic utility of the molecules to treat leishmaniasis.

5.2. Introduction

Leishmaniasis is an emerging disease with about 12 million infected people worldwide (World Health Organization, 2010). Several species and sub-species of *Leishmania*, an intra-macrophage obligate parasite causes clinically heterogeneous leishmaniasis. The fatal visceral leishmaniasis is linked to *Leishmania donovani*, *L. infantum* and *L. chagasi* while *L. donovani* is particularly rampant in India (Desjeux, 2004; Davies, 2003). Leishmaniasis is widespread in 98 countries with 350 million people at risk. Unfortunately, leishmaniasis has been underreported for several decades and still has a neglected status (Clem, 2000; Alvar *et al.*, 2012). The currently available therapeutics is associated with high toxicity, high cost and low efficacy (Table 5.1) (Alvar *et al.*, 2008; Chakravarty, 2010). Emergence of HIV coinfection, drug-resistance and disease spreading worldwide worsens the current scenario and urges the need for novel drugs to combat leishmaniasis. Few research organizations stress on the control measures along with the identification of new drug targets and drugs (Chawla and Madhubala, 2010; Shukla, 2000).

Table 5.1: Current drugs for the treatment of leishmaniasis and their disadvantages.

Molecule	Disadvantages*
Pentostam	Intravenous/ intramuscular infusion; Phlebotoxic; Gastro-intestinal disturbances
Amphotericin B	Intravenous/ intramuscular infusion; Overdose leads to cardio-respiratory arrest; Nephrotoxicity; Hepatotoxicity
Aminosidine	Intramuscular injection; Poorly absorbed after oral administration
Pentamidine	Leukopenia; Hypotension; Nephrotoxicity; Hepatitis; Seizures; Hypoglycemia; Bronchospasm; Effective on specific forms of cutaneous leishmaniasis
Miltefosine	Teratogenicity

* several sub-species of *Leishmania* have already gained drug resistance

Leishmania is a facultative lipid scavenging pathogen with dimorphic life cycle. The parasite is transmitted into mammalian host by female phlebotomine sand flies and it relies highly on its lipid metabolism to survive in different milieu of the hosts (Chappuis *et al.*, 2007) (Fig. 5.1). The lipid metabolism of pathogen plays a vital role in energy storage and pathogenesis (Smith *et al.*, 2012). Additional roles of the lipid metabolism have been implicated in membrane organization and dynamics, membrane trafficking and cellular signaling (Melo and Dvorak, 2012). The systemic metabolic analysis in *Leishmania major* revealed that (i) most of the intracellular reactions participate in lipid metabolism and (ii) 38% of the essential genes are associated with

lipid metabolism (McConville and Naderer, 2011). Several studies have also highlighted that lipid metabolism was elevated in amastigotes when compared with promastigotes (Rosenzweig *et al.*, 2008; Naderer and McConville, 2008; Tielens and Hellemond, 2009). Miltefosine, a FDA-approved anti-leishmanial drug is believed to target lipid metabolism of *Leishmania* (Rakotomanga *et al.*, 2007). Lipid metabolism of *Leishmania* gained lot of attention in the last decade and of our interest for the discovery of new drug targets (Opperdoes and Coombs, 2007).

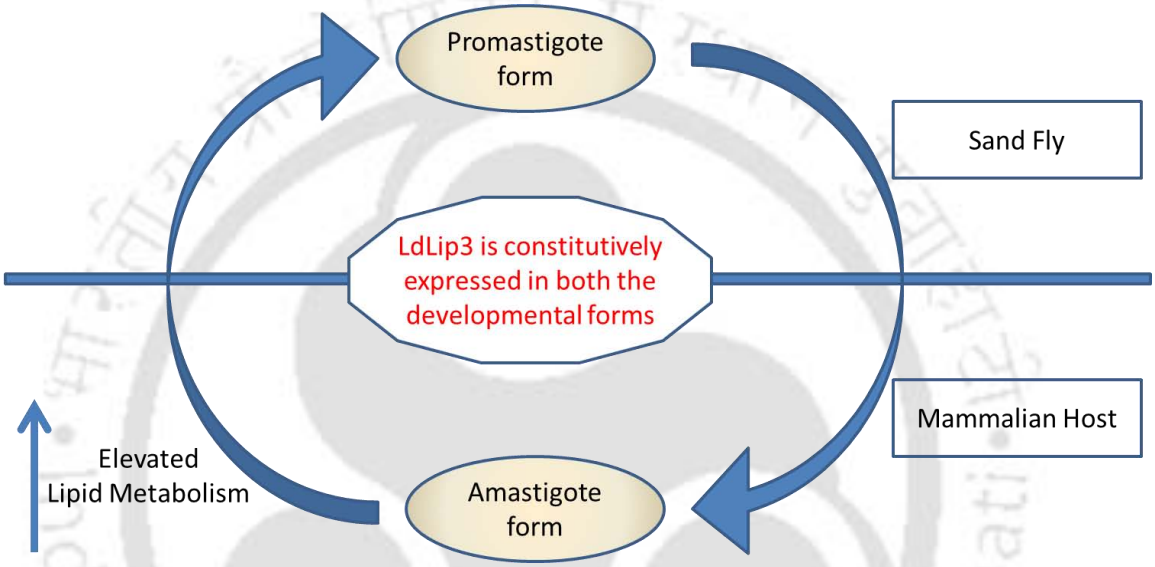


Figure 5.1: LdLip3 lipase is constitutively expressed in both amastigote and promastigote forms while lipid metabolism is elevated in amastigote form.

Lipases are one of the key enzymes in lipid metabolism which act on triglycerides and phospholipids of the host. These macromolecules are abundant in mammalian phagolysosomes where lipid-scavenging pathogens reside (McConville *et al.*, 2007). Lipases of pathogens have multitude of functions in many pathophysiological processes including virulence, transmission, life cycle development, modulation of host lipids and host immune responses (Stehr *et al.*, 2004; Russell *et al.*, 2010; Dhouib *et al.*, 2010; West *et al.*, 2011). Lipases from human pathogens notably *Mycobacterium tuberculosis*, *Candida albicans*, *Cryptococcus neoformans* and *Malassezia* species are proven for their role in virulence and survival (Singh *et al.*, 2010, Park *et al.*, 2010).

LdLip3, a lipase from *L. donovani*, is believed to participate in key biological processes that include host lipid degradation, alteration of phagolysosomes membrane and structural

remodeling of membrane lipids for the survival and infection of pathogen. LdLip3 also favors the parasite growth and development in the host (Shakarian *et al.*, 2010). Hence, LdLip3 could be an attractive drug target to combat leishmaniasis. Targeting lipases from lipid metabolism of the pathogen is considered as therapeutic strategy in the present study and attempts were made to identify novel potential inhibitors that might specifically kill leishmanial cells. The three dimensional (3D) model of LdLip3 was predicted which was further utilized for the structure-based virtual screening. The similarity-based virtual screening of representative chemical scaffolds was opted for the identification of potential inhibitors (Tanrikulu *et al.*, 2013; Shen *et al.*, 2012; Oprea and Matter, 2004). Furthermore, the anti-leishmanial activity of the screened hits was validated with experimental assay, resulting in the identification of four novel lead-like molecules. The identified novel inhibitors could serve as potential candidates for drug development against leishmaniasis and even for infectious diseases caused by Trypanosomatida protozoans.

5.3. Materials and Methods

5.3.1. Homology Modeling

The protein sequence of LdLip3 (UniProt ID: D7P7V3) and its representative homologs from leishmanial species were collected from UniProtKB (www.uniprot.org). The retrieved sequences were subjected to multiple sequence alignment with ClustalW and subsequently analyzed with ESPript to identify functionally important residues, motifs and protein domains (Thompson *et al.*, 1994; Gouet *et al.*, 1999) (Table 5.2).

The appropriate template for homology modeling was identified by searching the closest structural homologs using 'BLASTP' program against Protein Data Bank (PDB) database. Lipase from *Rhizomucor miehei* (Rm-TGL) in its inhibitor bound conformation (PDB id: 4TGL) was used as template for structure modeling based on its sequence and functional homology (Derewenda *et al.*, 1992). The first 20 residues of LdLip3 were not modeled due to lack of structural information on its template Rm-TGL and as the region are far from the active site. Initially hundred models were generated with MODELLER9v7 which were then ranked on the basis of their discrete optimized protein energy (DOPE) and restraint violation (molpdf) scores (Marti-Renom *et al.*, 2000; Shen and Sali, 2006). The best model having reasonable DOPE and

molpdf scores with acceptable statistics from Ramachandran plot was validated with PROCHECK (Laskowski *et al.,* 1993) and Errat plot (Colovos and Yeates, 1993) from structure validation package using NIH SAVES server. Protein fold of LdLip3 model was evaluated using knowledge-based energy profile of ProSA program (Sippl, 1993; Wiederstein and Sippl, 2007). The best model of LdLip3 from structure validation studies was analyzed for stability using GROMOS96 43a1 force field under periodic boundary conditions in GROMACS4.5.5 package (Berendsen *et al.,* 1995; van der Spoel *et al.,* 2005; Hess *et al.,* 2008). The system was solvated with simple point charge (SPC) water model. The net charge of system was neutralized by replacing water molecules that are at least 3.50 angstroms (Å) from the protein surface with four chlorine ions. The system was energy minimized to remove bad contacts using steepest descent algorithm. The solvent was equilibrated for 500 ps in NVT ensemble by restraining the solute atoms through a harmonic force constant of 1000 kJ mol^{-1} nm^{-2}. Then the system was equilibrated by restraining all the bonds for 500 ps in NPT ensemble which was followed by 500 ps of equilibration without any restraints. Production run was carried out for 20 ns with NPT ensemble using 2 fs of integration time. Simulations were carried out with velocity rescaling thermostat at 300K in which protein and non-protein atoms were coupled to separate temperature coupling baths and pressure was controlled at 1 atm using Parrinello-Rahman barostat. The linear constraint solver (LINCS) algorithm was used to constrain the bonds involving hydrogen atoms (Hess *et al.,* 1997; Hess *et al.,* 2008). Particle Mesh Ewald (PME) summation method was used for calculating the long-range electrostatic interactions with 12 Å cut-off. Conformations of LdLip3 model from the production run were analyzed as time-dependent function to check the conformational stability of energy minimized LdLip3 model in the solvent system.

5.3.2. Structure-based Virtual Screening

Virtual Screening: The success of structure-based virtual screening in the identification and enrichment of lead and lead-like molecules are well reported (Lyne, 2002; Kellenberger, 2007; Nicola *et al.,* 2007; Khan *et al.,* 2009; Leung *et al.,* 2011). Virtual screening of small molecules was performed using the energy minimized model of LdLip3 with AutoDock4.2 using lamarckian genetic algorithm (LGA) (Morris, 1998; Cosconati *et al.,* 2010). NCI diversity set II containing 1880 structurally diverse chemical molecules (2654 molecular structures) was used to identify representative molecular scaffolds. The choice of NCI diversity set greatly reduces the

computational time as it contains only the representative molecules of wide and diverse scaffolds. LdLip3 model was preprocessed by adding hydrogen atoms to all the residues followed by addition of Gasteiger-Marsili charges. Then non-polar hydrogen atoms were merged into their respective heavy atoms and atom types were fixed using AutoDockTools (Morris *et al.,* 2009). All the small molecules were similarly processed using 'prepare_ligand4' python script available with AutoDockTools distribution. For docking simulations, a grid box of 90 x 90 x 90 points was constructed with grid spacing of 0.375 Å keeping catalytic serine (Ser168) at the center of grid box that covers the entire active site of LdLip3 model. Grid maps were generated for all the atom types present in the grid box along with electrostatic and desolvation maps using the AutoGrid utility. The whole docking procedure was then repeated with the closest human structural ortholog - monoglyceride lipase (MGL), to find the potential inhibitors that could be effective towards LdLip3 than human MGL (PDB id: 3HJU) (Bertrand *et al.,* 2010). Molecular docking of NCI diversity set on LdLip3 was performed with an initial population size of 300 for 20 independent LGA runs. In each run, the best individual from each generation was propagated to the next generation and remaining docking parameters were set to default. Each molecule was clustered on the basis of positional root-mean-square deviation (RMSD) between the docked positions to predict the optimal binding of the molecule with LdLip3.

The ten top hits were chosen as initial data set based upon their difference in binding free energy better than -2 kcal/mol and interactions of the top hits with leishmania-specific residues of LdLip3 to carry out similarity screening. Similar molecules (tanimoto coefficient > 0.6) of 10 top hits from ZINC and its natural product database resulted in 22919 molecules including isomers which were collected and preprocessed (Irwin and Shoichet, 2005; Irwin *et al.,* 2012). Virtual screening process was repeated with these similar molecules on both LdLip3 model and human MGL with an initial population size of 300 and 20 independent LGA runs with stringent evaluations of 30000000. The final top 20 hits were re-docked with 100 independent LGA runs to increase the reliability of the docking predictions (reliability test). The best conformation of the largest cluster from the reliability test was considered as optimal conformation of the final top 20 hits with LdLip3. Docking results were visually analyzed with AutoDockTools and PyMOL to understand the docking poses and interactions between the potential inhibitors and LdLip3 (DeLano, 2002).

Cluster Footprinting Analysis: In order to improve the reliability of the docking studies, clustering analysis of the docked poses has been carried out on the screened 20 top hits. AuPosSOM (Automated analysis of Poses using self-organizing map) was implemented to identify correct conformations of the hits as well as to discriminate the active hits from the non-active hits from docking studies (Bouvier *et al.,* 2010). AuPosSOM classify the docked poses using drug-protein contact descriptors with SOM followed by unsupervised cluster analysis. Default parameters of AuPosSOM were used while the four top hits which have been chosen to carry out *in vivo* studies is considered as active compounds. The top hits were clustered according to the similar binding mode to LdLip3 and generated a Newick tree file which was visualized by Dendroscope. The contact analysis was investigated to determine key interactions and complementarity with the binding site of LdLip3. The studies have been calibrated in the tuberculosis chapter for efficient clustering.

5.3.3. Validation Studies of Structure based Drug Discovery with in vivo assay

Parasites, cell lines and chemicals: The *Leishmania donovani* (BHU-1081) promastigote culture was obtained from Prof. Shyam Sundar, Banaras Hindu University, India and cultivated in M199 liquid media supplemented with 15% heat-inactivated fetal bovine serum (FBS), 100 U/ml penicillin and 100 µg/ml streptomycin. Mouse macrophage cell line J774A.1 used in the study was procured from "National Centre for Cell Science" (NCCS), Pune, India and cultured in RPMI 1640 media supplemented with 10% heat inactivated FBS, 2mM glutamine, 100 U/ml penicillin and 100 µg/ml streptomycin. The identified top hits were procured from MolPort, Russia. All the chemicals used in the experiments were of the highest grade procured from Sigma-Aldrich and Merck.

Anti-leishmanial activity assay: Anti-leishmanial effect of the procured top hits were investigated by optimized MTT (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide) assay (Mosmann, 1983; Hussain *et al.,* 1993; Saudagar and Dubey, 2011). For the assay, test molecules (5 mM in DMSO) serially diluted in media were added to 96-well culture plates with a concentration range of 0.75 – 100 µm in a total volume of 200 µl. Mouse macrophage cell line J774A.1 were cultured in complete Dulbecco's modified Eagle's medium (DMEM), 2000 cells/well were seeded and were allowed to adhere overnight. 0.5% DMSO served as a negative

control, whereas 20% DMSO served as positive control for both parasite and human cell lines. Exponentially growing *L. donovani* promastigotes (2×10^6 cells/ml) were freshly transformed in M199 media and incubated with graded concentrations of compounds at 25°C for 48 h. After treatment, cells were centrifuged and subsequently washed with 1× ice-cold phosphate-buffered saline (PBS) and incubated in fresh M199 media with 15% heat-inactivated FBS and 0.5 mg/ml MTT for 4 h. The cell viability was determined by measuring the optical density at 570 nm for reduced formazan, compared to untreated cells. The IC_{50} value for each identified novel inhibitors was calculated by plotting percent cell viability vs. concentration.

5.4. Results and Discussion

5.4.1. Homology Modeling

The non-availability of the experimentally determined structure for lipases from pathogenic source hinders the use of structure-based drug discovery approach for several deadly infectious diseases. Homology modeling of LdLip3 was carried out for the first time to address the same with the aid of lipase from *Rhizomucor miehei* in its inhibitor bound conformation (PDB id: 4TGL, Resolution: 2.60 Å) to identify potential anti-leishmanial molecules. Sequence analysis revealed that LdLip3 is highly conserved throughout the leishmanial species. Conservation between LdLip3 and Rm-TGL is satisfactory in terms of fold and functional residues which is necessary for the reliability of the LdLip3 model. The catalytic residues were conserved and correspond to Ser168, Asp232 and His288 in LdLip3 (depicted as star in Fig. 5.2).

Table 5.2: LdLip3 and its homologs used in the multiple sequence alignment study.

UniProtKB Accession	Name	Organism	Identity (%)*	Similarity (%)*
D7P7V3	LdLip3	*L. donovoni*		
A4I6H7	Li-Lip	*L. infantum*	99	99
Q4Q6I0	Lm-Lip	*L. major*	89	91
P19515	Rm-TGL	*R. miehei*	35	45

* identity and similarity of the corresponding protein against LdLip3.

The modeled structure confirmed that LdLip3 is a member of serine hydrolases with α/β fold (Fig. 5.3A). The 3D model of LdLip3 (modelled extends from 21 to 302 residues with 30% identity and 44% similarity towards Rm-TGL) was significantly conserved with leishmanial homologs and Rm-TGL. LdLip3 has α/β hydrolase domain with five-stranded parallel β-sheet

covered by two-stranded anti-parallel β-sheet at the N-terminus as shown in Fig. 5.3A. The nucleophile serine (Ser168) was found at the bend of the tight turn in the 'nucleophilic elbow' and other catalytic residues (Asp232 and His288) are positioned adjacent to each other within the active site. Taken together, structural features of LdLip3 are similar to that of Rm-TGL and other triacylglycerol lipases. Apart from conserved catalytic residues, strict conservation was observed between LdLip3 and Rm-TGL at the active site cavity (4 Å around catalytic residues) namely His167, Gly170, Ala172, Gly202, Pro204, Val234, Pro235, Val283, Asp287 and Tyr291 which might play a role in the enzymatic activity. The oxyanion hole which is involved in substrate binding and tetrahedral intermediate stabilization was identified as Ser102 and Val169 from sequence and structure analysis.

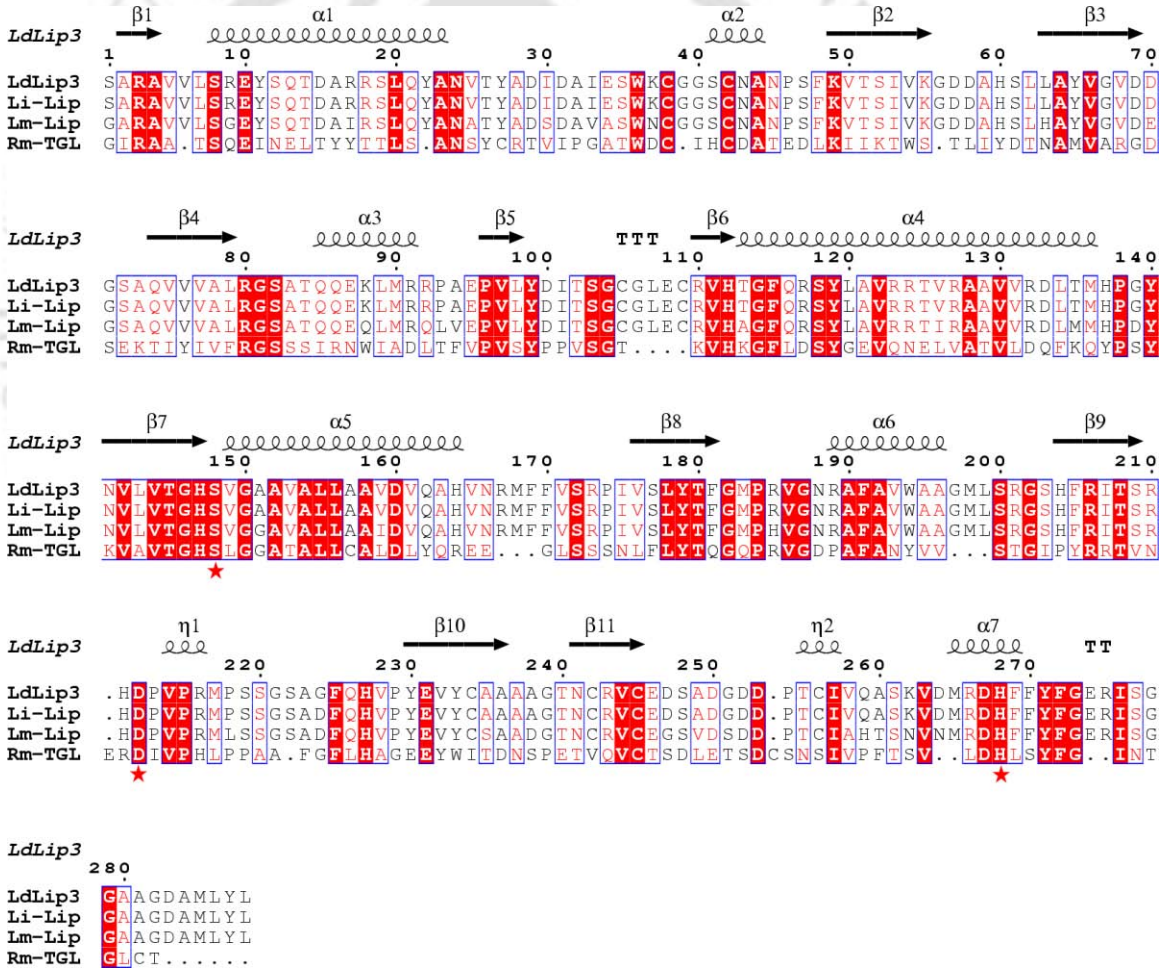


Figure 5.2: Multiple sequence alignment of LdLip3 with its homologs. The catalytic resides have been highlighted with star. Numbering of amino acids is denoted with respect to LdLip3. This figure was produced using ESPript.

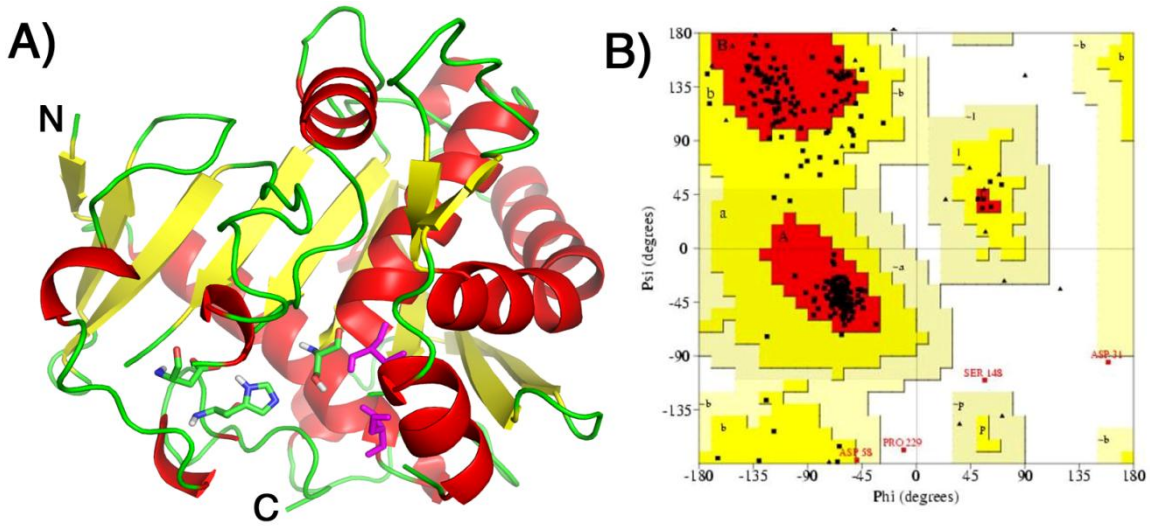


Figure 5.3: A) The 3D model of LdLip3 shown in cartoon representation. Catalytic residues and oxyanion hole (magenta) residues were shown in sticks. This figure was produced using PyMOL. B) Ramachandran plot of LdLip3 revealing the acceptable statistics of dihedral angle distribution of amino acids.

The energy minimized structure of LdLip3 model has acceptable statistics of backbone dihedral angle distribution of amino acids in Ramachandran plot with 86.6% in core, 12.2% in additionally allowed and 0.4% in generously allowed region as presented in Fig. 5.3B. Along with Asp51 and Pro249, the catalytic nucleophile, Ser168 was observed in disallowed region of Ramachandran plot, a typical feature found in most of the energy minimized serine lipases/esterases. G-factor of LdLip3 was -0.19 which also indicates acceptable statistics of distribution of phi and psi along with chi1, chi2 and chi3 angles (G-factor less than -0.5 is considered to be unusual). Errat plot of the final model of LdLip3 showed an overall score for structural quality of 81.319 with little steric hindrance between few amino acids (Fig. 5.4). As expected, Verify-3D also revealed that 79.86% of the amino acids in the current structure of LdLip3 have compatible 1D-3D score greater than 0.2. The LdLip3 model has Z-score of-5.97 in the range of native conformations of crystal structures and also comparable with Rm-TGL (Z-score: -7.01) with ProSA-web. The fold quality of LdLip3 from ProSA was comparable with experimentally determined Rm-TGL indicating the acceptability of the modelled LdLip3 structure (Fig. 5.5). Molecular dynamics (MD) was carried out to assess stability of the energy-minimized LdLip3 model to ensure its reliability for structure-based virtual screening. Time evolution of RMSD of the backbone Cα atoms of LdLip3 showed that it undergoes a significant change in the initial 5ns of simulations and converged with fluctuations less than 0.1 Å thus

indicating stable conformation of LdLip3 model (Fig. 5.6). MD simulations suggested that the energy-minimized LdLip3 model is satisfactory for virtual screening process.

Structural comparison of the energy minimized LdLip3 model with Rm-TGL on Cα-backbone atoms (161 out of 219 aligned atoms) shows overall RMSD of 0.176 Å. Comparison of residues in the catalytic site region (9 out of 12 Cα-backbone atoms around 4 Å of catalytic residues) was 0.337 and residues Arg38, Ser102, Arg111, Met285 and Arg286 in LdLip3 were observed in the catalytic site region. The leishmania-specific residues at the catalytic pocket have been utilized to screen potential leishmania-specific lipase inhibitors.

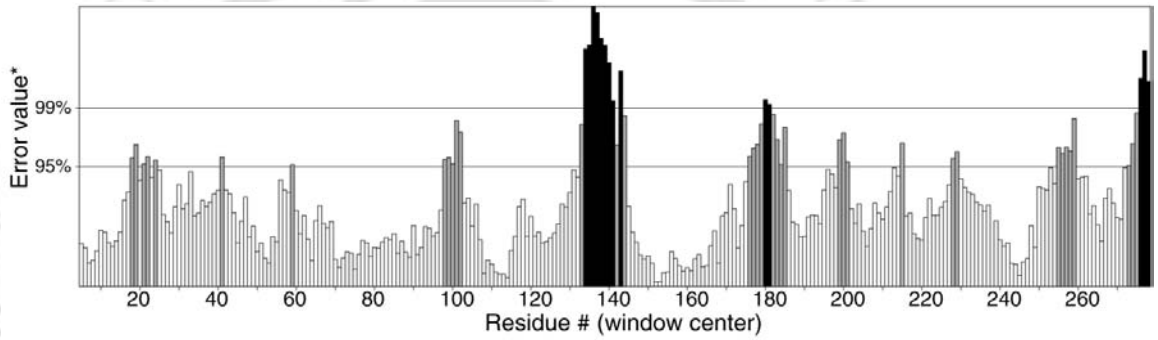


Figure 5.4: Errat plot of LdLip3 model signifying that the structure has structural quality of 8.39 indicating low steric hindrance.

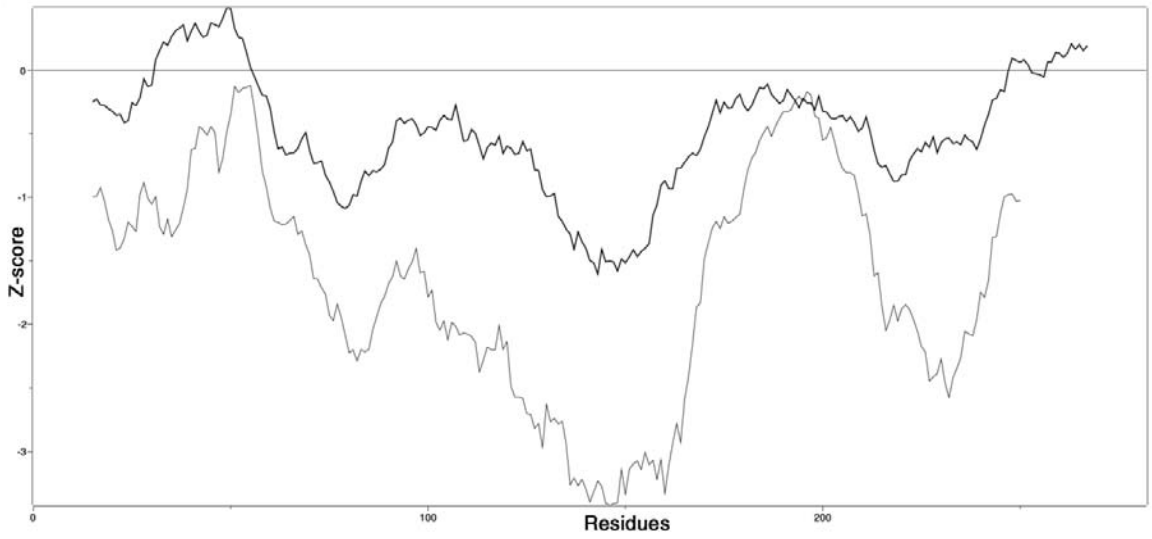


Figure 5.5: Comparative ProSA energy profile of LdLip3 and its template, Rm-TGL in thick and thin line respectively reflects that both have similar fold.

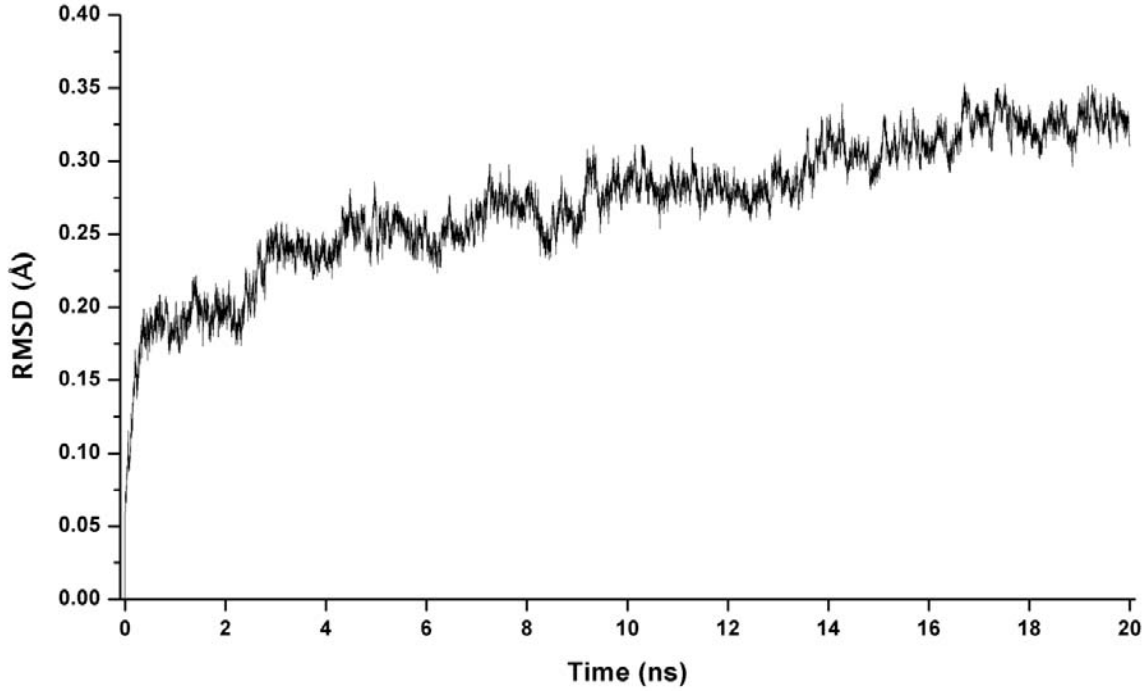


Figure 5.6: RMSD plot from the production run of 20ns from MD simulation reflects that the LdLip3 model has stable conformation.

5.4.2. Structure-based Virtual Screening

Structure-based virtual screening was used to identify potential inhibitors using LdLip3 modelled structure. The rigid docking protocol of AutoDock was implemented for virtual screening (Sousa *et al.*, 2006). In our earlier studies, docking protocol was validated for lipases with help of human MGL (Saravanan *et al.*, 2012). Human MGL was considered as the anti-target from the human proteome for the present study. Similarity-based virtual screening was carried out in a step-wise manner to identify potential inhibitors utilizing NCI diversity set followed by small molecules from ZINC database including natural products (Fig. 5.7).

Since the structural features required for its inhibition is not yet known, the active site of LdLip3 was considered to screen structurally diverse molecules (NCI diversity dataset II) to identify potential hits with different scaffolds. Screening of diversity set was initially carried out against human MGL to reduce the possibility of potential inhibitors inhibiting human MGL. The molecules with higher binding affinity towards LdLip3 than human MGL (based on the difference in predicted free energy of binding) were considered as the top hits of initial screening (Table 5.3). Subsequently, similarity screening of the ten top hits from initial screening was

carried out on ZINC and its natural product database to identify potential inhibitors that could be selective towards LdLip3. Top hits of computational studies were visually inspected and analyzed for all possible hydrogen bond and hydrophobic interactions with LdLip3 using PyMOl. The twenty top hits of similarity screening are listed in Table 5.4. The contact analysis of the top hits were studied with LdLip3 and compared with Lm-Lip, Rm-TGL and Human MGL. The non-conserved amino acids of LdLip3 in contact with the top hits may lead to the specific inhibition of LdLip3 (Table 5.5).

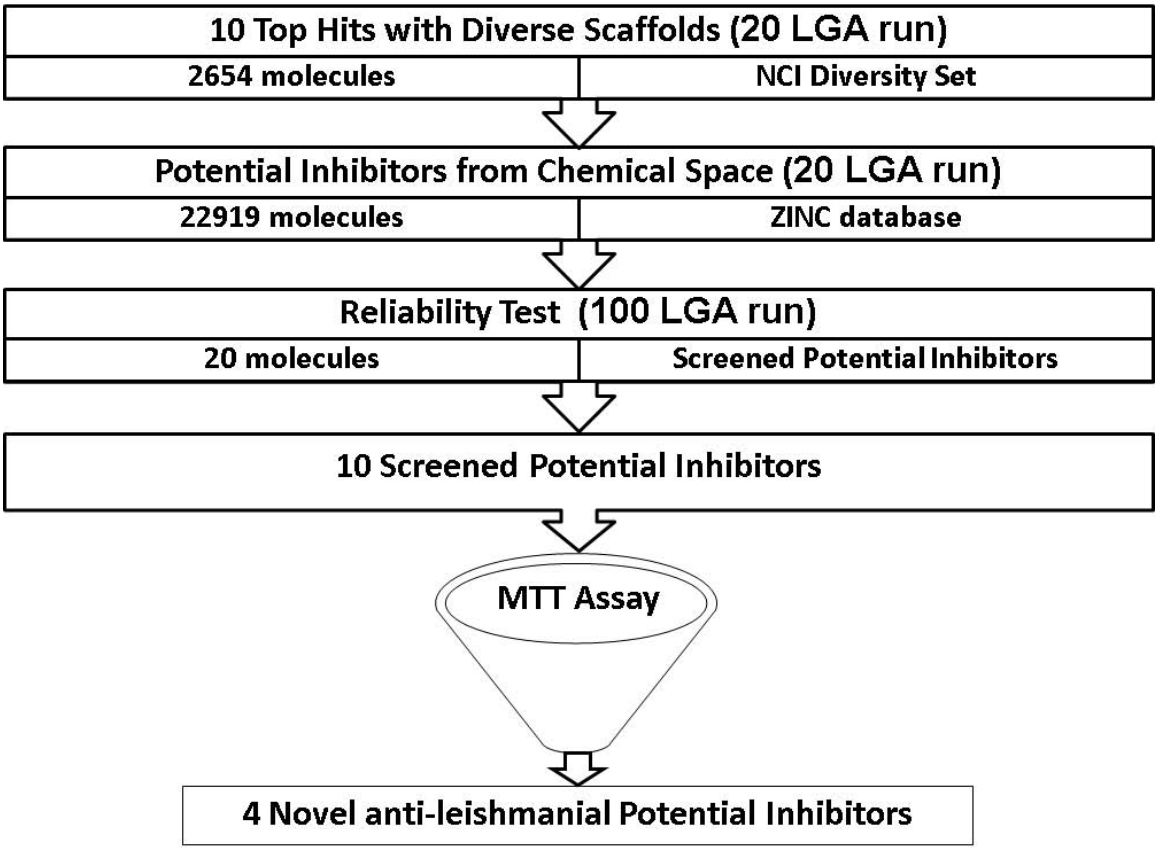
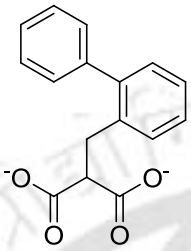
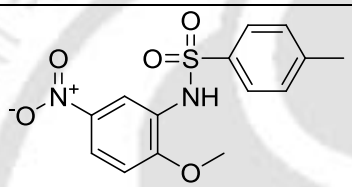
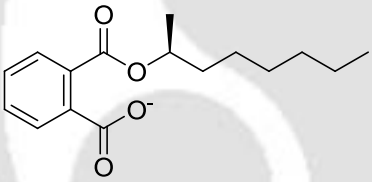
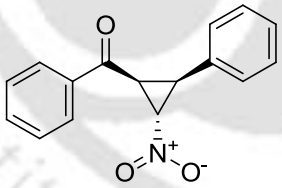
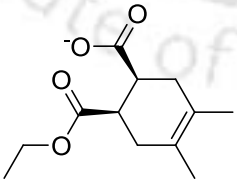
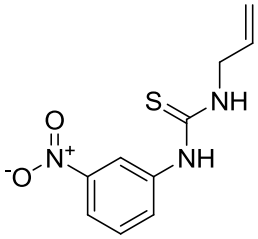
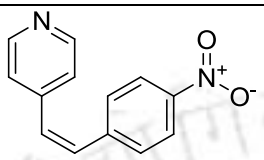
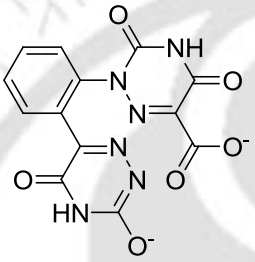
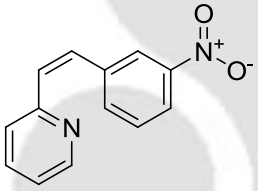
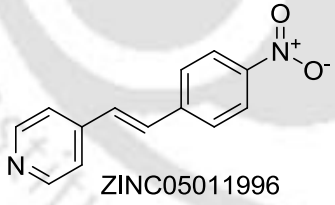


Figure 5.7: Flow chart of the similarity-based virtual screening used to identify the potential inhibitors that might specifically kill leishmanial cells.

Table 5.3: Initial virtual screening on NCI diversity set. Top 10 hits were selected based on the difference in free energy of binding to be greater than or equal to -2 kcal/mol.

S. No.	Molecule	Free Energy of Binding (kcal/mol)		
		LdLip3	Human MGL	Difference
1.	 ZINC01615121	-8.86	-6.39	-2.47
2.	 ZINC00304325	-9	-6.73	-2.27
3.	 ZINC01605775	-7.73	-5.54	-2.19
4.	 ZINC13154324	-8.4	-6.22	-2.18
5.	 ZINC05086225	-7.25	-5.1	-2.15

6.	 ZINC05137909	-8.18	-6.1	-2.08
7.	 ZINC05011996*	-7.97	-5.89	-2.08
8.	 ZINC29590292	-8.25	-6.18	-2.07
9.	 ZINC12340238	-8.05	-5.98	-2.07
10.	 ZINC05011996	-8.16	-6.18	-1.98

***** Molecule at different protonation or tautomeric state.

Table 5.4: The twenty top hits identified from similarity-based virtual screening. Number of conformations in largest cluster is given in parenthesis.

S. No.	Molecule	Free Energy of Binding (kcal/mol)		
		LdLip3	Human MGL	Difference
1.	ZINC04701128	-9.23 (51)	-6.00 (18)	-3.23
2.	ZINC04008765*	-8.83 (93)	-6.11 (25)	-2.72
3.	ZINC13409341	-8.33	-5.66	-2.67

		(87)	(29)	
4.	ZINC01821375*	-9.78 (61)	-7.18 (15)	-2.6
5.	ZINC13997735	-9.19 (100)	-6.64 (61)	-2.55
6.	ZINC06117316*	-8.60 (100)	-6.19 (15)	-2.41
7.	ZINC12653571*	-8.60 (84)	-6.43 (46)	-2.17
8.	ZINC47367575	-9.76 (64)	-7.60 (29)	-2.16
9.	ZINC13406344	-8.51 (91)	-6.38 (24)	-2.13
10.	ZINC18707267	-10.37 (50)	-8.35 (29)	-2.02
11.	ZINC00472657	-9.78 (30)	-7.70 (52)	-2.08
12.	ZINC01165396	-10.04 {41)	-7.94 (39)	-2.1
13.	ZINC01609336	-9.32 (40)	-7.38 (11)	-1.94
14.	ZINC02141591	-10.26 (22)	-7.55 (20)	-2.71
15.	ZINC02873758	-9.60 (23)	-7.49 (12)	-2.11
16.	ZINC03844930	-9.81 (32)	-6.93 (13)	-2.88
17.	ZINC04280995	-9.46 (33)	-6.33 (8)	-3.13
18.	ZINC05260149	-9.55 (48)	-7.30 (35)	-2.25
19.	ZINC07053328	-10.28 (22)	-8.22 (36)	-2.06
20.	ZINC12296526	-8.35 (37)	-7.54 (22)	-0.81

Contact footprinting through clustering analysis of docked poses from screened hits: Contact based clustering using the docking poses which has been investigated in earlier chapter can provide insight on the binding mode of small molecules with the protein of interest. The screened hits were clustered into 5 clusters which have been investigated at the level of sub-clusters to discriminate the contacts of each sub-cluster (Fig. 5.8). Clustering revealed that the top hits were clustered based on the possible interactions as well as the contacts with LdLip3 according to the orientation of the top hits at the active site of LDLip3. The sub-clusters observed within the clusters showed the minor variation in the mode of hydrogen bond interactions (Table 5.6).

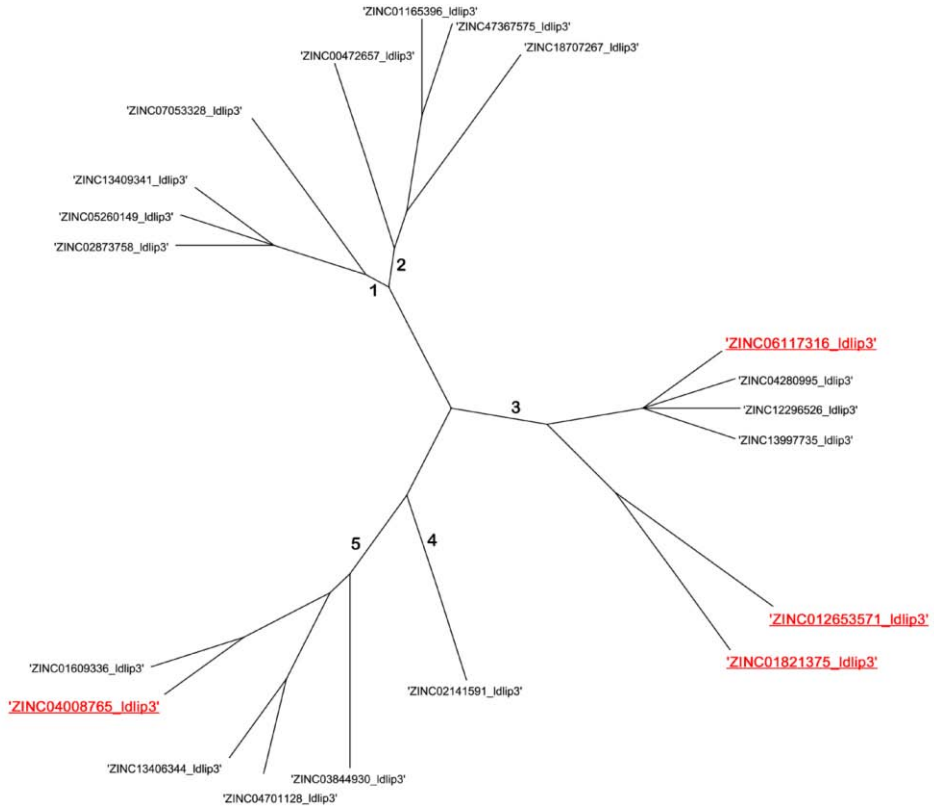


Figure 5.8: Tree representation of contact footprints clustering for 20 screened hits with LdLip3. Numbers on branches represent each of the five clusters with ZINC database accession of each screened hits. The top hits screened for *in vivo* assay are represented in bold.

Table 5.6: Conserved contacts of screened twenty top hits at the ligand binding pocket of LdLip3 within clusters and various sub-clusters in each cluster, and screened hits belonging to the sub-cluster.

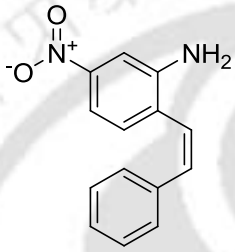
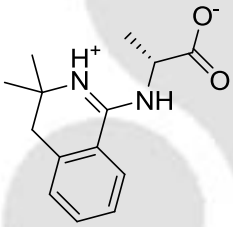
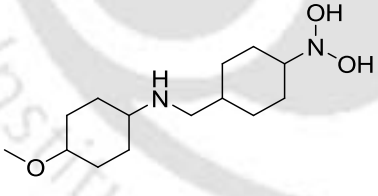
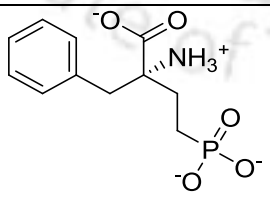
Cluster	Contact residues conserved within cluster	Hydrogen bond interactions conserved with LdLip3	Sub-cluster (Molecule ZINC DB IDs)	contact residues conserved within subcluster	Hydrogen bond interactions within subcluster
1	Lys108, Arg111, Arg112, Ala114, His132, Phe135, Ser168, Pro204, Val234, Met237, Pro238	Arg111[HE], Arg111[HH21]	1. [ZINC13409341, ZINC02873758, ZINC05260149] 2. [ZINC07053328]		- - - -
2	Lys108, Arg111, Arg112, Ala114, His132, Phe135, Pro204, Val234, Met237, Pro238		1. [ZINC00472657] 2. [ZINC01165396, ZINC47367575] 3. [ZINC18707267]	 Pro93	Arg111[HE], Arg111[HH21], Arg111[O] - - Arg111[HE], Arg111[HH21], Arg112[O]
3	Lys108, Arg111, Arg112, Met237	Arg111[HH21]	1. [ZINC04280995, ZINC06117316*, ZINC12296526, ZINC13997735] 2. [ZINC01821375*, ZINC12653571*]	 His112, Phe115, Pro184, Pro218	- - - - Arg111(HE) -
4	Lys108, Arg111, Arg112, Ala114, His132, Phe135, Ser168, Pro204, Val234, Met237, Pro238	Arg111[HE], Arg111[HH21]	1. [ZINC02141591]		-
5	Lys108, Arg111, Arg112, Phe135, Pro204, Val234, Met237, Pro238	Arg111[HE], Arg111[HH21]	1. [ZINC03844930] 2. [ZINC04701128, ZINC13406344] 3. [ZINC01609336, ZINC04008765*]	 Ala94, His112 Ala94, His112,	Arg112[HH11] - - - -

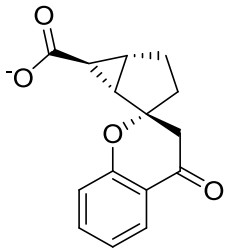
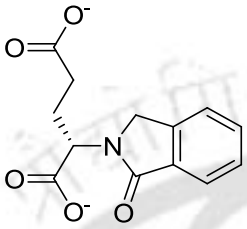
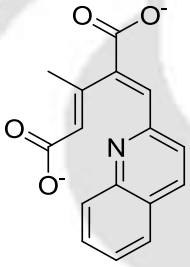
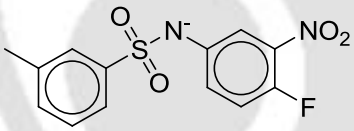
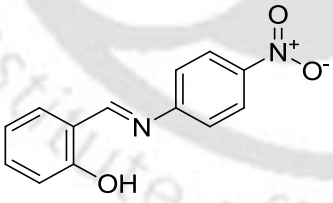
* indicates the potential inhibitors that were validated by MTT assay.

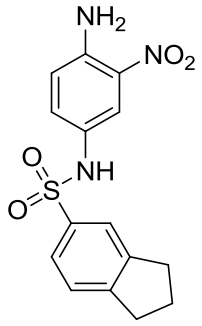
The screened hits from similarity screening were prioritized with set of customized parameters to increase the reliability of the docking studies: (a) number of conformations in the largest cluster for LdLip3 is set to be more than 50 and (b) difference between the free energy of binding

between ldlip3 and human MGL was set to be less than 2 kcal/mol. The screened potential inhibitors from SBDD are listed in Table 5.7.

Table 5.7: The screened potential inhibitors from SBDD study. Number of conformations of top hits in cluster with best energy of LdLip3 and human MGL is given in parenthesis. Hydrogen bond interactions were predicted with AutoDockTools. Atom of the residues involved in hydrogen bonds was indicated in the square brackets.

S. No.	Molecule	Free Energy of Binding (kcal/mol)			Residues forming hydrogen bond
		LdLip3	Human MGL	Difference	
1.	 ZINC04701128	-9.23 (51)	-6.00 (18)	-3.23	Arg111[HE], Arg111[HH21]
2.	 ZINC04008765*	-8.83 (93)	-6.11 (25)	-2.72	Arg111[HE], Arg111[HH21]
3.	 ZINC13409341	-8.33 (87)	-5.66 (29)	-2.67	Arg111[HE], Arg111[HH21]
4.	 ZINC01821375*	-9.78 (61)	-7.18 (15)	-2.6	Arg111[HE], Arg111[HH21]

5.	 ZINC13997735	-9.19 (100)	-6.64 (61)	-2.55	Arg111[HE], Arg111[HH21],
6.	 ZINC06117316*	-8.60 (100)	-6.19 (15)	-2.41	Arg111[HE], Arg111[HH21], Arg112[HH12], Arg112[HH22]
7.	 ZINC12653571*	-8.60 (84)	-6.43 (46)	-2.17	Arg111[HE], Arg111[HH21], Arg111[HH22]
8.	 ZINC47367575	-9.76 (64)	-7.60 (29)	-2.16	Arg111[HE], Arg111[HH21]
9.	 ZINC13406344	-8.51 (91)	-6.38 (24)	-2.13	Arg111[O], Arg111[HE], Arg111[HH21]

10.	 ZINC18707267	-10.37 (50)	-8.35 (29)	-2.02	Arg111[HE], Arg111[HH21], Arg112[O]
Substrate	4-lumbelliferyl stearate	-6.5 (9)	-8.9 (4)	2.40	nil
Inhibitor	Orlistat	-4 (7)	-7.64 (2)	-0.64	Arg111[2]
Inhibitor	ZYH	-7.27 (66)	-11.14 (46)	3.87	nil

* indicates the potential inhibitors which were validated by MTT assay.

Comparative analysis of the top hits with the substrate as well as the inhibitors of human MGL reveals that the top hits bind well with LdLip3 and could be effective against LdLip3. The differences in the 'lid' region which covers the catalytic triad between LdLip3 and human MGL gives rise to different binding modes of top hits. Difference in the residues comprising binding pockets of the top hits in LdLip3 were exploited in the identification of potential inhibitors that makes them specific towards LdLip3 than human MGL. Best hit ZINC04701128 forms two hydrogen bond interactions with side-chain atoms of Arg111 (Fig. 5.9). Most of the top hits form hydrogen bonds with Arg111 and Arg112. The interaction mode of the screened molecules which were selected for experimental validation has been shown in Fig. 5.10. The polar moiety of top hits interacts well with the charge amino acids of LdLip3 notably Lys108, Arg111, Arg112 which is not present in human MGL. The ring moiety of top hits helps in stabilizing them in the hydrophobic pocket of LdLip3 notably Ala114, Phe115, Val169, Val234, Met237, Val283.

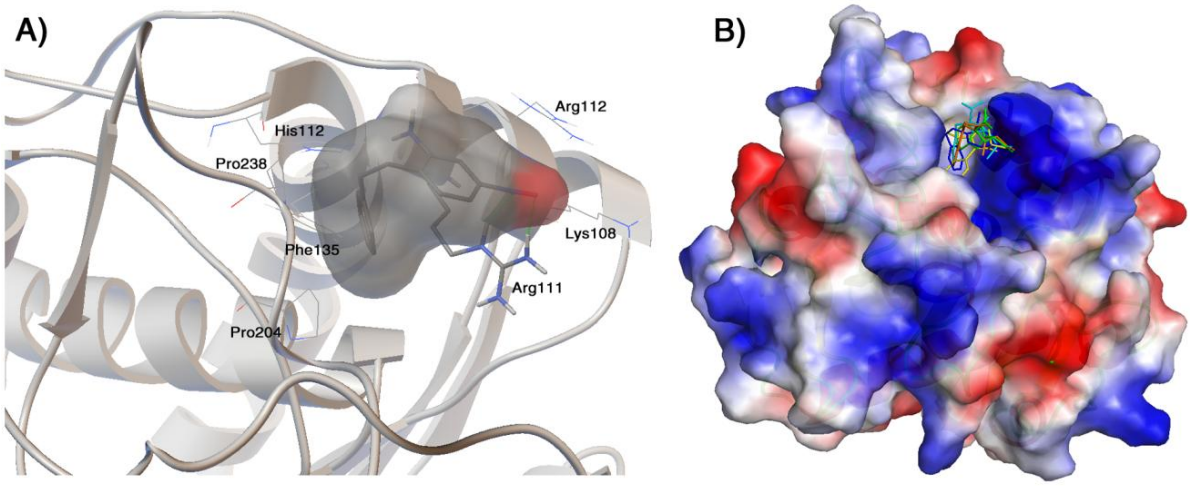


Figure 5.9: A) Hydrogen bond interactions of LdLip3 and the best hit ZINC04701128 predicted by docking studies. B) Top Hits interact with the leishmania-specific residue Arg111 in the active site pocket of LdLip3.

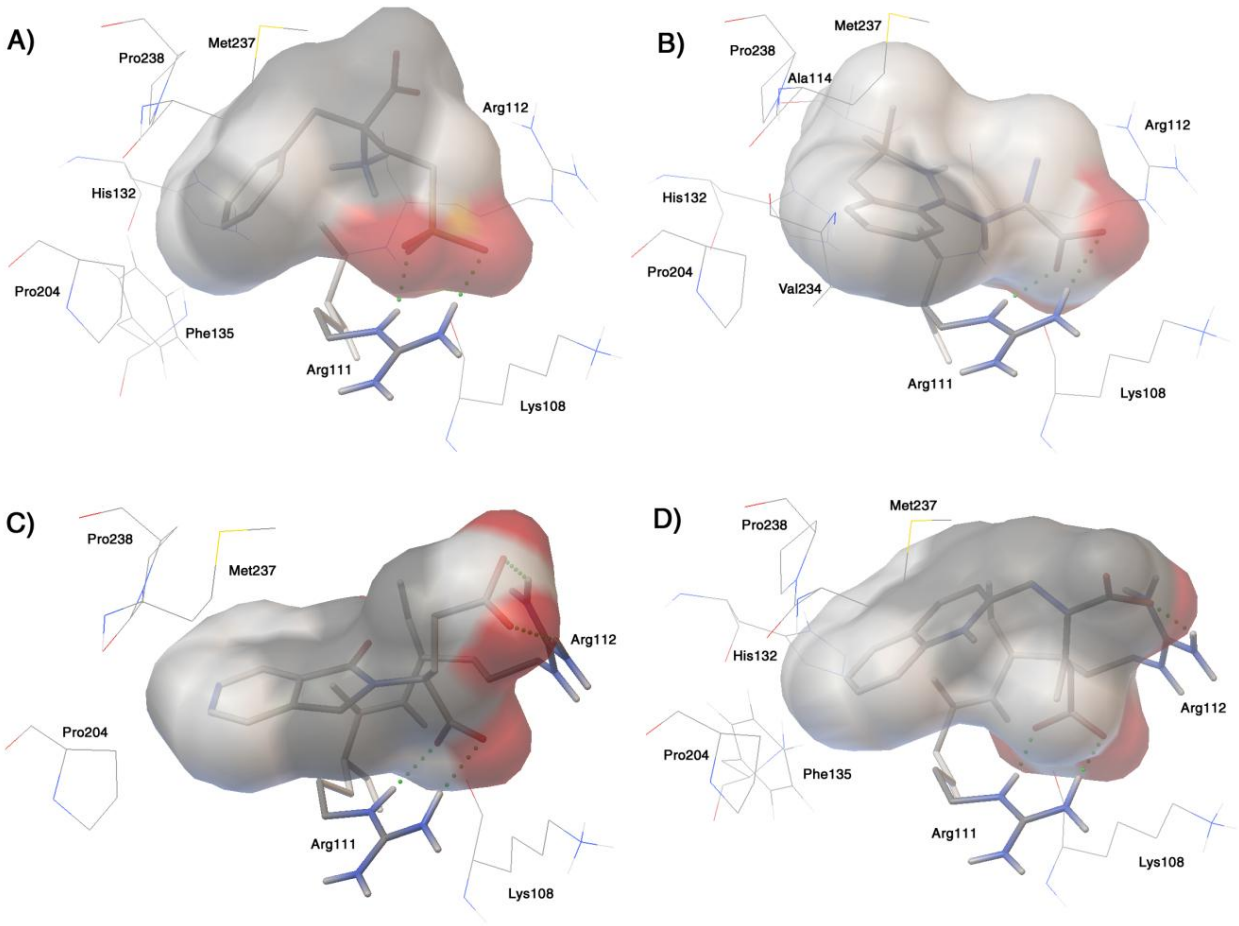


Figure 5.10: Hydrogen bond interactions in the active site pocket of LdLip3 with the screened molecules A) ZINC01821375, B) ZINC04008765 C) ZINC06117316 and D) ZINC12653571.

5.4.4. Anti-leishmanial activity of the screened molecules

The *in vitro* anti-leishmanial activity of commercially available potential inhibitors from the top 20 hits was evaluated by MTT assay (Fig. 5.11). A dose-dependent death profile of *Leishmania donovani* (BHU-1081strain) promastigote cells was observed with all the tested molecules in the range of 0.75-100 µM. The molecules ZINC01821375 and ZINC06117316 were more effective as anti-leishmanial compared to ZINC04008765 and ZINC12653571. The IC_{50} values for each of the molecule were calculated by plotting viable promastigotes in percentage *vs* concentration. The IC_{50} were found to be 5.2 ± 1.8 µM for ZINC01821375, 13.1 ± 2.6 µM for ZINC04008765, 9.4 ± 2.6 µM for ZINC06117316 and 17.3 ± 3.1 µM for ZINC12653571. In addition, the effect of identified novel inhibitors on mouse macrophage cell line J774A.1 was also tested. No significant toxicity was shown on the macrophage cells even at higher concentration up to 100 µM, indicating the identified novel inhibitors selectively inhibit leishmanial cells.

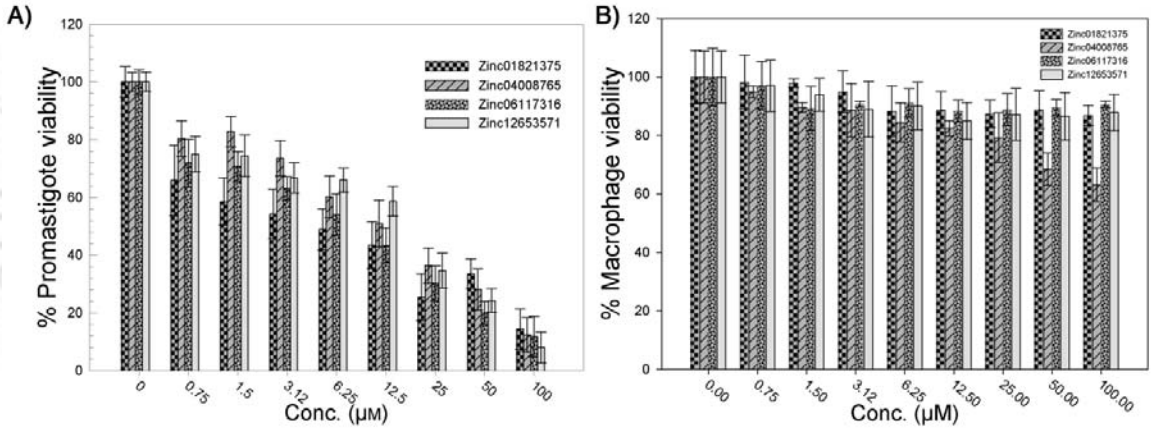
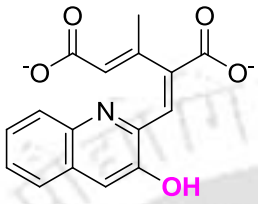
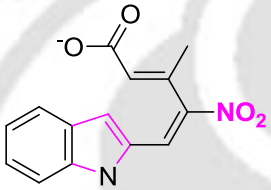
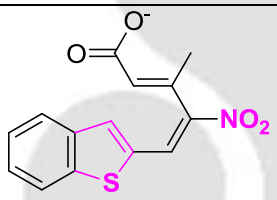
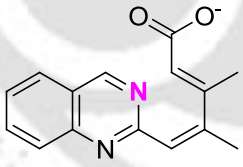


Figure 5.11: MTT cell proliferation assay showing promising anti-leishmanial activity of the commercial available potential inhibitors in a dose-dependent manner. A) Effect of the identified novel inhibitors on *L. donovani* promastigotes. B) Effect of the identified novel inhibitors on mouse macrophage cell line. Values are mean ± SD of four determinations

5.4.5. Design of new molecules

Contact footprinting analysis from SBDD provided insights on the binding mode of interaction between the best hits with LdLip3 which aid to design new molecules with improved potential anti-leishmanial agents. The visual inspection of binding mode of the validated leishmanial agents concurrent with contact footprinting analysis revealed that the possibility of functional groups that favor the ionic interactions as well as fills the interior pocket which the best hits failed to do so.

Table 5.8: The new molecules designed based on SBDD and *in vivo* studies. Number of conformations of top hits in cluster with best energy of LdLip3 and human MGL is given in parenthesis. Hydrogen bond interactions were predicted with AutoDockTools. Atom of the residues involved in hydrogen bonds was indicated in the square brackets.

S. No.	Molecule*	Free Energy of Binding (kcal/mol)			Residues forming hydrogen bond
		LdLip3	Human MGL	Difference	
1.	 LdLip3-1	-11.67 (68)	-8.36 (32)	-3.31	Arg111[O], Arg111[HH21], Arg112[HH11]
2.	 LdLip3-2	-9.93 (78)	-7.35 (24)	-2.58	Arg111[O], Arg111[HH21], Arg112[HH12]
3.	 LdLip3-3	-9.57 (85)	-7.13 (18)	-2.44	Arg111[HE], Arg111[HH21], Arg112[HH12]
4.	 LdLip3-4	-9.17 (78)	-6.88 (13)	-2.29	Arg111[HE], Arg111[HH21], Arg112[HH12]

* the inclusion/replacement of functional group on the top hits are depicted with magenta color.

The detailed information of new molecules is summarized in Table 5.8 and Figure 5.12. The NH_2, NO_2 and OH have been incorporated in the quinolone ring in the 'meta' position relative to the nitrogen atom. As expected, the incorporation of OH group improvised the free energy of binding of the new molecule (increase in free energy of binding by -3.07 kcal/mol). The change in conformation of binding mode was observed with the incorporation of hydroxy group which lead to the formation of hydrogen bond with leishmania-specific residue Arg112. Replacement

of the six-membered ring with five-membered ring as well as carboxyl group with NO_2 group avoided steric clashes and improved the efficiency of the new molecules in terms of predicted free energy of binding. The study provides insight on the design of lead molecules which could be effective new anti-leishmanial agents.

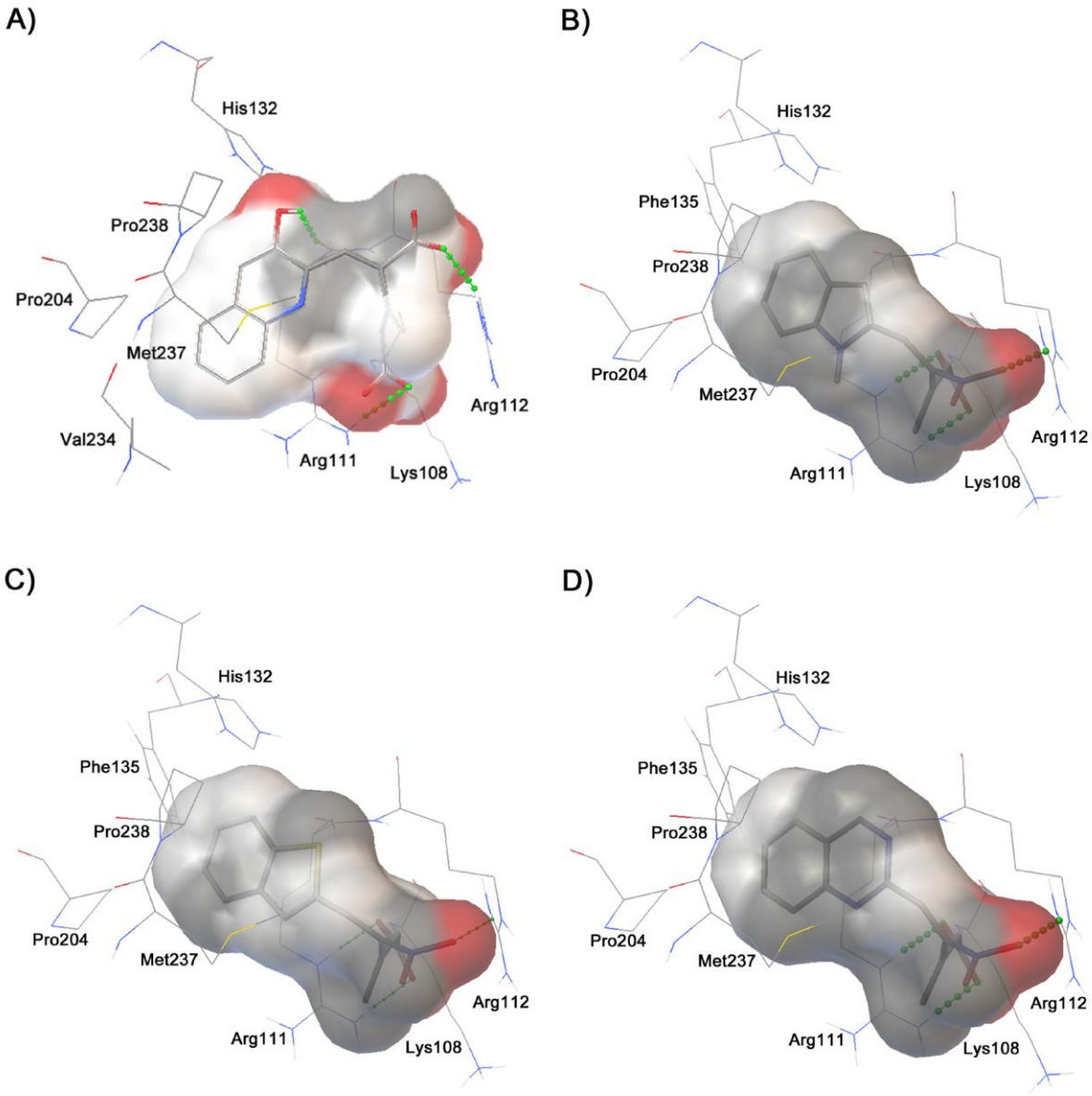


Figure 5.12: Hydrogen bond interactions in the active site pocket of LdLip3 with the designed molecules A) LdLip3-1, B) LdLip3-2, C) LdLip3-3 and D) LdLip3-4.

5.5. Conclusions

The present chapter reports LdLip3 lipase as a target enzyme towards leishmaniasis therapeutics. The modeled structure of LdLip3 has been deposited in protein model database with accession number PM0078869 (www.mi.caspur.it/PMDB). The discovery of novel inhibitors of LdLip3 lipase by structure-based (similarity-based) drug discovery approach was carried out. Our results suggest that the potential inhibitors bind at the entrance of LdLip3 active site, notably forms hydrogen bond with the leishmanial-specific residue Arg111. The interaction mode of potential inhibitors restrains the critical 'Lid' region which might result in the inhibition of LdLip3. The anti-leishmanial activity of the four screened molecules (ZINC01821375, ZINC06117316, ZINC04008765 and ZINC12653571) was validated through MTT assay in promastigote leishmanial cells. The screened molecules had anti-leishmanial activity with IC_{50} (% viable promastigotes *vs* concentration) of 5.2 ± 1.8 µM, 13.1 ± 2.6 µM, 9.4 ± 2.6 µM and 17.3 ± 3.1 µM for ZINC01821375, ZINC04008765, ZINC06117316 and ZINC12653571 respectively. The screened molecules showed negligible toxicity in mouse macrophage and therefore, highly selective towards the pathogen. Integration of computational and experimental analysis provides valuable insights on the identified new potential inhibitors to facilitate the design of novel and more effective leads to combat leishmaniasis.

Chapter VI

CONCLUSIONS

6.1. Summary

Lipases are ubiquitous in nature and efficiently catalyze hydrolysis reactions. The applications of lipases as biocatalysts are ever-growing which qualifies them as emerging commercial enzyme. Unfortunately very few such lipases are available for commercial applications. A consummate understanding of lipases with appropriate classification and structural information can therefore promote exploring their hidden potential for commercial applications. The thesis intends use bioinformatics in an integrated manner is a rational step forward in this direction. The first chapter discusses about the introduction and significance of lipases. The literature review emphasizes on the understanding and exploiting the potential of lipases. The second chapter pinpoints the need and importance of a database for lipases. A database has been created for triacylglycerol lipases which provide comprehensive information about the sequence, structure and bibliographic information which can be utilized for further research on lipases. The established framework of our database will be instrumental to understand lipases and can be extended to other enzymes. An effort also has been made to classify them into sub-families with bio-physiochemical perspective. The sub-family classification successfully reflected and correlated phylogeny, biochemical and functional features of lipases. From our database, lipases were chosen as representative for industrially, biologically and pharmacologically important families to venture into their applicability in the following chapters of the thesis. In the third chapter, lipases from *Staphylococcus* and *Pseudomonas* source were chosen for their industrial emulation by increasing their structural stability. Sequence and structural analysis on candidate lipases led to the identification of mutagenesis hotspots as well as to perform *in silico* mutagenesis. The mutants were investigated on atomic level to shed light on the structural features with the help of molecular dynamics simulations. The chapter provided a rational platform to emulate their structural stability facilitating its industrial use. Our database and classification schema can also be credited for bringing out the physiological importance of lipases in pathogens. This has been explored in the fourth and fifth chapter. Pathogens that rely highly on its lipid metabolism have been considered for the current research. Effort has been made to target the lipases to combat infectious diseases notably tuberculosis and leishmaniasis. Structure-based drug discovery approach has been implemented with integrated computational approach to

investigate the drug targets and to design new molecules with therapeutic potentials. Lipases Rv0183 and Rv3802c from *Mycobacterium tuberculosis* were identified as drug targets for the current study. Similarity-based virtual screening was the method of choice to identify potential inhibitors. The study came up with potential dual inhibitors targeting lipases against tuberculosis for the first time. The computational approach of tuberculosis has been extended to identify potential inhibitors to combat leishmaniasis. Cell culture studies validated the potentiality of the screened hits as well as identified small molecules with anti-leishmanial activity and negligible toxicity. This framework can thus provide insights to design new lead molecules for the therapeutics of tuberculosis and leishmaniasis, even other diseases. To summarize the research, it can be said that the thesis has ventured deeper into the existing understating and application of lipases as well as laid a framework for further exploration of lipases by the creation of the database. The work has also shown the utility of classifying lipases as it could correlate their physicochemical and functional properties. Most importantly, the thesis has laid a platform to design lipases for industrial applications through *in silico* studies and to develop lead molecules against lipases to combat pathogenic diseases.

6.2. Future Perspective

The thesis provides valuable insights on the structure, function and dynamics of lipases. It is a rational step toward understanding and exploiting lipases for their industrial potential. Furthermore, the potential of lipases as drug target to combat infectious diseases was also presented. The future perspective on the current research can be explored in several directions. To list few:

- The framework of TLDB can be used to create a database for all known lipases, even extend to other enzyme classes.
- Identified *in silico* mutants of the representative lipases can be investigated in detail with experimental techniques.
- Apart from the studied representative lipases, investigations can be carried out on lipases that have been brought out by sub-family classification study.
- The SBDD framework could be implemented to identify potential inhibitors for key lipases for other infectious diseases caused by lipid scavenging pathogens.
- The respective potential inhibitors identified for drug targets Rv0183 and Rv3802c as well as LdLip3 can be validated further which might validate the therapeutic utility of the drug targets.
- The potential dual inhibitors for lipases was presented, its therapeutic utility could be validated experimentally that might open up new horizon for better human life.
- The new molecules designed based on SBDD approach can be synthesized and taken to identify new leads thereby new drugs for tuberculosis and leishmaniasis.

BIBLIOGRAPHY

BIBLIOGRAPHY

1. Accelrys DS Visualizer, v2017347; Accelrys Software Inc: San Diego, CA, 2009.
2. Adindla S, Guruprasad L. Sequence analysis corresponding to the PPE and PE proteins in Mycobacterium tuberculosis and other genomes. J Biosci. 2003; 28:169-179.
3. Alberghina L, Schmid RD, Verger R, editors. Lipases: structure, mechanism and genetic engineering. Weinheim:VCH, 1991.
4. Ali YB, Verger R, Abousalham A. Lipases or Esterases: Does it really matter? Toward a new Bio-Physico-Chemical Classification. Methods in Molecular Biology. 2012; 861:31-51.
5. Alvar J, Aparicio P, Aseffa A, Den Boer M, Cañavate C, Dedet JP, Gradoni L, Ter Horst R, López-Vélez R, Moreno J. The relationship between leishmaniasis and AIDS: the second 10 years. Clin Microbiol Rev. 2008; 21:334-359.
6. Alvar J, Ve´lez ID, Bern C, Herrero M, Desjeux P, Cano J, Jannin J, den Boer M. WHO Leishmaniasis Control Team Leishmaniasis Worldwide and Global Estimates of Its Incidence. PLoS One. 2012; 75:e35671.
7. Armstrong JA, Hart PD. Response of cultured macrophages to *Mycobacterium tuberculosis*, with observations on fusion of lysosomes with phagosomes J Exp Med, 1971; 134:713-740.
8. Arpigny JL, Jaeger KE. Bacterial lipolytic enzymes: classification and properties. Biochem. J. 1999; 343:177-183.
9. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. (Webserver) 2009; 37:202-208.
10. Bailey TL, Gribskov M. Methods and statistics for combining motif match scores. J Comput Biol. 1998; 5:211-221.
11. Bailey TL, Williams N, Misleh C, Li WW. MEME: discovering and analyzing DNA and protein sequence motifs. Nucleic Acids Res. (Webserver) 2006; 34:369-373.
12. Balcao VM, Paiva AL, Malcata FX. Bioreactors with immobilized lipases: state of the art. Enzyme and Microbial Technology. 1996; 18(6):392-416.
13. Barry CE. Interpreting cell wall 'virulence factors' of *Mycobacterium tuberculosis*. Trends Microbiol. 2001; 9:237-241.
14. Beatty WL, Russell DG. Identification of mycobacterial surface proteins released into subcellular compartments of infected macrophages. Infect Immun. 2000; 68:6997-7002.
15. Benjamin S, Pandey A. Candida rugosa lipases: molecular biology and versatility in biotechnology. Yeast. 1998; 14:1069–1087.
16. Berendsen HJC, Postma JPM, van Gunsteren WF, Hermans J. Interaction models for water in relation to protein hydration. *In*: Pullman B eds Intermolecular forces. Reidel, Dordrecht. 1981; pp. 331-342.
17. Berendsen HJC, Postma JPM, Vangunsteren WF, Dinola A, Haak JR. Molecular dynamics with coupling to an external bath. J Chem Phys. 1984; 81:3684-3690.
18. Berendsen HJC, Vanderspoel D, Vandrunen R. GROMACS: A message-passing parallel molecular dynamics implementation. Comput. Phys. Commun. 1995; 91:43-56.
19. Berry M, Kon OM. Multidrug- and extensively drug-resistant tuberculosis: an emerging threat. Eur Respir Rev. 2009; 18:195-197.

20. Bertrand T, Auge F, Houtmann J, Rak A, Vallee F, Mikol V, Berne PF, Michot N, Cheuret D, Hoornaert C, Mathieu M. Structural basis for human monoglyceride lipase inhibition. J Mol Biol. 2010; 396:663-673.
21. Bornscheuer UT, Bessler C, Srinivas R, Hari Krishna S. Optimizing lipases and related enzymes for efficient application. Trends in Biotechnology. 2002; 20:433-437.
22. Boshoff HI, Barry CE. Is the mycobacterial cell wall a hopeless drug target for latent tuberculosis? Drug Discov Today: Disease Mechanisms 2006; 3:237-245.
23. Bouvier G, Evrard-Todeschi N, Girault JP, Bertho G. Automatic clustering of docking poses in virtual screening process using self-organizing map. Bioinformatics. 2010; 26:53-60.
24. Brennan PJ, Crick DC. The cell-wall core of Mycobacterium tuberculosis in the context of drug discovery. Curr Top Med Chem. 2007; 7:475-488.
25. Brennan PJ. Structure, function, and biogenesis of the cell wall of Mycobacterium tuberculosis. Tuberculosis. 2003; 83:91-97.
26. Brocca S, Secundo F, Ossola M, Alberghina L, Carrea G, Lotti M. Sequence of the lid affects activity and specificity of Candida rugosa lipase isoenzymes. Protein Sci. 2003; 12:2312-2319.
27. Brown DP, Krishnamurthy N, Sjölander K. Automated protein subfamily identification and classification. PLoS Comput Biol. 2007; 3:e160.
28. Brzozowski AM, Derewenda U, Derewenda ZS, Dodson GG, Lawson DM, Turkenburg JP, Bjorkling F, Huge-Jensen B, Patkar SA, Thim LA. A model for interfacial activation in lipases from the structure of a fungal lipase–inhibitor complex. Nature. 1991; 351:491-494.
29. Canaan S, Maurin D, Chahinian H, Pouilly B, Durousseau C, Frassinetti F, Scappuccini-Calvo L, Cambillau C, Bourne Y. Expression and characterization of the protein Rv1399c from Mycobacterium *tuberculosis*. A novel carboxyl esterase structurally related to the HSL family. Eur J Biochem. 2004; 27:13953-13961.
30. Carter P, Wells JA. Dissecting the catalytic triad of a serine protease. Nature. 1988; 332:564-568.
31. Chakravarty J, Sundar S. Drug Resistance in Leishmaniasis. J Glob Infect Dis. 2010; 2:167-176.
32. Chappuis F, Sundar S, Hailu A, Ghalib H, Rijal S, Peeling RW, Alvar J, Boelaert M. Visceral leishmaniasis: what are the needs for diagnosis, treatment and control? Nat Rev Microbiol. 2007; 5:873-882.
33. Chawla B, Madhubala R. Drug targets in Leishmania. J Parasit Dis. 2010 341:1-13.
34. Clem A. A current perspective on leishmaniasis. J Glob Infect Dis. 2010; 2:124-126.
35. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, *et al*. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature. 1998; 393:537-544.
36. Cole ST. Learning from the genome sequence of *Mycobacterium tuberculosis* H37Rv. FEBS Lett. 1999; 452:7-10.
37. Colovos C, Yeates TO. Verification of protein structures: patterns of nonbonded interactions. Protein Sci. 1993; 2:1511-1519.
38. Corbett EL, Watt CJ, Walker N, Maher D, Williams BG, Raviglione MC, Dye C. The growing burden of tuberculosis: global trends and interactions with the HIV epidemic. Arch Intern Med. 2003; 163:1009-1021.

39. Cosconati S, Forli S, Perryman AL, Harris R, Goodsell DS, Olson AJ. Virtual Screening with AutoDock: Theory and Practice. Expert Opin Drug Discov. 2010; 56:597-607.
40. Cotes K, Bakala N'goma JC, Dhouib R, Douchet I, Maurin D, Carriere F, Canaan S. Lipolytic enzymes in *Mycobacterium tuberculosis*. Appl Microbiol Biotechnol. 2008; 785:741-749.
41. Côtes K, Dhouib R, Douchet I, Chahinian H, de Caro A, Carrière F, Canaan S. Characterization of an exported monoglyceride lipase from Mycobacterium tuberculosis possibly involved in the metabolism of host cell membrane lipids. Biochem J. 2007; 4083:417-427.
42. Crellin PK, Vivian JP, Scoble J, Chow FM, West NP, Brammananth R, Proellocks NI, Shahine A, Le Nours J, Wilce MC, Britton WJ, Coppel RL, Rossjohn J, Beddoe T. Tetrahydrolipstatin Inhibition, Functional Analyses, and Three-dimensional Structure of a Lipase Essential for Mycobacterial Viability. J Biol Chem. 2010; 285:30050-30060.
43. Dahiyat BI. *In silico* design for protein stabilization. Curr Opin Biotech. 1999; 10:387-390.
44. Daleke MH, Cascioferro A, de Punder K, Ummels R, Abdallah AM, van der Wel N, Peters PJ, Luirink J, Manganelli R, Bitter W. Conserved Pro-Glu (PE) and Pro-Pro-Glu (PPE) protein domains target LipY lipases of pathogenic mycobacteria to the cell surface via the ESX-5 pathway. J Biol Chem. 2011; 286:19024-19034.
45. Daniel J, Deb C, Dubey VS, Sirakova TD, Abomoelak B, Morbidoni HR, Kolattukudy PE. Induction of a novel class of diacylglycerol acyltransferases and triacylglycerol accumulation in *Mycobacterium tuberculosis* as it goes into a dormancy-like state in culture. Journal of Bacteriology. 2004; 186:5017-5030.
46. Darden T, York D, Pedersen L. 1993. Particle mesh Ewald: An N-logN method for Ewald sums in large systems. J Chem Phys. 1993; 98:10089-10092.
47. Davies CR, Kaye P, Croft SL, Sundar S. Leishmaniasis: new approaches to disease control Brit Med J. 2003; 326:377-382.
48. DeLano WL. The PyMOL Molecular Graphics System DeLano Scientific, San Carlos, CA, USA, 2002.
49. Derewenda U, Brzozowski AM, Lawson DM, Derewenda ZS. Catalysis at the interface: the anatomy of a conformational change in a triglyceride lipase. Biochemistry. 1992; 31:1532-4151.
50. Derewenda Z S, Sharp AM. News from the interface: the molecular structures of triacyl-glyceride lipases. Trends Biochem Sci. 1993; 18:20-25.
51. Desjeux P. Leishmaniasis: current situation and new perspectives Comp Immunol Microbiol Infect Dis. 2004; 27:305-318.
52. Dhouib R, Laval F, Carrière F, Daffé M, Canaan S. A monoacylglycerol lipase from Mycobacterium smegmatis involved in the bacterial cell interaction. J Bacteriol. 2010; 19:4776-4785.
53. Discala C, Benigni X Barillot E, Vaysseix G. DBcat: a catalog of 500 biological databases. Nucleic Acids Res. 2000; 28:8-9.
54. Douchet I, de Haas G, Verger R. Lipase regio- and stereoselectivities toward three enantiomeric pairs of didecanoyl-deoxyamino-O methyl glycerol: A kinetic study by the monomolecular film technique. Chirality. 2003; 15:220-226.

55. Ducati RG, Ruffino-Netto A, Basso LA, Santos DS. The resumption of consumption -- a review on tuberculosis. Mem Inst Oswaldo Cruz. 2006; 101:697-714.
56. Essmann U, PereraL, Berkowitz M L, Darden T, Lee H, Pedersen L G A. Smooth particle mesh Ewald method. J Chem Phys. 1995; 103:8577- 8592.
57. Finn RD, Mistry J, Tate J, Coggill P, Heger A, Pollington JE, Gavin OL, Gunasekaran P, Ceric G, Forslund K, Holm L, Sonnhammer EL, Eddy SR, Bateman A. The Pfam protein families database. Nucleic Acids Res. (Database) 2010; 38:211-222.
58. Fischer M, Pleiss J. The Lipase Engineering Database: a navigation and analysis tool for protein families. Nucleic Acids Res. 2003; 31:319-321.
59. Fojan P, Jonson PH, Petersen MTN, Petersen SB. What distinguishes an esterase from a lipase: a novel structural approach. Biochimie. 2000; 82:1033-1041.
60. Frieden TR, Munsiff SS. The DOTS strategy for controlling the global tuberculosis epidemic. Clin Chest Med. 2005; 26:197-205.
61. Fujii T, Hata Y, Ooseki M, Moriyama H, Wakagi T, Tanaka N, Oshima T. The crystal structure of zinc-containing ferredoxin from the thermoacidophilic archaeon Sulfolobus sp. strain 7. Biochem. 1997; 36:1505-1513.
62. Garton N, Christensen H, Minnikin D, Adegbola R, Barer M. Intracellular lipophilic inclusions of mycobacteria in vitroand in sputum Microbiology. 2002; 148:2951-2958.
63. Ghosh D, Wawrzak Z, Pletnev V Z, Li N, Kaiser R, Pangborn W, Jörnvall H, Erman M, Duax W L. Structure of uncomplexed and linoleate-bound *Candida cylindracea* cholesteol esterase. Structure 1995; 3:279-288.
64. Gouet P, Courcelle E, Stuart DI, Metoz F. ESPript: multiple sequence alignments in PostScript. Bioinformatics. 1999; 15:305-308.
65. Gough J. Genomic scale sub-family assignment of protein domains. Nucleic Acids Res. 2006; 34:3625-3633.
66. Guenin-Macé L, Siméone R, Demangel C. Lipids of pathogenic Mycobacteria: contributions to virulence and host immune suppression. Transbound Emerg Dis. 2009; 56:255-268.
67. Gupta R, Gupta N, Rathi P. Bacterial lipases: an overview of production, purification and biochemical properties. Appl. Microbiol. Biotechnol. 2004; 64:763-781.
68. Haki G D, Rakshit S K. Developments in industrially important thermostable enzymes: A review. Bioresour. Technol. 2003; 89:17-34.
69. Hasan F, Shah AA, Hameed A. Industrial applications of microbial lipases. Enzyme and Microbial Technology. 2006; 39:235-251.
70. Heikinheimo P, Goldman A, Jeffries C, Ollis D L. Of barn owls and bankers: a lush variety of alpha/beta hydrolases. Struct. Fold. Des. (Reviews) 1999; 7:141-146.
71. Henke E, Pleiss J, Bornscheuer UT. Activity of lipases and esterases towards tertiary alcohols: insights into structure-function relationships. Angew Chem Int Ed Engl. 2002; 41:3211-3213.
72. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. LINCS: A linear constraint solver for molecular simulations. J Comput Chem. 1997; 18:1463-1472.
73. Hess B, Kutzner C, van der Spoel D, Lindahl E, GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation J Chem Theory Comput. 2008; 4:435-447.
74. Hess B. P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation J Chem Theory Comput. 2008; 4:116-122.

75. Horchani H, Chaâbouni M, Gargouri Y, Sayari A. Solvent-free lipase-catalyzed synthesis of long-chain starch esters using microwave heating: Optimization by response surface methodology. Carbohydrate Polymers. 2010; 79:466-474.
76. Horchani H, Mosbah H, Salem N B, Gargouri Y, Sayari. A Biochemical and molecular characterisation of a thermoactive, alkaline and detergent-stable lipase from a newly isolated *Staphylococcus aureus* strain. Journal of Molecular Catalysis B: Enzymatic 2009; 56:237-245.
77. Hornak V, Simmerling C. Targeting structural flexibility in HIV-1 protease inhibitor binding. Drug Discov Today. 2007; 123:132-138.
78. Hotelier T, Renault L, Cousin X, Negre V, Marchot P, Chatonnet A. ESTHER, the database of the alpha/beta-hydrolase fold superfamily of proteins. Nucleic Acids Res. (Database) 2004; 32:145-147.
79. Houde A, Kademi A, Leblanc D. Lipases and Their Industrial Applications. Applied Biochemistry and Biotechnology. 2004; 118:155-170.
80. Hussain RF, Nouri AM, Oliver RT. A new approach for measurement of cytotoxicity using colorimetric assay. J Immunol Method. 1993; 160:89-96.
81. Illanes A. Stability of biocatalysts. Electron. J. Biotechnol. 1999; 21:7-15.
82. Irwin JJ, Shoichet BK. ZINC- a free database of commercially available compounds for virtual screening. J Chem Inf Model. 2005; 45:177-182.
83. Irwin JJ, Sterling T, Mysinger MM, Bolstad ES, Coleman RG. ZINC: A Free Tool to Discover Chemistry for Biology. J Chem Inf Model. 2012; 52:1757-1768.
84. Ivanov AA, Baskin II, Palyulin VA, Piccagli L, Baraldi PG, Zefirov NS. Molecular modeling and molecular dynamics simulation of the human A2B adenosine receptor. The study of the possible binding modes of the A2B receptor antagonists. J Med Chem. 2005; 4822:6813-6820.
85. J Shen, J Jiang, G Kuang, C Tan, G Liu, J Huang, Y Tang. Discovery and structure-activity analysis of selective estrogen receptor modulators via similarity-based virtual screening. Eur J Med Chem. 2012; 54:188-196.
86. Jaeger K E, Dijkstra B W, Reetz M T. Bacterial biocatalysts: molecular biology, three-dimensional structures, and biotechnological applications of lipases. Annu. Rev. Microbiol. 1999; 53:315-351.
87. Jaeger K E, Eggert T. Lipases for biotechnology. Current Opinion in Biotechnology 2002; 13:390-397.
88. Jaenicke R, Boehm G. The stability of proteins in extreme environments. Curr. Opin. Struct. Biol. 1998; 8:738-748.
89. Jeong S T, Kim H K, Kim S J, Pan J G, Oh TK, Ryu SE. Crystallization and preliminary X-ray analysis of a thermoalkalophilic lipase from *Bacillus stearothermophilus* L1. Acta Crystallog. D. 2001; 57:1300-1302.
90. Robinson WE, Cordeiro M, Abdel-Malek S, Jia Q, Chow SA, Reinecke MG, Mitchell WM. Dicaffeoylquinic acid inhibitors of human immunodeficiency virus integrase: inhibition of the core catalytic domain of human immunodeficiency virus integrase. Mol Pharmacol. 1996; 504:846-855.
91. Kang H, Jihyun F K, Myung H K, Seung-Hwan P, Tae-Kwang O, Cheol-Goo H. MELDB: A database for microbial esterases and lipases. FEBS Letters. 2006; 580:2736-2740.

92. Karlsson M, Contreras JA, Hellman U, Tornqvist H, Holm C. cDNA cloning, tissue distribution, and identification of the catalytic triad of monoglyceride lipase Evolutionary relationship to esterases, lysophospholipases, and haloperoxidases. J Biol Chem. 1997; 272:27218-27223.
93. Karlsson M, Tornqvist H, Holm C. Expression, purification, characterization of histidine-tagged mouse monoglyceride lipase from baculovirus-infected insect cells. Protein Expr Purif. 2000; 18:286-292.
94. Kelil A, Wang S, Brzezinski R, Fleury A. CLUSS: clustering of protein sequences based on a new similarity measure. BMC Bioinformatics. 2007; 8:e286.
95. Kelil A, Wang S, Brzezinski R. CLUSS2: an alignment-independent algorithm for clustering protein families with multiple biological functions. Int J Comput Biol Drug Des. 2008; 1:122-140.
96. Kellenberger E, Springael JY, Parmentier M, Hachet-Haas M, Galzi JL, Rognan D. Identification of nonpeptide CCR5 receptor agonists by structure-based virtual screening. J Med Chem. 2007; 50:1294-1303
97. Khan MT, Fuskevåg OM, Sylte I. Discovery of potent thermolysin inhibitors using structure based virtual screening and binding assays. J Med Chem. 2009; 52:48-61.
98. Kourist R, Brundiek H, Bornscheuer UT. Protein engineering and discovery of lipases. Eur. J. Lipid Sci. Technol., 2010; 112:64-74.
99. Labar G, Bauvois C, Borel F, Ferrer JL, Wouters J, Lambert DM. Crystal structure of the human monoacylglycerol lipase, a key factor in endocannabinoid signaling Chembiochem. 2010; 112:218-227.
100. Lang DA, Mannesse MLM, De Haas GH, Verheij HM, Dijkstra BW. Structural basis of the chiral selectivity of Pseudomonas cepacia lipase. European Journal of Biochemistry. 1998; 254:333-340.
101. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check stereo chemical quality of protein structures. J Appl Crystallogr. 1993; 26:283-291.
102. Lassmann T, Frings O, Sonnhammer EL. Kalign2: high-performance multiple alignment of protein and nucleotide sequences allowing external features. Nucleic acids res. 2009; 37:858-865.
103. Lenfant N, Hotelier T, Bourne Y, Marchot P, Chatonnet A. Proteins with an alpha/beta hydrolase fold: Relationships between subfamilies in an ever-growing superfamily. Chem Biol Interact. 2013; 203:266-268.
104. Leung CH, Chan DS, Kwan MH, Cheng Z, Wong CY, Zhu GY, Fong WF, Ma DL. Structure-based repurposing of FDA-approved drugs as TNF-α inhibitors. Chem Med Chem. 2011; 6:765-768.
105. Lexa KW, Carlson HA, Binding to the open conformation of HIV-1 protease Proteins. 2011; 797:2282-2290.
106. Lexa KW, Damm KL, Quintero JJ, Gestwicki JE, Carlson HA. Clarifying allosteric control of flap conformations in the 1TW7 crystal structure of HIV-1 protease. Proteins. 2009; 744:872-880.
107. Lindahl E, Hess B, van der Spoel D. GROMACS 30: a package for molecular simulation and trajectory analysis. J Mol Mod. 2001; 7:306-317.
108. Liu J, Barry CE, Besra GS, Nikaido H. Mycolic acid structure determines the fluidity of the mycobacterial cell wall. J Biol Chem. 1996; 271:29545-29551.

109. Long JZ, Li W, Booker L, Burston JJ, Kinsey SG, Schlosburg JE, Pavón FJ, Serrano AM, DE Selley, LH Parsons, AH Lichtman, BF Cravatt. Selective blockade of 2-arachidonoylglycerol hydrolysis produces cannabinoid behavioral effects. Nat Chem Biol. 2009; 5:37-44.
110. Long JZ, Nomura DK, Cravatt BF. Characterization of monoacylglycerol lipase inhibition reveals differences in central and peripheral endocannabinoid metabolism. Chem Biol. 2000; 16:744-753.
111. Longhi S, Czjzek M, Lamzin V, Nicolas A, Cambillau C. Atomic resolution 10 Å crystal structure of Fusarium solani cutinase: stereochemical analysis. J Mol Biol. 1997; 268:779-799.
112. Looger L L, Dwyer M A, Smith J J, Hellinga H W. Computational design of receptor and sensor proteins with novel functions. Nature. 2003; 423:185-190.
113. Lyne PD. Structure-based virtual screening: an overview. Drug Discov Today. 2002; 7:1047-1055.
114. Magrane M, Uniprot Consortium. UniProt Knowledgebase: a hub of integrated protein data. Database (Oxford). 2011; 2011:bar009.
115. Målen H, Pathak S, Søfteland T, de Souza GA, Wiker HG. Definition of novel cell envelope associated proteins in Triton X-114 extracts of *Mycobacterium tuberculosis H37Rv*. BMC Microbiol. 2010; 10:e132.
116. Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, Fong JH, *et al*. CDD: a Conserved Domain Database for the functional annotation of proteins. Nucleic Acids Res. (Database) 2011; 39:225-229.
117. Marmiesse M, Brodin P, Buchrieser C, Gutierrez C, Simoes N, Vincent V, Glaser P, Cole ST, Brosch R. Macro-array and bioinformatic analyses reveal mycobacterial 'core' genes, variation in the ESAT-6 gene family and new phylogenetic markers for the Mycobacterium tuberculosis complex. Microbiology. 2004; 150:483-496.
118. Marti-Renom M A, Stuart A C, Fiser A, Sánchez R, Melo F, Sali A. Comparative protein structure modeling of genes and genomes. Annu. Rev. Biophys. Biomol. Struct. 2000; 29:291-325.
119. Massova I, Pirkle H, Edwards BFP, Mobashery S. Insights into the three-dimensional structure of crotalase: implications for biological activity and substrate specificity. Bioorg. Med. Chem. Lett. 1997; 7:3139-3144.
120. McConville M, de Souza D, Saunders E, Likic VA, Naderer T. Living in a phagolysosome; metabolism of Leishmania amastigotes. Trends Parasitol. 2007; 23:368-374.
121. McConville MJ, Naderer T. Metabolic pathways required for the intracellular survival of Leishmania. Annu Rev Microbiol. 2011; 6:543-561.
122. McKinney JD, Honer zu Bentrup K, Munoz-Elias EJ, Miczak A, Chen B, Chan WT, Swenson D, Sacchettini JC, Jr Jacobs WR, Russell DG. Persistence of *Mycobacterium tuberculosis* in macrophages and mice requires the glyoxylate shunt enzyme isocitrate lyase. Nature. 2000; 406:735-738.
123. McWilliam H, Valentin F, Goujon M, Li W, Narayanasamy M, Martin J, Miyar T, Lopez R. Web services at the European Bioinformatics Institute – 2009. Nucleic Acids Res. (Webserver) 2009; 37:6-10.

124. Melo RCN, Dvorak AM. Lipid body–phagosome interaction in macrophages during infectious diseases: host defense or pathogen survival strategy? PLoS Pathog. 2012; 8:e1002729.
125. Meniche X, Labarre C, de Sousa-d'Auria C, Huc E, Laval F, Tropis M, Bayan N, Portevin D, Guilhot C, Daffé M, Houssin C. Identification of a stress-induced factor of Corynebacterineae that is involved in the regulation of the outer membrane lipid composition. J Bacteriol. 2009; 191:7323-7332.
126. Messaoudi A, Belguith H, Ghram I, Hamida JB. LIPABASE: a database for 'true' lipase family enzymes. Int J Bioinform Res Appl. 2011; 7:390-401.
127. Morley K L, Kazlauskas R J. Improving enzyme properties: when are closer mutations better? TRENDS in Biotechnology. 2005; 23:231-237.
128. Morris GM, Goodsell DS, Halliday RS. Automated docking using a lamarckian genetic algorithm and empirical binding free energy function. J Comput Chem. 1998; 19:1639-1662.
129. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. J Comput Chem. 2009; 3016:2785-2791.
130. Mosmann T. Rapid colorimetric assay for cellular growth and survival: application to proliferation and cytotoxicity assays. J Immunol Methods. 1983; 65:55-63.
131. Murzin A G, Brenner S E, Hubbard T, Chothia C. SCOP: a structural classification of proteins database for the investigation of sequences and structures. J. Mol. Biol. 1995; 247:536-540.
132. Nachega JB, Chaisson RE. Tuberculosis drug resistance: a global threat. Clinical Infectious Diseases. (Supplement) 2003; 36:24-30.
133. Naderer T, McConville MJ. The Leishmania-macrophage interaction: a metabolic perspective. Cell Microbiol. 2008; 10:301-308.
134. Nicola G, Smith CA, Lucumi E, Kuo MR, Karagyozov L, Fidock DA, Sacchettini JC, Abagyan R. Discovery of novel inhibitors targeting enoyl-acyl carrier protein reductase in *Plasmodium falciparum* by structure-based virtual screening. Biochem Biophys Res Commun. 2007; 358:686-691.
135. Noble M E, Cleasby A, Johnson L N, Egmond M R, Frenken L G. The crystal structure of triacylglycerol lipase from *Pseudomonas glumae* reveals a partially redundant catalytic aspartate. FEBS Letters. 1993; 331:123-128.
136. Ogston A. On Abscesses. Classics in Infectious Diseases. Rev. Infect. Dis. 1984; 61:122-128.
137. Ollis D L, Cheah E, Cygler M, Dijkstra B, Frolow F, Franken S M, Harel M, Remington S J, Silman I, Schrag J, Sussman J L, Verschueren K H G, Goldman A The alpha/beta hydrolase fold. Protein Engineering. 1992; 5:197-211.
138. Opperdoes FR, Coombs GH. Metabolism of Leishmania: proven and predicted Trends Parasitol. 2007; 23:149-158.
139. Opperdoes FR, Michels PAM. The metabolic repertoire of Leishmania and implications for drug discovery In: Myler P, Fasel N Eds Leishmania, after the genome, 1st edn Caister Academic Press, Norfolk, UK. 2008; pp 123-158.
140. Oprea TI, Matter H. Integrating virtual screening in lead discovery. Curr Opin Chem Biol. 2004; 8:349-358.

141. Paiva AL, van Rossum D, Malcata FX. Kinetics of lipase-mediated synthesis of butyl butyrate in n-hexane. Biocatalysis and Biotransformation. 2002; 20:43-51.
142. Park H, Lee J, Lee S. Critical assessment of the automated AutoDock as a new docking tool for virtual screening Proteins: Struct, Funct, and Bioinf. 2006; 65:549-554.
143. Park M, Do E, Jung W H. Lipolytic Enzymes Involved in the Virulence of Human Pathogenic Fungi. Mycobiology. 2013; 41:67-72.
144. Parker SK, Barkley RM, Rino JG, Vasil ML, *Mycobacterium tuberculosis* Rv3802c Encodes a Phospholipase/Thioesterase and Is Inhibited by the Antimycobacterial Agent Tetrahydrolipstatin. PLoS One. 2009; 4:e4281.
145. Patil KJ, Chopda MZ, Mahajan RT. Lipase biodiversity. Indian Journal of Science and Technology. 2011; 4:971-982.
146. Patkar S, Vind J, Kelstrup E, Christensen MW, Svendsen A, Borch K, Kirk O. "Effect of mutations in *Candida Antarctica* B lipase." Chemistry and physics of lipids. 1998; 93:95-101.
147. Pereira EB, Zanin GM, Castro HF. "Immobilization and catalytic properties of lipase on chitosan for hydrolysis and esterification reactions." Brazilian Journal of Chemical Engineering. 2003; 20:343-355.
148. Pieters J. Evasion of host cell defense mechanisms by pathogenic bacteria. Curr Opin Immunol. 2001; 13:37-44.
149. Pleiss J, Fischer M, Schmid RD. Anatomy of lipase binding sites: the scissile fatty acid binding site. Chem. Phys. Lipids. 1998; 93:67-70.
150. Rakotomanga M, Blanc S, Gaudin K, Chaminade P, Loiseau PM. Miltefosine affects lipid metabolism in *Leishmania donovani* promastigotes. Antimicrob Agents Chemother. 2007; 51:1425-1430.
151. Ray A. Application of Lipase in Industry. Asian J. Pharm. Tech. 2012;2(2): 33-37.
152. Ritz N, Curtis N. Mapping the global use of different BCG vaccine strains. Tuberculosis. 2009; 89:248-251.
153. Rose PW, Bi C, Bluhm WF, Christie CH, Dimitropoulos D, Dutta S, Green RK, Goodsell DS, Prlic A, Quesada M, Quinn GB, Ramos AG, Westbrook JD, Young J, Zardecki C, Berman HM, Bourne PE. The RCSB Protein Data Bank: new resources for research and education. Nucleic Acids Res. (Database) 2013; 41:475-482.
154. Rosenzweig D, Smith D, Opperdoes F, Stern S, Olafson RW, Zilberstein D. Retooling Leishmania metabolism: from sand fly gut to human macrophage. FASEB J. 2008; 22:590-602.
155. Russell DG, VanderVen BC, Lee W, Abramovitch RB, Kim M, Homolka S, Niemann S, Rohde KH. *Mycobacterium tuberculosis* wears what it eats. Cell Host & Microbe. 2010; 8:68-76.
156. Russell DG. *Mycobacterium tuberculosis*: here today, and here tomorrow. Nat Rev Mol Cell Biol. 2001; 2:569-577.
157. Saario S M, Laitinen J T. Monoglyceride lipase as an enzyme hydrolyzing 2-arachidonoylglycerol. Chem. Biodivers. 2007; 4:1903-1913.
158. Salihu A and Alam MZ. Production and applications of microbial lipases: A review. Scientific Research and Essays. 2012; 7:2667-2677.
159. Salter H. Teaching bioinformatics. Biochemical Education. 1998; 26:3-10.
160. Sansom CE, Smith CA. Computer applications in biomolecular sciences. Part 2: bioinformatics and genome projects. Biochem Educ. 2000; 28:127-131.

161. Saravanan P, Avinash H, Dubey VK, Sanjukta P. Targeting essential cell wall lipase Rv3802c for potential therapeutics against tuberculosis. J Mol Graphics Modell. 2012; 38:235-242.
162. Saravanan P, Venkatesan SK, Mohan CG, Patra S, Dubey VK. Mitogen-activated protein kinase 4 of Leishmania parasite as a therapeutic target. Eur J Med Chem. 2010; 4512:5662-5670.
163. Sassetti CM, Boyd DH, Rubin EJ. Genes required for mycobacterial growth defined by high density mutagenesis Mol Microbiol. 2003; 48:77-84.
164. Saudagar P, Dubey VK. Cloning, expression, characterization and inhibition studies on Trypanothione Synthetase, a drug target enzyme, from *Leishmania donovani.* Biol Chem. 2011; 392:1113-1122.
165. Saven J G. Combinatorial protein design. Curr. Opin. Struct. Biol. 2002; 12:453-458.
166. Saxena R K, Ghosh P K, Gupta R, Bradoo S, Gulati R. Microbial lipases: Potential biocatalysts for the future industry. Current Science. 1999; 77:101-115.
167. Schalk-Hihi C, Schubert C, Alexander R, Bayoumy S, Clemente JC, Deckman I, DesJarlais RL, et al. Crystal structure of a soluble form of human monoglyceride lipase in complex with an inhibitor at 1.35 Å resolution. Protein Sci. 2011; 20:670-683.
168. Scheer M, Grote A, Chang A, Schomburg I, Munaretto C, Rother M, Söhngen C, Stelzer M, Thiele J, Schomburg D. BRENDA, the enzyme information system in 2011. Nucleic Acids Res. (Database) 2011; 39:670-676.
169. Scheib H, Pleiss J, Stadler P, Kovac A, Potthoff AP, Haalck L, Spener F, Paltauf F, Schmid RD. "Rational design of Rhizopus oryzae lipase with modified stereoselectivity toward triradylglycerols." Protein engineering 11, no. 8 (1998): 675-682.
170. Schmidt-Dannert C, Rua M, Atomi H, Schmid R D.. Thermoalkalophilic lipase of *Bacillus thermocatenulatus*. I. Molecular cloning, nucleotide sequence, purification and some properties. Biochim. Biophys. Acta. 1996; 1301:105-114.
171. Schuettelkopf AW, van Aalten DMF. PRODRG - a tool for high-throughput crystallography of protein-ligand complexes. Acta Crystallogr D Biol Crystallogr. 2004; 60:1355-1363.
172. Shah NS, Wright A, Bai GH, Barrera L, Boulahbal F, Martín-Casabona N, Drobniewski F, *et al*. Worldwide Emergence of Extensively Drug-resistant Tuberculosis. Emerg Infect Dis. 2007; 13:380-387.
173. Shahlaei M, Madadkar-Sobhani A, Mahnam K, Fassihi A, Saghaie L, Mansourian M, Homology modeling of human CCR5 and analysis of its binding properties through molecular docking and molecular dynamics simulation. Biochim Biophys Acta. 2011; 180:802-817.
174. Shakarian AM, McGugan GC, Joshi MB, Stromberg M, Bowers L, Ganim C, Barowski J, Dwyer DM. Identification, characterization, and expression of a unique secretory lipase from the human pathogen *Leishmania donovani*. Mol Cell Biochem. 2010; 341:17-31.
175. Shan Y, Kim ET, Eastwood MP, Dror RO, Seeliger MA, Shaw DE. How does a drug molecule find its target binding site? J Am Chem Soc. 2011; 13324:9181-9183.
176. Sharma R, Chisti Y, Banerjee UC. Production, purification, characterization, and applications of lipases. Biotechnology Advances. 2001; 19:627-662.
177. Shekhar S, Chawla S. Spatial Databases: A Tour. Prentice-Hall, 2002.

178. Shen MY, Sali A. Statistical potential for assessment and prediction of protein structures. Protein Sci. 2006; 15:2507-2524.
179. Shukla AK, Singh BK, Patra S, Dubey VK. Rational approaches for drug designing against leishmaniasis. Appl Biochem Biotechnol. 2010; 160:2208-2218.
180. Sih, C. J. and Wu, S.H. Resolution of Enantiomers via Biocatalysis. *In:* Topics of stereochemistry. John Wiley & Sons, Inc. 1998; 19: pp. 63-125.
181. Simons J W F A, Götz F, Egmond M R, Verheij H M. Biochemical properties of staphylococcal phospholipases. Chem. Phys. Lipids. 1998; 93:27-37.
182. Sims PA, Wong CF, McCammon JA. A computational model of binding thermodynamics: the design of cyclin-dependent kinase 2 inhibitors. J Med Chem. 2003; 4615:3314-3325.
183. Singh G, Jadeja D, Kaur J. Lipid hydrolizing enzymes in virulence: *Mycobacterium tuberculosis* as a model system. Crit Rev Microbiol. 2010; 36:259-269.
184. Sippl M J. Recognition of Errors in Three-Dimensional Structures of Proteins. Proteins: Struct. Funct. Genet. 1993; 17:355-362.
185. Smith CA, Toogood H S, Baker H M, Daniel R M, Baker E N. Calcium-mediated thermostability in the subtilisin superfamily: the crystal structure of Bacillus Ak.1 protease at 1.8 Å resolution. J. Mol. Biol. 1999; 294:1027-1040.
186. Smith TK, Reynolds TB, Denny PW. Lipid metabolism as a therapeutic target. Biochemistry Res Int. 2012; e158139.
187. Song H, Sandie R, Wang Y, Andrade-Navarro MA, Niederweis M. Identification of outer membrane proteins of *Mycobacterium tuberculosis* Tuberculosis Edinb. 2008; 88:526-544.
188. Song X, Qi X, Hao B, Qu Y. Studies of substrate specificities of lipases from different sources. Eur. J. Lipid Sci. Technol. 2008; 110:1095-1101.
189. Sousa SF, Fernandes PA, Ramos MJ. Protein–ligand docking: Current status and future challenges. Proteins. 2006; 65:15-26.
190. Stehr F, Felk A, Gácser A, Kretschmar M, Mähnss B, Neuber K, Hube B, Schäfer W, Expression analysis of the Candida albicans lipase gene family during experimental infections and in patient samples. FEMS Yeast Res. 2004; 4:401-408.
191. Stevens R C, Gouaux JE, Lipscomb W N. Structural consequences of effector binding to the T-state of aspartate carbamoyltransferase: crystal structures of the unligated and ATP-complexed and CTP-complexed enzymes at 2.6 Å resolution. Biochemistry. 1990; 29:7691-7701.
192. Street A G, Mayo S L. Computational protein design. Structure. (Review) 1999; 7:105-109.
193. Subba Rao G, Vijayakrishnan R, Kumar M. Structure-based design of a novel class of potent inhibitors of InhA, the enoyl acyl carrier protein reductase from *Mycobacterium tuberculosis*: a computer modelling approach. Chem Biol Drug Des. 2008; 725:444-449.
194. Sultana R, Tanneeru K, Guruprasad L. The PE-PPE domain in mycobacterium reveals a serine α/β hydrolase fold and function: an in-silico analysis. PLoS One. 2011; 6:e16745.
195. Takayama K, Wang C, Besra GS, Pathway to synthesis and processing of mycolic acids in *Mycobacterium tuberculosis*. Clin Microbiol Rev. 2005; 18: 81-101.

196. Tanrikulu Y, Krüger B, Proschak E. The holistic integration of virtual screening in drug discovery. Drug Discov Today. 2013; 187:358-364.
197. Teplyakov A V, Gros P, Hol WG. Crystallographic study of eglin-C binding to thermitase. Advan. Expt. Med. Biol. 1996; 379:5-9.
198. Thompson J D, Higgins D G, Gibson TJ. Clustalw: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 1994; 22:4673-4680.
199. Tielens AGM, van Hellemond JJ. Surprising variety in energy metabolism within Trypanosomatidae. Trends Parasitol. 2009; 25:482-490.
200. Tiesinga J J W, van Pouderoyen G, Nardini M, Ransac S, Dijkstra BW. Structural Basis of Phospholipase Activity of Staphylococcus hyicus lipase. J. Mol. Biol. 2007; 371:447-456.
201. Todar K, Online Textbook of Bacteriology, "The Good, the Bad, and the Deadly" Science Magazine. 2004; 304:1421-1632.
202. Tyndall J D A, Sinchaikul S, Fothergill-Gilmore L A, Taylor P, Walkinshaw M D. Crystal Structure of a Thermostable Lipase from *Bacillus stearothermophilus P1*. J. Mol. Bio. 2002; 323:859-869.
203. van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ. GROMACS: fast, flexible, and free. J Comput Chem. 2005; 26:1701-1718.
204. van Gunsteren, Billeter W F, Eising S R, Hunenberger AA, Kruger P H, Mark P, Scott AE, Tironi IG. Biomolecular simulation: the GROMOS96 manual and user guide. Vdf Hochschulverlag AG, Zurich. 1996.
205. van Kampen M D, Rosenstein R, Gotz F, Egmond M R Cloning, purification and characterisation of *Staphylococcus warneri* lipase 2. Biochim. Biophys. Acta. 2001; 1544:229-241.
206. van Oort MG, Deveer AMTJ, Dijkman R, Tjeenk M L, Verheij H M, de Haas G. H, Wenzig E, Gotz F. Purification and substrate specificity of *Staphylococcus hyicus* lipase. Biochemistry. 1989; 28:9278-9285.
207. Verger R. Interfacial activation of lipases: Facts and artifacts. Trends Biotechnol. 1997; 15:32-38.
208. Vieille C, Zeikus G. Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. Microbiol. Mol. Biol. Rev. 2001; 65:1-43.
209. Viso A, Cisneros JA, Ortega-Gutierrez S. The medicinal chemistry of agents targeting monoacylglycerol lipase. Curr. Top. Med. Chem. 2008; 8:231-246.
210. Vissa VD, Brennan PJ. The genome of *Mycobacterium leprae*: a minimal mycobacterial gene set. Genome Biol. (Reviews) 2001; 2:10231-10238.
211. Vulfson E N. Industrial applications of lipases. *In*: Wooley P and Petersen SB, eds Lipases-their structure, biochemistry and application. Cambridge University Press, Cambridge. 1994; pp 271-288.
212. Wang DF, Helquist P, Wiech NL, Wiest O. Toward selective histone deacetylase inhibitor design: homology modeling, docking studies, and molecular dynamics simulations of human class I histone deacetylases. J Med Chem. 2005; 4822:6936-6947.
213. Wang L, Veenstra D L, Radmer R J, Kollman P A. Can one predict protein stability? An attempt to do so for residue 133 of T4 lysozyme using a combination of free energy

derivatives, PROFEC, and free energy perturbation methods. Proteins. 1998; 32:438-458.
214. Watanabe S, Kodaki T, Makino K. Complete reversal of coenzyme specificity of xylitol dehydrogenase and increase of thermostability by the introduction of structural zinc. J Biol Chem. 2005; 280:10340-10349.
215. West NP, Cergol KM, Xue M, Randall EJ, Britton WJ, Payne RJ. Inhibitors of an essential mycobacterial cell wall lipase Rv3802c as tuberculosis drug leads. Chem Commun Camb. 2011; 4718:5166-5168.
216. West NP, Chow FME, Randall EJ, Wu J, Chen J, Ribeiro JMC, Britton WJ. Cutinase-like proteins of *Mycobacterium tuberculosis*: characterization of their variable enzymatic functions and active site identification. FASEB J. 2009; 23:1694-1704.
217. West NP, Wozniak TM, Valenzuela J, Feng CG, Sher A, Ribeiro JMC, Britton WJ. Immunological diversity within a family of cutinase-like proteins of Mycobacterium tuberculosis. Vaccine. 2008; 26:3853-3859.
218. Wicker N, Perrin GR, Thierry JC, Poch O. Secator: a program for inferring protein subfamilies from phylogenetic trees. Mol Biol Evol. 2001; 18:1435-1441.
219. Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res. (Webserver) 2007; 35:407-410.
220. Wilson JF. More cases of drug-resistant tuberculosis complicate care but spur global research and innovation. Ann Intern Med. 2009; 1506:433-436.
221. World Enzymes, http://www.reportlinker.com/p0148002/World-Enzymes-Market.html. 2009.
222. World Health Organization. Control of the Leishmaniasis: Report of a meeting of the WHO Expert Committee on the Control of Leishmaniasis, Geneva, 22–26 March 2010 WHO Technical Report Series, no 949 Geneva: World Health Organization. 2010.
223. World Health Organization. Leishmaniasis and leishmania/HIV co-infection. WHO report on global surveillance of epidemic-prone infectious diseases 2000 Report no WHO/CDS/CSR/ISR/20001 Geneva: World Health Organization. 2000.
224. World Health Organization, Global Tuberculosis Control 2009 Epidemiology, Strategy, Financing Report No: WHO/HTM/TB/2009411 Geneva, 2009 World Health Organization.
225. World Health Organization. Global Tuberculosis Control. 2011.
226. Wu CH, Nikolskaya A, Huang H, Yeh LS, Natale DA, Vinayaka CR, Hu ZZ, Mazumder R, Kumar S, Kourtesis P, Ledley RS, Suzek BE, Arminski L, Chen Y, Zhang J, Cardenas JL, Chung S, Castro-Alvear J, Dinkov G, Barker WC. PIRSF: family classification system at the Protein Information Resource. Nucleic Acids Res. (Database) 2004; 32:112-114.
227. Xiong J. Essential bioinformatics. Cambridge University Press, 2006.
228. Zhang Y. Persistent and dormant tubercle bacilli and latent tuberculosis. Front Biosci. 2004; 9:1136-1156.

APPENDIX

ANNEXURE I

List of UniProt Accession number of triacylglycerol lipases in our Triacylglycerol Lipase Database

A0AZ26	A0B3W5	A0EJ12	A0KEH6
A0KFL9	A0MTM1	A0Q0G8	A0Q394
A0RAN9	A0RK82	A1C6D6	A1DH10
A1E152	A1EMR8	A1F3S2	A1JLY3
A1KN95	A1T8V3	A1TUU1	A1U9Z2
A1UCL0	A1UHL6	A1UXK3	A2BQ73
A2BVQ5	A2C150	A2CAS9	A2P8X1
A2PS04	A2QE77	A2QF64	A2QGD9
A2QHE2	A2QHE9	A2QL90	A2QM14
A2QN29	A2QS46	A2QTI0	A2QTI9
A2QTZ0	A2QUC1	A2QZX4	A2R199
A2R1N7	A2R1X8	A2R2I5	A2R502
A2R709	A2R775	A2R8Z3	A2RZR8
A3EIQ3	A3GG80	A3GSF6	A3H0H8
A3LUV6	A3LV34	A3MH06	A3NMC8
A3NPH3	A3P7T9	A3PA07	A3PBW6
A3PTJ7	A3PW99	A3Q139	A3RQM7
A3RSM7	A3RTR2	A4F5W2	A4F7W2
A4F8A9	A4FC79	A4FGM5	A4HJN0
A4I6H7	A4I6H8	A4I752	A4JM54
A4JM55	A4LEY1	A4LP38	A4T9L6
A4WT30	A4WW65	A4XIZ1	A4XVF4
A4XW51	A4Y0Q3	A4YTP9	A4YW76
A4YWT9	A5ABZ1	A5DC90	A5E567
A5EAZ6	A5EJD5	A5ELJ2	A5EQU9
A5EYU1	A5G0F2	A5G974	A5I055
A5I3I2	A5IJ88	A5IPI7	A5IW97
A5JBL5	A5PK46	A5TEM3	A5TX95
A5V2S2	A5W0K8	A5WGV1	A5XMN8
A6A105	A6AE39	A6AS17	A6B1H2
A6LLI9	A6M186	A6N2U6	A6REI4
A6TRP1	A6TYA4	A6U555	A6U643

A6VCA5	A6W0V8	A6W8L3	A6WEQ4
A6WQH1	A6X6P9	A6XWR7	A6ZTP2
A7FSB2	A7FVB3	A7GNJ4	A7H8B2
A7HKW6	A7HW94	A7J993	A7K4N2
A7KAJ9	A7KAM5	A7TG13	A7UM95
A7Z124	A8D5W2	A8DYT7	A8EAU0
A8ENY2	A8F4V9	A8FGA4	A8FPC3
A8FXT7	A8G7T5	A8GDX0	A8GG18
A8GYN1	A8H2B1	A8KD69	A8KGU9
A8MHF3	A8QYB2	A8WEN5	A8YZE4
A8Z5H0	A9ACD8	A9AJ98	A9AM27
A9AMF2	A9APA8	A9BAX7	A9BHY3
A9BQ79	A9GHA3	A9GRJ0	A9K6S5
A9KNE3	A9KWN6	A9QXC9	A9VNH8
A9VZU3	A9W0D5	A9YY76	B0CPL4
B0DKP6	B0DSX5	B0EC71	B0FTZ8
B0K1R5	B0K9T2	B0KQY1	B0LW76
B0RCM8	B0RFW0	B0SL98	B0SLA0
B0SLV9	B0SPZ3	B0SZK4	B0TMF5
B0TPH2	B0UKB7	B0ULB5	B0V8J3
B0V9K7	B0VSR2	B1FB11	B1FNV3
B1FS11	B1G4I6	B1H4G3	B1H807
B1IFP8	B1IH63	B1IPN5	B1IZW1
B1JC81	B1JMS4	B1JNZ3	B1K1E8
B1K2U8	B1K3A3	B1K3P3	B1K8M1
B1KLI9	B1KUW0	B1KXI1	B1L809
B1LV02	B1M7F3	B1PDN0	B1PF33
B1PN84	B1PN85	B1QMN8	B1QNK6
B1SYQ6	B1T0Y4	B1TA42	B1TBB3
B1TH09	B1Y394	B1YI89	B1YQ26
B1YYF5	B1Z1J5	B1ZJ99	B1ZL40
B1ZR32	B2A362	B2CX98	B2H4N5
B2HBU4	B2IGV3	B2JJ10	B2K0M0
B2K2B3	B2L2K1	B2T787	B2TQP5
B2U848	B2UJU0	B2UZD8	B2UZE0
B2V6Q0	B3EAW3	B3GVV8	B3GVV9

B3QIA9	B3R4H4	B3VLA7	B3Y9N3
B4A6G2	B4ANV6	B4BJ29	B4EF94
B4EVM3	B4SHZ7	B4SXS2	B4TAC1
B4TP51	B4U881	B4UG28	B5A5A8
B5AEN3	B5AEN4	B5C678	B5CCT2
B5EEN0	B5EYG8	B5IHW5	B5IM43
B5JRY4	B5JS88	B5JVC2	B5JVC3
B5JX61	B5MLH0	B5MWD2	B5NHY9
B5NQ86	B5NYC2	B5PDK2	B5PSW9
B5QCM6	B5WVU9	B5ZM88	B5ZQE4
B6A8S6	B6AAA7	B6DAC2	B6KGG1
B6KKB1	B6KMV7	B6KR35	B6R031
B6VK93	B7CRH5	B7CWN1	B7DMH0
B7GVS2	B7GX66	B7GXX5	B7IAS8
B7KY62	B7L0B9	B7P112	B7P113
B7P312	B7PB92	B7PF83	B7PF84
B7PFS6	B7PIT8	B7PJN0	B7PM63
B7PN82	B7PNP9	B7PQN1	B7PU37
B7PX98	B7PXA4	B7Q0Y0	B7Q378
B7Q3Y8	B7Q6Z0	B7QBJ6	B7QBS8
B7QCR6	B7QCR7	B7QDU6	B7QDU8
B7QFC5	B7QFF2	B7QFF3	B7QML9
B7QNE0	B7UDC5	B7VF67	B7WX86
B7Z150	B7Z151	B8CYQ3	B8DQN9
B8DTF8	B8EA00	B8EL21	B8FK35
B8FTQ4	B8GQZ6	B8GYS5	B8H4U7
B8HEJ2	B8IP64	B8J3Q5	B8JD79
B8K3C5	B8K5B1	B8KFQ3	B8KQY9
B8KXA8	B8YIE6	B9B7G1	B9BEG9
B9BJY8	B9C3Z3	B9CSH3	B9CT83
B9CTC0	B9DIH4	B9LZR3	B9MGA9
B9MM09	B9RE31	B9RY83	B9SHQ0
B9SSJ6	B9SSJ7	B9SSJ8	B9U198
B9VRH2	B9W1M8	B9W7C3	B9W8W2
B9W8X6	B9W8X7	B9W8X8	B9W905
B9WD18	B9WII2	B9WIP0	B9WJE5

B9WKF1	B9WKJ1	B9WMT8	B9Z4S3
B9Z791	C0A799	C0GJG6	C0K3M4
C0KTZ9	C0LM31	C0LW36	C0VWM3
C0WHG9	C0Y3A2	C0YDP7	C0ZPM4
C0ZPM5	C0ZW13	C0ZZD7	C1AGL6
C1AWK0	C1AZZ1	C1B4K0	C1BA82
C1K6G4	C2C742	C2EM83	C2GL67
C2HYF2	C2I9U6	C2IF41	C2ITZ7
C2LFD0	C3K592	C3KBE9	C3KY70
C3L2E1	C3LUP1	C3MET7	C3NY10
C3VIF1	C4B002	C4I4I8	C4I862
C4W7I2	C4WBK9	C4WDU3	C5D9A0
C5N0N0	C5N1K6	C5NKA5	C5P2U8
C5PEB3	C5PF68	C5Q1M7	C5Q508
C5QAK5	C5QB30	C5QB84	C5QBB7
C5QNR2	C5QNV8	C5QPA8	C5QYQ4
C5QZ46	C5QZA7	C5QZE3	C5T1Z6
C5TAG4	C5TEA6	C5UWM9	C5UWN1
C5VNH6	C5VSR8	C5ZNU0	C5ZT46
C6BES3	C6BLI1	C6RQB4	C6RQB5
C6S0G0	C6U4P0	C6U825	C6WFI2
C6WJP6	C7BZE6	C7IPG8	C7LTP1
C7M226	C7Q5V1	C7QZV3	C7R0I5
C7RLB1	C8NNR9	C8NRL1	C8NRL2
C8PW62	C8RX47	C8STT3	C9NYE7
C9P528	C9PGS2	C9QJX7	C9QSM9
C9RTJ8	C9XVA9	C9Y4F1	C9YDZ4
C9YEH0	C9ZIJ5	D0H260	D0HU78
D0I108	D0IJN7	D0LBA3	D0MGA7
D0VTR1	D1A1K1	D1A2G9	D1A2H1
D1A646	D1A8A5	D1A9G5	D1C198
D1GU03	D1RQD7	D1RXL8	D1WPX0
D1WQ04	D1XGB7	D1XKT2	D2AXW6
D2BFB4	D2CGD5	D2KE05	D2N423
D2N490	D2NAQ3	D2PMS0	D2Q2R5
D2Q844	D2R540	D2Z3K5	D3CSD1

D3CTK4	D3D110	D3D2K0	D3D6G2
D3EST5	D3ETZ1	D3EZZ4	D3FB47
D3FC90	D3PZ28	D3Q6I6	D3Q953
D3R5P1	D3V150	D3V7P0	D3VEQ2
D3XB96	D3YC89	D3YH12	D3YKV3
D4DZP7	D4FLA8	D4FLE1	D4FLJ4
D4FLU5	D4H9R7	D4H9R8	D4HAV4
D4MP07	D4PHA8	D4XNS7	D4XPN5
D4XU47	D4YRV0	D5AP46	D5BS18
D5DJ21	D5E2W8	D5HFC3	D5MM05
D5P2S7	D5PA62	D5RBH3	D5UPT4
D5UVJ3	D6DDV7	D6SBP8	D6SDM6
D6UE07	D6UEY3	D6ZKF0	D7AYZ5
D7D605	D7SFB5	D7URU5	D7VN99
D8GZ72	D8H4C5	D8KJU6	D8L2D9
D8LPK4	D8ND21	D9ID42	D9SZ22
D9T622	D9T737	D9TAI6	D9TEX9
E0KCZ3	E0KGK4	E0KS38	E0KVE1
E0MUJ8	E0N3J4	E0P5G1	E0P691
E0Q925	E0QHH4	E0V9B5	E0V9B6
E0V9B8	E0VAX8	E0VE21	E0VEM4
E0VGW4	E0VGW7	E0VIM4	E0VJB4
E0VL85	E0VQZ5	E0VRR9	E0VUI9
E0VV11	E0W3V6	E0W3V9	E0W3W0
E1B2U7	O08369	O13444	O13934
O28511	O46559	O59952	O66015
O68310	O68551	O88354	O94091
P00591	P04634	P04635	P07098
P07867	P07882	P08658	P0C0R3
P0C0R4	P0C548	P10335	P11150
P15493	P16233	P17573	P17892
P19515	P19833	P19835	P20261
P22088	P22394	P22760	P24484
P24640	P25275	P25641	P26504
P26876	P26877	P27656	P27657
P29183	P29605	P30122	P32946

P32947	P32948	P32949	P36165
P37957	P40308	P40600	P40601
P41247	P41365	P41773	P50903
P54317	P54318	P54857	P61870
P61871	P61872	P65288	P65289
P77909	P79066	P80035	P81139
P83629	Q02104	Q02157	Q05489
Q09BX9	Q09KJ5	Q0BAH5	Q0CXU6
Q0E8R5	Q0JRM8	Q0KC09	Q0KI77
Q0MVP2	Q0MVP3	Q0P5B7	Q0PIK6
Q0PW43	Q0PW44	Q0QJZ3	Q0S2T7
Q0S7H9	Q0SDB8	Q0SE70	Q0SFH8
Q0V1P1	Q0VN25	Q0ZAX7	Q12043
Q12614	Q12616	Q17RR3	Q1B7F7
Q1BCD4	Q1BEY9	Q1BJT0	Q1BM22
Q1DZE0	Q1I484	Q1I849	Q1IBE9
Q1L776	Q1LGR0	Q1M315	Q1XBG1
Q1ZZV0	Q21T36	Q22KE1	Q22LP7
Q22RL6	Q22WB7	Q22Z77	Q22Z78
Q23FD6	Q24I21	Q29458	Q2FDJ1
Q2FJU4	Q2FUU5	Q2G155	Q2H6M8
Q2KI18	Q2KTB3	Q2KTB4	Q2KTB5
Q2LV59	Q2PQ79	Q2QFX1	Q2SS43
Q2SS44	Q2T7L1	Q2TPV1	Q2YVD0
Q2YZ74	Q38ZY5	Q390B6	Q393T2
Q397N4	Q39AV1	Q39B99	Q39BF9
Q39NZ2	Q3EJZ2	Q3EJZ3	Q3EK80
Q3EKS5	Q3EL23	Q3EL41	Q3EPJ6
Q3ETC5	Q3EVM6	Q3EX93	Q3IF07
Q3JIH9	Q3JKC4	Q3KBP5	Q3KIU1
Q3M991	Q3R236	Q3SZ79	Q3T5M2
Q3YJM9	Q45FB3	Q45VN4	Q47RJ6
Q47RJ7	Q4A627	Q4GYU8	Q4JL88
Q4MLA6	Q4PHZ2	Q4Q615	Q4Q6I0
Q4V566	Q4V6L4	Q4W9R3	Q4WAG0
Q4WAP6	Q4WBJ7	Q4WE77	Q4WEP8

Q4WHB3	Q4WNF9	Q4WW22	Q4WYQ5
Q4X180	Q58MF5	Q59260	Q59537
Q59538	Q59539	Q59644	Q59798
Q59811	Q59932	Q59933	Q59E63
Q5A4N0	Q5B0L1	Q5DPX0	Q5DRN8
Q5FZZ8	Q5GMI6	Q5HCM7	Q5HJ48
Q5HKF8	Q5HKP6	Q5I4I3	Q5KGQ6
Q5KL13	Q5KYG5	Q5MAI6	Q5NPT6
Q5S8F1	Q5U9G1	Q5U9G2	Q5VKJ7
Q5WDN0	Q5XTQ4	Q601F4	Q629P4
Q632H6	Q63AX5	Q63EX9	Q63HV1
Q63JH5	Q64285	Q64424	Q64425
Q670W0	Q67ZU1	Q6A6T8	Q6A8X9
Q6B4I1	Q6B4I3	Q6BLM0	Q6C2N7
Q6CU02	Q6F7I3	Q6FD56	Q6FPI6
Q6G604	Q6GCF1	Q6GDD3	Q6GJZ6
Q6HC20	Q6HIB3	Q6HME3	Q6KI90
Q6KIM4	Q6M5A0	Q6MPV3	Q6MT79
Q6MT80	Q6MTN3	Q6S5M9	Q6TRY9
Q6WER3	Q71DJ5	Q72GR1	Q75EN3
Q76D26	Q76D27	Q76D28	Q76D29
Q76D31	Q76D32	Q79F14	Q79SZ7
Q7A7P2	Q7JR83	Q7K3Z8	Q7LST4
Q7M370	Q7M4U7	Q7NUI4	Q7P4E7
Q7RYY1	Q7TX94	Q7U656	Q7UIK0
Q7UQZ0	Q7V291	Q7V8C5	Q802C5
Q816H0	Q818B9	Q81D60	Q81E50
Q81H00	Q82WB6	Q842J9	Q84EK3
Q86B55	Q872L3	Q872L4	Q872L5
Q872L6	Q875G8	Q875G9	Q875H0
Q894N1	Q896Q1	Q8AYE8	Q8BJ56
Q8IMS3	Q8L1V2	Q8L5T0	Q8L6B0
Q8MR12	Q8NIN8	Q8NJ51	Q8NQF4
Q8NR85	Q8NU59	Q8NU60	Q8NUI5
Q8NYC2	Q8QGW1	Q8RC83	Q8RG34
Q8RKT7	Q8U3I6	Q8VBX1	Q8VQP2

Q91UY0	Q91WW7	Q938A9	Q93A71
Q93J06	Q93MV2	Q93MW7	Q96AD5
Q96VC9	Q97VT8	Q97VW1	Q99156
Q99PG0	Q99WQ6	Q9BHD5	Q9CPP7
Q9EV86	Q9F0L9	Q9HDQ8	Q9HFW6
Q9HXQ8	Q9K5F4	Q9KJG6	Q9KX30
Q9L514	Q9L6C7	Q9L6D3	Q9LZA6
Q9NST1	Q9P451	Q9P4E5	Q9P4E6
Q9P4E7	Q9P4E8	Q9P8F7	Q9P8V9
Q9P8W0	Q9P8W1	Q9P8W2	Q9P8W5
Q9QZH8	Q9R4V3	Q9R4W9	Q9RBY1
Q9TS81	Q9TSB3	Q9V9N3	Q9VB88
Q9VB89	Q9VB90	Q9VB91	Q9VB92
Q9VB93	Q9VB94	Q9VBK6	Q9VG46
Q9VG47	Q9VG48	Q9VG50	Q9VJH1
Q9VKR3	Q9VKR5	Q9VKS5	Q9VKT1
Q9VKT2	Q9VKT9	Q9VLU1	Q9VPE9
Q9VQQ5	Q9VUH7	Q9VX01	Q9VZN6
Q9W195	Q9W448	Q9WVG5	Q9X2S3
Q9X5I2	Q9X6Z3	Q9Y5X9	Q9Z4M7

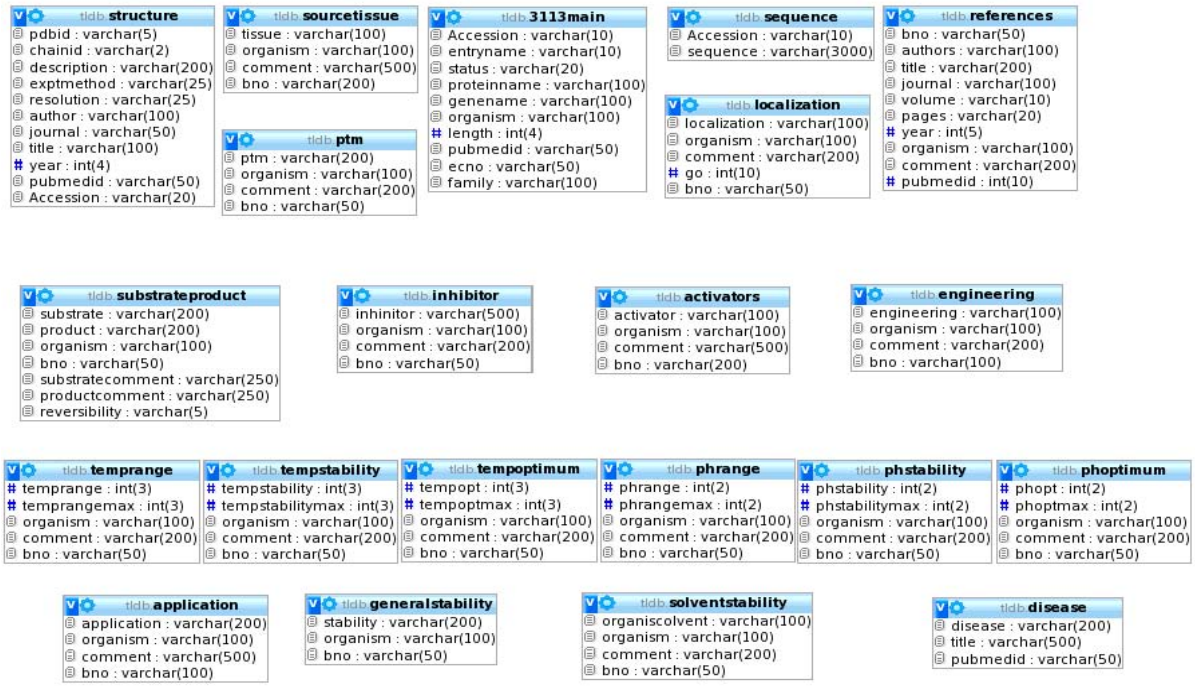
ANNEXURE II

List of Protein Data Bank Accession number of triacylglycerol lipases in our Triacylglycerol Lipase Database

1AKN	1BU8	1CRL	1CUA
1CUB	1CUC	1CUD	1CUE
1CUF	1CUG	1CUH	1CUI
1CUJ	1CUU	1CUV	1CUW
1CUX	1CUY	1CUZ	1CVL
1DT3	1DT5	1DTE	1DU4
1EIN	1ETH	1EX9	1F6W
1FFA	1FFB	1FFC	1FFD
1FFE	1GPL	1GT6	1GZ7
1HLG	1HPL	1HQD	1I6W
1ISP	1JI3	1JMY	1K8Q
1KU0	1LBS	1LBT	1LGY
1LLF	1LPA	1LPB	1LPM
1LPN	1LPO	1LPP	1LPS
1N8S	1OIL	1QGE	1R4Z
1R50	1RP1	1T2N	1T4M
1TAH	1TCA	1TCB	1TCC
1TGL	1THG	1TIA	1TIB
1TIC	1TRH	1YS1	1YS2
2DSN	2ES4	2FX5	2HIH
2LIP	2NW6	2ORY	2OXE
2PPL	2PVS	2QUA	2QUB
2QXT	2QXU	2W22	2Z5G
2Z8X	2Z8Z	2ZJ6	2ZJ7
2ZVD	2ZYH	2ZYI	2ZYR
2ZYS	3A6Z	3A70	3D2A
3D2B	3D2C	3G7N	3ICV
3ICW	3LIP	3TGL	4LIP
4TGL	5LIP	5TGL	

ANNEXURE III

Database schema and the organization of data implemented in Triacylglycerol Lipase Database.



ANNEXURE IV

Hyperlinks to Databases and Software used in the Thesis

- AutoDock (http://autodock.scripps.edu/)
- GROMACS package (http://www.gromacs.org/)
- MODELLER (http://salilab.org/modeller/)
- NCBI BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi)
- NCI diversity set II (http://dtp.nci.nih.gov/branches/dscb/diversity_explanation.html)
- NIH SAVES server (http://nihserver.mbi.ucla.edu/SAVES/)
- ProSA (www.came.sbg.ac.at/prosa.php)
- Protein model database (http://mi.caspur.it/PMDB/)
- Tetrahydrolipstatin (http://www.drugbank.ca/drugs/DB01083)
- UniprotKB database (http://www.uniprot.org/)

PUBLICATIONS

PUBLICATIONS

A. Publications from Thesis Work

1. **Saravanan P**, Avinash H, Dubey VK and Sanjukta P. (2012) Targeting essential cell wall lipase Rv3802c for potential therapeutics against tuberculosis. *J Mol Graphics Modell*. 38:235-242.
2. **Saravanan P**, Dubey VK and Sanjukta P. (2012) Potential selective inhibitors against Rv0183 of *Mycobacterium tuberculosis* targeting host lipid metabolism. *Chem Biol Drug Des*. 79(6):1056-1062.
3. **Saravanan P**, Dubey VK and Sanjukta P. (2011) *In silico* characterization of thermoactive, alkaline and detergent-stable lipase from *Staphylococcus aureus* strain. *In Silico Biol*. 10(5):265-276.
4. **Saravanan P**, Alpana AT and Sanjukta P. (2010) Deciphering Role of Amino Acids for the Stability of *Staphylococcus aureus* Lipase (SAL3). *Interdiscip Sci Comput Life Sci*. 2(3):271-279.

***Manuscript in Preparation*:**

1. **Saravanan P**, Debamitra C and Sanjukta P. Triacylglycerol Lipase Database.
2. **Saravanan P** and Sanjukta P. Structural investigations of potential lipases: Computational studies on *in silico* mutants to emulate structural stability.
3. **Saravanan P** and Sanjukta P. Identification of potential multi-target inhibitors targeting both cell wall and lipid metabolism to combat tuberculosis.
4. **Saravanan P**, Prakash S, Dubey VK and Sanjukta P. Potential inhibitors of LdLip3 lipase targeting pathogens lipid metabolism to combat Leishmaniasis.

B. Publications from Research Collaborations

1. Arvind A, Jain V, **Saravanan P** and Mohan CG. (2013) Uridine Monophosphate Kinase as Potential Target for Tuberculosis: From Target to Lead Identification. *Interdiscip Sci Comput Life Sci*. [*Accepted*].
2. Arvind A, Kumar V, **Saravanan P** and Mohan CG. (2012) Homology modeling, molecular dynamics and inhibitor binding study on MurD ligase of *Mycobacterium tuberculosis*. *Interdiscip Sci Comput Life Sci*. 4(3):223-238.
3. Khan AT, Lal M, Bagdi PR, Basha RS, **Saravanan P**, Sanjukta P. (2012) Synthesis of tetra-substituted pyrroles, a potential phosphodiesterase 4B inhibitor, through nickel(II) chloride hexahydrate catalyzed one-pot four-component reaction. *Tetrahedron Letters* 53(82):4145–4150.
4. Debamitra C, Sumit KS, **Saravanan P**, Sanjukta P. (2012) Design of lead peptide drugs from mushroom targeting cysteine proteases. *Med Chem Res*. 22(4): 2038-2049.
5. Debamitra C, **Saravanan P**, Dubey VK and Sanjukta P. (2012) Unraveling the Rationale behind Organic Solvent Stability of Lipases. *Appl Biochem Biotech*. 167(3):439-461.
6. Chakraborty D, **Saravanan P**, Sanjukta P and Dubey VK. (2012) Studies on ornithine decarboxylase of *Leishmania donovani*: Structure modeling and inhibitor docking. *Med Chem Res*. 22(1): 466-478.
7. Debamitra C, **Saravanan P**, Dubey VK and Sanjukta P. (2011) *In Silico* Characterization of Thermostable Lipases. *Extremophiles* 15(1):89-103.
8. Jain V, **Saravanan P**, Arvind A and Mohan CG. (2011) First Pharmacophore Model of CCR3 Receptor Antagonists and its Homology Model-Assisted, Stepwise Virtual Screening. *Chem Biol Drug Des*. 77(5):373-387.
9. Shikhar G, Adyary F, Mikko JV, **Saravanan P**, Santeri P, Paivi J, Mark JS, Pia V, Mohan, CG. (2011) Molecular Docking Guided Comparative GFA, G/PLS, SVM and ANN models of Structurally Diverse Dual Binding Site Acetylcholinesterase Inhibitors. *Mol Inform*. 30:689–706.

10. **Saravanan P**, Santhosh KV, Mohan CG**,** Sanjukta P and Dubey VK. (2010) Mitogen-activated protein kinase 4 of Leishmania parasite as a therapeutic target. *European J Med Chem*. 45(12):5662-5670.

11. Kumar V, **Saravanan P**, Arvind A, Mohan CG. (2010) Identification of hotspot regions of MurB oxidoreductase enzyme using homology modeling, molecular dynamics and molecular docking techniques. *J Mol Model*. 17(5):939-953.

12. Awale M, Kumar V, **Saravanan P**, Mohan CG. (2009) Homology modeling and atomic level binding study of Leishmania MAPK with inhibitors. *J Mol Model*. 16(3):475-88.

C. Participations in Conferences/Seminars/Meetings

1. **Saravanan P** and Sanjukta P. "Identification of potential multi-target inhibitors for tuberculosis" presented in International conference on Biomolecular forms and functions organized by Indian Institute of Science in 8-11 Jan 2013. [***Poster presentation***]

2. **Saravanan P** and Sanjukta P. "True Lipase Database" presented in European Conference on Computational Biology, Basel, Europe organized by Swiss Institute of Bioinformatics in 9-12 Sep 2012. [***Poster presentation***]

3. **Saravanan P** and Sanjukta P. "Targeting Essential Cell Wall Lipase Rv3802c to Combat Tuberculosis: Homology Modeling, Virtual Screening and Comparative Docking Studies" in International Conference for Open Source Computer-Aided Translational Medicine organized by Institute of Microbial Technology in 22-25 Feb 2012. [***Oral presentation***]

4. **Saravanan P** and Sanjukta P. "Potential inhibitors targeting LdLip3 to combat Leishmaniasis" presented in Indian Biophysical Meeting organized by University of Madras in 19-21 Jan 2012. [***Poster presentation***]

5. **Saravanan P** and Sanjukta P. "In silico characterization and virtual screening of Rv0183, a monoglyceride lipase from Mycobacterium tuberculosis" presented in Society of Biological Chemists (India) meeting organized by Indian Institute of Science in 13-15 Dec 2010. [***Poster presentation***]

6. **Saravanan P** and Sanjukta P. "In Silico Characterization and Structural Modeling of Thermoactive and Alkaline Staphylococcus Lipase" presented in Asia-Pacific Bioinformatics Conference organized by National Centre for Biological Sciences in 18-21 Jan 2010. [***Poster presentation***]