Journal of Molecular Graphics and Modelling 38 (2012) 235–242

Contents lists available at SciVerse ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/JMGM

Targeting essential cell wall lipase Rv3802c for potential therapeutics against tuberculosis

Parameswaran Saravanan, Hindupur Avinash, Vikash Kumar Dubey, Sanjukta Patra*

Department of Biotechnology, Indian Institute of Technology Guwahati, Guwahati 781039, Assam, India

ARTICLE INFO

Article history:
Received 16 January 2012
Received in revised form 15 June 2012
Accepted 19 June 2012
Available online 4 August 2012

Keywords:
Rv3802c
Tuberculosis
Lipase
Cell wall
Lipid metabolism
Homology modelling
Virtual screening

ABSTRACT

Cell wall and lipid metabolism plays a vital role in the survival and infection of *Mycobacterium tuberculosis*. Increase in the incidences of life-threatening multidrug-resistant (MDR) and extreme drug-resistant (XDR) tuberculosis worsens the existing scenario and urge the need of new druggable targets and new drugs. Targeting Rv3802c, an essential cell wall lipase, can open up a new arsenal to fight the dreadful opportunistic pathogen. Our current study highlights the essentiality of Rv3802c. Its 3D structure is predicted for the first time which provides insight in identifying the ligand binding sites. Our analysis showed Rv3802c is highly conserved throughout mycobacterial species with no significant sequence homolog found in human proteome. Virtual screening followed by comparative docking studies of Rv3802c with its closest human structural homolog has been carried out to identify potential inhibitors effective towards mycobacterial proteins. Two diverse molecules from ZINC database, ZINC26726377 and ZINC43866786 have been identified as potential inhibitors effective towards Rv3802c based on the difference in predicted binding free energy of −3.99 and −3.28 kcal/mol respectively. Rv3802c is a promising drug target and also a step towards understanding and targeting the pathogen's cell wall and lipid metabolism simultaneously to combat tuberculosis.

1. Introduction

It has been 90 years since BCG was first administered in humans as a vaccine for tuberculosis [1], caused by *Mycobacterium tuberculosis*. Despite long and widespread usage of the vaccine and the drugs for tuberculosis across the globe, one-third of world's population is still infected [2]. Causing 1.45 million deaths and 8.8 million fresh cases in 2010, tuberculosis continues to be one of the deadliest diseases [3]. The highly complex and unique cell wall is very critical for the survival and infection of mycobacteria within the host cells [4]. It is reinforced by the fact that 15.5% of the total coding region of its genome is related to cell wall and related proteins [5,6]. The cell wall core – mycolyl arabinogalactan peptidoglycan (mAGP) complex found beyond the plasma membrane is insoluble when cell wall is disrupted [7]. The insoluble core is essential for the viability of the cell and thus targeted in the context of discovery of new drugs [8]. Mycolic acids in mAGP complex are long-chain α-alkyl β-hydroxy fatty acids, the signature lipids of the hydrophobic mycobacterial cell wall [9]. Mycolic acids are essential for the survival of the microorganism and also contribute to the pathogenesis of *M. tuberculosis* through a variety of mechanisms [10,11]. They are also the determining factor for permitting molecules through the cell wall [12]. 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 [13].

The genetic cluster Rv3799c–Rv3804c is dedicated to the synthesis of core structures of the mycobacterial cell wall, including mycolic acids and arabinogalactan and the genetic cluster is found to be conserved in actinomycetes including *Corynebacterium glutamicum* [14]. Rv3802c ortholog in *C. glutamicum* is Ncgl2775 which plays a key role in the regulation of the outer lipid composition under heat stress conditions. Rv3802c with their orthologs were proposed to be envelope lipids regulation factor (ElrF) family [15]. Rv3802c of *M. tuberculosis* is a secreted protein found as an integral part of the cell wall and could participate in cell wall metabolism, suggesting its distinct role in mycobacteriology [16–18]. Rv3802c is characterized as phospholipase/thioesterase and is active towards all fatty acid chain lengths of nitrophenyl esters [19]. Secreted lipases of *M. tuberculosis* like Rv3802c need to be studied in detail to understand the mechanisms of pathogenesis, as they are secreted at key stages of infection and the pathogen depends on its lipid metabolism extensively for many of its needs [20–23].

Rv3802c gained importance because of its location in a conserved mycolic acid biosynthesis gene cluster and its essentiality, which was validated from mutagenesis experiment [24]. Of the seven putative cutinases present in *M. tuberculosis*, Rv3802c is the only one reported to be essential and translocated to

* Corresponding author. Tel.: +91 361 2582213; fax: +91 361 2582249.
E-mail address: sanjukta@iitg.ernet.in (S. Patra).

1093-3263/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jmgm.2012.06.016

cell wall [25,26]. Rv3802c is also conserved in *Mycobacterium leprae*, which is considered a minimal genome [27]. Conditional disruptions of Rv3802c counterpart in *Mycobacterium smegmatis* leads to the loss of cell wall integrity and internal structure of the cell [28]. All these evidences point to the essentiality of Rv3802c. The human lipase inhibitor, tetrahydrolipstatin (http://www.drugbank.ca/drugs/DB01083), inhibits both MSMEG_6394 and Rv3802c with varied degree of inhibition drives us to search for potential inhibitors effective towards mycobacterial proteins [19,28,29].

Considering the recent increase in incidences of life-threatening multidrug resistant (MDR) and extreme drug resistant (XDR) tuberculosis [30,31], finding a new drug target and new drug to cripple pathogen's survival and infection is of high importance which may be achieved by targeting both cell wall and lipid metabolism.

2. Methods

2.1. Sequence analysis

The Rv3802c sequence [UniProt ID: O53581] and its representative homologs from mycobacterial species were collected from Uniprot database [http://www.uniprot.org]. The retrieved sequences were analysed using ClustalW [32] [http://www.ebi.ac.uk/Tools/msa/clustalw2] to identify the conservation of amino acids, motifs and protein domain.

2.2. Homology modelling

Template for homology modelling was identified using 'BLASTP' program (http://blast.ncbi.nlm.nih.gov/Blast.cgi) against Protein Data Bank (PDB) database to identify the closest structural homologs. MSMEG_6394 of *M. smegmatis* [PDB ID: 3AJA] was selected as the template for structure modelling studies based on its sequence and functional homology. Homology modelling of Rv3802c was carried out with MODELLER9v7 [33] which generated 100 initial models. These were ranked based upon their Discrete Optimized Protein Energy (DOPE) [34] and MOLPDF scores. The best models having the least DOPE and MOLPDF scores with acceptable statistics from Ramachandran plot were selected. Validation studies were carried out on selected Rv3802c models using PROCHECK [35] and ERRAT plot [36] from structure validation package using NIH SAVES server (http://nihserver.mbi.ucla.edu/SAVES/). Native protein fold evaluation on validated Rv3802c model was carried out using the knowledge-based energy profile of ProsaII program [37]. The best Rv3802c model from structure validation studies was energy minimized with GROMACS4.0.7 package (http://www.gromacs.org/) using steepest descent algorithm with GROMOS96 43a1 force field and simple point charge (SPC) water model [38–40].

2.3. Molecular dynamics

Molecular dynamics (MD) simulation on the energy minimized Rv3802c model was carried out with the periodic boundary conditions applied in three dimensions to analyse the stability of the protein model. The net charge of system was neutralized by the addition of eleven sodium ions by replacing water molecules that are at least 3.50 Å from the protein surface. The solvent was equilibrated for 100 ps by restraining the solute atoms through a harmonic force constant of 1000 kJ nm^{-2} in NVT ensemble followed by NPT ensemble. NPT production run was carried out for 10 ns with no restraints using 2 fs of integration time. All MD simulations were carried out with temperature of 300 K 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. 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 [41,42]. As a whole, 2000 conformations from production run were analysed to check whether the Rv3802c model is energetically stable in the solvent system.

2.4. Virtual screening

Two varied approaches were followed to find out the potential inhibitors for Rv3802c using structure-based virtual screening which may enrich lead molecules for drug discovery in pharmaceutical industry for translational research [43]. In both approach, the screening was carried out with AutoDock4.2 (http://autodock.scripps.edu/) using Lamarckian genetic algorithm (LGA) [44,45] considered to be one of the best for virtual screening studies [46]. In the first approach, the initial small molecule dataset used in the study was the NCI diversity set (http://dtp.nci.nih.gov/branches/dscb/div2_explanation.html) containing 2654 structurally diverse chemical molecules including isomers. The choice of NCI diversity set greatly reduces the computational time as it contains only the representative molecules of wide and diverse kinds. Hydrogen was added to all the residues of Rv3802c model 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 [47]. All ligands were similarly processed using 'prepare_ligand4.py' script available with AutoDockTools distribution. The entire active site region of Rv3802c model was covered by a grid box with $90 \times 90 \times 90$ points and 0.375 Å of grid spacing keeping catalytic serine at the centre of grid box. Electrostatic and desolvation maps along with grid maps for all the atom types of ligand and Rv3802c model were generated with AutoGrid program. The whole docking procedure was then repeated with the closest human structural ortholog – monoglyceride lipase (MGL) [48] to find the potential inhibitors effective towards Rv3802c. Docking simulations on diversity set were performed with an initial population size of 300 and 10 independent LGA runs. In each run, the best individual from each generation was propagated to the next generation (elitism) and remaining docking parameters were set to default.

Following the rigid docking protocol, the results were analysed based on the predicted free energy of binding. The molecules with highest difference in binding energies of Rv3802c and human MGL were chosen for further analysis. The top 10 compounds were searched in ZINC database [49] based on SMILES for similar molecules with a cut off of 60%. This resulted in 23,479 molecules including isomers which were used in docking analysis on both Rv3802c and human MGL. Docking simulations on similar molecules were carried out with an initial population size of 300 and 20 independent LGA runs with stringent long evaluations.

The second approach was to consider the recently identified leads (four potent molecules than known inhibitor, tetrahyrolipstatin) that have been reported for its inhibitory action against Rv3802c in laboratory experiments by West et al. [25] as the initial dataset. Similarity search of the initial dataset was carried out on ZINC database with a cut-off of 60%. This resulted in 8023 molecules including isomers. Docking calculations for second approach were carried out with an initial population size of 300 and 20 independent LGA runs with stringent long evaluations. Docking calculations were validated with the help of the co-crystallized structure of human MGL with reversible inhibitor [PDB ID: 3PE6] [50] to check whether docking can predict/mimic the native pose of the inhibitor.

3. Results and discussion

3.1. Sequence analysis

M. tuberculosis Rv3802c comprises of 336 amino acids with a molecular mass of 35,448 Da, a lipase with significant phospholipase A and thioesterase activity [19]. Based on the experimental studies on Rv3802c essentiality and its role in the organism's cellular integrity from *M. smegmatis*, Rv3802c is considered to be good drug target for treating tuberculosis [28,29]. 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 S1). 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. 1). Fig. 1 was generated using ESpript [51]. 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 [52] 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 which has predicted secondary structure of alpha helices and beta strands [53,54]. Interestingly, MSMEG_6394 has been used by Sultana et al. [55] 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 [56] which might have role in translocating Rv3802c to mycobacterial cell wall.

3.2. Homology modelling

From sequence analysis and NCBI search against PDB database, the functional homolog of Rv3802c in *M. smegamatis*, MSMEG_6394 with a resolved crystal structure of 2.9 Å was identified as the

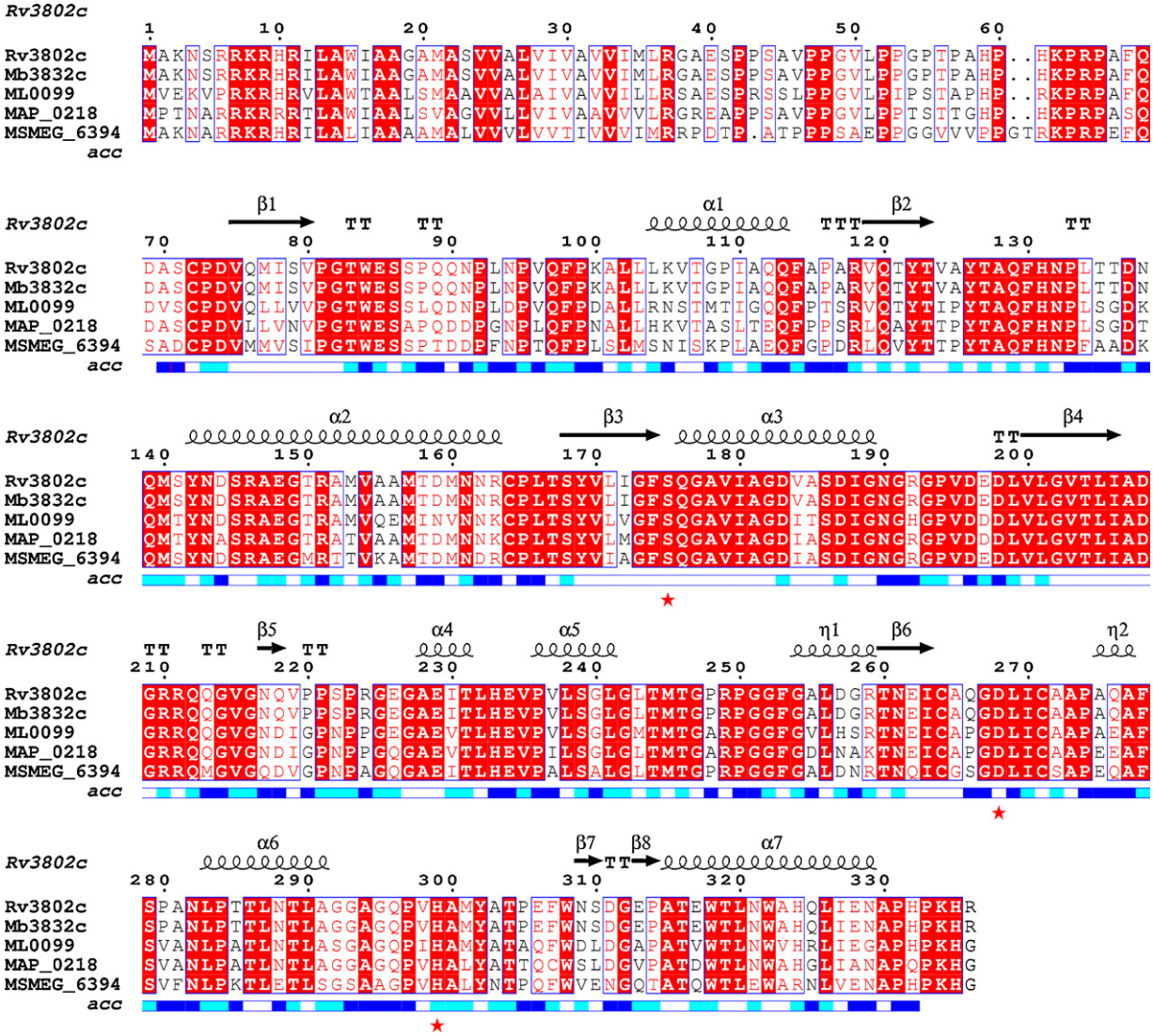

Fig. 1. Multiple sequence alignment of Rv3802c of *M. tuberculosis* with its mycobacterial homologs reflects that the amino acids are conserved throughout the protein. Numbering of amino acids is denoted with respect to Rv3802c and its secondary structures are also shown. The catalytic residues are shown with star symbol. Surface accessibility is depicted for the modeled region with white, cyan and blue colour for buried, intermediate and accessible residues respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of the article.)

This figure was produced using ESPript [51].

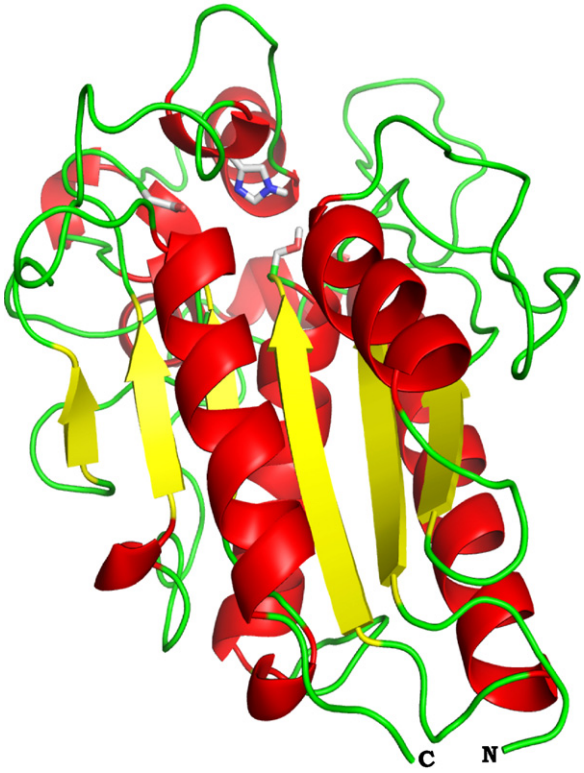

Fig. 2. The modeled structure of Rv3802c in cartoon representation reflects the α/β hydrolase fold and the catalytic residues are shown in stick representation. This figure was produced using PyMOL [61].

template for homology modelling studies. Another contributing factor for its selection was its inhibition by tetrahydrolipstatin. The modelled structure confirmed that Rv3802c is a member of cutinase family of α/β hydrolases (Fig. 2). 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) is represented with "acc" legend which reflects surface accessibility of the amino acids, is depicted in Fig. 1. The putative function of the non-modelled region could be anchoring Rv3802c in the cell wall so that it could participate in cell wall remodelling. 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. 2. 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] [57] identifies Thr83 and Gln176 as the putative oxyanion hole residues [55].

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. S1. 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. S2). As expected, 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 [58]. The fold quality of Rv3802c from ProsaII was comparable with experimentally determined MSMEG_6394 indicates the acceptability of the modelled Rv3802c structure (Fig. S3).

Structural comparison of Rv3802c model with MSMEG_6394 on Cα-backbone atoms shows overall Root Mean Square Deviation (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 also be the reason that tetrahydrolipstatin inhibits Rv3802c effectively when compared with MSMEG_6394 and more experimental studies would provide insights in specific inhibiting potential of tetrahydrolipstatin.

3.3. Molecular dynamics

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. 3. The graph clearly indicates that there is significant change in RMSD for the initial 5 ns 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.

3.4. Virtual screening

Structure-based virtual screening (SBVS) is the most preferred technique to identify novel potential inhibitors against protein of interest. The Rv3802c structure was modelled for first time, with little known about its inhibitors or the structural features required for better inhibition. SBVS was used to identify potential inhibitors targeting Rv3802c active site to serve as possible starting points for translational research. To summarize, structurally diverse compounds were virtually screened to identify potential inhibitors for alleviated drug therapy against tuberculosis. Docking protocol was validated using human MGL (PDB ID: 3PE6), the structural homolog of Rv3802c which is co-crystallized with reversible inhibitor (PDB ligand id: ZYH). Docking calculations could mimic the pose of reversible inhibitor reasonably well. Flexible docking by keeping the side chain of catalytic serine flexible was not satisfactory. Superimposition of the inhibitors pose between crystal structures and docking calculations are shown in Fig. S4. So docking protocol used in our virtual screening studies would be effective and preferable to screen large number of ligands.

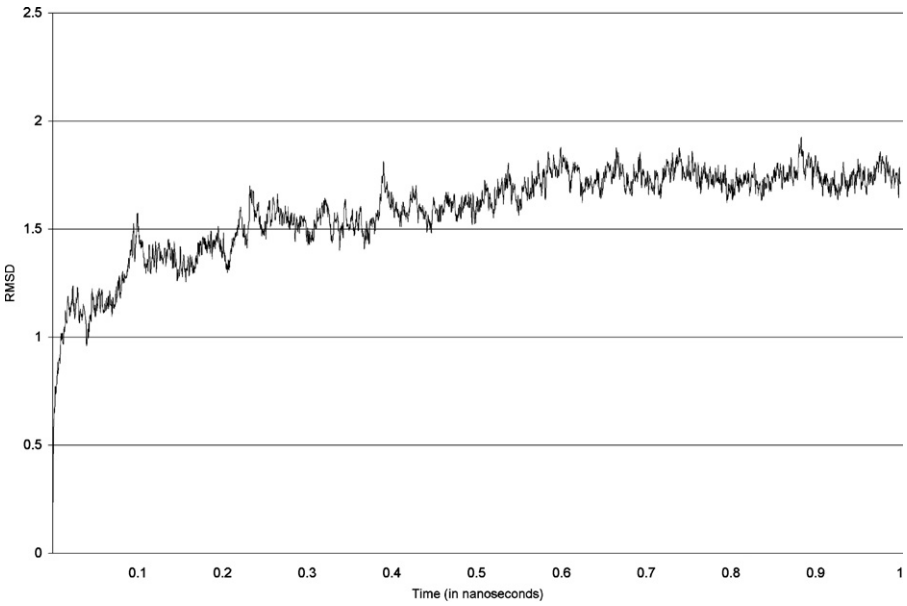

Fig. 3. MD simulation of modeled Rv3802c shows that the energy-minimized structure is energetically stable.

Most of the known lipase inhibitors are irreversible inhibitors which form covalent bond with catalytic serine, for instance SAR269 [48], JZL184 [59,60] and tetrahydrolipstatin [21]. 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. Screening was carried out again with 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 specific inhibitors against Rv3802c and the top hits of initial screening have been listed in Table S2. Similarity search (60% similarity using tanimotto coefficient) of 10 top hits from initial screening was performed on ZINC database and the virtual screening with above protocol was performed again with 20 GA runs on the similar molecules to find potential inhibitors which are specific towards Rv3802c (secondary screening). The top hits of secondary screening are listed in Table 1. Second approach identified the top hits from similarity screening considering the potent compounds reported by West et al. as initial dataset are listed in Table 2. Top hits of computational studies were visually inspected and analysed for all possible hydrogen bond and hydrophobic interactions with Rv3802c using PyMOL [61]. The hydrogen bond interaction mode of the best hits from approach I and II are shown in Fig. 4 respectively which are energetically favourable and may provide effective inhibition. Best hit ZINC26726377 from approach I forms hydrogen bond interactions with backbone NH atom of Thr83, and also with side chain atoms of Ser175, Lys100 and His299. Lys100 is also forming a charge-π interaction with the aromatic ring of the ZINC26726377. The binding pattern of putative oxyanion hole Thr83 and catalytic Ser175 with ZINC26726377 was also observed in the human MGL with an irreversible inhibitor (PDB ID: 3PE6) makes us speculate that the binding of ZINC26726377 could provide better inhibition. Best hit ZINC43866786 from approach II forms hydrogen bond interactions with backbone O atom of Lys100, and also with side chain atom of Glu85 which also has charge interaction with the ZINC43866786. The details of the hydrogen bond of the best hits have been listed in Table S3. The best hits were also

Table 1
Top 10 hits selected on secondary screening using similar molecules from ZINC database based on the difference in predicted free energy of binding (ΔG) according to AutoDock. Number of conformations in cluster (CL) with best energy top hits is given in square brackets.

S. No.	Molecule zinc database ID	Free energy of binding (kcal/mol)			Hydrogen bond	Contact residues
		Rv3802c ΔG [CL]	Human MGL ΔG [CL]	ΔG Difference		
1.	ZINC26726377	−10.5 [5]	−6.51 [1]	−3.99	T83, S175, H299, A300, K100	G82, F98, L102, F174, G292, G293, M301, A303
2.	ZINC43422783	−11.55 [4]	−8.29 [1]	−3.26	H88	G82, W84, A101, F174, G292, A294, P297, A300, M301, A303,
3.	ZINC34801822	−10.9 [7]	−7.97 [1]	−2.93	K100, G293	G82, W84, F174, G292, M301, Y302
4.	ZINC58522851	−12.18 [7]	−9.26 [1]	−2.92	W84, E85	G82, F98, K100, A101, D137, Q139, F174, G292, A291
5.	ZINC34801818	−11.05 [4]	−8.17 [1]	−2.88	K100, G293	G82, W84, F174, G292, M301
6.	ZINC36089082	−10.62 [5]	−7.81 [1]	−2.81	K100	G82, W84, F174, G292, M301
7.	ZINC43860875	−10.37 [3]	−7.74 [1]	−2.63	K100, G293	G82, W84, F174, G292, M301, Y302
8.	ZINC34801820	−11.52 [1]	−8.91 [2]	−2.61	K100, G293	G82, W84, F174, S175, G292
9.	ZINC28262919	−9.49 [7]	−6.89 [1]	−2.6	–	G82, F98, A101, F174, A291, G292, G293, M301, A303
10.	ZINC06616133	−9.9 [19]	−7.38 [7]	−2.52	–	T83, W84, A101, F174, G292, M301, T316

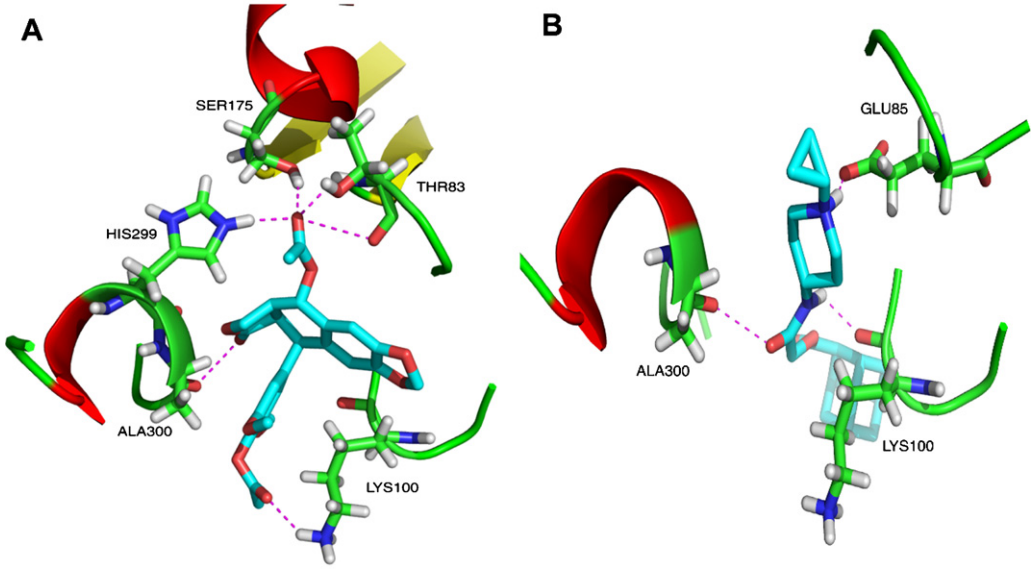

Fig. 4. Hydrogen bond interaction mode of the best hits ZINC26726377 and ZINC43866786 with Rv3802c from approaches I and II respectively. This figure was produced using PyMOL [61].

Table 2
Top 10 hits selected on secondary screening using similar molecules using *in vitro* inhibitors derived from tetrahydrolipstatin based on the difference in predicted free energy of binding (ΔG) according to AutoDock. Number of conformations in cluster (CL) with best energy top hits is given in square brackets.

S. No.	Molecule zinc database ID	Free energy of binding (kcal/mol)			Hydrogen bond	Contact residues
		Rv3802c ΔG	Human MGL ΔG	ΔG difference		
1.	ZINC43866786	−10.82 [19]	−7.54 [2]	−3.28	E85, K100, A300	G82, W84, P99, L104, K105, F174, S175, G293, H299, A303, T316
2.	ZINC48658902[a]	−9.15 [20]	−6.59 [2]	−2.56	E85, K100, S175, H299	G82, W84, A101, K105, F174, G292, A303
3.	ZINC48658902[b]	−9.15 [20]	−6.62 [7]	−2.53	E85, K100, S175, H299	G82, W84, A101, K105, F174, G292, A303
4.	ZINC42157073[a]	−8.4 [20]	−6.41 [12]	−1.99	E85, S175, H299	G82, W84, A101, K105, F174, S175, G292
5.	ZINC00052222[a]	−9.6 [14]	−7.67 [9]	−1.93	T83, E85, Y302	W84, K100, A101, F174, S175, G292, G293, M301
6.	ZINC48658901[a]	−8.55 [20]	−6.63 [2]	−1.92	E85, S175, H299, K100	G82, W84, A101, K105, F174, G292, M301, A303
7.	ZINC48658901[b]	−8.55 [19]	−6.65 [4]	−1.9	S175, E85, H299, K100	G82, W84, A101, K105, F174, G292, A303
8.	ZINC42157073[b]	−8.4 [20]	−6.51 [2]	−1.89	E85, S175, H299	G82, W84, A101, K105, F174, S175, G292
9.	ZINC00052222[b]	−9.74 [12]	−7.88 [2]	−1.86	T83, E85, Y302	W84, K100, A101, F174, S175, G292, G293, M301
10.	ZINC48658902[c]	−8.23 [20]	−6.57 [7]	−1.66	K100, S175, H299	G82, W84, A101, K105, F174, G292, M301, A303

[a,b,c]Molecule at different protonation or tautomeric state.

stabilized by the hydrophobic residues of the cavity mainly by Lys100 and Ala300 through aromatic–aliphatic hydrophobic interaction (Fig. S5).

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, Lys100, Ala101, Leu102, Lys105, Phe174, Ser175, Gly292, Gly293, H299, Ala300, Met301 and Tyr302. It is noteworthy to mention that Ser175, Asp268 and His299 are conserved catalytic residues. 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 pharmacophore of the top hits were visually inspected and it suggests that these pharmacophores and the interactions of the ligands 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 [62] which may prevent the pathogen enter dormancy, may reduce the therapy time and may be active on natural variant strains of *M. tuberculosis*.

4. Conclusions

Targeting new drug targets which participates in both cell wall and lipid metabolism of *M. tuberculosis* like Rv3802c might be effective way to fight against tuberculosis. Being a cutinase-like protein, Rv3802c is significantly divergent from human proteome. Several evidences suggest that essential cell wall lipase, Rv3802c can be attractive drug target and worth pursuing to identify potential inhibitors. Molecular modelling and structure based

virtual screening strategy targeting the substrate binding pocket of Rv3802c identified two structurally diversed molecules as potential inhibitors. Further, the identified inhibitors may have desirable inhibitory effect as anti-tuberculosis drugs against active, MDR and XDR tuberculosis which can be validated by biological tests. Our current study on Rv3802c might be another step towards understanding and spur novel drug design against tuberculosis. Research on Rv3802c and their potential inhibitors and synergistic effects with DOTS therapy needs to be intensified to eradicate tuberculosis.

Conflicts of interest

None declared.

Acknowledgements

Infrastructural facilities provided by Indian Institute of Technology Guwahati and Financial support by Department of Information Technology to PS in the form of research grants are acknowledged. Homology-modelled structure of Rv3802c is submitted to Protein Model Data Base (http://mi.caspur.it/PMDB/) with the identification number PM0077735.

Ethical approval: None required.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jmgm.2012.06.016.

References

[1] N. Ritz, N. Curtis, Mapping the global use of different BCG vaccine strains, Tuberculosis 89 (2009) 248–251.
[2] E.L. Corbett, C.J. Watt, N. Walker, D. Maher, B.G. Williams, M.C. Raviglione, C. Dye, The growing burden of tuberculosis: global trends and interactions with the HIV epidemic, Archives of Internal Medicine 163 (2003) 1009–1021.
[3] World Health Organization, Global Tuberculosis Control 2011 (2011).
[4] C.E. Barry III, Interpreting cell wall 'virulence factors' of *Mycobacterium tuberculosis*, Trends in Microbiology 9 (2001) 237–241.
[5] S.T. Cole, R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S.V. Gordon, K. Eiglmeier, S. Gas, C.E. Barry, F. Tekaia, K. Badcock, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. Davies, K. Devlin, T. Feltwell, S. Gentles, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, A. Krogh, J. McLean, S. Moule, L. Murphy, K. Oliver, J. Osborne, M.A. Quail, M.A. Rajandream, J. Rogers, S. Rutter, K. Seeger, J. Skelton, R. Squares, S. Squares, J.E. Sulston, K. Taylor, S. Whitehead, B.G. Barrell, Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence, Nature 393 (1998) 537–544.
[6] S.T. Cole, Learning from the genome sequence of *Mycobacterium tuberculosis* H37Rv, FEBS Letters 452 (1999) 7–10.
[7] P.J. Brennan, Structure, function, and biogenesis of the cell wall of *Mycobacterium tuberculosis*, Tuberculosis 8 (3) (2003) 91–97.
[8] P.J. Brennan, D.C. Crick, The cell-wall core of *Mycobacterium tuberculosis* in the context of drug discovery, Current Topics in Medicinal Chemistry 7 (2007) 475–488.
[9] K. Todar, Online Textbook of Bacteriology, The Good, the Bad, and the Deadly, vol. 304, Science Magazine, 2004, pp. 1421–1632.
[10] L. Guenin-Macé, R. Siméone, C. Demangel, Lipids of pathogenic Mycobacteria: contributions to virulence and host immune suppression, Transboundary and Emerging Diseases 56 (2009) 255–268.
[11] J. Pieters, Evasion of host cell defense mechanisms by pathogenic bacteria, Current Opinion in Immunology 13 (2001) 37–44.
[12] J. Liu, C.E. Barry, G.S. Besra, H. Nikaido, Mycolic acid structure determines the fluidity of the mycobacterial cell wall, Journal of Biological Chemistry 271 (1996) 29545–29551.
[13] H.I. Boshoff, C.E. Barry, Is the mycobacterial cell wall a hopeless drug target for latent tuberculosis? Drug Discovery Today: Disease Mechanisms 3 (2006) 237–245.
[14] K. Takayama, C. Wang, G.S. Besra, Pathway to synthesis and processing of mycolic acids in *Mycobacterium tuberculosis*, Clinical Microbiology Reviews 18 (2005) 81–101.
[15] X. Meniche, C. Labarre, C. de Sousa-d'Auria, E. Huc, F. Laval, M. Tropis, N. Bayan, D. Portevin, C. Guilhot, M. Daffé, C. Houssin, Identification of a stress-induced factor of Corynebacterineae that is involved in the regulation of the outer membrane lipid composition, Journal of Bacteriology 191 (2009) 7323–7332.
[16] H. Song, R. Sandie, Y. Wang, M.A. Andrade-Navarro, M. Niederweis, Identification of outer membrane proteins of *Mycobacterium tuberculosis*, Tuberculosis (Edinburgh) 88 (2008) 526–544.
[17] H. Målen, S. Pathak, T. Søfteland, G.A. de Souza, H.G. Wiker, Definition of novel cell envelope associated proteins in Triton X-114 extracts of *Mycobacterium tuberculosis* H37Rv, BMC Microbiology 10 (2010) 132.
[18] M. Marmiesse, P. Brodin, C. Buchrieser, C. Gutierrez, N. Simoes, V. Vincent, P. Glaser, S.T. Cole, R. Brosch, Macro-array bioinformatic analyses reveal mycobacterial 'core' genes, variation in the ESAT-6 gene family and new phylogenetic markers for the *Mycobacterium tuberculosis* complex, Microbiology 150 (2004) 483–496.
[19] S.K. Parker, R.M. Barkley, J.G. Rino, M.L. Vasil, *Mycobacterium tuberculosis* Rv3802c encodes a phospholipase/thioesterase and is inhibited by the antimycobacterial agent tetrahydrolipstatin, PLoS ONE 4 (2009) e4281.
[20] S. Canaan, D. Maurin, H. Chahinian, B. Pouilly, C. Durousseau, F. Frassinetti, L. Scappuccini-Calvo, C. Cambillau, Y. Bourne, Expression and characterization of the protein Rv1399c from *Mycobacterium tuberculosis*. A novel carboxyl esterase structurally related to the HSL family, European Journal of Biochemistry 271 (2004) 3953–3961.
[21] W.L. Beatty, D.G. Russell, Identification of mycobacterial surface proteins released into subcellular compartments of infected macrophages, Infection and Immunity 68 (2000) 6997–7002.
[22] J.A. Armstrong, P.D. Hart, Response of cultured macrophages to *Mycobacterium tuberculosis* with observations on fusion of lysosomes with phagosomes, Journal of Experimental Medicine 134 (1971) 713–740.
[23] D.G. Russell, *Mycobacterium tuberculosis*: here today, and here tomorrow, Nature Reviews Molecular Cell Biology 2 (2001) 569–577.
[24] C.M. Sassetti, D.H. Boyd, E.J. Rubin, Genes required for mycobacterial growth defined by high density mutagenesis, Molecular Microbiology 48 (2003) 77–84.
[25] N.P. West, F.M.E. Chow, E.J. Randall, J. Wu, J. Chen, J.M.C. Ribeiro, W.J. Britton, Cutinase-like proteins of *Mycobacterium tuberculosis*: characterization of their variable enzymatic functions and active site identification, FASEB Journal 23 (2009) 1694–1704.
[26] N.P. West, T.M. Wozniak, J. Valenzuela, C.G. Feng, A. Sher, J.M.C. Ribeiro, W.J. Britton, Immunological diversity within a family of cutinase-like proteins of *Mycobacterium tuberculosis*, Vaccine 26 (2008) 3853–3859.
[27] V.D. Vissa, P.J. Brennan, The genome of *Mycobacterium leprae*: a minimal mycobacterial gene set, Genome Biology 2 (Reviews 1023) (2001) 1–1023.8.
[28] P.K. Crellin, J.P. Vivian, J. Scoble, F.M. Chow, N.P. West, R. Brammananth, N.I. Proellocks, A. Shahine, J. Le Nours, M.C.J. Wilce, W.J. Britton, R.L. Coppel, J. Rossjohn, T. Beddoe, Tetrahydrolipstatin inhibition, functional analyses, and three-dimensional structure of a lipase essential for mycobacterial viability, Journal of Biological Chemistry 285 (2010) 30050–30060.
[29] N.P. West, K.M. Cergol, M. Xue, E.J. Randall, W.J. Britton, R.J. Payne, Inhibitors of an essential mycobacterial cell wall lipase (Rv3802c) as tuberculosis drug leads, Chemical Communications 47 (2011) 5166–5168.
[30] N.S. Shah, A. Wright, G. Bai, L. Barrera, F. Boulahbal, N. Martín-Casabona, F. Drobniewski, C. Gilpin, M. Havelková, R. Lepe, R. Lumb, B. Metchock, F. Portaels, M.F. Rodrigues, S. Rüsch-Gerdes, A. Van Deun, V. Vincent, K. Laserson, C. Wells, J.P. Cegielski, Worldwide emergence of extensively drug-resistant tuberculosis, Emerging Infectious Diseases 13 (2007) 380–387.
[31] M. Berry, O.M. Kon, Multidrug- and extensively drug-resistant tuberculosis: an emerging threat, European Respiratory Reviews 18 (2009) 195–197.
[32] J.D. Thompson, D.G. Higgins, T.J. Gibson, W. Clustal, Improving the sensitivity of progressive multiple sequence alignment through sequence weighting position-specific gap penalties and weight matrix choice, Nucleic Acids Research 22 (1994) 4673–4680.
[33] M.A. Marti-Renom, A.C. Stuart, A. Fiser, Comparative protein structure modelling of genes and genomes, Annual Review of Biophysics and Biomolecular Structure 29 (2000) 291–325.
[34] M.Y. Shen, A. Sali, Statistical potential for assessment and prediction of protein structures, Protein Science 15 (2006) 2507–2524.
[35] R.A. Laskowski, M.W. MacArthur, D.S. Moss, J.M. Thornton, PROCHECK: a program to check stereo chemical quality of protein structures, Journal of Applied Crystallography 26 (1993) 283–291.
[36] C. Colovos, T.O. Yeates, Verification of protein structures: patterns of nonbonded interactions, Protein Science 2 (1993) 1511–1519.
[37] M.J. Sippl, Recognition of errors in three-dimensional structures of proteins, Proteins: Structure, Function, and Bioinformatics 17 (1993) 355–362.
[38] H.J.C. Berendsen, D. Van der Spoel, R. Van Drunen, GROMACS: a message passing parallel molecular dynamics implementation, Computer Physics Communications 91 (1995) 43–56.
[39] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A.E. Mark, H.J. Berendsen, GROMACS: fast, flexible, and free, Journal of Computational Chemistry 26 (2005) 1701–1718.
[40] B. Hess, C. Kutzner, D. van der Spoel, E. Lindahl, GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation, Journal of Chemical Theory and Computation 4 (2008) 435–447.
[41] B. Hess, H. Bekker, H.J.C. Berendsen, J.G.E.M. Fraaije, LINCS. A linear constraint solver for molecular simulations, Journal of Computational Chemistry 18 (1997) 1463–1472.
[42] B. Hess, P-LINCS. A parallel linear constraint solver for molecular simulation, Journal of Chemical Theory and Computation 4 (2008) 116–122.
[43] P.D. Lyne, Structure-based virtual screening: an overview, Drug Discovery Today 7 (2002) 1047–1055.

[44] G.M. Morris, D.S. Goodsell, R.S. Halliday, Automated docking using a Lamarckian genetic algorithm and empirical binding free energy function, Journal of Computational Chemistry 19 (1998) 1639–1662.
[45] S. Cosconati, S. Forli, A.L. Perryman, R. Harris, D.S. Goodsell, A.J. Olson, Virtual screening with AutoDock: theory and practice, Expert Opinion on Drug Discovery 5 (2010) 597–607.
[46] H. Park, J. Lee, S. Lee, Critical assessment of the automated AutoDock as a new docking tool for virtual screening, Proteins: Structure, Function, and Bioinformatics 65 (2006) 549–554.
[47] G.M. Morris, R. Huey, W. Lindstrom, M.F. Sanner, R.K. Belew, D.S. Goodsell, A.J. Olson, AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility, Journal of Computational Chemistry 30 (2009) 2785–2791.
[48] T. Bertrand, F. Augé, J. Houtmann, A. Rak, F. Vallée, V. Mikol, P.F. Berne, N. Michot, D. Cheuret, C. Hoornaert, M. Mathieu, Structural basis for human monoglyceride lipase inhibition, Journal of Molecular Biology 396 (2010) 663–673.
[49] J.J. Irwin, B.K. Shoichet, ZINC – a free database of commercially available compounds for virtual screening, The Journal of Chemical Information and Modeling 45 (2005) 177–182.
[50] C. Schalk-Hihi, C. Schubert, R. Alexander, S. Bayoumy, J.C. Clemente, I. Deckman, R.L. DesJarlais, K.C. Dzordzorme, C.M. Flores, B. Grasberger, J.K. Kranz, F. Lewandowski, L. Liu, H. Ma, D. Maguire, M.J. Macielag, M.E. McDonnell, T.M. Haarlander, R. Miller, C. Milligan, C. Reynolds, L.C. Kuo, Crystal structure of a soluble form of human monoglyceride lipase in complex with an inhibitor at 1.35 Å resolution, Protein Science 20 (2011) 670–683.
[51] P. Gouet, E. Courcelle, D.I. Stuart, F. Metoz, ESPript: multiple sequence alignments in PostScript, Bioinformatics 15 (1999) 305–308.
[52] R.D. Finn, J. Mistry, J. Tate, P. Coggill, A. Heger, J.E. Pollington, O.L. Gavin, P. Gunesekaran, G. Ceric, K. Forslund, L. Holm, E.L. Sonnhammer, S.R. Eddy, A. Bateman, The Pfam protein families database, Nucleic Acids Research Database Issue 38 (2010) D211–D222.
[53] S. Adindla, L. Guruprasad, Sequence analysis corresponding to the PPE and PE proteins in *Mycobacterium tuberculosis* and other genomes, Journal of Biosciences 28 (2003) 169–179.
[54] A. Marchler-Bauer, S. Lu, J.B. Anderson, F. Chitsaz, M.K. Derbyshire, C. DeWeese-Scott, J.H. Fong, L.Y. Geer, R.C. Geer, N.R. Gonzales, M. Gwadz, D.I. Hurwitz, J.D. Jackson, Z. Ke, C.J. Lanczycki, F. Lu, G.H. Marchler, M. Mullokandov, M.V. Omelchenko, C.L. Robertson, J.S. Song, N. Thanki, R.A. Yamashita, D. Zhang, N. Zhang, C. Zheng, S.H. Bryant, CDD: a Conserved Domain Database for the functional annotation of proteins, Nucleic Acids Research 39 (D) (2011) 225–229.
[55] R. Sultana, K. Tanneeru, L. Guruprasad, The PE-PPE domain in mycobacterium reveals a serine α/β hydrolase fold and function: an in-silico analysis, PLoS ONE 6 (2011) e16745.
[56] H. McWilliam, F. Valentin, M. Goujon, W. Li, M. Narayanasamy, J. Martin, T. Miyar, R. Lopez, Web services at the European Bioinformatics Institute–2009, Nucleic Acids Research 37 (2009) W6–W10.
[57] S. Longhi, M. Czjzek, V. Lamzin, A. Nicolas, C. Cambillau, Atomic resolution (1.0 Å) crystal structure of *Fusarium solani* cutinase: stereochemical analysis, Journal of Molecular Biology 268 (1997) 779–799.
[58] M. Wiederstein, M.J. Sippl, ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins, Nucleic Acids Research 35 (2007) W407–W410.
[59] J.Z. Long, W. Li, L. Booker, J.J. Burston, S.G. Kinsey, J.E. Schlosburg, F.J. Pavón, A.M. Serrano, D.E. Selley, L.H. Parsons, A.H. Lichtman, B.F. Cravatt, Selective blockade of 2-arachidonoylglycerol hydrolysis produces cannabinoid behavioral effects, Nature Chemical Biology 5 (2009) 37–44.
[60] J.Z. Long, D.K. Nomura, B.F. Cravatt, Characterization of monoacylglycerol lipase inhibition reveals differences in central and peripheral endocannabinoid metabolism, Chemistry and Biology 16 (2009) 744–753.
[61] W.L. DeLano, The PyMOL Molecular Graphics System, DeLano Scientific, San Carlos, CA, USA, 2002.
[62] T.R. Frieden, S.S. Munsiff, The DOTS strategy for controlling the global tuberculosis epidemic, Clinics in Chest Medicine 26 (2005) 197–205.