Modeling, molecular docking, probing catalytic binding mode of acetyl-CoA malate synthase G in *Brucella melitensis* 16M

Pradeepkiran Jangampalli Adi [a], Nanda Kumar Yellapu [b], Bhaskar Matcha [a,*]

[a] *Division of Animal Biotechnology, Department of Zoology, Sri Venkateswara University, Tirupati 517502, Andhra Pradesh, India*
[b] *Biomedical Informatics Centre, Vector Control Research Centre, Puducherry 605006, India*

ABSTRACT

There are enormous evidences and previous reports standpoint that the enzyme of glyoxylate pathway malate synthase G (MSG) is a potential virulence factor in several pathogenic organisms, including *Brucella melitensis* 16M. Where the lack of crystal structures for best candidate proteins like MSG of *B. melitensis* 16M creates big lacuna to understand the molecular pathogenesis of brucellosis. In the present study, we have constructed a 3-D structure of MSG of *Brucella melitensis* 16M in MODELLER with the help of crystal structure of *Mycobacterium tuberculosis* malate synthase (PDB ID: 2GQ3) as template. The stereo chemical quality of the restrained model was evaluated by SAVES server; remarkably we identified the catalytic functional core domain located at 4th cleft with conserved catalytic amino acids, start at ILE 59 to VAL 586 manifest the function of the protein. Furthermore, virtual screening and docking results reveals that best leadmolecules binds at the core domain pocket of MSG catalytic residues and these ligand leads could be the best prospective inhibitors to treat brucellosis.

© 2016 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

An important enzyme of the glyoxylate pathway, malate synthase G (MSG) which involves catalyzation and condensation process to produce subsequent hydrolysis of acetyl-coenzyme A (acetyl-CoA) and glyoxylate to form malate and CoA. This reaction is part of the glyoxylate cycle, which allows certain organisms to derive their carbon requirements from two-carbon compounds by bypassing the two carboxylation process steps of the citric acid cycle [1]. MSG is an important metabolic enzyme which helped in various pathogenic bacteria through bypassing the TCA cycle rather than using glyoxylate cycle [2]. This cycle is essential for growth on two-carbon compounds such as ethanol and acetate, and plays an anaplerotic role in the provision of main precursors for biosynthesis. MSG is one of the two enzymes of glyoxylate cycle, which is essential for the persistence of many bacteria [3]. This is an important functional gene in many pathogenic aerobe bacteria like *M. tuberculosis* and facultative *B. melitensis* 16M [4,5]. The sequence analysis of MSG of *B. melitensis* 16M is known to contain two domains such as malate synthase G and beta subunit domain that are involved in binding of acetyl-CoA and glyoxylate and are responsible for the *Brucella* pathogenesis by bypassing the

aerobic conditions. In the present study, we modeled MSG structure of *B. melitensis* 16M and compared with previously determined crystal structures of substrate and product complexes from the database.

Protein modeling is a challenge in drug discovery, because predicting the accurate 3-D structure of proteins has always been and remains a complicated assignment [6]. In Template based protein modeling (TBM), the accuracy of protein structures, particularly their binding sites, is essential for the success of modeling protein complexes. Overall, approximately 50% of complexes with their interfaces modeled by high-throughput techniques had accuracy suitable for meaningful docking experiments. This percentage will grow with the increasing availability of co-crystallized protein-protein complexes [7]. TBM structure prediction techniques rely on the study of principles that dictate the 3-D structure of proteins from the theory of evolution viewpoint [8], recently; this type of modeling becomes a most popular modeling. TBM involves several steps; identification of homologous (templates), alignment of target to template, structure building, refinement and validation. Moreover, as molecular docking and virtual screening becomes more predictive and prevalent; the possibility of interfacing such tools with functional genomics via threading or homology modeling becomes increasingly tempting. MSG in *Brucella* has been reported without 3-D structure from the previous screening reports [5]. Hence the lack of crystal structures for best candidate protein like MSG in our previous studies continues by predicting the 3-D structure of MSG by using comparative

* Correspondence to: Department of Zoology, Sri Venkateswara University, Tirupati 517502, India.
E-mail address: matchabhaskar2010@gmail.com (B. Matcha).

modeling in MODELLER v9.12. Hence the potential drug target, MSG was robust by 3-D structure, evaluated, and deposited in Protein Model Data Base (PMDB) which stores manually built 3-D models of proteins.

2. Materials and methods

2.1. Homology modeling and optimization

Homology modeling of protein MSG matures in MODELLER 9.12 by using python scripts. Protein sequence was subjected to Blast-P against PDB to find out suitable template for homology modeling. The constructed model was optimized using variable target function method (VTFM) and tuned by adjusting 'automodel.library' schedule, 'automodel.max' var-iterations and 'automodel.max'. The molecular dynamics (MD) with simulated annealing (SA) step was tuned by adjusting 'automodel.md' level with conjugate gradients (CG), residue range, and energy scaling factor, and then refined with SA parameterization. The whole optimization can be repeated multiple times if desired (by default it is run only once) by adjusting 'automodel.repeat' optimization, whereas in the present case the optimization was not attained in the first run with default parameters and hence moved for the second run. The VTFM optimization with maximum 500 iterations and MD optimization with slow level mode was carried out and the whole cycle was repeated for two times to generate an optimized conformation of the model with a gradients of 0.2 Å to 0.1 Å [9]. The optimized model was evaluated by Ramachandran plot PROCHECK (for conformations of the ψ and φ angles are possible for an amino-acid residues in a protein), verify_3D (Determines the compatibility of an atomic model (3D) with its own amino acid sequence (1D), ERRAT (verifying protein structures determined by crystallography and modeled proteins) and WHATIF (checking of many sterochemical parameters of the residues in the model).

2.2. Molecular dynamics of MSG protein

In the anticipation of attaining the stabilized conformation of the MSG, the structure was subjected to MD simulations in the Discovery Studio 4.0. The software uses the standard dynamic cascade tool. The structure was loaded into the graphical window of Discover Studio and initially refined with a two steps of energy minimization using steepest descent and conjugate gradient algorithm to an RMSD of 0.1 Å in a total of 3000 maximum steps. The system was heated from 50 K to 300 K in 100 ps and equilibrated for another 100 ps under constant pressure. Further the system was progressed into production phase for 20,000 ps with NPT ensembles under generalized born implicit solvent model. The energy levels and RMSD were observed throughout the production phase and the stabilized conformation was trapped and saved for further studies.

2.3. Active site prediction

Active site prediction is necessitate to find out the specific functional part of the protein therefore, it seems apparent that prediction of protein domains is crucial to understand the protein function. Thus, 3D structural domains and active site residues were identified by a computed atlas of surface topography of proteins server (CASTp) [10].

2.4. Ligands screening from ZINC database

Ligands were screened from the ZINC database on the basis of 90% structural similarity of known inhibitor of MSG of *M.*

tuberculosis i.e. 4-(2-hydroxyethyl)-1-piperazine ethanesulfonic acid [11]. Based on such structural resemblance, 2285 analogs were identified and subsequently prepared as a ligand spread sheet for further virtual screening.

2.5. Virtual screening

PyRx virtual screening software [12] was used for the virtual screening of ligands that includes AutoDock [13] and AutoDock Vina [14] with the Lamarckian genetic algorithm (LGA) as scoring function. Active site dimensions were set as grid size of center X=54.41 Å, center Y=30.78 Å, center Z=79.94 Å (XYZ axis) to dock the ligands where 10 maximum exhaustiveness was calculated for each ligand. Before initiation of docking operation, charges were assigned to protein and ligand structures by AutoDock Vina. The resulted ligands have high potential for being used as drug candidates.

2.6. Pharmacokinetic analysis

FAFDrugs3 tool can perform computational prediction of some ADME-Tox properties (Adsorption, Distribution, Metabolism, Excretion and Toxicity) [15] and admetSAR provides the latest and most comprehensive manually curated data for diverse chemicals associated with known ADMET profiles [16] both web servers were used to analyze the ADMET properties of top 5 Virtual screening hits. ADMET properties of top successive hits were checked in optimal descriptors. Moreover the oral toxicity properties were also analyzed with this server and the probable accessorial MSG protein targets were checked for every successive lead.

3. Results

3.1. Homology modeling and optimizatio

The Blast-P results revealed a perfect reliable template, i.e. crystal structure of a malate synthase family protein (P9WK17) from *M. tuberculosis* with 2.3 Å resolutions (PDB ID: 2GQ3) shared 99% query coverage and 61% identity with 0.0 *E*-Value. Align 2-D results showed a high-level sequence similarity between target and template. Based on alignment, 100 basic models were generated for target protein by using MODELLER 9.12 comparative modeling and the missing side chains were added and aligned from WHATIF server.

3.2. Structure validation from all the validation servers

Structure validation SAVES results stated that homology modeled protein MSG possesses reasonable 3-D structure with good stereo-chemical quality of Ramachandran plot where PROCHECK analysis showed most favored regions [a,b,l] with 592 residues among 728 with 93.4% and additional allowed regions [a,b,l,p] of 39 residues with 6.2% and identified only one residue located at disallowed regions but maximum amino acids were present in most favored regions proven that modeled protein showing the good reliable structure. Protein compatibility of an atomic model (3-D) with its amino acid sequence (1-D) of window of 21 residues. verify_3D results stated that all amino acids reliable to convert to form 3-D structure. ERRAT Error values of modeled protein plotted as a function of the position of a sliding 9-residue window resulted that overall quality factor with 88.88% indicating that reliable high-resolution modeled structure.

The resultant models were sorted by discrete optimized protein energy (DOPE) scoring function. The model number 72 (M72),
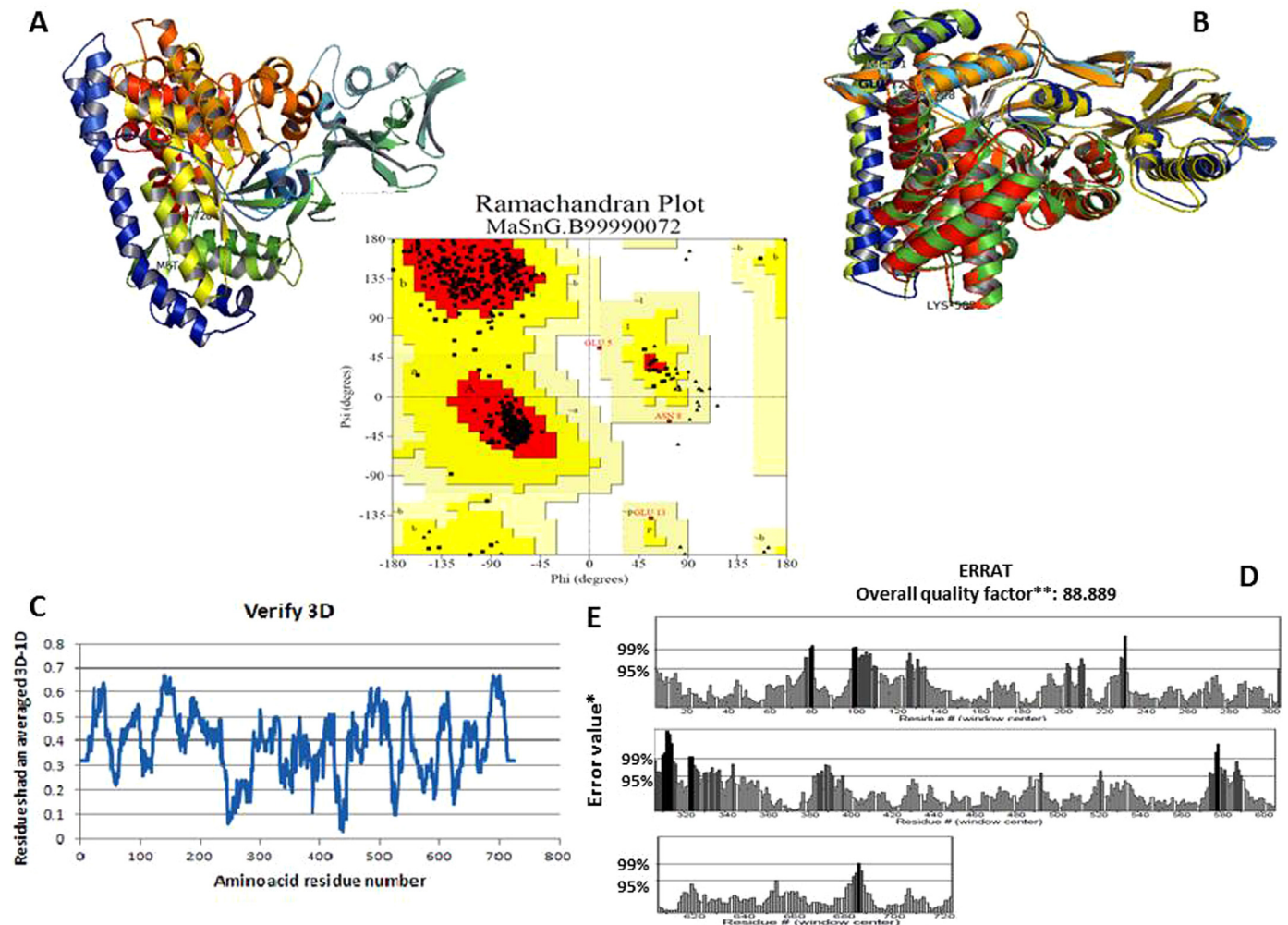
Fig. 1. Malate synthase G stereo chemical quality of the optimized protein. (A) Secondary structure of modeled protein from PyMOL. (B) The superimposition of target and template were generated by PyMOL. (C) Verify3D of Malate synthase G modeled protein plot shows that all the residues are with positive compatibility score indicating that they are reasonably folded. (D) ERRAT (v 2.0) values showed reasonable scores from SAVES (Structural Analysis and Verification Server). (E) Ramachandran plot.

consists low dope score, i.e. 86197.39 with 3868.75 kcal/mol energy. Hence, we selected M72 for further studies and all amino acid residues were shown to be at allowed regions. The protein secondary confirmation and superimposed structure with the template are shown in (Fig. 1A and B), and the modeled protein stereo-chemical quality and accurate results of Ramachandran plot PROCHECK 93.4% accuracy, Protein atomic model alignment and sequence residues 3-D-1D, verify_3D with decent alignment, Error values of modeled protein ERRAT 88.88% results illustrated in (Fig. 1C–E).

3.3. Molecular dynamic simulations

The validated structure of MSG was further subjected to MD simulations to get a reliable and perfect conformation so as to proceed for molecular docking studies. The MD results were analyzed after successful 20,000 ps of production time and it was observed that a stabilized nature was attained at the end of the simulation period (Fig. 2A). The potential energy levels were observed where the structure showed maximum energy levels above 5000 kcal/mol and fluctuated for a period of 10,000 ps and thereafter stabilized in the further production phase around 4000 kcal/mol. Further, the RMSD plot showed that there exists fluctuation in the RMSD for a period of 17,000 ps and thereafter stabilized around 6.5 Å (Fig. 2B). These results from MD simulations showed that stabilized conformation was obtained by the

end of simulation period and best suitable for further investigations and hence utilized the same for prediction of binding sites and molecular docking studies.
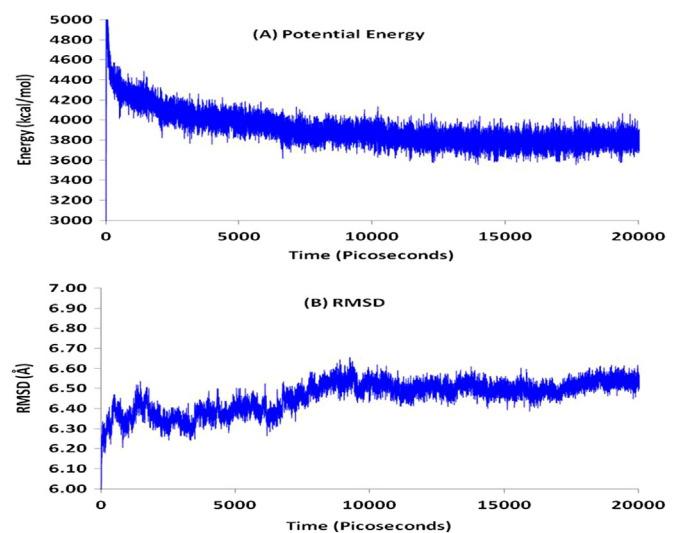

Fig. 2. (A) Potential Energy levels of *Brucella melitensis* 16M protein Malate synthase G. (B) RMSD plots of protein Malate synthase G obtained from molecular dynamics simulations showing the stabilized nature of the structure.

3.4. Prediction of Active site

CastP active site results showed all reasonable binding pockets of the protein model. The core domain ID 137 showed good prospective active residues with area 332.8 Å² and volume 427 Å² for ligand binding and functional modifications. The overall cassette binding cleft cavity will start with ILE 69 and ends at VAL 684. Thus, an active site of *B. melitensis* MSG starts from Isoleucine 69 to Valine 586 amino acids actively involved in the binding fit of lead molecules (Fig. 3A and B). Hence, we have selected the N-terminal binding pocket of the target protein for further docking simulations.

3.5. Virtual screening and docking

This process involves the template modeled protein existing ligand used as template structure and prepared ligands of reliable lead molecules. Where the template protein contains the various ligands coenzyme A, magnesium ion, D-malate where we had omitted these complex ions and other molecules and selected 4-(2-hydroxyethyl)-1-piperazine ethanesulfonic acid related ligands screened from ZINC database. The structure, based virtual screening was done using 2285 structural analogs of the existing ligand of MSG crystal structure of *M. tuberculosis* through Auto-Dock Vina in PyRx software and the docking simulation energies were determined for each ligand with 10 exhaustiveness. Root mean square deviation (RMSD) calculations were carried out through lamarkin geometric algorithm (LGA). Active site grid dimensions were set at X=54.41 Å, Y=30.78 Å and Z=79.94 Å for the center and total size dimensions were set as X=69.23 Å, Y=76.35 Å and Z=81.81 Å. All the best leads superimposed core cavity interactions were visualized through PyMOL software (Fig. 3C).

4. Discussion

The homology model of MSG, which a key enzyme of the glyoxylate pathway in *Brucella* provides valuable information to understand its function that how the opportunistic bacterial pathogens use glyoxylate pathway during pathogenesis and how metabolic pathways employed by pathogenic bacteria during infection this pathways essential for the survival of pathogens inside the host could be helpful to develop new therapeutic strategies
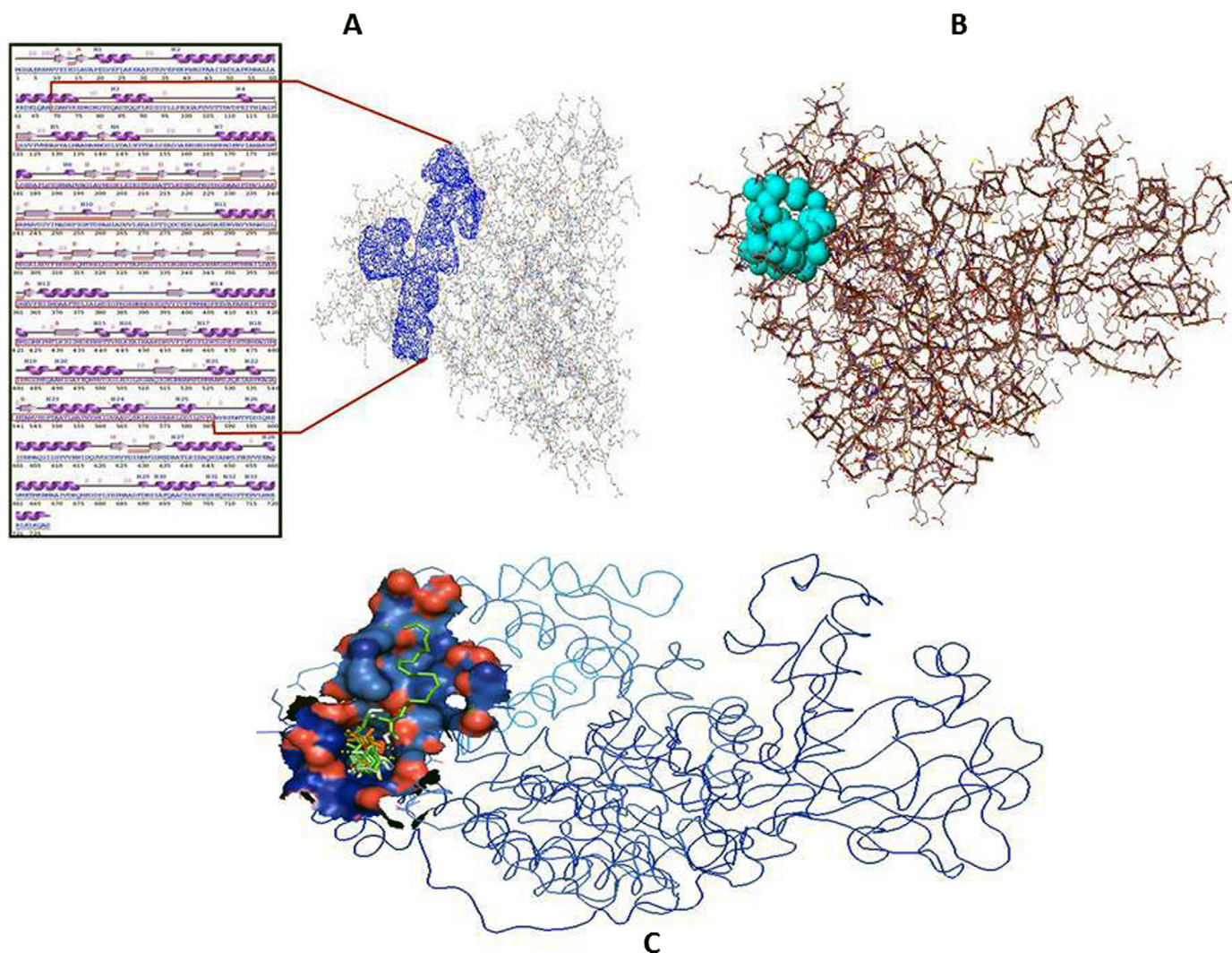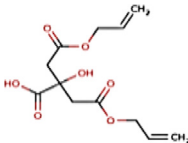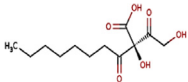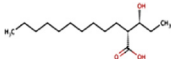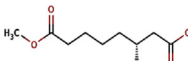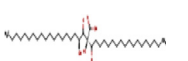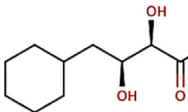

Fig. 3. (A) Malate synthase secondary sequence analysis composed sheets turns, and coils and helix turn helices and hairpins and the sequence highlighted with red box functional domain region. The wire frame of malate synthase G and blue color highlighted region is catalytic cleft of functional domain. (B) Surface structure of modeled protein malate synthase G protein domain predicted by CastP server shown as spheres. (C) All best ligand molecules superimposed where binding at core domain of the MSG protein.

Table 1
The 5 4-(2-hydroxyethyl)-1-piperazine ethanesulfonic acid analogs showed good binding energies than 4-(2hydroxyethyl)-1-piperazine ethanesulfonic acid (positive control). Reliable H-bond interactions with conserved motif residues and bond distances were illustrated.

Rank	Ligand ID	Name of compound	2D structure	Binding Energy (Kcal/Mol)	Interactions		Bond length (Å)
					Protein residues	Ligand atoms	
1	ZINC78173801	4-allyloxy-2-(2-allyloxy-2-oxo-ethyl)-2-hydroxy-4-oxo-butanoic		−7.2	**Tyr81**CA-N	C12O6	2.89
					Gln80 CA-N	C12O6	3.08
					Val584CA-N	C11O2	3.24
2	ZINC78690149	(2R)-2-hydroxy-2-(2-hydro-xyacetyl)-3-oxo-decanoic		−7.1	**Tyr81**CA-N	C4O4	2.94
					Tyr81CA-O	C12O3	2.72
					Leu582C-O	C9O2	2.93
					Val584CA-	C9O2	2.80
3	ZINC79573345	(2R)-2-[(1R)-1-hydroxypropyl] dodecanoic		−7.0	**Asp78**CG-OD1	C14O3	3.03
					Gly80CA-N	C14O3	3.15
					Tyr81CA-N	C14O3	3.29
					Val584CA-N	C15O1	3.24
4	ZINC79646184	(3R)-3-hydroxy-8-methoxy-8-oxo-octanoic		−6.9	**Tyr81**CA-O	C5O5	2.87
					Gln83CD-NE2	C8O4	3.12
					Ser487CB-OG	C9O3	2.94
					Val584CA-N	C8O2	3.04
5	ZINC79673354	(2S,4S)-2,4-dihydroxy-2-octadeca-noyl-3-oxo-henicosanoic		−6.9	Val584CA-O	C35O2	2.80
					Val586C-O	C38O2	2.83
					Val586C-O	C38O4	2.89
					Ala587C-O	C38O4	3.07
6	ZINC78283617 (positive control)	4-(2-hydroxyethyl)-1-piperazine ethanesulfonic acid		−6.8	**Tyr81**C-O	C10O2	2.89
					Gln83CD-NE2	C8O3	3.03
					Leu582C-O	C3O3	2.99
					Val584CA-N	C9O4	3.09

against brucellosis. The Ramachandran plot showed only one amino acid GLU 5, in disallowed region, and all amino acids were present in most favored regions. The other parameters such as protein error values from ERRAT showed an increased quality factor of 88.88% after optimization, and Verify_3D showed a compatibility score above zero (92.46% ^{3}D–^{1}D values) indicating that the protein contains favored side chain environments and good fold regions. The robust modeled of *Brucella* MSG is a huge protein composed of 728 amino acids consisting 7 sheets, 4 beta alpha-beta units, 10 beta hairpins, 9 beta bulges, 26 strands, 33 helices, 50 helix-helix interactions, 55 beta turns and 8 gamma turns (Fig. 3A). Among 10 different clefts of the structure, the 4th cleft is considered as conserved with 10.54 Å average depths and is located from ILE 69 to 586 VAL which is forming the functional domain of the structure (Fig. 3B). The amino acids such as ASP 72, SER 474, GLU 476, TYR 73, THR 473, and SER 583 were found to be playing a catalytic role by forming inter-molecular interactions with the ligands. Recent reports evidenced that glyoxylate shunt enzymes have been implicated as virulence factors in several different pathogens such as *M. tuberculosiss* [17–19], *C. albicans* [20]

and *P. aeruginosa* [21]. All these results strongly encouraging the protein model as good and more reliable target for the docking studies.

Thus *in silico* approaches have been developed as an important part of several drug discovery programmes, from lead finding to its optimization [22–24]; methodologies such as ligand or targeted based computational screening [25] procedures are broadly used in many drug discovery studies. Molecular docking procedures were utilized in the current study to predict how small molecules bind to MSG structure and also optionally enabled to model binding parameters of ligand with a number of distinct conformational clusters and to find all possible minimum binding energy. PyRx virtual screening was used to screen libraries of compounds against best potential drug target of *B. melitensis* 16M. In the present study, *M. tuberculosis* malate synthase family protein (PDB: 2GQ3) was selected as a template to build the model *Brucella* MSG structure (accession number: NP_539297.1). Moreover, a ligand 4-(2-hydroxyethyl)-1-piperazine ethanesulfonic acids present within the active site of 2GQ3 structure paved clues to identify its analogs from Zinc database and used for virtual
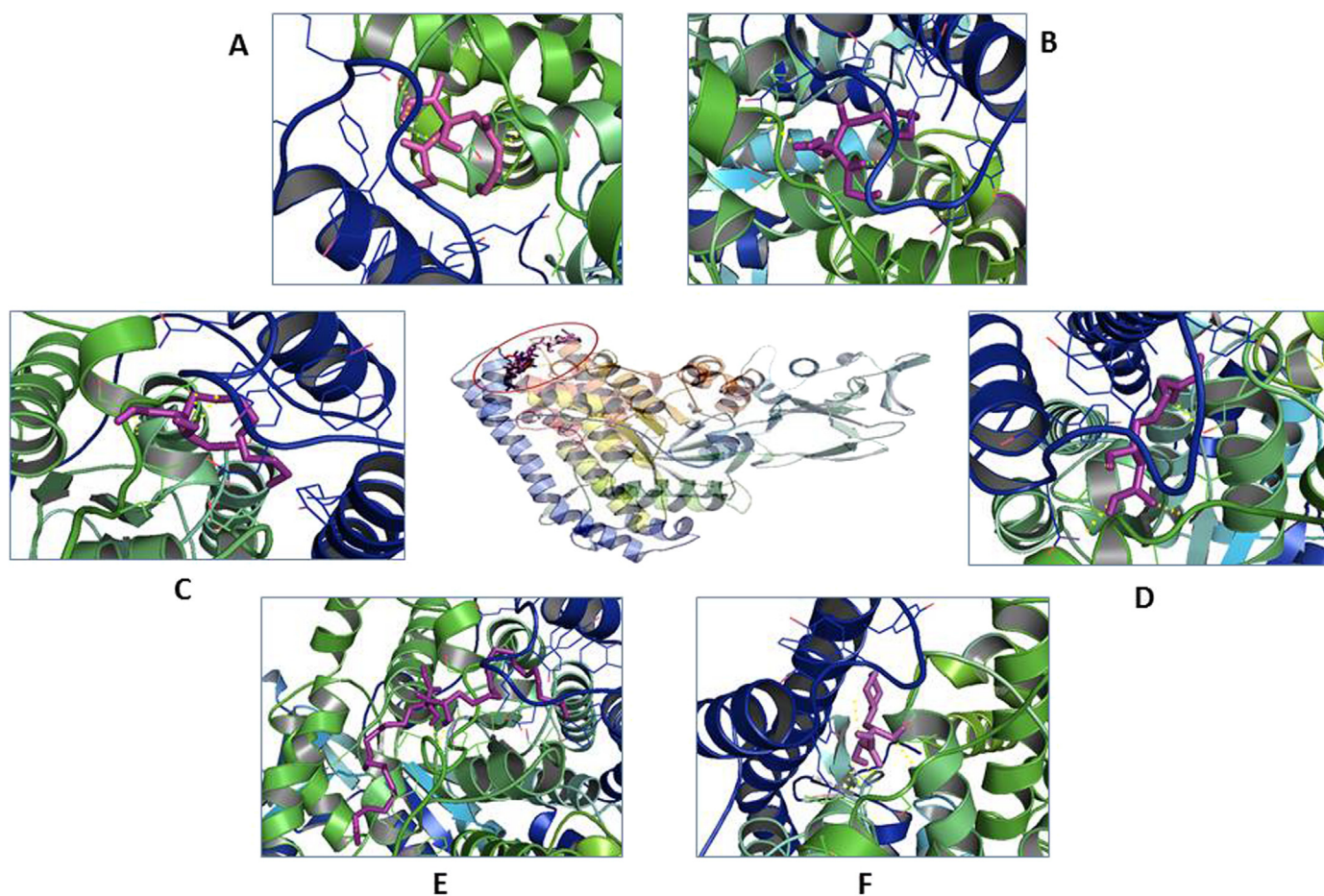
Fig. 4. The predicted docking simulations of leads with the target protein based on lamarkin geometric algorithm and PyRx analyses: the five best leads were compared with positive control which include (A) 4-allyloxy-2-(2-allyloxy-2-oxo-ethyl)-2-hydroxy-4-oxo-butanoic (ZINC78173801), (B) (2R)-2-hydroxy-2-(2-hydroxyacetyl)-3-oxo-decanoic (ZINC78690149), (C) (2R)-2-[(1R)-1-hydroxypropyl] dodecanoic (ZINC79573345), (D) (3R)-3-hydroxy-8-methoxy-8-oxo-octanoic (ZINC79646184), (E) (2S, 4S)-2,4-dihydroxy-2-octadecanoyl-3-oxo-henicosanoic (ZINC79673354) and (F) 4-(2-hydroxyethyl)-1-piperazine ethanesulfonic acid (ZINC78283617) (positive control). These leads are showing good H-bond interactions that are indicated with yellow dashed lines. The superimposition of above leads was distributed in only N-terminal region shown as surface with purple sticks and polar contrasts with yellow dashed lines.

screening studies. The errors in the identified leads were solved by lead optimization in PyRx which include Open Bable, ligand energy minimization interface with UFF (United Force Field) with a limit of 500 iterations for each ligand. The energy-minimized ligands were converted into AutoDock ligand format and prepared as a data-set (a).

The virtual screening of 4-(2-Hydroxyethyl)-1-piperazine ethanesulfonic acid related analogs screened from ZINC database revealed that, 2285 compounds with 90% similarity cut off screens and prepared ligand spreadsheets, and from among that only five compounds showed best affinity with positive control i.e. 4-(2-hydroxyethyl)-1-piperazine ethanesulfonic acid. The ligands 4-allyloxy-2-(2-allyloxy-2-oxo-ethyl)-2-hydroxy-4-oxo-butanoic (ZINC78173801), (2R)-2-hydroxy-2-(2-hydroxyacetyl)-3-oxo-decanoic (ZINC78690149), (2R)-2-[(1R)-1-hydroxypropyl] dodecanoic (ZINC79573345), (3R)-3-hydroxy-8-methoxy-8-oxo-octanoic (ZINC79646184) and (2S,4S)-2,4-dihydroxy-2-octadecanoyl-3-oxo-henicosanoic (ZINC79673354) showed binding affinities of − 7.2, − 7.1, − 7.0, − 6.9 and − 6.9 (Kcal/Mol) respectively, and 4-(2-hydroxyethyl)-1-piperazine ethanesulfonic acid (ZINC78283617) showed − 6.8 (kcal/Mol) respectively. Docking results and hydrogen bond interactions with these ligands and their bond angles, bond lengths and atoms involved in these interactions were analyzed and illustrated in (Table 1). The complete core site analysis also revealed that all the leads have the ability to interact with 132 amino acid residues were considered as binding pocket loop. From the complete core binding pocket catalytic amino acid residues that

actively participated in the interaction with the best ligand molecules were as follows ILE 69, **ASP 70**, TRP 72, **TYR 73**, **THR 473**, **SER 474**, GLU 476, ALA 477, LEU582, **SER 583** and VAL 586 by means of C——O, O——H, N——O, N——H, CA——N atoms. The hydrogen bond formation was majorly observed among all the analogs with almost equal bond distances. 4-allyloxy-2-(2-allyloxy-2-oxo-ethyl)-2-hydroxy-4-oxo-butanoic (ZINC78173801), (2R)-2-hydroxy-2-(2-hydroxyacetyl)-3-oxo-decanoic (ZINC78690149), (2R)-2-[(1R)-1-hydroxypropyl] dodecanoic (ZINC79573345), (3R)-3-hydroxy-8-methoxy-8-oxo-octanoic (ZINC79646184) and (2S,4S)-2,4-dihydroxy-2-octadecanoyl-3-oxo-henicosanoic (ZINC79673354) showed three, four, six and two hydrogen bond interactions respectively. Residues such as ASP 72, GLY 80, TYR 81, VAL 584 with 4-allyloxy-2-(2-allyloxy-2-oxo-ethyl)-2-hydroxy-4-oxo-butanoic, TYR 81, LEU 582 and VAL584 with (2R)-2-hydroxy-2-(2-hydroxyacetyl)-3-oxo-decanoic, (2R)-2-[(1R)-1-hydroxypropyl] dodecanoic with GLY 80, TYR 81, and VAL 584, (3R)-3-hydroxy-8-methoxy-8-oxo-octanoic with TYR 81, GLN 83, SER 474, VAL 584, and (2S,4S)-2,4-dihydroxy-2-octadecanoyl-3-oxo-henicosanoic with VAL 586, VAL 584 and ALA 587 sequence residues were associated with the hydrogen bond formation. Moreover, all the above 4-(2-hydroxyethyl)-1-piperazine ethanesulfonic acid analogs showed reliable affinities with MSG protein (Fig. 4A–F). Other active site amino acid residues were observed to participate in the formation of hydrophobic interactions with ligands. All these interactions reflected a good affinity levels among the target and ligands than with the positive control 4-(2-hydroxyethyl)-1-piperazine ethanesulfonic acid (Fig. 4F).

Table 2
(A) FAF-Drugs3 results tabulated for top 5 successive hits and of control ligand ZINC78283617. (B) admetSAR toxicity results of top 5 successive hits and of control ligand ZINC78283617 ligand virtual screening data.

T2. A	**FAF-Drugs3 results**					
S.No	**Property**	**ZINC78173801**	**ZINC78690149**	**ZINC79573345**	**ZINC79646184**	**ZINC79673354**	**ZINC78283617**
1.	**Log P**	− 0.07	1.17	5.26	0.00	15.46	1.65
2.	**Log aqueous solubility (SW)(dm − 3)**	− 0.76	− 1.53	− 3.96	− 0.58	− 5.17	− 1.87
3.	**Molecular Weight (kDa)**	272.25	260.28	258.40	204.22	439.00	202.25
4.	**H Bond donors**	1	2	1	1	2	2
5.	**H Bond Acceptors**	7	6	3	5	6	4
6.	Solubility(mg/l)	127595.82	56312.33	4906.87	114552.52	11906.87	31188.76
7.	**Oral Bioavailability (VEBER)**	Good	Good	Good	Good	Good	Good
8.	**Oral Bioavailability (EGAN)**	Good	Good	Good	Good	Good	Good
9.	**Phospo lipidosis**	NonInducer	NonInducer	NonInducer	NonInducer	NonInducer	NonInducer
10	**Status**	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted
T2. B	**admetSAR toxicity**						
1.	**Blood-Brain Barrier**	BBB+	BBB+	BBB+	BBB+	BBB+	BBB+
2.	**Human Ether-a-go-go-Related Gene Inhibition**	Weak inhibitor Non-inhibitor	Weak inhibitor Non-inhibitor	Weak inhibitor Non-inhibitor	Weak inhibitor Non-inhibitor	Weak inhibitor Non-inhibitor	Weak inhibitor Non-inhibitor
3.	**AMES Toxicity**	Non AMES toxic	Non AMES toxic	Non AMES toxic	Non AMES toxic	Non AMES toxic	Non AMES toxic
4.	**Carcinogens**	Non-carcinogens	Non-carcinogens	Non-carcinogens	Non-carcinogens	Non-carcinogens	Non-carcinogens
5.	**Fish Toxicity**	High FHMT	High FHMT	High FHMT	High FHMT	High FHMT	High FHMT
6.	**Tetrahymena pyriformis Toxicity**	High TPT	High TPT	High TPT	High TPT	High TPT	High TPT
7.	**Biodegradation**	Not ready biodegradable	Ready biodegradable	Ready biodegradable	Ready biodegradable	Ready biodegradable	Ready biodegradable
8.	**Acute Oral Toxicity**	III	III	III	III	III	III

All the five best leads molecules obey the Lipinski rule of five and efficient ADMET properties. A comparative oral toxicity property analysis was carried out for the above best 5 lead molecules against positive control to confine FAFdrugs3 and admetSAR toxic properties. The Lipinski rule of five defines four simple physicochemical parameter ranges of orally active compounds like molecular weight; log-P, H-bond donors and H-bond acceptors (MWT_500, logP_5, H-bond donors_5, and H-bond acceptors_10) (Table 2A). So, prediction of drug-like non-toxic compounds is an important for modern drug discovery which are mainly obtained from the repositories of the modern drug data report (MDDR), comprehensive medicinal chemistry (CMC) and Derwent word drug index (WDI) [26,27]. Herein, the 4-(2-hydroxyethyl)-1-piperazine ethanesulfonic acid and best 5 lead compounds tested their non-toxic and druggability nature were identified based on the rule of five predictions like Blood-Brain Barrier, AMES Toxicity, carcinogens, acute oral toxicity from admetSAR results showed best druggability nature (Table 2B). The proposed 5 inhibitors with drug-like properties against the target were further optimized along with non-oral toxicity with confirmatory results (Table 2A). These results were strongly suggesting towards in vitro and in vivo further investigations to prove their potential as anti-brucellosis agents.

5. Conclusion

In conclusion we selected MSG as a target protein, further, its modeling studies followed by structure based virtual screening using PyRx AutoDock Vina, helped to identify five best potential inhibitors i.e. 4-allyloxy-2-(2-allyloxy-2-oxo-ethyl)-2-hydroxy-4-oxo-butanoic (ZINC78173801), (2R)-2-hydroxy-2-(2-hydroxyacetyl)-3-oxo-decanoic (ZINC78690149), (2R)-2-[(1R)-1-hydroxypropyl] dodecanoic (ZINC79573345), (3R)-3-hydroxy-8-methoxy-8-oxo-octanoic (ZINC79646184) and (2S,4S)-2,4-dihydroxy-2-octadecanoyl-3-oxo-henicosanoic (ZINC79673354) which showed

good binding orientations and strong affinities within the active site. Furthermore, in vivo studies are required to evaluate the prospective drug activity and efficacy of the proposed leads against MSG. Thus, the in silico strategies adopted in the present methodologies are standard where reduces time and the cost with a safe and effective. And these drugs prior to further the clinical trials. The present study we proposed five best effective inhibitors of Brucella MSG protein, and provide a root line to understand protein, ligand molecular interactions in drug discovery process to treat the Brucella infections.

Appendix A. Transparency document

Transparency document associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.bbrep.2016.08.020.

References

[1] A.J. Cozzone, Regulation of acetate metabolism by protein phosphorylation in enteric bacteria, Annu. Rev. Microbiol. 52 (1998) 127–164.
[2] B. Roucourt, N. Minnebo, P. Augustijns, K. Hertveldt, G. Volckaert, R. Lavigne, Biochemical characterization of malate synthase G of P. aeruginosa, BMC Biochem. 10 (2009) 1–7.
[3] R. Kumar, V. Bhakuni, A functionally active dimer of mycobacterium tuberculosis malate synthase G, Eur. Biophys. J. 39 (2010) 1557–1562.
[4] C.V. Smith, C.C. Huang, A. Miczak, D.G. Russell, J.C. Sacchettini, K. Höner zu Bentrup, Biochemical and structural studies of malate synthase from Mycobacterium tuberculosis, J. Biol. Chem. 278 (2003) 1735–1743.

[5] J.A. Pradeepkiran, S.B. Sainath, K.K. Kumar, M. Bhaskar, Complete genome-wide screening and subtractive genomic approach revealed new virulence factors, potential drug targets against bio-war pathogen *Brucella melitensis* 16M, Drug Des. Dev. Ther. 9 (2015) 1691–1706.

[6] P.R. Daga, R.Y. Patel, R.J. Doerksen, Template-based protein modeling: recent methodological advances, Curr. Top. Med. Chem. 10 (2010) 84–94.

[7] P.J. Kundrotas, I.A. Vakser, Accuracy of protein-protein binding sites in high-throughput template-based modeling, PLoS Comput. Biol. 6 (2010) e1000727.

[8] C.N. Cavasotto, S.S. Phatak, Docking methods for structure-based library design, Methods Mol. Biol. 685 (2011) 155–174.

[9] A. Sali, T.L. Blundell, Comparative protein modelling by satisfaction of spatial restraints, J. Mol. Biol. 234 (1993) 779–815.

[10] T.A. Binkowski, S. Naghibzadeh, J. Liang, CASTp: computed atlas of surface topography of proteins, Nucleic Acids Res. 31 (2003) 3352–3355.

[11] D. Mavrici, D.M. Prigozhin, T. Alber, *Mycobacterium tuberculosis* RpfE crystal structure reveals a positively charged catalytic cleft, Protein Sci. 23 (2014) 481–487.

[12] S. Dallakyan, A.J. Olson, Small-molecule library screening by docking with PyRx, Methods Mol. Biol. 1263 (2015) 243–250.

[13] G.M. Morris, R. Huey, A.J. Olson, Using AutoDock for ligand-receptor docking, Curr. Protoc. Bioinforma. (2008) (Chapter 8: Unit 8.14).

[14] O. Trott, A.J. Olson, AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multi-threading, J. Comput. Chem. 31 (2010) 455–461.

[15] D. Lagorce, O. Sperandio, H. Galons, M.A. Miteva, B.O. Villoutreix, FAF-Drugs2: free ADME/tox filtering tool to assist drug discovery and chemical biology projects, BMC Bioinform. 9 (2008) 1–9.

[16] F. Cheng, W. Li, Y. Zhou, J. Shen, Z. Wu, G. Liu, P.W. Lee, Y. Tang, admetSAR: a comprehensive source and free tool for assessment of chemical ADMET properties, J. Chem. Inf. Model. 52 (2012) 3099–3105.

[17] J.E. Graham, J.E. Clark-Curtiss, Identification of Mycobacterium tuberculosis RNAs synthesized in response to phagocytosis by human macrophages by selective capture of transcribed sequences (SCOTS), Proc. Natl. Acad. Sci. USA 96 (1999) 11554–11559.

[18] K. Honer zu Bentrup, A. Miczak, D.L. Swenson, D.G. Russell, Characterization of activity and expression of isocitrate lyase in Mycobacterium avium and *Mycobacterium tuberculosis*, J. Bacteriol. 181 (1999) 7161–7167.

[19] J.D. McKinney, K. Honer zu Bentrup, E.J. Munoz-Elias, A. Miczak, B. Chen, W. T. Chan, D. Swenson, J.C. Sacchettini, W.R.J. Jacobs, D.G. Russell, Persistence of *Mycobacterium tuberculosis* in macrophages and mice requires the glyoxylate shunt enzyme isocitrate lyase, Nature 406 (2000) 683–685.

[20] M.C. Lorenz, G.R. Fink, The glyoxylate cycle is required for fungal virulence, Nature 412 (2001) 83–86.

[21] U. Ha, S. Jin, Expression of the soxR Gene of *Pseudomonas aeruginosa* is inducible during infection of burn wounds in mice and is required to cause efficient bacteremia, Infect. Immun. 67 (1999) 5324–5331.

[22] W.P. Walters, M.T. Stah, M.A. Murcko, Virtual screening - an overview, Drug Discov. Today 3 (1998) 160–178.

[23] T. Langer, R.D. Hoffmann, Virtual screening: an effective tool for lead structure discovery, Curr. Pharm. Des. 7 (2001) 509–527.

[24] J. Bajorath, Integration of virtual and high-throughput screening, Nat. Rev. Drug Discov. 1 (2002) 882–894.

[25] H. Gohlke, G. Klebe, Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors, Angew. Chem. Int. Ed. 41 (2002) 2644–2676.

[26] D. Frishman, P. Argos, Knowledge-based protein secondary structure assignment, Proteins 23 (1995) 566–579.

[27] T.I. Oprea, Property distribution of drug-related chemical databases, J. Comput. Aided Mol. Des. 14 (2000) 251–264.