

Research Journal of Pharmaceutical, Biological and Chemical Sciences

Identification of Potential Inhibitors of CysteinyI tRNA Synthetase from *Mycobacterium leprae* TN Strain through *in silico* Virtual Screening.

S Mohana Parvatha Varddhini, R Sankaranarayanan, M Udayakumar and S Thamotharan*.

Department of Bioinformatics, School of Chemical and Biotechnology, SASTRA University, Thanjavur – 613 401, Tamil Nadu, India.

ABSTRACT

Aminoacyl tRNA synthetases are believed to be a novel antibacterial, antifungal and antiparasitics targets due to their involvement in the process of protein translation. In this work, we have performed computer aided virtual screening of 14,400 compounds taken from Maybridge Hitfinder database and identified fourteen potential lead molecules for CysteinyI tRNA synthetase of *Mycobacterium leprae* strain TN (*MICysRS*). We have also performed docking of first-line antileprosy drugs such as dapsons, rifampicin and clofazimine with *MICysRS*. The computational virtual screening and docking results show that these fourteen compounds have high binding affinity for *MICysRS*. The top ranked compound, HTS11201, 7-[(3,4-dichlorophenyl)methyl]-1,3-dimethyl-8-[2-(2-oxoindol-3-yl)hydrazinyl]purine-2,6-dione, has the binding score of -11.3 kcal/mol. Interestingly, eight strictly conserved residues from the catalytic domain of *MICysRS* are involving in the formation of hydrogen bonding interactions with the lead molecules.

Keyword: cysteinyI tRNA synthetase; *Mycobacterium leprae*; virtual screening; antibacterial drug target; molecular docking.

*Corresponding author

INTRODUCTION

Aminoacyl tRNA synthetases (aaRSs) play a crucial role in the process of protein translation by charging the tRNAs with the appropriate amino acids [1]. These enzymes are found in all living organisms and catalyze the aminoacylation of tRNAs in two-step reaction mechanism [2]. In the first step, an amino acid is condensed with an ATP molecule to form an enzyme-bound complex of aminoacyl adenylate intermediate. In the second step, the amino acid is transferred onto a cognate tRNA to produce the desired product [3]. Aminoacyl tRNA synthetases can be categorized into two distinct classes (I and II) based on their structural features [4-7]. Each class consists of ten aminoacyl tRNA synthetases. In general, class I aminoacyl tRNA synthetases have a Rossmann dinucleotide-binding fold in their catalytic domain with characteristic HIGH and KMSKS motifs [8, 9]. The class II aminoacyl tRNA synthetase enzymes contain anti-parallel six-stranded β -sheets packed between α -helices. The class II enzymes have completely different active site constitution when compared to class I enzymes.

The aminoacyl tRNA synthetases are known potential drug targets for the development of antibacterial drugs because of their involvement in protein synthesis and cell viability. The drug discovery against bacterial and fungal aminoacyl tRNA synthetases have been carried out by various groups [10; references therein]. In this work, our focus lies in identifying potential lead molecules for *Mycobacterium leprae* through computational virtual screening approach. Leprosy is a curable infectious disease for which the multidrug therapy (MDT) has been widely used world over. This disease is caused by *Mycobacterium leprae* and it has been eliminated successfully from a large part of the world. However, leprosy remains a serious health problem in major endemic countries like India and Brazil [11]. The MDT consists of first-line drugs namely dapsone, clofazimine and rifampicin which are designed to prevent the emergence and spread of the drug-resistance strains. Alarmingly, there are lots of evidences have emerged recently, that clearly show this bacterium have developed resistance to these first-line drugs. Hence, there is an urgent need for the development of new antileprosy compounds to overcome the problems of drug resistances, relapses and disabilities [11].

The present study focuses on identifying potential lead molecules against cysteinyl tRNA synthetase (CysRS) of *Mycobacterium leprae* strain TN (MICysRs). The CysRS which belongs to class Ia aminoacyl tRNA synthetase which acylates cysteinyl tRNA with cysteine. A search in protein data bank reveals the X-ray crystal structures of CysRS from various organisms such as *Escherichia coli* [12], *Escherichia coli K-12* [13], *Coxiella burnetii* [PDB ID: 3TQO] and *Borrelia burgdorferi* [PDB ID: 3SP1]. All these structures contain zinc ion which is situated at the base of the active site that is approximately at a distance of 8 Å from the phosphate oxygens of adenosine monophosphate. The zinc ion plays an essential role for amino acid discrimination without editing mechanism. The small molecule compounds for virtual screening have been retrieved from Maybridge Hitfinder (www.maybridge.com). There are roughly 14,400 compounds present in this database. The molecular docking of first-line drugs with CysRS of *Mycobacterium leprae* has also been carried out to identify the potential lead molecules.

MATERIALS AND METHODS

Homology modeling

The protein sequence of Cysteinyl tRNA synthetase of *Mycobacterium leprae* TN strain (MICysRS) (Uniprot ID: P57990) was retrieved from the UniProtKB database (<http://www.uniprot.org>) and it consisted of 473 amino acids. Since the three dimensional structure of cysteinyl tRNA synthetase of *Mycobacterium leprae* TN strain was not available, the homology modeling was carried out using Modeller v9.1 [14]. To start with, the homologous structure for MICysRS was identified using BLAST search protocol [15] against protein data bank. This protocol revealed *E.coli* CysRS (PDB ID: 1U0B, chain B) as the template. With this template, the three dimensional structure of MICysRS was built. The sequence alignment between *Ml* and *E.coli* CysRS enzymes showed that they shared 41% sequence identity. The sequence alignment was performed using CLUSTALW [16]. The final model of MICysRS that had 473 amino acid residues was subjected to structure validation. The program RAMPAGE server [17] was used to assess the stereochemical quality of the MICysRS modelled structure. From Ramachandran map, the number of residues in the favoured and allowed regions were found to be 96 and 3.8%, respectively. This validated model was further used for virtual screening.

Active site prediction

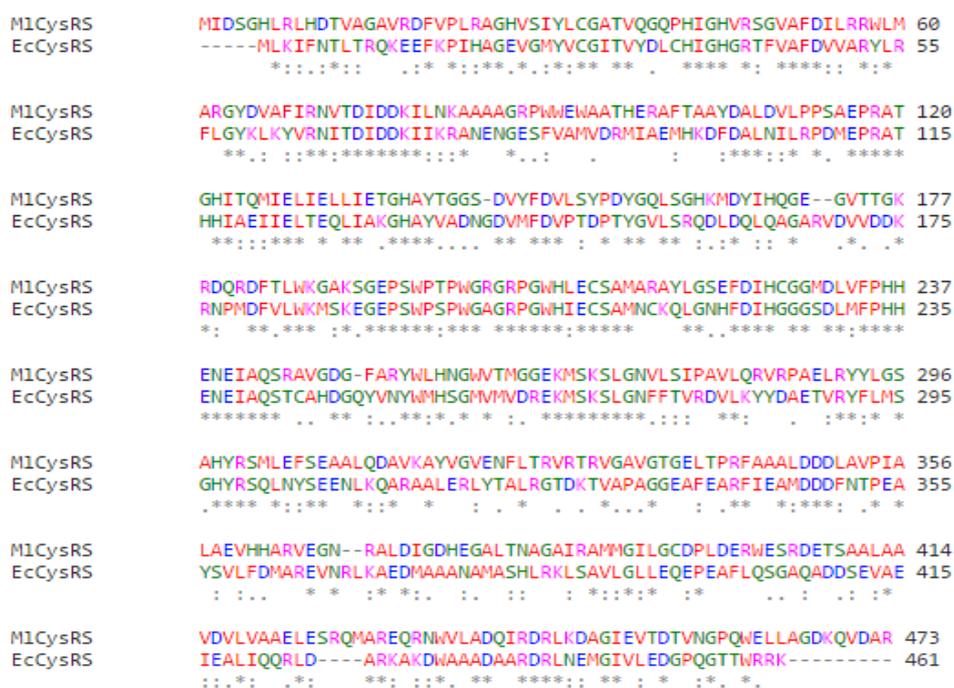
The active site of the *MCysRS* structure was predicted based on the crystal structure complex of CysRS-adenosine monophosphate (AMP) from *Borrelia burgdorferi* (PDB ID: 3SP1). To identify the location and size of the grid box on *MCysRS*, structure of *MCysRS* was superimposed onto the crystal structure of CysRS-adenosine monophosphate. Using PyRx-0.8 [18], a grid box was centered with coordinates $x = 74.6123$, $y = 18.9526$ and $z = 1.1800$ Å on *MCysRS* receptor molecule and this grid was further used for docking study. Moreover, CONSURF server [19, 20] was used for the identification of functional regions as well as active site residues in the protein. The modelled structure was used as an input to the CONSURF server. The context-specific iterated-basic local alignment search tool (CSI-BLAST) was used to identify the homologous sequences with the maximum percentage identity between sequences being 95 and the minimum percentage identity between sequences being 35. In CSI-BLAST the default E-value cut off was taken as 0.0001 and 3 iterations of CSI-BLAST was carried out.

Database and Virtual screening

In this study, Maybridge HitFinder (www.maybridge.com) chemical libraries was used for virtual screening. This database consisted of 14,400 chemically diverse collection of compounds which also showed drug-likeness. All the ligand molecules (14,400) were retrieved as sdf format and used as inputs for ligand preparation in OPEN BABEL implemented in PyRx. The OPEN BABEL converted these ligands into pdbqt formats for virtual screening. During ligand preparation, only polar hydrogens were added and non-polar hydrogens were merged in each ligand structures. The receptor molecule was also converted as a pdbqt format. The first-line drugs of *Mycobacterium leprae* used as control (dapsons, clofazimine and Rifampicin) were also prepared in pdbqt format. For all these compounds, AutoDock Vina program [18] implemented in PyRx was used to calculate binding energies.

RESULTS AND DISCUSSION

Figure 1: Pairwise sequence alignment of cysteinyl tRNA synthetases (CysRS) of *M.leprae* and *E.coli*. This diagram was generated using CLUSTALW.



The homology model of *MCysRS* is built using the crystal structure of *E.coli* CysRS model. The pairwise sequence alignment between *MCysRS* and *E.coli* CysRS is shown in Figure 1 and the protein sequence identity between them is 41%. The cartoon representation of the final model of *MCysRS* is shown in Figure 2A. The

monomer of *MICysRS* consists of 473 amino acid residues. As in *E.coli* CysRS [13], the modelled structure of *MICysRS* also has four domains, the domain having Rossmann fold comprising of residues from 27 to 136 and 210 to 260. These two polypeptides having Rossmann fold are connected by a connective polypeptide (CP) domain that consists of residues from 137 to 209. The third domain namely the stem contact (SC) domain contains residues from 261 to 325, while the residues from 336 to 393 form the anticodon domain. A search in protein data bank for CysRS reveals, in the structures of all the unliganded form of this protein, the C-terminal end of the polypeptide is highly disordered. Hence the coordinates of atoms corresponding to this region are not present in PDB. However, this region is clearly defined in the ligand bound form of this protein [13]. Since, the liganded form of protein from *E.coli* is taken as template, this disordered region is clearly modelled in *MICysRS*.

The active site residues are highly conserved as predicted using CONSURF server. This server has picked 465 unique CysRS protein sequences from various organisms. The residues located within 4.0 Å from the compounds are Leu 32, Cys 33, Gly 34, Ala 35, Thr 36, Gln 38, His 42, Gly 44, His 45, Ser 48, Gly 49, Phe 52, Trp 60, Arg 70, Thr 73, Asp 76, Lys 78, Phe 101, Tyr 105, Trp 207, His 208, Glu 210, Gly 228, Gly 229, Met 230, Asp 231, Leu 232, His 236, Asn 258, Gly 259, Trp 260, Val 261, Met 268, Ser 269 and Lys 270.

Figure 2A: The cartoon representation of the monomer of cysteinyl tRNA synthetase (CysRS) of *Mycobacterium leprae* TN strain. The various domains present in the monomer are shown in different colours (Rossmann fold-red, CP domain-blue, SC fold-cyan, helical bundle domain-yellow and α/β C-terminal domain-orange). The location of the active site is indicated in circle. **2B:** Ramachandran map for modelled structure of CysRS of *Mycobacterium leprae* TN strain showing the residues in favoured, allowed and outlier regions.

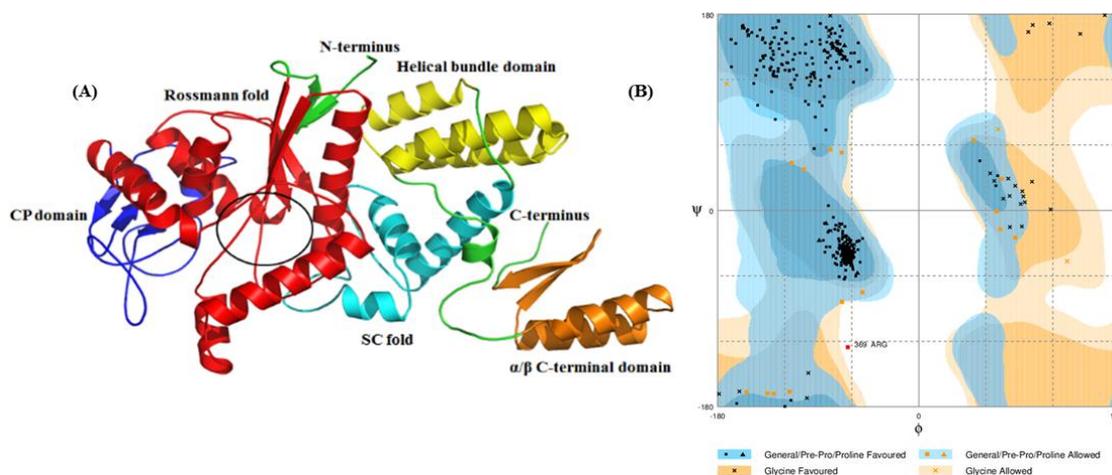
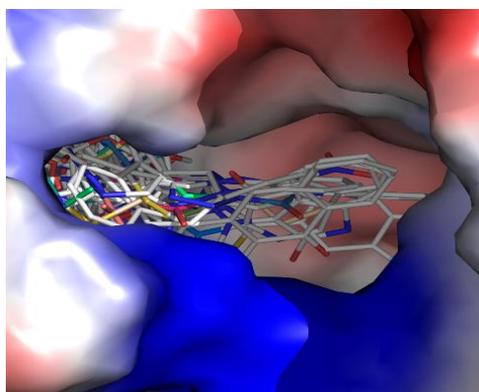


Figure 3: A close-up view of active site of *MICysRS*. The best docking poses of fourteen compounds are shown in stick model and the protein molecule is rendered in electrostatic charge representation.



Initially, docking of three first-line drug molecules (dapson, rifampicin and clofazimine) are carried out with *MICysRS* receptor molecule. The binding energies of the drug molecules are in the range between – 8.9 and –6.9 kcal/mol. Interestingly, dapson shows highest binding energy (–6.9 kcal/mol) with the protein

and clofazimine has the least value (−8.9 kcal/mol). Then, the computer aided virtual screening is carried out for 14,400 compounds. The virtual screening result suggests that the fourteen compounds are having good binding energies that are better than the first-line anti-leprosy drugs. The binding energies of these compounds including first-line drugs are listed in Table 1. The predicted docking poses of compounds identified from this study are shown in Figure 3.

Table 1: The binding energies of the compounds identified from this study along with the first-line drugs of anti-leprosy.

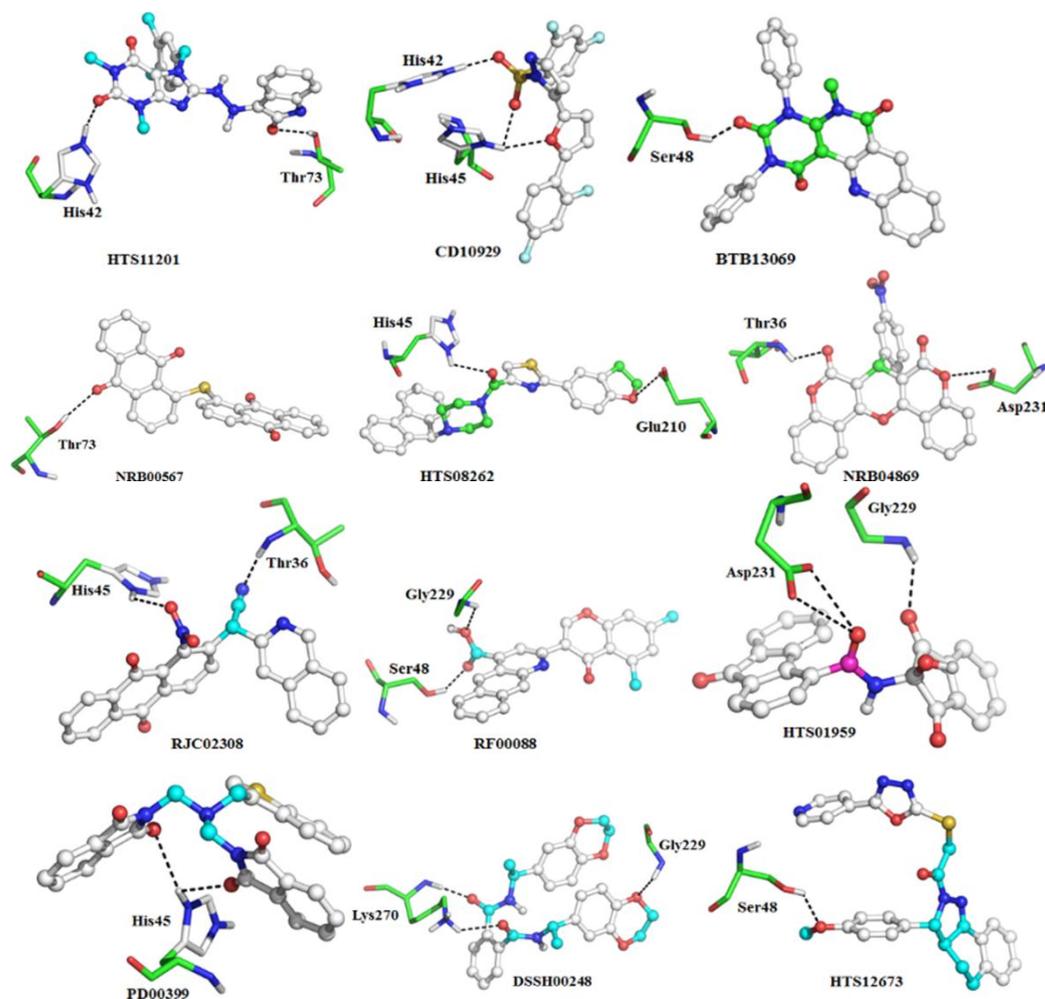
S.No.	Compound ID	IUPAC name	AutoDock Vina binding energy (in kcal/mol)
1	HTS11201	7-[(3,4-dichlorophenyl)methyl]-1,3-dimethyl-8-[2-(2-oxoindol-3-yl)hydrazinyl]purine-2,6-dione	−11.3
2	CD10929	5-[5-(2,4-difluorophenyl)furan-2-yl]-1-(2,4-difluorophenyl)sulfonylpyrazole	−11.2
3	HTS11989	1-[2-[3-(4-nitrophenyl)-1-phenyl-1H-pyrazol-4-yl]-5-(4-pyridinyl)-1,3,4-oxadiazole-3(2H)-yl]-1-ethanone	−11.2
4	BTB13069	5-methyl-2,4-diphenyl-1,2,3,4,5,6-hexahydrobenzo[b]pyrimido[4,5-h][1,6]naphthyridine-1,3,6-trione	−11.1
5	NRB00567	1-(9,10-dioxoanthracen-1-yl)sulfanylanthracene-9,10-dione	−11.1
6	CD12049	N2-phenyl-5-(1,3-dibenzylhexahydropyrimidin-5-yl)-2H-1,2,3,4-tetraazole-2-carboxamide	−11.0
7	HTS08262	[2-(2,3-dihydro-1-benzofuran-5-yl)-1,3-thiazol-4-yl] [4-(9H-fluoren-9-yl)piperazino] methanone	−10.8
8	NRB04869	7-(3-nitrophenyl)-6H, 7H,8H-chromeno[3',4':5,6]pyrano[3,2-c]chromene-6,8-dione	−10.8
9	RJC02308	2-isoquinolin-3-yl-2-(1-nitro-9,10-dioxoanthracen-2-yl)acetonitrile	−10.8
10	RF00088	2-(5,7-dimethyl-4-oxochromen-3-yl)benzo[g]quinolone-4-carboxylic acid	−10.8
11	HTS01959	N-(2-hydroxy-1,3-dioxo-2,3-dihydro-1H-inden-2-yl)-9-oxo-9H-fluorene-4-carboxamide	−10.6
12	PD00399	2-[[1-benzothiophen-3-ylmethyl-[(1,3-dioxisoindol-2yl)methyl]amino]methyl]isoindole-1,3-dione	−10.3
13	DSSH00248	N1,N2-di[1-(2,3-dihydro-1,4-benzodioxin-6-yl)ethyl]phthalamide	−10.2
14	HTS12673	1-[3-(4-methoxyphenyl)-3,3a,4,5-tetrahydro-2H-benzo[g]indazol-2-yl]-2-[[5-(4-pyridinyl)-1,3,4-oxadiazol-2-yl]sulfanyl]-1-ethanone	−10.2
First-line drugs (control)			
15	clofazimine	N-5-bis(4-chlorophenyl)-3-propan-2-yliminophenazin-2-amine	−8.9
16	Rifampicin	(7S,9E,11S,12R,13S,14R,15R,16R,17S,18S,19E,21Z)-2,15,17,27,29-pentahydroxy-11-methoxy-3,7,12,14,16,18,22-heptamethyl-26-[(1E)-[(4-methylpiperazin-1-yl)imino]methyl]-6,23-dioxo-8,30-dioxo-24-azatetracyclo[23.3.1.1 ⁴ ,7.0 ^{5,28}]triaconta-1(29),2,4,9,19,21,25,27-octaen-13-yl acetate	−7.0
17	Dapsone	Diaphenylsulfone	−6.9

In the modelled structure of *M/CysRS*, eight conserved residues, Thr36, His42, His45, Ser48, Thr73, Glu210, Gly229 and Asp231 of the catalytic domain are all taking part in hydrogen bonding interactions with the compounds of top 14 hits identified from this study. In addition, the conserved amino acid residue Lys270 from the SC fold is also making hydrogen bonding interactions with one of the compounds DSSH00248. The representative hydrogen bonding interactions are shown in Figure 4. The top ranking compound HTS11201 is making two hydrogen bonding interactions. The conserved residues His42 and Thr73 present at the N-terminus of Rossmann fold are also involving in the hydrogen bonding interactions with this compound. Interestingly, the compound, CD10929, forms bifurcated hydrogen bonding interactions in such a way that the first hydrogen bonding is formed between the side chain nitrogen (N^{δ1}) atom of His45 residue and one of the oxygen atoms of the O=S=O group of the compound and the second hydrogen bonding interaction is formed with the oxygen atom of the furan group. The N^{ε2} of His42 is making a hydrogen bonding interaction with one of the oxygens of O=S=O group.

For BTB13069, one of the oxygen atoms is acting as an acceptor for hydrogen bonding interaction with the side chain hydroxyl group of Ser48. The compound NRB00567 is also taking part in the formation of

hydrogen bonding with the side chain hydroxyl group of Thr73 residue. Moreover, the compound HTS08262 is forming a hydrogen bonding interaction with the side chain nitrogen ($N^{\delta 1}$) of His45 residue. The side chain carboxylate group of Glu210 residue is participating in hydrogen bonding interaction with the oxygen atom of benzofuran moiety. The interaction for the compound NRB04869 is mainly due to Thr36 residue. In addition to this, oxygen atom of one of the chromens moiety of NRB04869 acts as an acceptor for an intermolecular hydrogen bonding with the side chain carboxylate group of Asp231 residue. The residues Thr36 and His45 from the first halves of the Rossmann fold are participating in the hydrogen bonding interaction with the compound RJC02308.

Figure 4: The hydrogen bonding interactions between compounds and *M/CysRS*. The small molecule compounds are shown in ball and stick and the amino acid residues that make hydrogen bonds are displayed in stick representation.



In HTS01959, the residues Asp231 and Gly229 are involving in the hydrogen bonding interactions. The former interaction is through the side chain carboxylate group, while the latter is due to backbone nitrogen. Moreover, the conserved residues of Ser48 and Gly229 are forming hydrogen bonding interactions with the carboxylic group of RF00088. In PD00399, the side chain nitrogen ($N^{\delta 1}$) of His45 residue acts as a donor for two different hydrogen bonding interactions involving oxygen atoms of the ligand, thereby forming bifurcated hydrogen bonding interactions. Interestingly, the compound DSSH00248 is making three hydrogen bonding interactions with the residues Gly229 and Lys270. The latter residue is located in SC fold and making two interactions through nitrogen atoms of backbone and side chain with the compound. The highly conserved residue Ser48 is participating in the formation of hydrogen bonding interaction with the hydroxyl oxygen atom of HTS12673. The interactions for the other compounds are mainly stabilized by hydrophobic and van Waals contacts.

It is worth mentioning here that the compound HTS01959 which is characterized as one of the top ranking compounds through the process of virtual screening is also found to be an active inhibitor for N(5)-carboxyaminoimidazole ribonucleotide, N(5)-CAIR synthetase from *E.coli*. The puchem BioAssay information for this compound reveals the IC₅₀ value being 8.7 μ M for N(5)-CAIR enzyme (NIH Chemical Genomics Center; AID:350915).

CONCLUSIONS

The homology modelling of CysRS of *Mycobacterium leprae* TN strain has been carried out in the present study. The modelled structure of *MICysRS* reveals that the topological architecture is highly similar to *E.coli* CysRS. Moreover, the virtual screening that has been performed using AutoDock Vina in PyRx software identified 14 potential lead compounds for the inhibition of *MICysRS*. Interestingly, one of the 14 compounds HTS01959 is found to possess inhibitory activity against an another enzyme N(5)-CAIR synthetase of *E.coli*. All these compounds show better binding affinity than the first-line anti-leprosy drugs. However, this study has to be substantiated through *in vitro* and *in vivo* experiments.

ACKNOWLEDGEMENTS

The authors sincerely thank the support from the management of SASTRA University. We wish to express our gratitude to Dr. S. Swaminathan, Dean, Sponsored Research, SASTRA University for his constant encouragement.

REFERENCES

- [1] Pham JS, Dawson KL, Jackson KE, Lim EE, Pasaje CFA, Turner KEC, Ralph SA, Int J Parasitol: Drugs Drug Resist 2014;4:1–13.
- [2] Ibba M, Söll D, Annu Rev Biochem 2000; 69:617–650.
- [3] Fersht AR, Enzyme Structure and Function. Freeman and Co., New York, 1985.
- [4] Eriani G, Delarue M, Poch O, Gangloff J, Moras D, Nature 1990; 347:203–206.
- [5] Cusack S, Härtlein M, Leberman R, Nucleic Acids Res 1991; 19:3489– 3498.
- [6] Moras D, Trends Biochem Sci 1992; 17:159–164.
- [7] Schimmel P, Trends Biochem Sci 1991; 16:1–3.
- [8] Lee SW, Cho BH, Park SG, Kim S, J cell science 117:3725–3734.
- [9] Nagle GM, Doolittle RF, J Mol Biol 1995; 40:487–498.
- [10] Sathya R, Thamotharan S, Int J Quantum Chem 2015; 115:187–195.
- [11] Kar KH, Gupta R, Clinics in Dermatology 2015; 33: 55–65.
- [12] Newberry KJ, Hou YM, Perona JJ, EMBO J 2002; 21:2778–2787.
- [13] Hauenstein S, Zhang CM, Hou YM, Perona JJ, Nat Struct Mol Biol 2004; 11:1134–1141.
- [14] Eswar N, Marti-Renom MA, Webb B, Madhusudhan MS, Eramian D, Shen M, Pieper U, Sali A. Current Protocols in Bioinformatics, John Wiley & Soncs, Inc. Supplement 15, 2006; 5.6.1–5.6.30.
- [15] Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ, J Mol Biol 1990; 215:403–410.
- [16] Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG, Bioinformatics 2007; 23:2947–2948.
- [17] Lovell SC, Davis IW, Arendall III WB, De Bakker PIW, Word JM, Prisant MG, Richardson JS, Richardson DC, Proteins: Structure, Function & Genetics; 2002; 50:437–450.
- [18] Trott O, Olson AJ, J Comp Chem 2010; 31:455–461.
- [19] Celniker G, Nimrod G, Ashkenazy H, Glaser F, Martz E, Mayrose I, Pupko T, Ben-Tal N, Isr J Chem 2013; 53:199–206.
- [20] Ashkenazy H, Erez E, Martz E, Pupko T, Ben-Tal N, Nucleic Acids Res 2010; 38:W529–W533.