Chemogenomics Profiling of Drug Targets of Peptidoglycan Biosynthesis Pathway in *Leptospira interrogans* by Virtual Screening Approaches

Bhattacharjee, Biplab*, Rose Mary Simon2, Chaithra Gangadharaiah2, and Prashantha Karunakar1

1Department of Biotechnology, PES Institute of Technology, Bangalore, India
2Crucible of Research and Innovation (CORI Lab), PES Institute of Technology, Bangalore, India

Received: June 21, 2012 / Revised: February 1, 2013 / Accepted: February 13, 2013

Leptospirosis is a worldwide zoonosis of global concern caused by *Leptospira interrogans*. The availability of ligand libraries has facilitated the search for novel drug targets using chemogenomics approaches, compared with the traditional method of drug discovery, which is time consuming and yields few leads with little intracellular information for guiding target selection. Recent subtractive genomics studies have revealed the putative drug targets in peptidoglycan biosynthesis pathways in *Leptospira interrogans*. A ligand library for the murD ligase enzyme in the peptidoglycan pathway has also been identified. Our approach in this research involves screening of the pre-existing ligand library of murD with related protein family members in the putative drug target assembly in the peptidoglycan biosynthesis pathway. A chemogenomics approach has been implemented here, which involves screening of known ligands of a protein family having analogous domain architecture for identification of leads for existing druggable protein family members. By means of this approach, one murC and one murF inhibitor were identified, providing a platform for developing an anti-leptospirosis drug targeting the peptidoglycan biosynthesis pathway. Given that the peptidoglycan biosynthesis pathway is exclusive to bacteria, the *in silico* identified mur ligase inhibitors are expected to be broad-spectrum Gram-negative inhibitors if synthesized and tested in *in vitro* and *in vivo* assays.

Key words: Chemogenomics, antimicrobial, *Leptospira interrogans*, peptidoglycan biosynthesis

*Corresponding author.
Phone: +91-99-45532797; Fax: +91-80-26720886; E-mail: biplabbhattacharjee2010@gmail.com

Supplementary data for this paper are available on-line only at http://jmb.or.kr.
targets by relating the structures and activities of their ligands. It is related to the concepts of chemical genetics and chemical genomics in that all approaches investigate the perturbation of biological systems with small molecules. The hierarchical nature of the target class similarity has a profound influence on the way novel ligands are discovered. Although high-throughput screening has long been the method of choice for lead discovery in the pharmaceutical industry, knowledge about ligands of one target and the distance between targets in biological space facilitates educated guesses as to which molecules are suitable for novel targets. Thus, we can relate targets by the similarity of ligands to which they bind, a “central paradigm of chemogenomics.” By employing the “molecular similarity principle,” the target of similar compounds may also be the target of the query structure. Likewise, it is (within limitations) well established that similar targets bind to similar compounds. Thus, conventional similarity searching is inverted; whereas usually new ligands for a known target are desired, in this case, new targets for a known ligand are proposed. The rationale behind this approach was that similar ligands are more likely to bind not only to the same target, but also to the same protein folds or amino acid sequence when they occur in other proteins. For example, if ligands of a protein P1 with fold A are known, and this fold is shared by both protein P1 and protein P2, a target prediction of protein fold A for a test compound would encompass both protein P1 and protein P2, even though no ligands for protein P2 were present in our database [19].

In this study, the ligand library of murD was mapped to other members of the mur ligase family (murC, murE, murF, murA, and murG), so that new targets for this known ligand library can be proposed.

**Materials and Methods**

The process flow of the methodology is displayed in Fig. 1.

**Literature Retrieval and Annotation**

Putative drug targets in peptidoglycan biosynthesis in *L. interrogans* were retrieved from a literature survey [5]. The list of identified ligands for murD was also taken from pre-published work. The mur ligases of *L. interrogans* such as murC, murE, murF, murA, and murG were selected as drug targets for screening of the ligand library (murD) [2].

**Structure Generation for Putative Drug Targets**

The PDB structures of drug targets were not available for murC, murE, murF, murA, and murG in structural databases. The protein

---

**Fig. 1.** Flowchart of the overall methodology followed.
sequence of mur ligase family members (murC, murE, murF, murA, and murG) were retrieved from Uniprot. A BLASTp analysis was performed to find the template having high sequence identity with RCSB-PDB structures. The template PDB structures displayed very low sequence identity. Homology modeling was conducted with the SWISS Model webserver in Automatic Modelling Mode. A template search was performed through the BLAST and PSI-BLAST programs. The global alignment method was used for comparison between the target-template sequences. The target and template sequences are fed into the SWISS Model Automatic Modelling Mode Server. The pdb file generated from the SWISS Model Automatic Modelling Mode Server was validated for parameters like covalent bond distances and angles, stereochemical validation, and atom nomenclature using PROCHECK, and the overall quality factor of non-bonded interactions between different atoms types were measured by the ERRAT program and was found to be not a fit into a good theoretical model because of the lower sequence identity with the template. In order to derive a better model, Threading or Fold-based modeling was brought into practice. Threading or structural fold recognition predicts the structural fold of an unknown protein sequence by fitting the sequence into a structural database and selecting the best-fitting fold. The comparison emphasizes matching of secondary structures, which are most evolutionarily conserved. Therefore, this approach can identify structurally similar proteins even without detectable sequence similarity. The Phyre server utilizes a library of identified protein structures from the Structural Classification of Proteins (SCOP) database and updated depositions in the Protein Data Bank (PDB). A user-submitted query is scanned against a nonredundant sequence database and a profile is constructed, which is further deposited in a “fold library.” The known and predicted secondary structure of these proteins is also stored in the fold library. Five iterations of PSI-BLAST are used to gather both close and remote sequence homologs. Subsequent to profile construction, prediction of query secondary structure is done. Profile and secondary structure obtained in the preceding steps were scanned against the fold library using a profile–profile alignment algorithm, which ranks the alignment based on a calculated score. Further fitting of the calculated score to an extreme value distribution generates an E-value. The top 10 highest scoring alignments are subsequently employed for building up of full 3D models of the query. A loop library and reconstruction procedure is employed to repair missing or inserted regions caused by indels in the alignment in cases where it is feasible. Lastly, side-chains are positioned on the 3D model by means of a “fast graph-based algorithm” and “sidechain rotamer library” [10].

Structure Validation

The 3D structures obtained from Phyre2 were assessed using PROCHECK, ProSa, WHAT_CHECK, and ERRAT [7, 9, 11, 12, 18].

Domain Identification and Comparative Studies

The chemogenomics approach considers that members of a same protein family having a shared domain architecture should also have the same ligands binding to them. The murD ligand library was retrieved from past research, so the domain assembly of murD was compared with the domain architecture of murC, murE, murF, murA, and murG. The domains for the drug targets were predicted using PROSITE and the secondary structure was predicted using JPred. Comparative analysis was done among the mur ligases with respect to domain and secondary structure.

Active Site Prediction and Comparative Studies

The active site residues for the drug targets were identified using the Q-site finder. Pairwise alignment between active site residues of murD and the other mur family members (murC, murE, murF, murA, and murG) was accomplished using EMBOSS Pairwise Sequence Alignment to find the regions of conservation.

Structure Comparative Studies

The tertiary structures of murC, murE, murF, murA, and murG were compared individually with the murD structure in the DALI server, and interactive visualization of results was done to get a clear picture of the structural similarity. The drug targets that showed higher structural similarity and Z-score were taken for further analysis (Supplementary Table S3).

Mapping of Ligand Library to Mur Protein Family

The ligand library of murD in L. interrogans was retrieved from the literature survey. The mur ligases inhibitor chemical compound structures were drawn, and least energy conformers were generated using MarvinSketch. The lead compounds thus served as the ligand library for screening against murC, murE, and murF.

Docking Studies with Ligand Library Set

Protein ligand docking was performed using PyRx. The active site of an enzyme contains the catalytic and binding sites. The structure and chemical properties of the active site allow for the recognition and binding of the substrate. The binding energy scores from docking of the ligand library (of murD) with murC, murE, and murF were compared with their respective substrate molecules, and screening was done further to narrow our search for potent inhibitors. The IC50 of the ligand library against selected mur ligases were also established using AutoDock tools [13].

RESULTS AND DISCUSSION

Sixteen unique pathogen-specific pathways (Supplementary Table S2) were identified from the subtractive genomics approach in L. interrogans, and putative drug target assemblies of those pathways were listed [11]. In this research, we have taken peptidoglycan biosynthesis, because the murD ligand library was well characterized in prepublished work [5].

The BLASTp analysis of the drug targets with RCSB-PDB structures gave <35% identity (Supplementary Table S4). The structure generated from the SWISS MODEL server gave poor validation results because of low template-target identity. For structures with a low identity of template-target alignment, fold-recognition-based structure prediction is preferred. Structure prediction using the Phyre2 (Protein Homology/analogY Recognition Engine ver. 2.0) server was employed. Supplementary Figs. S2a and S2b display the Ramachandran plot analysis (by PROCHECK) of the model structure of murC and murF from Phyre2 server, showing that >90% of the residues is in favored and additional allowed regions.

Active site prediction using Q-Site Finder (http://www.modelling.leeds.ac.uk/qsitefinder) predicts a pocket
of murA (site volume 631 cubic Å³), murC (site volume 560 cubic Å³), murE (site volume of 339 cubic Å³), murF (site volume of 705 cubic Å³), and murG (site volume of 447 cubic Å³), which can be used as potential inhibitory sites.

Domain identification and comparative sequence analysis revealed stretches of high conservation and the presence of related domain architecture. The secondary structure prediction and analysis also revealed analogous secondary structural patterns in the region of active site of mur ligase enzymes. Literature review throws light on the topological and structural similarities of all four mur ligases, although they display low sequence identity [20]. Each of the mur ligases are constituted of three structural domains: (i) N-terminal Rossmann-fold domain, accountable for binding the UDPMurNAc substrate; (ii) Central domain, analogous to ATP-binding domains of several ATPases and GTPases; (iii) C-terminal domain (analogous to dihydrofolate reductase fold), probably connected to binding of an incoming amino acid. The conserved sequence motifs that are found in the four mur enzymes also resemble other members of the mur ligase family, counting folylpolyglutamate synthetase, cyanophycin synthetase, and the capB enzyme from Bacillales.

By using results of structure superimposition studies in the DALI Server, murC, murE, and murF were found to have higher structural similarity with murD based on RMSD and Z-score (Z-Score>5) (Supplementary Fig. S3 and Table S3).

Molecular docking studies of murC, murE, and murF against the murD ligand library was conducted using PyRx, so as to establish new targets for the known ligand library as per chemogenomics principles.

Based on the binding energy and IC₅₀ values, Ligand 1 (Table 1 and Fig. 2) was selected as the best performing lead for murC, and Ligand 2 was selected as the best performing lead for murF, and no ligand could be identified for murE, as all ligands in the murD ligand library gave more binding energy than the substrate UDP-N-acetylMuramoyl-1-alanyl-0-glutamate. The IC₅₀ values listed in Table 2 are predicted values from AutoDock PyRx. This being a putative ligand, experimental IC₅₀ values from bioassays are still not determined. Further in vitro experiments need to be done to ascertain the predicted inhibition potency of this ligand against murC and murF.

![Functional groups of Ligand 1, a lead molecule obtained after virtual screening of murC and murF.](image-url)
Fig. 3C depicts the best lead for murC having the highest binding affinity (-6.94 kcal/mol) and Fig. 3B depicts the best lead for murF having binding affinity (-7.18 kcal/mol).

In conclusion, drug design and discovery in the post-genomic era is breaking old standards and consistently restructuring the drug discovery protocol by incorporating the eons of information encoded in our genome and chemical space. The chemogenomics approach identifies new drug targets for pre-existing ligand libraries, thus presenting a prospect to deal with a manageable number of efficient data through supplementary pathway studies and experimental design. The mur ligases participating in the peptidoglycan biosynthesis pathway do not have any alternative mechanism to replace its catalytic activity. Our approach employing

<table>
<thead>
<tr>
<th>Property</th>
<th>murC</th>
<th>murE</th>
<th>murF</th>
<th>murD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Binding energy</td>
<td>-6.94</td>
<td>-7.14</td>
<td>-7.18</td>
<td>-7.17</td>
</tr>
<tr>
<td>Number of H bonds formed</td>
<td>2</td>
<td>5</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Amino acids in active site</td>
<td>ARG351, GLY119</td>
<td>LYS128, ARG352, GLY127, LYS128, THR125</td>
<td>GLU152</td>
<td>ASN136, HIS181, LYS318, LYS113</td>
</tr>
<tr>
<td>Predicted IC₅₀ (µM) from AutoDock</td>
<td>8.2</td>
<td>5.82</td>
<td>5.44</td>
<td>5.59</td>
</tr>
</tbody>
</table>

Table 2. Comparison of the binding energy of the lead candidate ligand with murC, murE, murF, and murD.

Fig. 3. Molecular docking studies of mur ligases with lead ligand molecules.
(A) Molecular docking interaction of murF with Ligand 2 (Sl. No. 2 in Table 1) having highest binding affinity (-7.18 kcal/mol). One H bond formation in the active site of murF with Ligand 2. (B) Molecular docking interaction of murD with Ligand 2 (Sl. No. 2 in Table 1) having binding affinity (-7.17 kcal/mol). Four H bond formation in the active site of murD with Ligand 2. (C) Molecular docking interaction of murC with Ligand 1 (Sl. No. 1 in Table 1) having binding affinity (-6.94 kcal/mol). Two H bond formation in the active site of murD with Ligand 2.
the chemogenomics methodology to map the murD ligand library has brought to light novel inhibitors for other mur ligase family members (murC and murF). Since the peptidoglycan biosynthesis pathway is unique to bacteria, the in silico identified mur ligase inhibitors can be further analyzed for their broad-spectrum Gram-negative inhibitory action in in vitro and in vivo studies.

Acknowledgments

This research has been carried out in the PES Institute of Technology, Bangalore, India. We would like to thank Dr. V. K. Agrawal, director of Crucible of Research and Innovation for moral and financial support all through the work.

REFERENCES