Structural analysis and molecular docking of potential ligands with chorismate synthase of Listeria monocytogenes: A novel antibacterial drug target

Md. Musharaf Hossain1,2, Pradip Kumar Roy1, A T M Jannatul Mosnaz1, Shahriar Kabir Shaki1,
Md. Mehedi Hasan1 and Shamsul H Prodhan1*

1Department of Genetic Engineering and Biotechnology, Shahjalal University of Science and Technology, Sylhet-3114, Bangladesh
2Department of Biology, University of Saskatchewan, Saskatoon, SK S7N5E2, Canada

Received 15 January 2014; revised 12 October 2014

Listeriosis, in particular that caused by Listeria monocytogenes, is a major foodborne pathogen, and its control is becoming difficult because of widespread emergence of drug resistance strains. Chorismate synthase (CS), an essential enzyme of shikimate pathway present only in bacteria, fungi, plant and some apiculplexan parasites, is a validated potential antimicrobial drug target. Antimicrobial development through the elucidation of essential structural features of the CS of L. monocytogenes (LmCS), identification and prioritization of potential lead compounds targeted against LmCS were done. Structure-based virtual screening and docking studies were performed using Autodock tools to retrieve potential candidates with high affinity binding against LmCS model from several ligand repositories. The potency of binding was also checked with other structurally similar CS from Streptococcus pneumoniae (SpCS) and Mycobacterium tuberculosis (McCS). The sequence and structural studies revealed LmCS was similar to other CS structures (1Q1L, 1QXO, 1R52, 1R53, 1SQ1, 1UM0, 1UMF, 1ZTB, 2011, 2012, 4ECD and 2G85) with each monomer presenting β-α-β sandwich topology with a central helical core. Molecular docking studies and ADME/Tox results revealed that ZINC03803450 and ZINC20149031 were most potent molecules binding into the active site of LmCS. Other two ligands ZINC13387711 and ZINC16052528 showed a strong binding affinity score against all three structures (LmCS, SpCS and McCS) and bind to LmCS with the predicted inhibition constant (Ki) values of 22.94 nM and 35.84 nM, respectively. A reported benzofuran-3-[2H]-one analog CHEMBL135212 with good ADME/Tox properties and experimental IC50 (nM) value of 7000 nM with SpCS could also be considered as a potential inhibitor of LmCS, as compared to previously reported 41 benzofuran-3-[2H]-one analogs against SpCS. This information will assist in discovering those compounds that may act as potent CS inhibitors. Further experimental studies and evaluation of structure-activity relationship could help in the development of potential inhibitors against listeriosis, as well as antibacterial chemotherapy.

Keywords: Virtual screening, Listeria monocytogenes, Chorismate synthase, Molecular docking

The emergence of multiple drug resistance (MDR) bacterial strains is increasing worldwide, partly due to the prolonged and uncontrolled use of drugs. Emerging antibiotic resistance in Listeria is reported recently pointing at overprescription of drugs in clinical treatments and heavy use of antibiotics as growth promoters in livestock husbandry as the major reasons for their emergence. Listeria monocytogenes is a Gram-positive facultative opportunistic intracellular foodborne pathogen often found in food and elsewhere in nature and has become an important cause of human foodborne infections worldwide. Listeria is more likely to cause death than other bacteria that cause food poisoning. These bacteria are well equipped to survive food processing technologies, as they are tolerant to extreme conditions, such as low pH, low temperature and high salt conditions. In 60% cases, human infection most commonly involves the central nervous system, where meningitis is usually present and infection in pregnant women may lead to abortion, stillbirth, or premature birth.

Human immunodeficiency virus (HIV) infection is also a significant risk factor for listeriosis. It has been estimated that the risk of contracting listeriosis is 300 to 1,000-times higher for AIDS patients than for the general population. The drug dose limit against Listeria spp. is also increased for currently prescribed antibiotics and more often combinational therapy is preferred. Therefore, there is a need for identification of new drug targets, unique and crucial to different essential pathways in microorganisms, particularly those not similar as pathways in human.

*For correspondence:
E-mail: shamsulhp@yahoo.com; shamsulgeb@gmail.com
1For Supplementary data see www.nopr.nicai.res.in
The shikimate pathway is essential to bacteria, fungi and plants as well as apicomplexan parasites which links carbohydrate biosynthesis pathway to aromatic compounds biosynthesis pathway. This seven-step metabolic route of shikimate pathway directs from phosphoenolpyruvate (PEP) and erythrose 4-phosphate to chorismate, a common precursor for the synthesis of aromatic amino acids and many aromatic compounds including, vitamins E and K, folic acid and ubiquinone. These aromatic compounds play essential roles in various functions i.e. UV protection, electron transport, signaling, iron uptake and plant defense etc.\(^7\) CS catalyzes the seventh and final step of the shikimate pathway by the trans-1, 4 elimination of phosphate from 5-enolpyruvylshikimate 3-phosphate (EPSP) to yield chorismate. Thus, inhibition of CS offers a lucrative target for the rational design of novel antimicrobials, as chorismate is the common metabolic branchpoint for the production of aromatic amino acids, vitamins K and E, co-enzyme Q, folate, enterobactin, chloramphenicol, plastoquinones, phenoxyazinones etc. A reduced flavin mononucleotide (FMN) works as co-factor of CS by donating an electron to facilitate the elimination of phosphate in generating chorismate from 5-enolpyruvylshikimate 3-phosphate (EPSP) and gets back to its normal state after the reaction\(^8\). CS is the site of action of the well-known sulphonamide series of antibacterial compounds and recently some benzofuran-3[2H]-one analogs have also been reported to have inhibitory action against CS\(^9\).

Computer-aided drug design (CADD) is an important part of the rational drug design process, in which extensive computer modeling methods are employed to reduce the costs and speed up the drug developing process\(^10\). The compounds that are virtually screened can stem from corporate or commercial compound collections, or from virtual compound libraries. Nowadays, high-throughput docking has become increasingly important in the drug discovery\(^11\). The docking studies have significant hit rate improvements with filtering schemes and other customized parameters compared to random screening, despite the technical challenges in precise prediction of the binding affinity and mode of their interaction with receptor molecule\(^12\).

In the present study, we have generated the model of CS enzyme of L. monocytogenes (LmCS) and have done an in silico structure-based inhibitors study of commercial compounds, natural benzofurans and in vitro characterized the inhibitors series. Besides, we have also done comparative docking analysis of potential inhibitors with structurally similar CS from Streptococcus pneumoniae and Mycobacterium tuberculosis. This study may provide a novel anti-bacterial target for structure-based drug design against listeriosis. Also, the ligands having high affinity bindings with the target may act as potent inhibitors of listeriosis and other diseases caused by S. pneumoniae and M. tuberculosis.

**Materials and Methods**

**Multiple sequence alignment and primary sequence analysis**

Chorismate synthase (CS), a gene product of AroC (YP_014550.1) of Listeria monocytogenes F2365 was retrieved from NCBI database. The CS sequences of L. monocytogenes and structurally similar PDB structures were aligned using ENBack 1.1 web server\(^13\) which requires a PDB file as an input. This web server enables the Basic Local Alignment Search Tool (BLAST) search against PDBAA database with 1e\(^-12\) and all considered CS sequences was then aligned using clustal multiple alignment program. Besides, distinct signature motifs and consensus patterns were identified using ScanProsite web server\(^14\).

**Comparative molecular modeling**

CS, an important enzyme of shikimate pathway which is not present in human, was selected as a drug target in this study. The modeling of three-dimensional (3D) structure of CS was done using homology modeling programs Easymodeller 4.0\(^15\) and Swiss-model server\(^16\). Homology modeling of this protein was done by using a suitable template structure from PDB (http://pdb.org/pdb/home/home.do) through blastP (Protein BLAST) search (http://blast.ncbi.nlm.nih.gov/Blast.cgi) against PDB database with default parameters. From the best hits, crystal structure of SpCS in complex with FMN and EPSP (PDB 1QXO) was selected as a template for model generation. The sequence identity of LmCS with the template structure was 61%. ClustalW program was used for the alignment of the query sequence with the template sequence.

**Model validation**

The stereochemical quality of each model was evaluated by online server Rampage\(^17\) and ERRAT\(^18\) plot that gives a measure of structural error at each residue in the protein. The refined model with least
number of residues in the disallowed region was further validated by VERIFY 3D19 of SAVES server (http://nihserver.mbi.ucla.edu/SAVES/) and ProQ20. ProQ, a neural network-based predictor based on a number of structural features predicts the quality of a protein model. ProQ is optimized to find correct models in contrast to other available softwares and servers those predict protein model quality. Two quality measures, LG score and MaxSub were predicated. ProSA-Web server (https://prosa.services.came.sbg.ac.at/prosa.php) and WHAT IF web server (http://nihserver.mbi.ucla.edu/SAVES/) and ProQ were also employed to evaluate the generated model of protein for potential error. WHAT IF web server calculates the RMS Z score of bond lengths and bond angles.

**Generation of LmCS dimer and tetramer**

CS from various microorganisms and other species exist as a tetramer in solution21. It is assumed that dimeric state of CS may also present in solution, as the dimer interface residues among species are conserved22. The dimer and tetrameric form of LmCS was generated by ClusPro 2.024, a protein-protein docking server (http://cluspro.bu.edu/home.php). For this feature, both the input structures must comprise the same sequence and the number of monomers that constitute the multimer must be defined. Users can select whether the structure will be a dimer, trimer, tetramer or pentamer25. The PyMOL was used to visualize the dimeric and tetrameric binding interfaces.

**Virtual screening setup**

**Preparation of the target protein and active site prediction**

The modeled structure of LmCS was prepared for structure-based virtual screening and molecular docking processes by removing all co-factors, iodide ions and water molecules. The binding site of the LmCS model was determined on the basis of previous studies and using the Computed Atlas of Surface Topography of proteins (CASTp) server26 at 1.4 Å radius probe. The internal cavity surface volume of the ligand binding site was calculated.

**Preparation of ligands for virtual screening**

Compound catalogs of Drugbank (experimental) from ZINC database and 1880 compounds with a rich structural and pharmacophore diversity (http://zinc.docking.org/catalogs/ncidiv) were chosen as the molecules in demand for virtual screening. The NCI diversity set III (NCIDS III) is a collection of approx. 1880 compounds that are structurally representative of a wide range of pharmacophores. Besides, some reported benzofuran-3(2H)-one analogs and natural benzofurans were also chosen for virtual screening. Prior to the docking, the small molecule libraries (ZINC database and NCIDS III) available in 3D-MOL2 format were converted into pdbqt file with the program raccoon.

**Virtual screening with AutoDock Vina**

AutoDock Vina27 is a new open-source program for drug discovery, molecular docking and virtual screening, offering multi-core capability, high performance and enhanced accuracy and ease of use. The preparation of the pdbqt file of receptor and determination of the grid box size were carried out using AutoDock Tools version1.5.6rc1 (The Scripps Research Institute, LaJolla, USA). Autodock Vina was used due to its accuracy and speed, which is approx. two orders of magnitude faster than its predecessor AutoDock428. AutoDock Tools version1.5.6rc1 was utilized to prepare the input pdbqt file for LmCS model, to set the size and the center of the grid box. Polar hydrogen atoms were added to the receptor. The grid box size at different dimension was set at 66 Å each and the spacing between grid points was kept 0.375 Å. AutoDock Vina requires the pdbqt input files of ligands to be prepared using raccoon29. The predicted binding affinity (kcal/mol), which indicates how strongly a ligand binds to the receptor, was calculated based on the scoring function used in AutoDock Vina. A more negative binding affinity indicated stronger binding.

**ADME/Tox properties**

The physico-chemical properties of the compounds, such as logP value, H-bond donors, H-bond acceptors, molecular weight and rotational bonds for ligands were calculated using the free ADME/tox filtering tool 230. The top hits from each screen were initially filtered for drug likeness by their adherence to Lipinski’s “rule of five”31 with no more than one violation. The analysis of World Drug Index (WDI), which led to Lipinski’s ‘rule-of-five’ identifies several critical properties that should be considered for compounds with oral delivery in mind. These properties, which are usually viewed more as guidelines rather than absolute cut-offs and it is recommended that compounds should conform to two or more of these rules32.

Absorption, distribution, metabolism, excretion and toxicity (ADME/Tox) are main five parameters to test
the drug likeness of a molecule because these properties account for the failure of most drug candidates in the clinical phases. After selecting potential ligands based on their binding affinity towards the target molecule, their ADME/Tox properties should be measured which influence the performance and pharmacological efficacy of the compounds as a drug. In addition, it is critical to simultaneously optimize binding affinity/selectivity, pharmacokinetic properties, while avoiding toxicity. Although these properties are usually measured in the lab, they can now be predicted with the help of bioinformatics\textsuperscript{33}. In an attempt to design new potential CS inhibitors, ADME/Tox properties of all the drug candidates were assessed by ACD/ADME Suite and ACD/Tox Suite platform maintained by Advanced Chemistry Development, Inc., Canada\textsuperscript{34}.

Molecular docking studies

**Molecular docking with autodock tools**

To carry out the docking simulation, we used the AutoDock 4.2 suite as molecular-docking tool\textsuperscript{35}. It is suitable software for performing automated docking of ligands to their macromolecular receptors. Typically, the ligands are substrates or drug candidates and the macromolecule is a protein of known three-dimensional structures. In this docking simulation, we used rigid docking protocols, in which the target protein \textit{Lm}CS and ligand molecules were kept as rigid. The graphical user interface program "AutoDock Tools" was used to prepare, run and analyze the docking simulations. Polar hydrogens were added into the receptor PDB file for the preparation of protein for docking simulation. Prior to docking, a pre-calculated grid map is required by AutoDock for each atom type that surrounds the region of interest in the receptor molecule.

In the present study, the binding site was selected based on the active site component and the substrate (EPSP) binding site of the \textit{Lm}CS (receptor). The substrate binding site residues were covered as grid box. The grid box size was set at 66 Å, 66 Å and 66 Å (coordinates at X, Y, and Z axis), though it was changed depending on the ligand size. AutoGrid 4.2 Program, supplied with AutoDock 4.2 was used to produce grid maps. The spacing between grid points was 0.375 Å. Docked conformations were clustered using a tolerance of 2.0 Å RMSD using AutoDock 4.2. The PyMOL Molecular Graphics System\textsuperscript{35} and DS Visualizer (Accelrys, Inc., USA) were employed to visualize and modify the receptor and ligand structures.

**Comparative molecular docking with Molegro Virtual Docker**

Best found hits and screened compounds along with reported Benzofuran-3[2H]-one analogs were also docked with structurally similar existing PDB structures of CS from \textit{S. pneumoniae} (\textit{Sp}CS) and \textit{M. tuberculosis} (\textit{Mt}CS) using Molegro Virtual Docker v4.0\textsuperscript{36,37} by considering the notion that they have almost similar active site residues and organization. The docking study of these CS PDB structures was performed with the molecular docking algorithm MolDock\textsuperscript{38}. For docking, the MolDock scoring function, which is based on a piecewise linear potential and a re-ranking procedure was applied to the highest ranked poses to increase docking accuracy. Water molecules with the protein structures were excluded from the docking experiments\textsuperscript{39}. The default setting of software parameters was used in this study and ligands were given as PDB format. Binding sites were restricted within a 15 radius circle centered at the observed binding of ligand in protein complex. Due to the stochastic nature of ligand-protein docking search algorithm, 10 runs were conducted and 5 docking poses were retained for each ligand. During docking run with the receptor, poses of the ligand with the highest ranked Moldock scores were compared and selected\textsuperscript{40}.

**Results and Discussion**

**Multiple alignment and primary sequence analysis**

The sequence of \textit{Lm}CS encoded by \textit{AroC} gene on chromosome no. 1 was obtained from NCBI with accession no YP_014550.1. The protein consists of 388 amino acids with a predicted molecular weight of 42,188 kDa and theoretical pI value 6.59. Sequence alignment of considered CS sequences revealed that \textit{Lm}CS had maximum homology (61% identity) with \textit{Sp}CS protein structure, PDB 1QXO which was used as a template for the generation of an \textit{Lm}CS model in this study. Besides, some other structures, such as 1Q1L from \textit{Aquiflex aeolicus} (52%), 4ECD from \textit{Bifidobacterium longum} (49%) and PDB structures 1ZTB, 2011 and 2012 from \textit{Mycobacterium tuberculosis} showed significant identities. A structure of CS from fungus \textit{Saccharomyces cerevisiae} (PDB1R52) was found to possess considerable identity (36%) with target \textit{Lm}CS. It has already been established that a sequence identity higher than 25% between two proteins is indicative of similar three-dimensional structures\textsuperscript{41}.

CS has different functional aspects. With regard to its activity, CS requires the reduced FMN as
co-factor. The category of CS which requires reduced FMN exogenously has been classified as mono-functional which is generally found in bacteria, algae, plants and protozoa, including *Plasmodium* species. On the contrary, CS enzymes from fungi e.g. *Neurospora crassa* and *S. cerevisiae* have an additional NADPH: FMN oxidoreductase activity and thus are categorized as bi-functional enzymes. CS represents a distinct class of a conserved enzyme family, as it lacks a primary sequence similarity to other classes of enzymes.

The *LmCS* model was found to exhibit three highly conserved signature motifs viz. sig.1: \texttt{\textbf{GESHG}}pGLT\texttt{\textbf{IIEGIP}}\textsuperscript{22}, sig.2: \texttt{\textbf{ErsSArE}}ttvr\texttt{VaaGAV}\textsuperscript{145} and sig. 3: \texttt{\textbf{RSDS}}Cavpa\texttt{AsVVaa}EAV\textsuperscript{353}, where small letter residues indicate a position that can be replaced by any amino acids in CS of different species and bold faced residues are highly conserved at given position. CS sequences known to belong to this class can be detected by the following consensus patterns:\textsuperscript{23}

(i) Signature 1: G-[DES]-S-H-[GC]-x\textsubscript{2}-[LIVM]-[GTIVLAMS]-x-[LIVTM]-[LIVM]-[DEST]-[GH]-x-[PV],
(ii) Signature 2: [GE]-x\textsubscript{2}-S-[AG]-R-x-[ST]-x\textsubscript{2}-[VT]-x\textsubscript{2}-[GA]-[STAVY]-[LIVMF],
and
(iii) Signature 3: R-[SHF]-D-[PSV]-[CSAVT]-x\textsubscript{4}-[SGAIVM]-x-[IVGSTAPM]-[LIVM]-x-E-[STAHNCG]-[LIVMA], where x denotes any amino acid, and amino acids in brackets depict the option at a given position. In Supplementary Fig. 1 (For details see footnote) residues marked with black star denote the binding residues of CS with EPSP. Of these, six highly conserved EPSP binding residues reside on three signature motifs.

**Generation and validation of 3D model of LmCS**

Homology search of *LmCS* sequence against PDB database yielded eight crystal structures better than threshold value. The high sequence identities between the target and template are shown in Supplementary Fig. 1, which ensured the quality of the generated model (Fig. 1). Ramachandran plot of 3D model, generated by Swissmodel showed 97.7% residues present in the favored region, 2.3% in allowed region and no residues in outlier region. On the other hand, model generated by the GUI version of modeler had one residue residing on the outlier region.

Furthermore, the overall quality factor of the 3D model generated by Swissmodel (98.421) was better than that of Easymodeller 4.0 which was about 87.895. VERIFY 3D, Pro-SA analysis showed that protein folding energy of modeled structure was in

![Fig. 1—Protein model of LmCS with β-α-β sandwich fold and substrate binding pocket [A. Cartoon diagram of LmCS monomer generated by Swissmodel showing the unique β-α-β sandwich fold; and B. Active site display of LmCS model (wireframe); Green balls represents amino acids residue around the active site and the name of the residues are also shown in green color]](image-url)
agreement with the validation of the model Supplementary Fig. 2 (For details see footnote). The RMS-Z scores for bond lengths and bond angles were calculated using “What If” web interface. This web server provides information about bond lengths and bond angles of a model that deviate more than 4 sigma from standard bond angles of Engh and Huber reference value for protein residues. The RMS-Z score for bond lengths and bond angles was expected to be near 1.0 for a normally restrained dataset. The RMS-Z scores for bond lengths and bond angles of the target model structure were found to be 0.624 and 0.986, respectively.

**Predicted monomeric structure of LmCS**

The monomeric structure of LmCS model consists of 10 α-helices (total 17 helices including α and 3 helices) and 17 β-strands. Generally, CS structure has been earlier predicted to be an α-β barrel as a result of secondary structure prediction efforts. However, each CS monomer consists of a helical core sandwiched by two beta sheets, which gives itself a characteristic β–α–β sandwich topology. The core of LmCS showed topological resemblance with other CS in terms of unique β–α–β sandwich fold. The noble fold was composed of two anti-parallel four stranded β-sheets (↑β1, ↓β2, ↑β6, ↓β3 and ↑β7, ↓β9, ↑β15, ↓β10) and four α-helices (↑α1, ↓α5, ↓α9, ↑α8), which were sandwiched between these two β-sheets (Fig. 1). These two major β-sheets may be referred to as N and C-terminal β-sheet layers, respectively. This 4β–4α–4β sandwich fold observed in this predicted model was in accordance with the reported folding characteristics of the template (1QXO) structure. This β–α–β sandwich conformation provides a scaffold for substrate (EPSP) and cofactor (FMN) binding to the protein. The amino acid residues reside on these sheets and helices show diverse functions, such as active site and oligomeric structure formation, dimer and tetramer stabilization, hydrophilic and hydrophobic interactions etc. which have been further discussed in the following sections.

Turns and loops play an important role in protein structures, connecting together beta-strands, strands to alpha-helices, or helices to each other. Beta-turn is a loop that links two β stands by 2-5 amino acids; more frequently, small and side chainless amino acid glycine (Gly) is found in loop region. The monomeric model of LmCS had four β-hairpins (β4–β5, β7–β8, β11–β12, β13–β14). The β11–β12 hairpin had three residues (Lys-Glu-Asp) in turn and rest of them (β4–β5, β7–β8, β13–β14) had two residue turns. The β7–β8 hairpin had type II turn, as it contained side chainless amino acid “Gly” in both positions, whereas the turns in other hairpins (β4–β5, β13–β14) were type I.

Many residues (Ser9, His10, Arg45, Arg48, His110, Ser132 and Arg337) in the loop regions were mainly involved in interaction with the substrate. Their flexible nature might allow EPSP to accommodate itself into the substrate binding pocket of LmCS. The amino acid sequences in turn regions may be variable. But, in some cases, when a loop has some specific functions, for example interaction with another protein or formation of the active site pocket of a protein, the sequence may be conserved. Earlier study has reported a model of CS of Plasmodium falciparum with predicted three β-hairpins, where four α-helices are found to be sandwiched by two anti-parallel five stranded β-sheet layers.

**Oligomeric form of LmCS**

CS dimer is the basic structure of its tetrameric form and the dimerization plays a functionally critical role. The crystal structure of CS from different species determined to be tetramer in solution and characterized as a dimer of dimers by gel filtration chromatography and dynamic light scattering analysis. So, the dimeric and tetrameric forms of the modeled structure were generated using ClusPro protein-protein docking server. The generated dimeric form of LmCS was found to have a characteristic eight anti-parallel β-strands (↑β7, ↓β9, ↑β15, ↓β10, ↑β10′, ↓β15′, ↑β9′, ↓β7′), formed by assembling two N-terminal four stranded β-sheet of two subunits (either between A and D subunits, or B and C subunits) like the previously solved CS structures (1QXO, 1Q1L and 1ZTB) (Fig. 2).

The dimer interface represented several interactions, mainly involving loops (β9–α8 and α8–β10) and α-helices (α3, α4 and α8). Two highly conserved residues (lys254 and phe258) present on β10 might be crucial for the formation of eight anti-parallel β-strands organization and dimer stabilization. About 31 inter-subunit hydrogen bonds assisted in the stabilization of the dimer. The main residues involved in the LmCS dimerization were Tyr3, Thr5, Lys118, Arg130, Leu229, Gly230, Lys238, Asp240, Val248, Ile250, Asn251, Lys254,
Phe258 and Glu357 (Fig. 3), where bold faced residues are strictly conserved and others are semi-conserved in almost all CS sequences from bacteria.

The tetrameric model of \( Lm \) CS was formed by two dimers interacting via their N-terminal faces with \( \beta \)-sheet stacking with the other as \( \beta-\beta \) sandwich topology. Several secondary structural features involved in the dimer-dimer interaction were N-terminal end \( \beta \) stands \( \beta_1 \) and \( \beta_2 \) and extremely short \( \beta_4 \), helix \( \alpha_9 \), and loop structures L1 within signature motif 1, L9 and L30. The N-terminal end \( \beta \) stands (\( \beta_1 \) and \( \beta_2 \)) of each monomeric subunit was the major component of the tetrameric conformation of \( Lm \) CS which interacted with the equivalent portion of an adjacent dimer to give \( \beta-\beta \) sandwich form. Those residues that constituted this N-terminal part, had major contribution to \( \beta-\beta \) stacking during tetrameric structure formation and these residues provided added stability to the multimeric state of \( Lm \) CS.

Besides, several other interactions (mostly hydrophobic) and interacting residues might provide added stability to the multimeric state of \( Lm \) CS.

However, predicting the structure of a macromolecular complex from their separate components is a challenging task, particularly when one or both proteins undergo conformational change(s) upon binding. So, the ultimate challenge is to adopt a scoring strategy amenable enough to tolerate these subtle changes and specific enough to mark a difference between natural and “misdocked” conformations. The docking algorithm of ClusPro uses atom-based potentials which may be extra-sensitive to the precise location of the interacting atoms and may not be flexible enough to tolerate changes upon interactions. So, it might unlikely give us the exact native look of protein-protein complex. Although it is improbable that rigid scoring function could provide the exact geometric picture of a macromolecular complex and the interactions. Nevertheless, for some cases, it does give high quality docking models that are perfectly suitable for biological and functional annotations.

Active site prediction of \( Lm \) CS and EPSP Binding

The amino acid residues involved in formation of the binding pocket are predicted and validated on the
basis of existing literature\textsuperscript{43,48,49}. Thus, all these residues are predicted as important residues for docking and later used to create grid during docking study. Fig. 1 shows the residues surrounding the binding pockets of \textit{LmCS}. This binding pocket comprised of both the substrate (EPSP) and cofactor (FMN) binding site, as there were some residues having interaction with both the compounds. In \textit{SpCS} tetramer, FMN molecules are bound to a deep pocket and the bulk surface area of it is buried within the protein. As such, only an edge of the isoalloxazine rings opposite to the ribityl chain being visible from the surface\textsuperscript{43}. The binding of EPSP and FMN is ordered, where the escape of FMN in FMN-bound CS is found to be totally blocked by the substrate binding\textsuperscript{50}. To have a precise idea about the substrate binding site and the residues involved, the docking of EPSP and \textit{LmCS} was performed with AutoDock 4.2.

It was expected that EPSP, a negatively charged molecule, will interact with a number of positively charged residues of \textit{LmCS}. In the final docked structure, interacting residues formed an extremely basic environment favorable for EPSP binding. The residues involved in polar interactions with EPSP were \textit{Ser9, His10, Arg39, Arg45, Arg48, His110, Ser132, Ala133, Arg134} and \textit{Arg337} (Fig. 4), where bold faced residues are strictly conserved and six of them resides on 3 signature motifs of CS sequences,
except Arg39 and His110. Seven basic interacting residues were His10, Arg39, Arg45, Arg48, His110, Arg134 and Arg337 which provided the basic environment which might favor EPSP binding.

Identification of putative inhibitors of LmCS

The global occurrence of resistance among emerging and re-emerging bacterial pathogens to several drugs is one of the extreme threats to and endangers public health. Thus, there is a real need of the search for a new antibacterial directed toward new targets. CS is the noble target for developing antibacterial drugs. Generation of a reliable model has opened the possibilities of computer-assisted inhibitor design in order to inhibit pathogens. To accomplish this objective within a feasible time-scale, we used AutoDock Vina for virtual screening of ligands and lead compounds prior to the docking study.

Virtual screening results

Virtual screening, used widely for the identification of lead compounds in drug discovery programs is generally divided into ligand-based and structure-based virtual screening. A total of 8811, out of 12070 (10190 from Experimental Drugbank and 1880 from NCIDS III) compounds passed the assigned Lipinski filter. However, all the ligands that failed to pass the rules were also virtually screened under a separate batch to see their binding affinity scores. A brief analysis of top binders, however, revealed that most compounds that failed the Lipinski rules were unlikely drug candidates. Although some failed compounds against the Lipinski rules also had good docking scores, this minimal criterion was imposed to focus the scope of our search to the more promising compounds.

EPSP as positive control for docking studies

In molecular docking, the size and center of coordinates of grid box needs to be validated in order to ensure that ligands bind to the binding pocket in the correct conformation. Our aim of this work had been to block the binding of the substrate (EPSP) to LmCS receptor molecule. So, we selected EPSP as the control molecule and the area of EPSP binding as grid box center. We docked the EPSP into the active site of the protein to analyze its binding pattern. The detailed view of docking of EPSP to the active site is shown in Fig. 4, which exhibited molecular docking with \( K_i \) value of 10.22 nM. EPSP formed 12 hydrogen bonds with 10 residues of the active site of LmCS. The residues involved in hydrogen bonding have been discussed earlier. The hydrogen bond donors and hydrogen bond acceptors of the receptor and ligands are tabulated in Supplementary Table 1. No residues were found to have van der Waals interaction with the substrate molecule in the active site.

Docking of best ligands from NCIDS III and Experimental Drugbank

AutoDock Vina used to screen large dataset of ligands, so the precision of docking found not comparable with usual docking software which focuses on one piece at a time. Thus, one should not be confused with the score obtained from virtual screening and usual docking. After virtual screening, the best 8 compounds with respective binding affinities to the active site of LmCS were identified from the ligand dataset studied (Fig. 5). Docking structures of these ligands and LmCS are shown in Supplementary Fig. 3 (For details see footnote). Two
schematic presentations of putative binding modes of the target protein structure with the compounds i.e. ZINC03803450, ZINC13387711, ZINC16052528 and the substrate (EPSP) are shown in Fig. 6.

The six compounds with strongest binding affinities identified from Experimental Drugbank were: ZINC03803450, ZINC00602427, ZINC20149031, ZINC13387711, ZINC16052528 and ZINC01587011 with the $K_i$ value of 2.1, 16.91, 20.31, 22.94, 35.84 and 38.09 nM, respectively. The binding energy showed that ZINC03803450 was the best ligand among them, considering its lowest energy score with $K_i$ value of 2.1 nM. The interacting residues of LmCS with all eight selected ligands from NCIDS III and Experimental Drugbank are listed at Table 1. It demonstrated that among all the binding residues His10, Arg45, Arg48 and Arg337 were almost common and conserved among almost all bacterial species. The conservation among different species indicated their importance at the active site of LmCS and interaction with the substrate molecule. The hydrogen bond donors and hydrogen bond acceptors during polar interactions between all eight ligands and LmCS are tabulated at Supplementary Tables 1 and 2.

Throughout the whole study, rigid body docking was followed, where both the receptor and ligand kept rigid. But, in natural cellular condition, macromolecules often undergo conformational change, or induced fit upon ligand binding in order to maximize energetically favorable interactions with the ligand or solvent,\(^5\) raising a question whether or not rigid-body docking is sufficiently good for most docking. On the other hand, flexible docking procedures which permit conformational change must intelligently select every possible subtle conformational change for consideration.\(^5\) Despite obvious limitations, rigid body docking methods are interesting because they are much faster than the flexible docking algorithms. The speed and relative

---

**Table 1**—Comparative docking results of eight best ligands with SpCS, McCS and LmCS

<table>
<thead>
<tr>
<th>Ligand name</th>
<th>Docking results with SpCS, McCS and LmCS</th>
<th>Interacting residues with LmCS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SpCS</td>
<td>MccS</td>
</tr>
<tr>
<td></td>
<td>Moldock rerank score</td>
<td>Moldock rerank score</td>
</tr>
<tr>
<td>ZINC03803450</td>
<td>-142.639</td>
<td>-96.272</td>
</tr>
<tr>
<td>ZINC01163259</td>
<td>-160.406</td>
<td>-134.477</td>
</tr>
<tr>
<td>ZINC00602427</td>
<td>-151.995</td>
<td>-121.154</td>
</tr>
<tr>
<td>ZINC20149031</td>
<td>-151.634</td>
<td>-62.894</td>
</tr>
<tr>
<td>ZINC16052528</td>
<td>-178.754</td>
<td>-123.766</td>
</tr>
<tr>
<td>ZINC01675857</td>
<td>-123.69</td>
<td>-100.293</td>
</tr>
<tr>
<td>ZINC01587011</td>
<td>-140.871</td>
<td>-113.327</td>
</tr>
</tbody>
</table>

---

\(^5\)For details see footnote.

---
Fig. 5—Compounds selected from different sources as potent CS inhibitors based on their binding affinities to LmCS by virtual screening and comparative molecular docking results with LmCS model, SpCS and MtCS.
accuracy of multiple conformation rigid body docking methods make them an attractive choice\textsuperscript{53}.

**Docking result of reported benzofuran-3[2H]-one analog and natural benzofurans**

Earlier, a series of 2(Z)-2-benzylidene-6,7-dihydroxybenzofuran-3[2H]-ones have been identified as potent inhibitors of bacterial CS of which 41 benzofuran-3[H]-one analogs reported as inhibitors of \textit{SpCS}\textsuperscript{5}. These reported 41 inhibitors of \textit{S. pneumoniae} were docked with the modeled structure of \textit{LmCS}. Four of these inhibitors with highest predicted binding efficiency and interacting residues are shown in Table 2. The experimental IC\textsubscript{50} value with \textit{SpCS} was obtained from the binding database (bindingdb.org). Besides, 100 natural benzofurans retrieved from ZINC database (http://zinc.docking.org/rings/54/substances/natural) were docked with \textit{LmCS} after virtual screening and Lipinski filtering process. Among them, best three natural benzofurans with lowest binding energy and \(K_i\) values much better than the reported compounds were selected (Table 2).
ligands was compared with (1QXO). The binding of EPSP and other selected best ligands was similar to its template structure obtained from castP analysis and the assumption that residues His11 and Arg139 with highest predicted binding efficiency into active site of LmCS, SpCS and MtCS were missing in reported CS model complex. Arg45 of LmCS formed two H-bonds to the substrate and cofactor binding interactions. The H-bonds of His10 and Arg48 with conserved residues His110, Asn251, His110, Asn251, Arg337, Ser9, Arg101, Arg107, Ser132, Asn251, Arg337, Ser9, Arg107, Arg337, Arg45, Ala342, Arg337, His10, Arg48, Ala342, Arg337, Ser9, Arg107, Arg337, Arg45, Asn251, His110, Ala333, Asp339 and natural benzofurans with LmCS, SpCS and MtCS.

To get an idea about EPSP binding interactions with LmCS, the substrate was docked into the active site of LmCS model based on the information obtained from castP analysis and the assumption that the binding is similar to its template structure (1QXO). The binding of EPSP and other selected best ligands was compared with SpCS and MtCS. The H-bonding of EPSP-LmCS model complex was very similar to other two CS homologs from S. pneumoniae and M. tuberculosis. Though they had a great extent of identity among them, they did have some distinctions in substrate and cofactor binding interactions. The H-bonds of His10 and Arg134 with both LmCS and SpCS were missing in reported EPSP-MtCS model complex with its equivalent residues His11 and Arg139. Furthermore, an additional amino acid residue Ala133 was found to have polar interaction with in LmCS model-EPSP complex which was not present in SpCS. Arg45 of LmCS formed a single H-bond with O9 of EPSP just like its equivalent Arg46 in MtCS model, whereas Arg45 of SpCS formed two H-bonds to the substrate EPSP. In case of co-factor binding, the H-bonds to FMN involving Asn251 in both LmCS and SpCS not reproduced in MtCS which was found replaced by another acidic amino acid Glu256.

The comparative study of best ligands from NCIDS III and Experimental Drugbank reported benzofuran-3[2H]-one analogs and natural benzofurans with highest predicted binding efficiency into active site of LmCS, SpCS and MtCS using Molegro virtual docker v4.0. The corresponding docking results of the best ligands with interacting residues of LmCS are shown in Tables 1 and 2. During docking study, the protein-ligand interacting site was restricted to the binding site of the EPSP. The predicted binding efficiency of considered ligands was estimated based on the values of inhibition constant $K_i$, docking score, docking energy and extent of interactions with active site residues. In docking study of eight best ligands from NCIDS III and Experimental Drugbank, ZINC03803450 and ZINC01163259 were found to be most potent molecules binding into the active site of LmCS. The other two ligands (ZINC13387711 and ZINC16052528) had the calculated lowest dockscore with SpCS, MtCS and LmCS (Table 2), indicating using these ligands having the predicted $K_i$ value of 22.94 nM and 35.84 nM with LmCS, respectively as the potent broad spectrum antimicrobials. Significantly, the drug likeness test also confirmed that these compounds followed “Lipinski rules of five” and possessed the criteria of standard drug like properties with good oral bioavailability.

A total of approximately 19% of the PDB ligands and 10% of the drugs, respectively violate more than one of the Lipinski rules. Compounds violating more than one of the Lipinski rules have assumed to have problems with bioavailability and are, therefore, presumably not suitable as drugs. In our study, CHEMBL335859 having experimental IC$_{50}$ value of 5500 nM with SpCS was found to be the most potent molecule binding into the active site residues of LmCS, among 41 reported benzofuran-3[2H]-one analogs. It had polar interactions with conserved substrate binding residues viz., His10, Arg48 and...
Arg337 in active site of LmCS. CHEMBL336282 having experimental IC$_{50}$ value of 1600 nM with SpCS was found to have strong binding into the active site of SpCS (moldock score -154.244), MtCS (moldock score -139.176) and LmCS (moldock score -147.116). The natural benzofuran compound ZINC13682976 from ZINC database gave the lowest binding energy of -10.61 Kcal/mol, as compared to other natural benzofurans and benzofuran-3[2H]-one analogs studied, which can also be considered as a potential inhibitor of LmCS.

**In Silico ADME/Tox prediction**

The bioavailability, bioactivity and toxicity of the candidate compounds were evaluated considering the scores of several component features, such as human oral bio-availability (F%), solubility (LogSw), first-pass metabolism (in liver and intestine), human intestinal absorption (maximum passive absorption in %), blood brain barrier permeability, P-glycoprotein substrate/inhibitor specificity, AMES test score (+/-), genotoxicity hazards (present/absent), endocrine system disruption (estrogen receptor specificity), probable hERG inhibitor (+/-), acute toxicity measure (LD$_{50}$) tested on mouse after oral administration etc. The results of ADME/Tox measured by ACD/ADME suite and ACD/Tox suite platform are given in Supplementary Table 3 ($^9$For details see footnote). All the ligands studied, including reported benzofuran-3[2H]-one analogs$^9$ showed different ADME/Tox measurements, some of them had poor oral bioavailability, some had more than 50% loss of drug candidates after absorption into the body, while other compounds were found carcinogenic and toxic to health because of being positive in Ames test and having genotoxicity hazards. Our ADME/Tox study showed that all the compounds possessed moderate (30-70%) human oral bioavailability, whereas three of four benzofuran-3[2H]-one analogs studied had shown poor (>30%) oral bioavailability, which might be due to poor passive absorption and first pass metabolism loss of the compound in liver and intestine (Supplementary Table 3. $^5$For details see footnote). Several selected ligands with good binding affinity score failed to achieve ADME/Tox pass because of having the probability of being mutagenic and carcinogenic to health.

AMES test is one of the most popular tests for assessing mutagenic properties of the compounds which is used worldwide as an initial screen to determine mutagenic properties of NCEs (new chemical entities) in drug and chemical industry.$^{34}$ Seven selected compounds i.e., ZINC00602427, ZINC01163259, ZINC01587011, ZINC01675857, ZINC09708077, ZINC13681793 and ZINC13682976 were found positive in Ames test and they belonged to genotoxicity hazard groups, such as aryltriazene, quinone type conformations and aromatic nitro groups. Overall, ZINC03803450, ZINC13387711, ZINC16052528 and ZINC20149031 had shown good score on almost all the parameters; ADME/Tox properties of the compounds and their good binding affinity score with CS indicated that they could be used as potential CS inhibitors. In toxicity prediction, lethal health effects of the compounds were measured by an *in silico* approach based on statistical model generated from previous experimental data on model organisms, which show that it is relatively easy to optimize the lethal doses of any molecule without killing any animal, which reduces both the cost and time.$^{38}$ Although the need of an experimental validation in drug development process can’t be replaced by an *in silico* approach, it assists in narrowing down the choices and select the right one as a drug after optimization.

**Conclusion**

In this study, we aimed to identify potential drug compounds against LmCS, the last enzyme of the shikimate pathway constitutes an excellent antimicrobial target. We generated a 3D model structure of LmCS (monomer, dimer and tetramer), which permitted virtual screening and the study of binding efficiency of selected screened ligands for the identification of putative inhibitors. All those selected ligands were also docked with the existing crystal structure of CS from *S. Pneumoniae* and *M. tuberculosis* to find their binding efficacy among other pathogenic species. This study can thus help to design and develop drug against listeriosis and the next generation of novel antibacterial drugs against pathogenic bacteria’s like *S. Pneumoniae* and *M. tuberculosis*.

**Acknowledgement**

The authors thank Dr. Abul Kalam Azad, Associate Professor, Dept. of Genetic Engineering and Biotechnology, Shahjalal University of Science and Technology, Sylhet-3114, Bangladesh for his valuable suggestions on the manuscript. We greatly acknowledge the co-operation of the Department of Genetic Engineering and Biotechnology, Shahjalal
University of Science and Technology, Sylhet-3114, Bangladesh during the research work.

References

10. The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC