The in-silico hunt for the inhibitor against LpxC enzyme from *Salmonella Typhi from* natural product derivatives

Sudhir Kumar Pal[1*], Tarun Kumar[2]

ABSTRACT

A large number of infections are caused by *Salmonella* worldwide at a remarkable pace. Salmonella enteric serotype *typhi* (*S. typhi*), gram negative bacteria (only cause disease in man) is the predominant causative agent of Typhoid fever. Typhoid fever is most common in poor and undeveloped countries of Asia and Africa. One of the major components of virulence factors produced during *salmonella* infection is Lipid A, which acts as a strong human immuno-modulator bacterial endotoxin. Regulation of Lipid A biosynthetic pathway occurs at second step, catalyzed by LpxC, a Zn^{2+} dependent metalloamidase. Systematic Screening of natural products library database fruitfully provided us a potent lead molecule {Pubchem CID: 1788783 (ZINC02133485) ((2S)-2-[[2-[3-(4-fluorophenyl)-5-methyl-7-oxofuro[3,2-g]chromen-6-yl]acetyl]amino]-3-(5-hydroxy-1H-indol-3-yl)propanoic acid)} which actively binds (binding energy: -10.7 kJ/mol and Kd: 14.35 nM) with LpxC enzyme and could be developed into a sound and potent inhibitor of LpxC enzyme after the application of drug development and processing strategies. Wet lab experimentation is required to validate these results for further use.

Keywords: *Salmonella Typhi, LpxC, MD Simulation, Inhibitors, Raetz pathway, Lipid A.*

1. Introduction

A large number of infections are caused by *Salmonella* worldwide. The most frequent Serotypes associated with invasive disease are *S.* Typhi, *S.* Typhimurium and *S.* Enteritidis (1,2). Prevalence of insidious non-typhoidal Salmonella (iNTS) disease is as high as 227 per 100,000 cases every year. Salmonella enteric serotype Typhi (*S.* Typhi), gram negative bacteria (only cause disease in man) is the predominant causative agent of Typhoid fever. Cases of 12% to 30% of untreated illness end up in the form of death. Reversion of fever can occur in about 10% of untreated people. Typhoid fever is most common in poor and undeveloped (low or middle-income) countries with rates greater than 100 per 100,000 persons per year in parts of Asia and Africa. US suffer outbreaks every year (3). The global burden of typhoid fever disease as estimated by WHO is 11-20 million cases annually, consequential casualties about 1,28,000-1,61,000 per year (www.who.int/news-room/fact-sheets/detail/typhoid). The *S. typhi* colonized gallbladder showed an asymptomatic chronic infection. It is a fact that a typhoid toxin with a carcinogenic potential is produced by *S. typhi*, that induces DNA damage and cell cycle alterations in intoxicated cells (4). This is also a reason to work on this bacterium as the subject of the study.

Most of the *Salmonella* infections originate from contaminated water and food and in hospitals. The situation has been aggravated due to increased incidents of drug resistance of *Salmonella* strains towards a broad range of antibiotics. *Salmonella* is resistant to a number of antibiotics viz, ampicillin, ciprofloxacin, Streptomycin, furazolidone, sulfonamides, tetracyclines and fluoro-quinolones (5,6). Although, two FDA approved vaccines are also available but with limited functionality. Children less than two years of age cannot be treated with these vaccines (3), putting children at the highest risk. Therefore, it is imperative to search for new drugs and their profound targets in bacteria which would have minimum homology and are functionally unrelated to human proteins.

One of the major components of virulence factors produced during *salmonella* infection is lipopoly-saccharide (LPS). LPS layer around outer membrane confers a first line of barrier to prevent entry of harmful substances like antibiotics and other small molecules into the cell (7). Three parts of LPS are an outermost immunodominant and highly variable repeating oligosaccharide known as the [O-antigen] which is linked to the [core oligosaccharide domain] which, in turn is anchored to outer membrane through glucosamine containing phosphorylated lipid i.e., [Lipid A]. Apart from

its function as a hydrophobic membrane anchor of LPS, Lipid A is a strong human immuno-modulator bacterial endotoxin (8). Takayama and colleagues in 1983 elucidated the first complete chemical structure of lipid A from *salmonella typhimurium* (9). It has been observed that *E.coli* mutants lacking lipid A, either do not survive or are highly sensitive to antibiotics (10). Lipid A biosynthesis involves nine conserved enzymes which all are needed for the viability of the cell. Regulation of Lipid A biosynthetic pathway occurs at second step, catalyzed by UDP-3-O-acyl-N-acetylglucosamine deacetylase (LpxC) (10). LpxC is a Zn^{2+} dependent metalloamidase which catalyzes the release of acetyl group from UDP-(3-O-(R-3-hydroxymyristoyl)-N-acetylglucosamine (myr- UDP- GlcN) to form UDP-(3-O-(R-3-hydroxymyristoyl))-N-glucosamine and acetate (**Fig.1**). LpxC catalyzed reaction is irreversible and therefore is a committed step in lipid A synthesis (10–13).
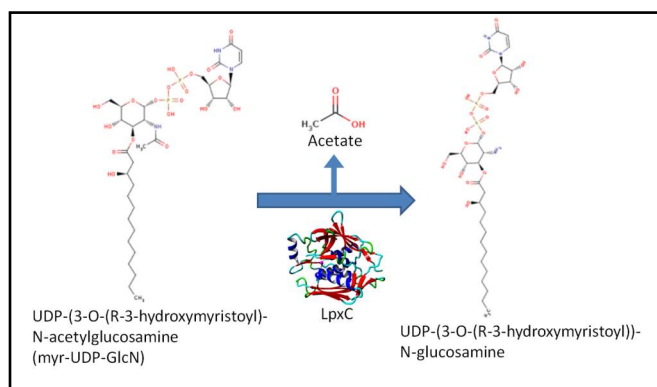

Fig. 1: *Action mechanism of LpxC enzyme*

Consequently, LpxC has been considered significant as a drug target. Structure elucidation of LpxC from *Escherichia Coli*, *Aquifex Aeolicus*, *Pseudomonas Aeruginosa* and *Yersinia Enterocolitica* has provided the substantial understanding of catalytic site topology and catalytic mechanism of the enzyme. Structural studies have prepared the ground for designing novel inhibitors against the enzyme (14–17). A number of LpxC small molecule inhibitors with hydroxamate moiety have been synthesized. One such hydroxamate based L-161, 240 inhibitor which inhibits LpxC from *E.coli* has been shown inactive against LpxC from *Pseudomonas aeruginosa.* This kind of differential inhibition can be attributed to subtle structural differences of the LpxC enzymes from the two organisms (18). In case of *E coli* lpxC, βa-βb loop is positioned away from the catalytic site resulting in enlarged catalytic site as compared to that of *Aa*LpxC and *Pa*LpxC. Consequently, inhibitor molecule BB-78485 with bulky naphthalene groups can be accommodated in *E.coli* LpxC catalytic site but not in *Aa*LpxC and *Pa*LpxC (19). The differences in inhibitor binding to the orthologs of LpxC in response to slight variations in their catalytic sites, necessitates structural

characterization of a greater number of LpxC orthologs. Moreover, the presence of zinc-ion in the active site further enhances the activity of LpxC and any lead molecule which directly interacts with the zinc-ion would be the best choice for further drug development process.

2. Material and Methods

2.1 Protein antigenicity and allergenicity

Antigenic nature of the protein is calculated using VaxiJen v2.0 web-server which helps in the prediction of potent antigens using default parameters (20). In the same manner, the allergenicity prediction was carried out using AllerTop v.2.0 webserver (21).

2.2 Homology modeling

The protein, UDP-3-O-(R-3-hydroxymyristoyl)-N-acetylglucosamine deacetylase (LpxC), which is 305 amino acids in length, was chosen as the target sequence (UniprotKB – Q8Z9G5 (LPXC_SALTI) to be modelled. The sequence was downloaded from Uniprot (22) and was subjected to protein-protein blast. 19 sequences were found to match with the target sequence. Out of those, Chain A of the Crystal Structure of the *Escherichia Coli* Lpxc/lpc-138 Complex (PDB ID: 4MQY_A) (query coverage=100%, E-value= 0.0, and sequence identity=98% (as calculated by the NCBI protein-BLAST program (http://blast.st-va.ncbi.nlm.nih.gov/Blast.cgi))) was selected as the best template for full length homology model preparation. The target and template sequences are aligned by clustal omega server (23,24) using defaults settings. To construct the protein model for the target sequence, Modeller 9.12 suite (25–27) was used. The secondary structure was predicted using PsiPred (28). Alignments of target and template sequences are adjusted to avoid big gaps in the secondary structure domain. After the side-chains had been built, optimized and fine-tuned, all newly modelled parts were subjected to a combined steepest descent and simulated annealing minimization (i.e. the backbone atoms of aligned residues were kept fixed to avoid potential damage). The original ligand and other molecules in the template structure were kept intact to get the absolute idea about the binding and active site of the newly made 3D structure of LpxC. As the template is good, the resultant structure is also good in terms of energy and stability [29]. For a comparative analysis, web-servers i.e. PS(2), Modweb, Phyre2, CPHModels-3, and Swiss Model were used to model 3D structures of the desired protein. The stereochemical quality and validation of the generated models were confirmed during the quality check of model geometry using Ramachandran plot (29,30), ProSA web-server (31), ERRAT (32), Verify3D (33), and G-factor calculations in PROCHECK (34,35).

2.3 Ligand Library preparation for Virtual Screening

Library of Natural Compounds (43,458 compounds) was downloaded from ZINC-12 database (https://zinc12.

docking.org/browse/catalogs/natural-products) and subjected to PyRX along-with two known LpxC inhibitors [3P3 (N-[(1S,2R)-2-hydroxy-1-(hydroxycarbamoyl) propyl]-4-(4-phenylbuta-1, 3-diyn-1-yl)benzamide) and L53 (4-[4-(4-aminophenyl) buta-1, 3-diyn-1-yl]-N-[(2S, 3S)-3-hydroxy-1-nitroso-1-oxobutan-2-yl] benzamide)] for compound cleaning, energy minimization, and conversion from 'sdf' to 'pdbqt' format for the purpose of Virtual Screening applying protocol of auto-dock vina. Known ligands were used as a benchmark and control for validation of the docking protocol. This prepared library was used to dock with the protein to get the best inhibitory compounds based on binding energy for further analysis. Maximum number of allocated degrees of freedom (DOF) as allowed by autodock is 10 for each ligand. Therefore, DOF are assigned automatically by the software according to compound's structure.

2.4 Molecular Docking

Initial blind docking of 43,458 compounds was performed using PyRx with default settings. The 5 best scoring compounds with lowest energy of binding or binding affinity was extracted and aligned with receptor structure for further analysis (subjected to AutoDock4). In order to carry out the molecular docking, we used the AutoDock 4.0 suite (12). Intermediary steps, such as pdbqt files for protein and ligands preparation and grid box creation were completed using Graphical User Interface program AutoDock Tools (ADT). ADT assigned polar hydrogens, united atom Kollman charges, solvation parameters and fragmental volumes to the protein. AutoDock saved the prepared file in PDBQT format. AutoGrid was used for the preparation of the grid map using a grid box. The grid size was set to 62 × 70 × 88 xyz points with grid spacing of 0.375 Å and grid centre was designated at X= 190.404, Y= 101.754, Z= -0.753 dimensions.

A scoring grid is calculated from the ligand structures to minimize the computation time. The Lamarckian Genetic Algorithm (LGA) was chosen to search for the best conformers. During the docking process, a maximum of 10 conformers was considered for each compound. The population size was set to 150 and the individuals were initialized randomly. Maximum number of energy evaluation was set to 2500000, maximum number of generations 27000, maximum number of top individuals that automatically survived set to 1, mutation rate of 0.02, and crossover rate of 0.8. Rest of the docking parameters were set to be default with 10 LGA runs. During the docking procedure, both the protein and ligands are considered as rigid. The results less than 1.0 Å in positional root-mean square deviation (RMSD) was clustered together and represented by the result with the most favorable free energy of binding. The pose with lowest energy of binding or binding affinity was extracted and aligned with receptor structure for further analysis.

2.5 Molecular Dynamics Simulations

The docked 5 complexes viz, LpxC_ZINC27215482, LpxC_ZINC70707494, LpxC_ZINC70707502, LpxC_ZINC018754, and LpxC_ZINC02133485 were subjected to molecular dynamics (MD) simulations to understand the conformational and structural changes during protein-ligand complex formation. Simulations also provide information regarding the strength of the interaction and also the binding partners took part in molecular bonding. MD simulations were performed using YASARA, version 15.10.18 (36,37), with the AMBER03 force field (38), which is a balanced force field that works well with both macro- and micro-molecules. The protein-ligand complex was placed in a water box that is 10 Å larger than each side of the protein. Hydrogen atoms were added to the protein structure at the appropriate ionizable groups according to the computed pKa in relation to the simulation pH, thus a hydrogen atom will be added if the computed pKa is higher than the pH. The pKa is computed for each residue according to the Ewald method (39,40). The structure was then minimized using steepest-descent method followed by simulated annealing. The simulation was performed at pH 7.0 in a 0.9% NaCl solution at 300 K temperature for 50 ns. A cut-off of 7.86 Å was used for van der Waals forces while Particle Mesh Ewald algorithm (41) were used for electrostatic forces. A multiple time step of 1.25 and 2.5 fs was used for intra-molecular and inter-molecular forces, respectively. All calculations were conducted on an Intel Core i5 2.50 GHz with 4 GB of RAM. Approximately, 4-5 ns/day of speed was obtained using the said specifications.

3. Results & Discussion

The selection of distant homologs of *St*LpxC protein was performed using the web tool Protein Pathtracker (42), which allows to track the evolutionary history of a query protein by locating homologs in selected proteomes along several evolutionary paths. Such sets of homologs are often applied to investigate the function of a protein and the degree to which experimental results can be transferred from one organism to another.

Utilizing default settings, we found 2 ortholog sequences in taxon Cellular Organisms, 10 orthologs in taxon Viridiplantae, 5 orthologs in taxon Bacteria, and 2 orthologs and 4 homologs in taxon Arthropoda. No sequences were found in Primates and Fungi taxa. Results and figures are provided in **supplementary data Appendix 2.**

Using FastaHerder2 (43), 241 similar sequences which are very similar or identical to StLpxC were clustered and aligned to make a small dataset of sequences which are functionally and evolutionarily close to each other. Weblogo 3 (44) was used to make the amino acid propensity diagram of these 241 clustered LpxC protein sequences (**Fig. 2**).

Fig. 2: Amino acid propensity diagram of 241 clustered sequences of LpxC (from residues 1 to 305 numbering as of StLpxC) made by Weblogo 3 (44).

The numbering of figure starts and ends according to the numbering of *St*LpxC. Colouring is set to be according to the charge of the amino acids.

All the 241 sequences are provided (as fasta text) in the **supplementary data Appendix 1** in aligned format. Thirty residues out of 305 are found to be identical in all the 241

sequences taken into consideration for propensity match.

VaxiJen server (20) which was used for antigenicity prediction gives overall protective antigen prediction score of 0.4307 (data not shown) which tells that the protein is a probable antigen. AllerTop server (21) predicted our protein sequence as a probable non-allergen.

3.1 Construction of *S. typhi* LpxC model structure

For modelling the LpxC structure, Uniprot sequence: Q8Z9G5 was chosen as the target sequence. For the construction of the protein, the *Escherichia Coli* Lpxc/lpc-138 Complex 3D structure (PDB ID: 4MQY) was chosen as a template. The 4MQY structure shared a sequence identity of 98% with the target sequence. The sequence alignment of the target to template sequence is shown in **supplementary data Appendix 3**.

YASARA based MODELLERv.9.12 suite is used for model creation along-with other web-based servers like (PS)2-3.0, Modweb, Phyre2, CPHModels-3.0, and Swiss Model. Parameters for the quality assessment of the generated models are listed in **Table 1**. Ensuing comparison of the model quality and stereochemistry guided the selection of the best model. None of the web-based servers except Swiss model supported the inclusion of Zn^{2+} cofactor. With the PROCHECK statistics (92.3% residues in the core favourable region), good ERRAT quality (93.266), Verify-3D score of 100%, and RMSD of 0.375 Å, model obtained from MODELLER seems to be the most reliable and high-quality representative 3D structure of StLpxC.

The knowledge-based energy graph calculated by ProSA for the top modelled LpxC is provided in **Fig. 3**.

Table 1: Stereochemical evaluation parameters for StLpxC structure models made by Modeller and different web servers.

| Models	PROCHECK								ERRAT	Verify-3D	ProSA	YASARA
	No. of non-glycine and non-proline residues				No. of proline, glycine and end residues	Residue coverage	Bad contacts	G factor	Quality factor (%)	3D-1D Scores>0.2 (%)	Z-Score	RMSD
	Most favoured region [A,B,L]	Additionally allowed region [a,b,l,p]	Generously allowed region [~a,~b,~l,~p]	Dis-allowed region								
YASARA	251 (92.3%)	20 (7.4%)	1 (0.4%)	0	33	305	3	0.12	93.266	100	-7.05	0.375
(PS)2-3.0	254 (93.4%)	17 (6.2%)	1 (0.4%)	0	33	305	1	0.02	82.8283	93.77	-7.07	0.123
Modweb	252 (94.0%)	15 (5.6%)	1 (0.4%)	0	32	300	1	0.05	84.589	99.00	-7.11	0.125
Phyre2	248 (91.2%)	21 (7.7%)	2 (0.7%)	1 (0.4%)	33	305	2	-0.05	76.0943	96.39	-7.16	0.147
CPHModels-3.0	234 (87.3%)	33 (12.3%)	0	1 (0.4%)	33	301	0	0.04	87.6712	99.67	-7.16	0.520
Swiss Model	246 (91.8%)	21 (7.8%)	1 (0.4%)	0	36	304	1	-0.12	97.561	98.33	-6.94	0.063
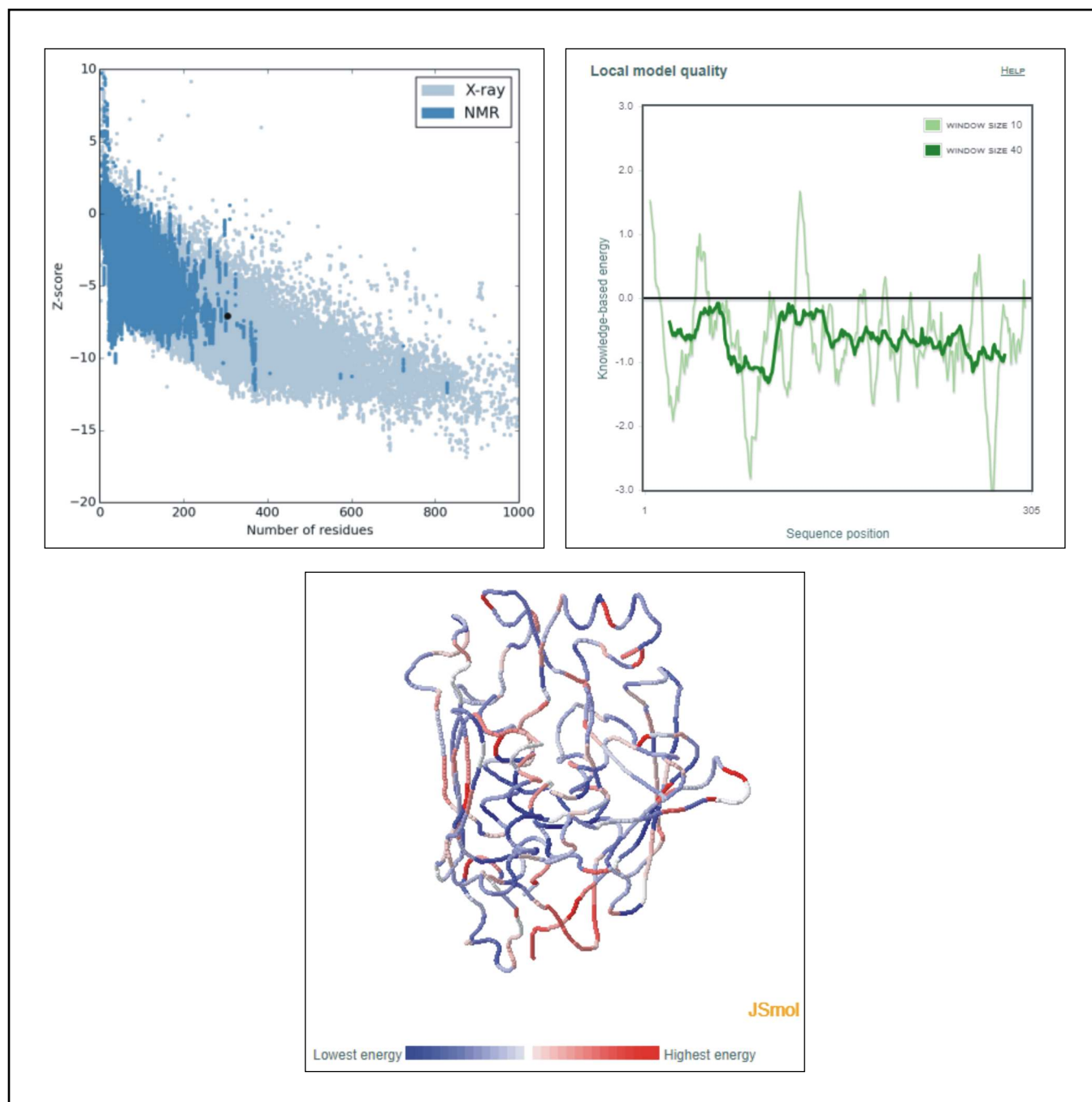
Fig. 3: *Knowledge-based energy profile of StLpxC obtained from ProSA-webserver. Black dot in the left image represents the homology model. Middle graph represents the energy profile of the protein.Right image shows the Energy distribution in the model according to the type of amino acids.*

The graph shows local model quality by plotting energy as a function of amino acid. As can be seen in the Fig. 3 (upper right), there are two types of lines: thick and thin. The thick line represents the average energy over each 40 residue fragments while the thin line depicts window size of 10 residues and be seen in the background of the graph. In general, the positive values correspond to erroneous part of the input structure. As can be seen in the graph, majority portion of the top modelled structure has negative value and hence can be considered as the most dependable model for the protein. Ramachandran Plot (29,30,45) shows the assignment of amino acids on the basis of phi-psi angles (**Fig. 4**). Number of residues fall in favoured region are 297 (98.0%), Number of residues fall in allowed region are 6 (2.0%). No outlier residues were present in the model.
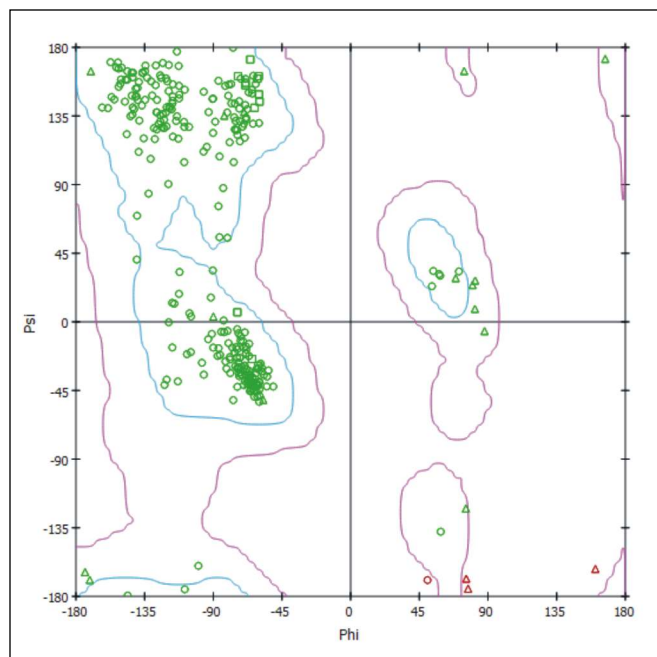
Fig. 4: *Ramachandran plot for StLpxC. Green cubes are prolines, green and red triangles are glycine, red circle is cysteine. Green circles are other amino acids. Figure is obtained from DSVisualizer.*

ModWeb and Swiss Model generated models were having stereochemistry and quality comparable to that of MODELLER model. However, incorporation of Zn^{2+} ion in the structural model (LpxC is a zinc-dependent metalloenzyme) was a feature supported by MODELLER alone, thereby guiding the selection of the most structurally relevant as well a good quality model. *St*LpxC 3D structure acquired in this study is having the same typical β1–α–α–β2 geometry of LpxC protein with characteristic slight changes accordig to the sequence of the protein.

3.2 Virtual Screening and molecular docking

Virtual screening and molecular docking was performed with Autodock vina embedded in the PyRX (46,47). All compounds were screened and out of them initial 20 compounds were selected which were subjected to AutoDock 4 for site specific docking. This results in 5 most preferred compounds viz. ZINC27215482, ZINC70707494, ZINC70707502, ZINC018754, and ZINC02133485 (**Table 2**) which were selected on binding energy and coefficient of inhibition parameters.

Polar and non-polar interactions (**Table 3**) were recorded using DS visualizer (48). It is pretty clear that all the 5 selected natural compounds bind in the active site of the protein and Zinc-metal is playing crucial role in stabilizing the molecules in most of the cases. In ZINC70707494, ZINC70707502, and ZINC02133485 ((2S)-2-[[2-[3-(4-fluorophenyl)-5-methyl-7-oxofuro[3,2-g]chromen-6-yl]acetyl]amino]-3-(5-hydroxy-1H-indol-3-yl)propanoic

acid) zinc metal directly forms polar or non-polar interaction. Leu-62, Phe-192, and Arg-196, His-238 also plays role in binding the small molecule via bonded or non-bonded interactions (**Fig. 5**).

3.3 Toxicity Prediction

Finalized 5 natural compound inhibitors are checked for their toxicity using online tool named ProTox-II (49). **Table 4** shows the results for toxicity prediction which depicted that compound ZINC01875407 shows mutagenicity. ZINC02133485 belongs to the class 6 of predicted toxicity and found to be non-toxic out of all five ligands.

Toxicity table and radar chart are provided as **supplementary data Appendix 4**. In ProTox-II (49) there are 6 classes for toxicity (1 to 6) in which class 1 has LD50 => 5 which is fatal in nature on the other hand class 6 shows LD50 > 5000 which means compound is non-toxic.

3.4 Molecular Dynamics Simulations

Five protein-ligand complexes obtained from virtual screening is subjected to molecular dynamics simulation studies using mdrun macro of YASARA suite for 50ns time frame. **Fig. 6** demonstrates the difference between the initial and final conformations of the protein-ligand complexes before and after the MD simulation runs of 50ns each. Initial and final poses are aligned and superimposed to understand the dynamicity of the complex.

ZINC27215482-LpxC has an RMSD of 1.188 Å over 300 aligned residues with 100.00% sequence identity. ZINC018754-LpxC has an RMSD of 1.253 Å whereas ZINC02133485-LpxC has an RMSD of 1.250 Å. In the case of ZINC70707494-LpxC the RMSD of initial to final conformation is 1.288 Å. On the other hand, ZINC 70707502-LpxC shows an RMSD of 1.293 Å. Overall RMSD for C-alpha atom and RMSF per residue based in all five complexes are shown in **Fig. 7**.

Total 200 frames are captured in 50 ns run at a rate of one frame per 250ps. LpxC-ZINC018754 complex shows non-equilibrated increasing RMSD after 100 frames up-to the end of simulation. This suggested that maybe system is not equilibrated enough to bring down the rmsd and stabilize the protein-ligand complex and needs to be run with an extended time-frame. All other complexes shows quite equilibrated trajectories with one or two in-between spikes of high energy. As shown in **Fig. 7 (left panel)**, the complex of ZINC02133485-LpxC showed a beautiful, equilibrated trajectory symbolizes the stable conformation and nice fit of ligand with the active site of the protein.

In ZINC70707502-LpxC complex area near residue number 68-72 and 160-164 shows higher fluctuations. Similar high fluctuating spikes can be observed in LpxC-ZINC018754 complex in the block of 200-240 residue numbers (**Fig. 7 (right panel)**). These fluctuations may

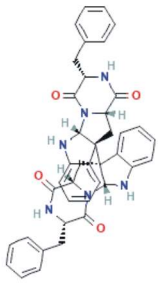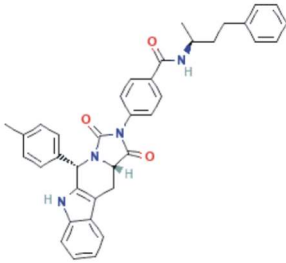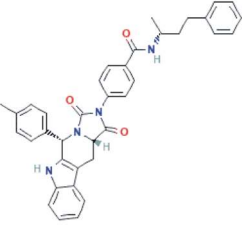***Table 2:** Physicochemical properties of the best five inhibitory screened natural compounds.*
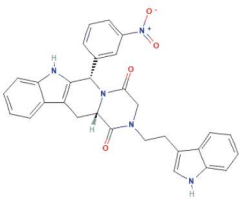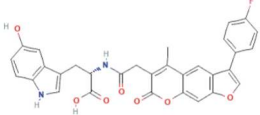
S.No.	Name of the Natural Compound	Structure	Molecular weight (g/mol)	No. of H-bond donors	No. of H-bond acceptors	Log p
1.	ZINC27215482 (CID: 92846501) [benzyl-(benzyl-dioxo-BLAHyl)BLAHdione]		664.766	4	10	3.39
2.	ZINC70707494 (CID: 95374393) [4-[dioxo(p-tolyl)BLAHyl]-N-[(1S)-1-methyl-3-phenyl-propyl]benzamide]		582.704	2	7	7
3.	ZINC70707502 (CID: 95374395) [4-[dioxo(p-tolyl)BLAHyl]-N-[(1R)-1-methyl-3-phenyl-propyl]benzamide]		582.704	2	7	7
4.	ZINC01875407 (CID: 1624198) [(6S,12aS)-2-[2-(1H-indol-3-yl)ethyl]-6-(3-nitrophenyl)-6,7,12,12a-tetrahydro-3H-pyrazino[1,2-b]$b-ca]		519.561	2	9	4.36
5.	ZINC02133485 (CID: 1788783) [(2S)-2-[[2-[3-(4-fluorophenyl)-7-keto-5-methyl-furo[3,2-g]chromen-6-yl]acetyl]amino]-3-(5-hydroxy-1H]		553.522	3	9	3.00

Table 3: Interactions table between Enzyme and inhibitors before and after MD simulations.
Binding affinity and inhibition constant for the respective
complexes after docking are also reported.

Complex	Binding Affinity (kcal/mol)	Inhibition Constant (nM)	Interactions			
			Before MD Simulation		After MD Simulation	
			Polar	Non-Polar	Polar	Non-Polar
ZINC27215482 [benzyl-(benzyl-dioxo-BLAHyl)BLAHdione]	-10.3	26.67	Met-61, Phe-192, Cys-63	Asp-197, Leu-62, Leu-18, Ala-215, His-265	Phe-194	Phe-192, Ile-159, Met-61, Arg-196, Asp-197
ZINC70707494 [4-[dioxo(p-tolyl)BLAHyl]-N-[(1S)-1-methyl-3-phenyl-propyl]benzamide]	-8.51	579.45	Nil	His-265, Lys-239, Phe-192, Leu-62, Thr-60, Met-61, Asp-242		
ZINC70707502 [4-[dioxo(p-tolyl)BLAHyl]-N-[(1R)-1-methyl-3-phenyl-propyl]benzamide]	-8.41	689.27	N/A	Leu-62, Asp-242, His-265, Lys-239, Phe-192, Thr-60, Met-61		
ZINC01875407 [(6S,12aS)-2-[2-(1H-indol-3-yl)ethyl]-6-(3-nitrophenyl)-6,7,12,12a-tetrahydro-3H-pyrazino[1,2-b]$b-ca]	-9.27	161.51	Lys-239, His-265	Ile-198, Cys-207, Cys-214, Ala-215, Leu-18, Leu-62, Phe-192, Asp-242, Gly-264		
ZINC02133485 [(2S)-2-[[2-[3-(4-fluorophenyl)-7-keto-5-methyl-furo[3,2-g]chromen-6-yl]acetyl]amino]-3-(5-hydroxy-1H]	-10.7	14.35	Asp-242, Phe-192, Lys-239	Cys-207, Gly-210, Leu-18, Cys-214, Leu-62, Ala-215, Thr-191, Cys-63	Cys-63	Leu-18, Met-61, Leu-62, Phe-192, Phe-194, Cys-207, Zinc-306

Table 4: Toxicity prediction results for the selected 5 natural compounds
as calculated using online tool ProTox-II (49).

Natural Compound	Predicated LD50 (mg/kg)	Predicted Toxicity class	Average Similarity (%)	Prediction accuracy (%)
ZINC27215482	28	2	61.22	68.07
ZINC70707494	650	4	64.74	68.07
ZINC70707502	650	4	64.74	68.07
ZINC01875407	700	4	69.87	68.07
ZINC02133485	8700	6	50.42	67.38

ZINC27215482-LpxC

ZINC70707494-LpxC

ZINC70707502-LpxC
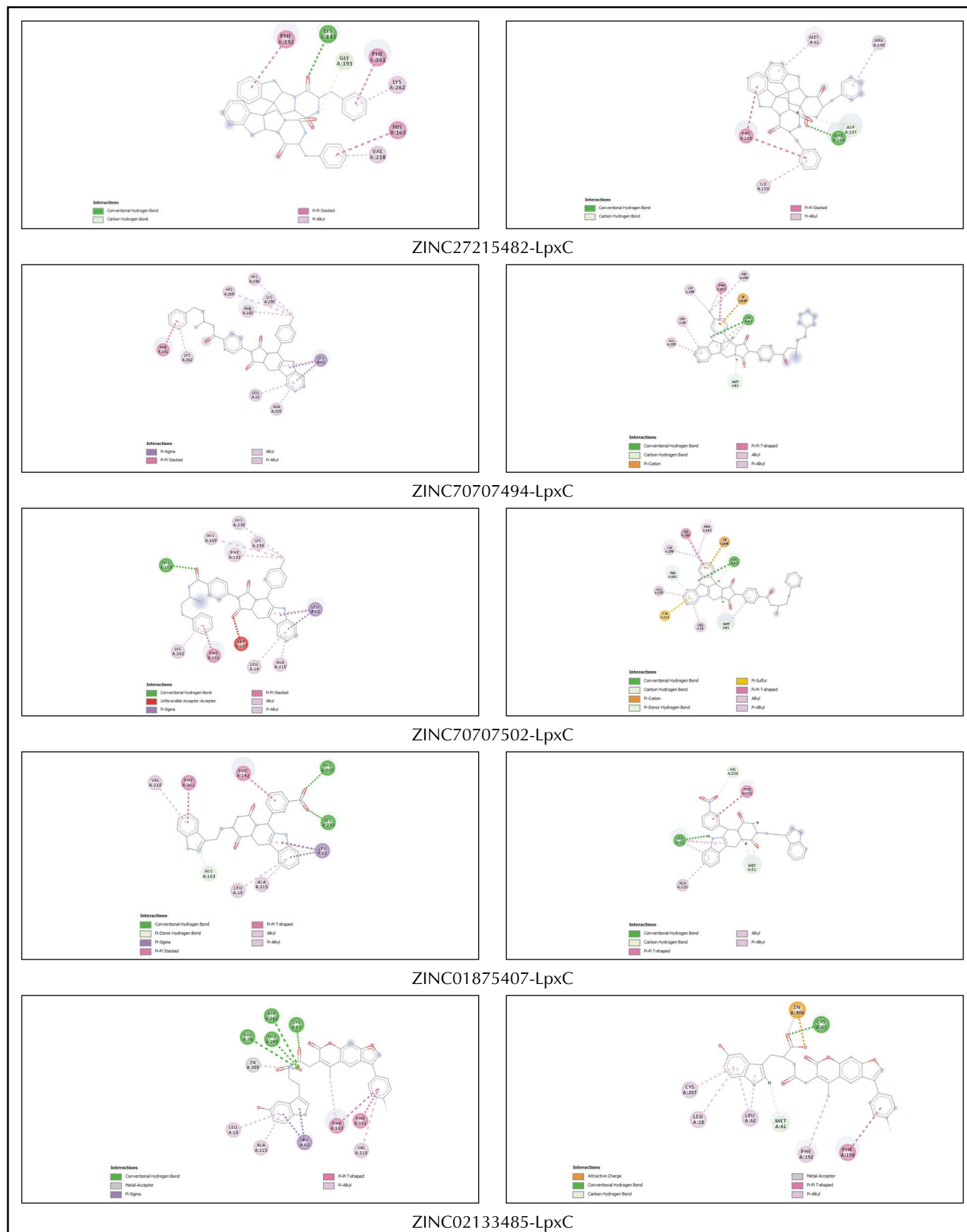
ZINC01875407-LpxC

ZINC02133485-LpxC

Fig. 5: *Interactions between Enzyme and 5 selected inhibitors before MD simulations (Left) and after MD simulations (Right).*
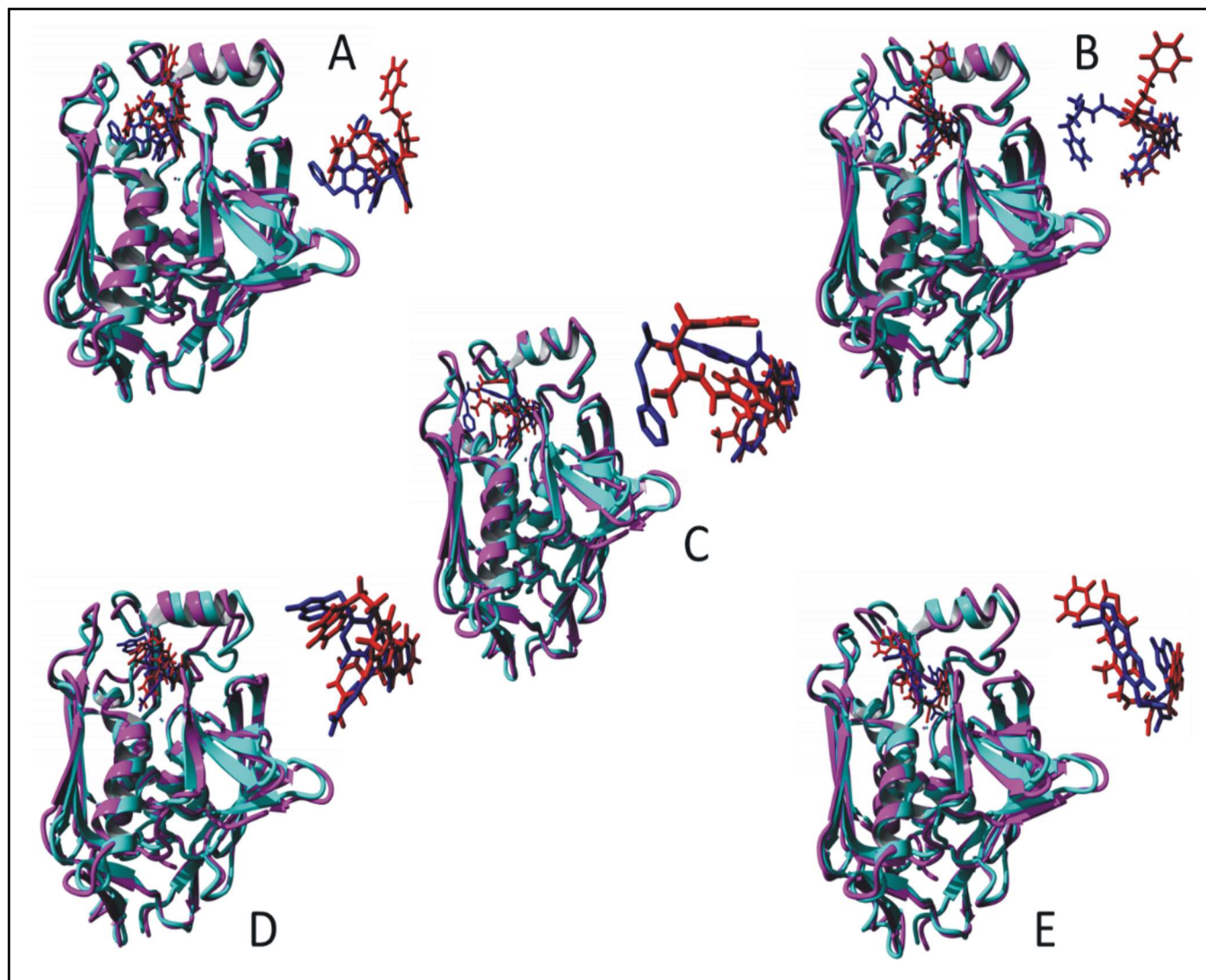
Fig. 6: *StLpxC conformations before (cyan) and after (magenta) 50 ns MD simulation run. Individual ligand conformations before (blue) and after (red) MD run are shown on the side zoomed leaflets in all the 5 complexes. (A) ZINC27215482-LpxC, (B) ZINC70707494-LpxC, (C) ZINC70707502-LpxC, (D) ZINC01875407-LpxC, and (E) ZINC02133485-LpxC.*
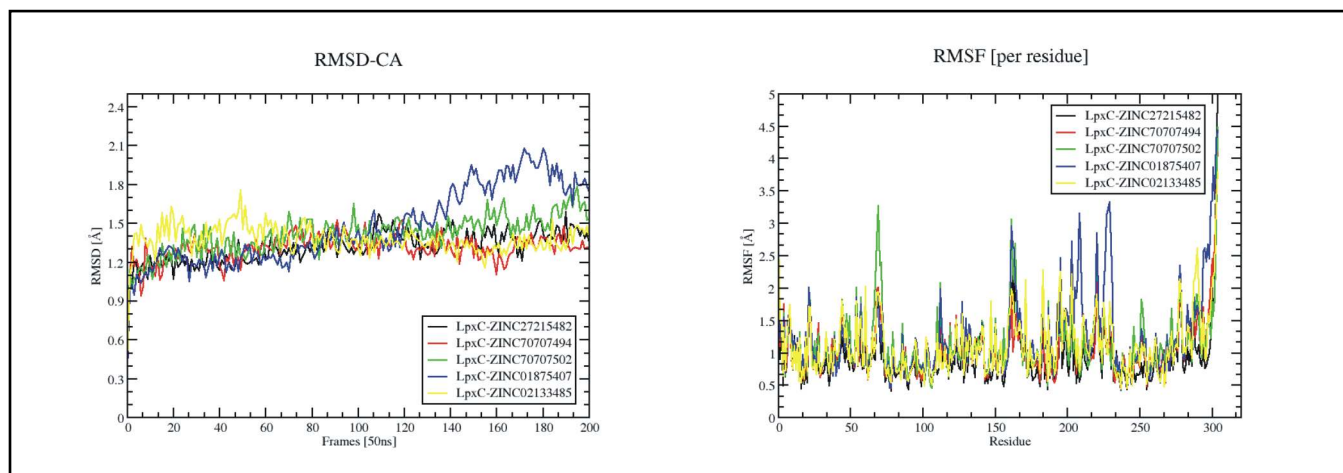

Fig. 7: *RMSD calculations in (Å) for C-alpha atom of the protein-ligand complex in 50ns. 200 frames are observed (1 frame per 250ps) (left panel). RMSF is calculated per residue based to understand the level of fluctuations imparted during the MD simulation.*
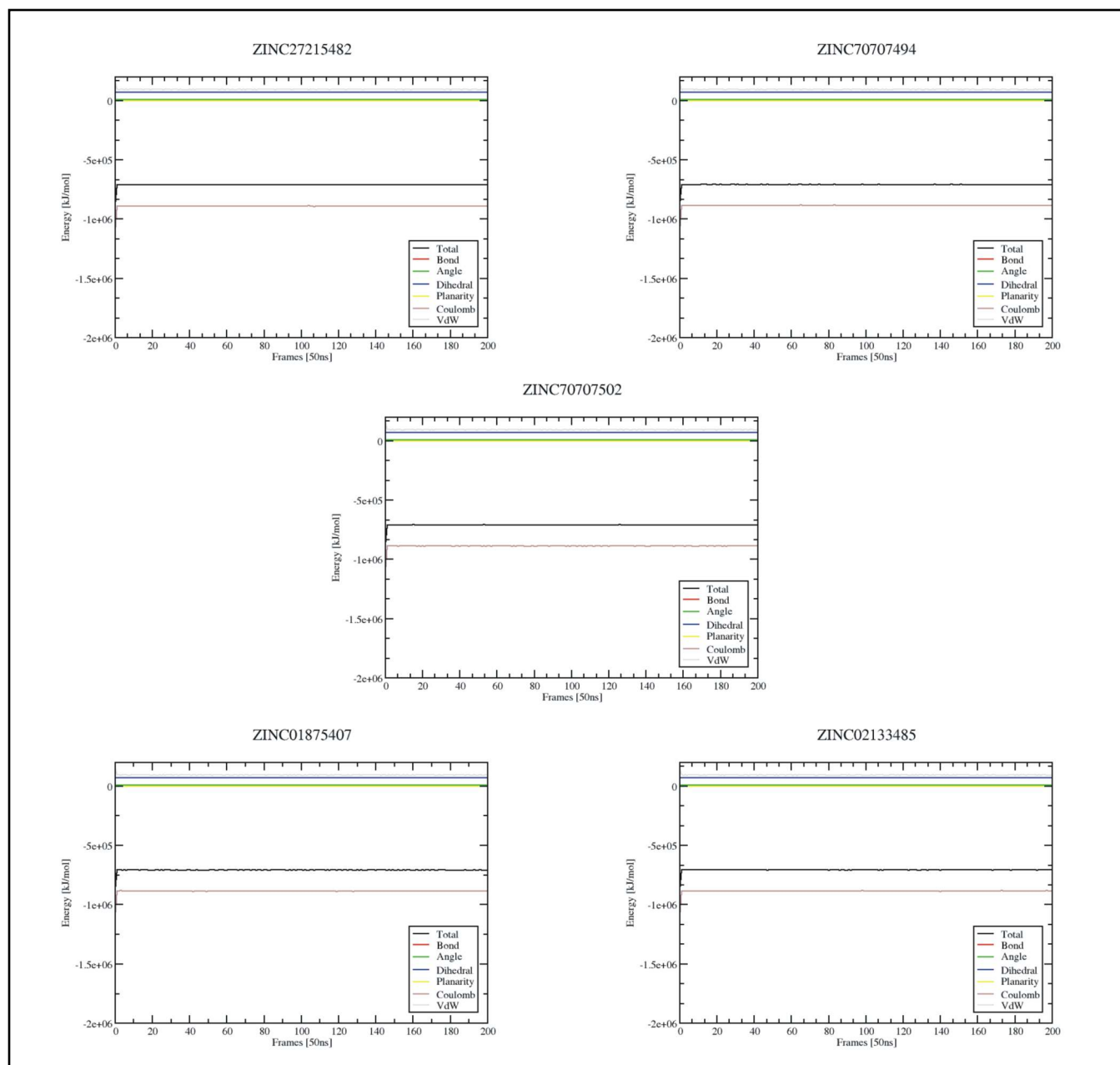
Fig. 8: *Energy plots for different energy contributors like Bond, Angle, Dihedral, Planarity, Coulomb, VdW to make up the total potential energy of all the 5 different complexes. In most of them bond energy is totally overlapped with the angle and planarity energies.*

address an important event (in the form of ligand interaction or a change in protein conformation) which takes place. Different energy contributors like Bond, Angle, Dihedral, Planarity, Coulomb, VdW to make up the total potential energy (**Fig. 8**) is quite equilibrated and normal during the whole 50ns course of simulation run.

4. Conclusion

Salmonella typhi, being the emerging opportunistic pathogen needs a check on its virulence in an urgent manner. Researchers throughout the world are working continuously to address this issue and finding targets and

drug molecules against the bacteria. This study is also a small contribution in the same direction. Starting from homology modelling of *St*LpxC to the docking and simulation of natural products ligands and till the prediction of toxicity issues, this study provides us a novel leading compound, ZINC02133485 (CID: 1788783) which utilizing further wet-lab analysis and experimentation can be developed in a potent LpxC inhibitor.

5. Acknowledgement

Authors would like to acknowledge Dr. Sanjit Kumar, Associate Professor, Guru Ghasidas Vishwavidyalaya,

Bilaspur (A Central University), Chhattisgarh for his continuous support in manuscript writing and providing YASARA software for academic research.

6. Conflict of Interest

Authors declare that there is no conflict of interest.

7. References

1. Kariuki S, Gordon MA, Feasey N, Parry CM. Antimicrobial resistance and management of invasive Salmonella disease. Elsevier.

2. Feasey NA, Masesa C, Jassi C. Three Epidemics of Invasive Multidrug-Resistant Salmonella Bloodstream Infection in Blantyre, Malawi, 1998–2014. academic.oup.com.

3. Van Camp RO, Bhimji SS. Vaccine, Typhoid. StatPearls. Treasure Island (FL): StatPearls Publishing; 2018.

4. Domenico ED, Cavallo I, Pontone M, Toma L. Biofilm producing Salmonella typhi: chronic colonization and development of gallbladder cancer. mdpi.com.

5. Threlfall EJ. Antimicrobial drug resistance in Salmonella: problems and perspectives in food- and water-borne infections. academic.oup.com.

6. Piddock LJV. Fluoroquinolone resistance in Salmonella serovars isolated from humans and food animals1. Wiley Online Library.

7. Emiola A, George J, Andrews SS. A Complete Pathway Model for Lipid A Biosynthesis in Escherichia coli. PLoS One. 2014;10(4):e0121216.

8. Zhou P, Zhao J. Structure, inhibition, and regulation of essential lipid A enzymes. Elsevier.

9. Takayama K, Qureshi N, Mascagni P. Complete structure of lipid A obtained from the lipopolysaccharides of the heptoseless mutant of Salmonella typhimurium. ASBMB.

10. Raetz CRH, Guan Z, Ingram BO, DA Six, Song F. Discovery of new biosynthetic pathways: the lipid A story. ASBMB.

11. Jackman JE, Fierke CA, Tumey LN, Pirrung M. Antibacterial agents that target lipid a biosynthesis in gram-negative bacteria inhibition of diverse UDP-3-O-(R-3-hydroxymyristoyl)-N-acetylglucosamine deacetylases. ASBMB.

12. Zhou P, Barb AW. Mechanism and inhibition of LpxC: an essential zinc-dependent deacetylase of bacterial lipid A synthesis. ingentaconnect.com.

13. Gronow S, Brade H. Invited review: Lipopolysaccharide biosynthesis: which steps do bacteria need to survive? journals.sagepub.com.

14. Coggins BE, McClerren AL, Jiang L, Li X, Rudolph J. Refined Solution Structure of the LpxC–TU-514 Complex and pKa Analysis of an Active Site Histidine:@ Insights into the Mechanism and Inhibitor Design,. ACS Publications.

15. Gennadios HA, DA Whittington, Li X, Fierke CA. Mechanistic Inferences from the Binding of Ligands to LpxC, a Metal-Dependent Deacetylase,. ACS Publications.

16. Clayton GM, Klein DJ, Rickert KW, Patel SB. Structure of the bacterial deacetylase LpxC bound to the nucleotide reaction product reveals mechanisms of oxyanion stabilization and proton transfer. ASBMB.

17. Mochalkin I, Knafels JD, Lightle S. Crystal structure of LpxC from Pseudomonas aeruginosa complexed with the potent BB‐78485 inhibitor. Wiley Online Library.

18. Mdluli KE, Witte PR, Kline T, Barb AW. Molecular validation of LpxC as an antibacterial drug target in Pseudomonas aeruginosa. Am Soc Microbiol.

19. Lee C-J, Liang X, Chen X, Zeng D, Joo SH, Chung HS, et al. Species-specific and inhibitor-dependent conformations of LpxC: implications for antibiotic design. Chem Biol. 2011 Jan 28;18(1):38–47.

20. Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC Bioinformatics. 2007 Jan 5;8:4.

21. Dimitrov I, Bangov I, Flower DR, Doytchinova I. AllerTOP v.2 – a server for in silico prediction of allergens. J Mol Model. 2014 Jun;20(6):2278.

22. The UniProt Consortium. UniProt: the universal protein knowledgebase. Nucleic Acids Res. 2017 Jan 4;45(D1):D158–D169.

23. Sievers F, Higgins DG. Clustal Omega for making accurate alignments of many protein sequences. Protein Sci. 2018 Jan;27(1):135-145.

24. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol. 2011 Oct 11;7:539.

25. Webb B, Sali A. Protein structure modeling with MODELLER. Protein Structure Prediction. 2014;

26. Fiser A, Sali A. Modeller: generation and refinement of homology-based protein structure models. Meth Enzymol. 2003;374:461-491.

27. Sali A, Potterton L, Yuan F, van Vlijmen H, Karplus M. Evaluation of comparative protein modeling by MODELLER. Proteins. 1995 Nov; 23(3):318–326.

28. Jones DT. Protein secondary structure prediction based on position-specific scoring matrices. J Mol Biol. 1999 Sep 17;292(2):195–202.

29. Lovell SC, Davis IW, Arendall WB, de Bakker PIW, Word JM, Prisant MG, et al. Structure validation by Calpha geometry: phi,psi and Cbeta deviation. Proteins. 2003 Feb 15;50(3):437–450.

30. Carugo O, Djinovic-Carugo K. Half a century of Ramachandran plots. Acta Crystallogr Sect D, Biol Crystallogr. 2013 Aug;69(Pt 8):1333–1341.

31. Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res. 2007 Jul;35(Web Server issue):W407–10.

32. Colovos C, Yeates TO. Verification of protein structures: patterns of nonbonded atomic interactions. Protein Sci. 1993 Sep;2(9):1511–1519.

33. Eisenberg D, Lüthy R, Bowie JU. VERIFY3D: assessment of protein models with three-dimensional profiles. Meth Enzymol. 1997;277:396–404.

34. Laskowski RA, MacArthur MW, Moss DS. PROCHECK: a program to check the stereochemical quality of protein structures. scripts.iucr.org.

35. Laskowski RA, MacArthur MW, Smith DK, Jones DT. PROCHECK v. 3.0. Program to Check the Stereochemistry Quality of Protein Structures. Operating Instructions.

36. Krieger E, Vriend G. New ways to boost molecular dynamics simulations. J Comput Chem. 2015 May 15;36(13):996–1007.

37. Krieger E, Joo K, Lee J, Lee J, Raman S, Thompson J, et al. Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Four approaches that performed well in CASP8. Proteins. 2009;77 Suppl 9:114–122.

38. Cornell WD, Cieplak P, Bayly CI, Gould IR. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. ACS Publications.

39. Krieger E, Nielsen JE, Spronk C. Fast empirical pKa prediction by Ewald summation. Elsevier.

40. Cheatham T, Miller JL, Fox T. Molecular dynamics simulations on solvated biomolecular systems: the particle mesh Ewald method leads to stable trajectories of DNA, RNA, and proteins. ACS Publications.

41. Essmann U, Perera L, Berkowitz ML. A smooth particle mesh Ewald method. aip.scitation.org.

42. Mier P, Pérez-Pulido AJ, Andrade-Navarro MA. Automated selection of homologs to track the evolutionary history of proteins. BMC Bioinformatics. 2018 Nov 19;19(1):431.

43. Mier P, Andrade-Navarro MA. FastaHerder2: Four Ways to Research Protein Function and Evolution with Clustering and Clustered Databases. J Comput Biol. 2016 Apr; 23(4):270–278.

44. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. Genome Res. 2004 Jun;14(6):1188–1190.

45. Ramachandran GN, Ramakrishnan C, Sasisekharan V. Stereochemistry of polypeptide chain configurations. J Mol Biol. 1963 Jul;7:95–99.

46. Dallakyan S, Olson AJ. Small-molecule library screening by docking with PyRx. Methods Mol Biol. 2015;1263:243–250.

47. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem. 2010 Jan 30;31(2):455–461.

48. Biovia D. BIOVIA Discovery Studio 2017 R2: A comprehensive predictive science application for the Life Sciences.

49. Banerjee P, Eckert AO, Schrey AK, Preissner R. ProTox-II: a webserver for the prediction of toxicity of chemicals. Nucleic Acids Res. 2018 Jul 2;46(W1):W257-W263.

❖ ❖ ❖