Original Article:

Molecular modelling, docking and dynamics studies of biotin carboxyl carrier protein of acetyl-CoA carboxylase to discover potential inhibitors

V. Priyadarshini,1 D. Pradhan,1 M. Munikumar,1 S. Sandeep,1 A. Umamaheswari,1 D. Rajasekhar2

Departments of 1 Bioinformatics, 2 Cardiology, Sri Venkateswara Institute of Medical Sciences, Tirupati

ABSTRACT

Background: Designing drug molecules targeting biotin carboxyl carrier protein of acetyl-CoA carboxylase (AccB), a novel common drug target among the pathogens of infective endocarditis (IE), would be imperative for IE therapy.

Methods: Homology modelling of AccB was performed in complex with biotin using Modeller9v8. DOPE score, PROCHECK, ProSA and ProQ analyses were used for model validation. Allosteric site residues of AccB were determined using PyMOL. The structural analogs of biotin were searched from the Ligand.Info metadatabase tool to compile an in-house library. Structure-based virtual screening for AccB was performed using Maestro v9.2. Molecular dynamic (MD) simulations for AccB-lead1 docking complex were performed using Desmond v3.0.

Results: The AccB structure was evaluated as reliable one for docking analysis. The prepared in-house library comprised of 357 biotin analogs. The structure-based virtual screening on AccB and in-house library led to identification of two lead molecules with better binding affinity compared to biotin (XP Gscore -2.216 kcal/mol). Lead1 showed the lowest XP Gscore of -5.847 kcal/mol with strong hydrogen bond network and good van der Waal interactions with AccB. Stable nature of the docking interactions was reproduced through MD simulations after 10 ns.

Interpretation: The binding orientations of lead1 were well in agreement with binding orientations of existing inhibitors. MD simulations shed light on natural dynamics of the docking complex in solution on different timescales.

Conclusions: Lead1 could, therefore, be considered for designing potent inhibitor of infective endocarditis.

Key words: Infective endocarditis, AccB, Ab initio modeling, Virtual screening, Molecular dynamic simulations

INTRODUCTION

Infective Endocarditis (IE) is a microbial infection of the endocardial surface, lining of the heart chambers and heart valves. IE is the fourth leading cause, after urosepsis, pneumonia and intra-abdominal sepsis, of life-threatening infectious disease syndromes with an annual incidence of 15,000 to 20,000 new cases despite advances in antimicrobial therapy, development of better diagnostic and surgical techniques.1 High morbidity and mortality rate is reported in IE.

The current in-hospital mortality rate for patients with IE is 15% to 20%,2-3 with annual mortality rate approaching 40%.3-5 This is in stark contrast to sustained and ongoing improvements observed in other cardiovascular diseases such as myocardial infarction.6 Definitive studies of IE have been limited by its relative infrequency, a problem compounded by the wide range of causative organisms such as Streptococcus mitis, Streptococcus mutans, Streptococcus salvarius, Enterococcus faecalis, Legionella pneumophila, Staphylococcus aureus subsp. aur MRSA, Staphylococcus aureus subsp. epidermidis, Brucella suis, Tropheryma whipplei, Coxiella burnetti, Bartonella quintana, Chlamydia pneumoniae, Neisseria meningitidis, HACEK organisms namely (Haemophilus parainfluenzae, Haemophilus aphrophilus, Actinobacillus [Haemophilus] actinomycetemcomitans, Cardiobacterium hominis, Eikenella species,
and *Kingella* species) and *Mycoplasma pneumoniae*.\textsuperscript{1,2,7,13,18}

Classically, the predominant causative agent of IE has been the viridans *Streptococcus*, which causes the subacute type of endocarditis. However, in many recently described patient populations, the most frequently isolated microorganism has been *Staphylococcus aureus*, manifested clinically as acute endocarditis.\textsuperscript{14} Congestive heart failure (CHF) complicates more than half of all episodes and it is the leading cause of death among IE patients.\textsuperscript{15} The second most common complication of IE is embolism, which occurs in about 40% of the IE patients.\textsuperscript{16} Frequently, this involves the central nervous system (CNS) with subsequent neurological complications and results in significant morbidity and mortality.

The high morbidity and mortality rate, antibiotic resistance and diverse group of causative pathogens necessitate implementation of alternative strategy for designing drug molecules against infective endocarditis. Systemic protocols were developed combining bioinformatics tools and databases, in other words, known as comparative genomic approach, subtractive genomic approach and pathway analysis method for drug target identification which were successfully applied against many diseases.\textsuperscript{17-22} Implementation of such systematic protocol led to identification of 18 common drug targets from eight selected bacterial pathogens of IE.\textsuperscript{18}

Fatty acid biosynthesis remains as an important pathway for drug target identification because the pathway is essential and cross-resistance to existing drugs that target other pathways is minimal.\textsuperscript{23} Disruption of fatty acid biosynthesis leads to severely impaired cellular functions and cell death.\textsuperscript{24-26} The first committed step in fatty acid synthesis is the carboxylation of acetyl-CoA by biotin carboxyl carrier protein of acetyl-CoA carboxylase (AccB) to form malonyl-CoA. Designing inhibitors targeting AccB of *Streptococcus mitis* would stop fatty acid biosynthesis in IE pathogens leading to their death. Thus, structure-based virtual screening was practiced in the present study to discover potent inhibitors of AccB.

### MATERIAL AND METHODS

#### Homology modelling

In the present study, *Streptococcus mitis*, *Staphylococcus aureus subsp. aur MRSA*, *Enterococcus faecalis*, *Brucella melitensis biovar Abortus*, *Coxiella burnetti*, *Bartonella quintana*, *Chlamydia pneumoniae*, and *Legionella pneumophila*, the eight major pathogens of IE,\textsuperscript{1,2,8-13,18} were selected for identification of common potential drug targets towards IE. AccB was identified as common drug target against the eight selected pathogens causing IE.\textsuperscript{18} *Streptococcus mitis*, one of the most predominant pathogens of IE, was selected as reference organism\textsuperscript{2,8} for common drug target identification. The protein sequence of AccB of *Streptococcus mitis* was retrieved from The Institute of Genomic Research (TIGR) (now called J. Craig Venter Institute), Rockville, MD, USA. The involvement of AccB metabolic pathway was analyzed at the Kyoto Encyclopaedia of Genes and Genome (KEGG) database.\textsuperscript{26} Structural templates for AccB were identified from the protein data bank (PDB). The structural templates 2C1G and 1BDO were considered for homology modeling of AccB. The homology model was generated in complex with biotin using Modeller 9 v8.\textsuperscript{27}

#### Model validation and allosteric site prediction

The predicted homology model for AccB of *S. mitis* was evaluated using the online tools PROCHECK,\textsuperscript{28} ProSA\textsuperscript{29} and ProQ.\textsuperscript{30} The allosteric site residues were visualized using PyMOL.\textsuperscript{31} Multiple sequence alignment of AccB from eight selected pathogens of IE was performed to see whether predicted allosteric site residues were conserved across these organisms.
High throughput virtual screening
Biotin was searched for structural analogs from Ligand.Info Meta-database tool. The tool performs a 2D structural geometry search from eight renowned small molecule databases such as Havard's ChemBank, ChemPDB, KEGG Ligand, Druglikeness National Cancer Institute (NCI), Anti-HIV NCI, Unannotated NCI, AkoS GmhB, Asinex Ltd etc for the queried small molecule. Maximum of 50 structural analogues of biotin were retrieved from each of eight structural databases of Ligand Info. Consequently, an in-house library of biotin structural analogues was compiled.

Molecular docking
Tertiary structure of AccB and in-house library of biotin structural analogs were imported to Maestro v9.0 environment for molecular docking to investigate binding affinity of the ligand dataset towards AccB. Each atom of protein and ligands must be fixed for any potential aberrations before molecular docking for accurate prediction of binding affinity and interactions. The AccB structure was preprocessed with the protein preparation workflow in the Maestro v9.2. All hydrogens were added to AccB which subsequently minimized with the OPLS 2005 force field and the impact molecular mechanics engine setting the maximum root mean square deviation (RMSD) of 0.30 Å. Minimization was performed constraining the heavy atoms with the hydrogen torsion parameters turned off, to allow free rotation of the hydrogen atoms. The ligands were prepared to expand protonation and tautomeric states at 7.0±2.0 pH units using LigPrep with Epik. High-energy ionization / tautomer states were removed during post LigPrep evaluations. The post LigPrep parameters were restrained to report at four stereo isomers for each compound. Pharmacological properties of the prepared ligands were assessed through Lipinski’s rule of five. The compounds with poor pharmacological properties were discarded. The compounds with reactive functional groups were eliminated by applying reactive filter parameters. After ensuring that the protein and ligands were in the accurate form for docking, the receptor-grid files were generated centered on the allosteric site residues. To soften the potential for non-polar parts of the receptor, van der Waal radii of receptor atoms were scaled to 1.00 Å with a partial atomic charge of 0.25. A grid box of size 20 Å × 20 Å ×20 Å was engendered on the primed protein structures by selecting allosteric site residues of AccB. A three phased subsequent docking protocol was implemented to the prepared protein and ligand dataset to rank the ligands based on their binding affinities towards AccB and to study interactions of the best lead. Glide high-throughput virtual screening (HTVS), standard precision (SP) and extra precision (XP) methods were applied. The XP docking method is highly accurate and generates 10000 poses for each ligand during docking and reports the best pose based on the energy term Emodel. The best poses of each ligand were further ranked based on XP Gscore. Lower XP Gscore for a ligand indicates better binding affinity towards protein. The cut-off XP Gscore parameter for XP docking was set to 0.0 kcal/mol, a constraint set to discard ligands with positive XP Gscore from the final docking output.

Molecular dynamic simulations
Molecular dynamic simulations were performed to docking complex of AccB-lead1 to evaluate the stability, conformational changes and to get insights into the natural dynamics on different timescales in solution. Simulations were carried out using Desmond v3.0 implemented in Maestro v9.0 graphical user interface. The system was embedded with simple point charge (SPC) water model and neutralized by replacing solvent molecules with Na+ and Cl ions. The final system containing approximately 30,570 atoms was set with periodic boundary conditions, the particle mesh Ewald (PME) method for electrostatics, a 10
Å cutoff for Lennard-Jones interactions, SHAKE algorithm for restricting motion of all covalent bonds involving hydrogen atoms and simulated through a multistep protocols devised in Maestro v9.2. In brief, the full system was minimized with maximum 200 iterations of a hybrid of the steepest descent and the limited memory Broyden-Fletcher-Goldfarb-Shanno (LBFGS) algorithms, with a convergence threshold of 50.0 kcal/mol/Å² followed by a similar unrestrained minimization with a convergence threshold of 5.0 kcal/mol/Å². The minimized system was relaxed in short span simulations with NVT ensemble (constant number of atoms N, volume V and temperature T) for 12 picoseconds (ps) restraining all non-hydrogen solute atoms (temperature = 10 K). Similar minimization in NPT (constant number of atoms N, pressure P and temperature T) ensemble was continued for a simulation time of 12 ps. Further, a 24 ps simulation in the NPT ensemble restrained with solute nonhydrogen atoms (temperature = 300 K) and a 24-ps simulation in the NPT ensemble with no restraints (temperature = 300 K) was performed. The temperatures and pressures in the short initial simulations were controlled using Berendsen thermostats and Berendsen barostats, respectively. The relaxed system was simulated for a simulation time of 10 ns with a time step of 2 fs, NPT ensemble using a Nose-Hoover thermostat at 300 K and Martyna-Tobias-Klein barostat at 1.01325 bar pressure. Trajectories of the docked complex in the system during simulations were recorded for every 4.8 ps.

RESULTS

AccB of S. mitis consists of 160 amino acid residues. Three-dimensional structure of AccB was not solved experimentally, hence, structural templates for the drug target were searched from the PDB. Two structural templates namely, 1BD0 (identity: 49%; similarity: 62%; query coverage 47%) and 2C1G (identity: 30%; similarity: 48%; query coverage 31%) were identified as suitable templates to build a reliable homology model of AccB.

The AccB homology model developed through comparative modeling was evaluated using diverse model validation techniques. Stereoochemistry assessment of the AccB showed 89.1% residues in most favorable regions of the Ramachandran plot (Supplementary Figure 1A). ProSA evaluation for the homology model revealed Z score of -4.08, which is well within the range of native conformations of crystal structures of similar length (Supplementary Figure 1B). ProQ tool assessment showed LG score of 1.236, is of fairly good quality that represents the model. The residues such as Tyr97, Ala99, Ala100, Gly101, Pro102, Lys127 and Met129 were determined to constitute biotin binding site (Figure 1); hence, were selected to define a grid for structure-based virtual screening.

The structure was optimized to add hydrogen atoms, remove water molecules and remove steric classes in 3D structure. Further, energy was minimized applying OPLS 2005 force field to obtain a structure with lower energetic conformation. A total of 1352 lower energetic protonation and tautomeric states were generated from 357 compounds of in-house library, 1131 conformers were passed in Lipinski’s filter from 1352 conformers. Removal of ligand conformations with reactive functional group further squeezed to the ligand dataset to 1048 conformations. The 1048 ligand conformations with good pharmacological property were selected for three stages of structure-based virtual screening. All 286 ligands were docked in HTVS docking and top 10% ligands (28) were selected for SP docking. Top 10% ligands (2) from SP docking were re-ranked in XP docking based on XP Gscore. Two lead molecules with good binding affinity towards AccB were obtained. Both the leads revealed better XP Gscores compared to biotin (–2.216 kcal/mol). Lead1 showed lowest XP Gscore of –5.847
The molecular interactions of docking complex of AccB - lead1 showed that Ala99, Ala100 and Lys127 were involved in intermolecular hydrogen bonding; Tyr97, Ala99, Ala100, Gly101, Ile122, Glu124, Lys127 and Met129 were involved in good van der Waal contacts (Figure 2). Similarly, AccB - lead2 showed that intermolecular hydrogen bonding involving Ala99 and Ala100, while Tyr97, Ala99, Ala100, Gly101, Pro102, Ile122, Glu124, Lys127, Met129, and Asn147 were involved in good van der Waal contacts (Figure 3).

The dynamical properties of AccB-lead1 docking complex were analyzed from trajectory data obtained from 10 ns MD simulations. The energy of the system was relatively stable throughout the simulations (Supplementary Figure 2A).
DISCUSSION

AccB catalyzes the carboxylation of acetyl-CoA to form malonyl-CoA where in biotin carboxyl carrier protein component of the enzyme carries the carboxyl group. AccB, a common protein among eight IE pathogens, was proposed as potential drug target for discovery of novel drug molecules against infective endocarditis. Comparative analysis against database of essential genes (DEG) suggested that AccB was critical to bacterial survival. Pathway analysis at KEGG\(^1\) had revealed that AccB plays pivotal role in fatty acid biosynthesis. There is no alternative pathway for fatty acid biosynthesis without the involvement of AccB; therefore, AccB is an essential and unique enzyme for fatty acid biosynthesis in pathogens of IE. The enzyme, being non-homologous to human, was proposed as a common drug target against pathogens of IE.

Predicting 3-dimensional protein structure from amino acid sequences becomes relatively easier if close homologous proteins have been solved, and high-resolution models can be built by aligning target sequences to the solved homologous structures. A suitable structural template was not identified for AccB to determine its complete structure; hence, two structural templates were selected for multiple template-based homology modeling incorporating ligand biotin. The stereo chemical quality of the model revealed that the model was reliable for structure-based drug designing. The computational engine used for the calculation of scores and plots was standard ProSA which uses knowledge-based potentials of mean force to evaluate model accuracy. Z-score of the AccB model was within the range of scores typically found for proteins of similar size belonging to experimentally-determined structure.\(^2\)

The Glide v5.7 software offers full spectrum of speed and accuracy for virtual screening from small molecule database to extremely accurate binding mode predictions, providing consistently high enrichment at every level. Accurate binding affinity prediction between protein and ligand in a computational environment requires careful optimization of their 3D structures. Optimization of protein and ligand in Maestro ensured inclusion of ligands with good pharmacological properties, lower energetic stereoisomers and tautomers of all structural analogs for accurate binding affinity and binding mode prediction. The three modes of virtual screening were used subsequently for fast screening of small molecules. The XP mode of docking is highly accurate and penalizes highly for minor steric classes. Hence, the compounds predicted to have good binding affinity were ranked based on XP Gscore. Lower XP Gscore represents higher binding affinity of the ligand towards protein. The result revealed that the two lead molecules could competitively inhibit AccB than biotin, suggesting that they are potential AccB inhibitors. It is noteworthy to mention that biotin interacted with AccB at the allosteric site through one intermolecular hydrogen bonds with Ala99 (Figure 1) while that of lead1 formed three intermolecular hydrogen bonds with Pro102 and Lys107 (Figure 2).
bonds (Figure 2), and lead2 formed two inter-molecular hydrogen bonds (Figure 3) including Ala99. The result revealed that the interactions of lead1 and lead2 were more stable compared to biotin.

Lead 1 and lead 2 showed the lowest XP Gscore compared to biotin, hence is of significant interest for designing potential inhibitors against AccB. The binding mode analysis revealed that lead1 and lead2 were blocking important amino acid residues through hydrogen bonds and van der Waal interactions. MD simulations for AccB-lead1 complex carried out closer to the physiological environment condition with the system embedded with water molecules, temperature and pressure. Therefore, the binding orientations of lead molecules obtained after simulations showed better correlation to their biologically active states. MD simulations also quantified stability of the docked complex.

RMSD measures overall dynamic fluctuations of docking complex in each trajectories during simulation time. Therefore, the lower RMSD of the system indicate small structural re-arrangement and lesser conformational changes. The energy plot, RMSD plot and RMSF plot analysis revealed that the AccB-lead1 docking complex was conformationally stable.

The conformational stability of the system was established and interaction stability of the system was monitored in each trajectories. The hydrogen bond monitoring results also revealed stable interaction of lead1 and AccB, indicating that lead1 could be considered as a potent AccB inhibitor.

Common drug targets of IE pathogens are of significant interest for structure-based drug designing. AccB is the first enzyme of fatty acid biosynthesis that catalyzes formation of malonyl-CoA from acetyl-CoA. The protein, essential for pathogen's survival, homologous among eight selected infective endocarditis pathogens and non-homologous to human, was identified as a common potential drug target. The reliable AccB 3D model and biotin of AccB were built for structure-based virtual screening. Two lead molecules were proposed based on comparative analysis of binding affinity (XP Gscore) towards AccB. Pharmacological properties of these two lead molecules correlated favorably with more than 95% approved drug molecules, indicating that they are the potential AccB inhibitors. MD simulations for the AccB-lead1 docking complex showed that the complex maintained conformational stability and interaction stability in physiological environment. Therefore, lead 1 is proposed as a potent inhibitor to start with experimental validation towards designing drug molecules for treatment of infective endocarditis.

ACKNOWLEDGEMENTS

This work was carried out at the Bioinformatics facility (BIF) sanctioned by the Department of Biotechnology (DBT), Ministry of Science and Technology, Government of India. One of the authors (VP) was supported by Senior Research Fellowship (SRF) awarded by Indian Council of Medical Research (ICMR), New Delhi.

REFERENCES


Supplementary figures are available at URL: http://svimstpt.ap.nic.in/jcsr/jhome.htm