_____

Molecular Characterization of SipC and SopD Proteins Involved in Salmonella Pathogenicity Andinteraction Studies with Potential Inhibitors.

Anuradha C H*[1], Shambhu M G[1], Anitha P.M[1], Kusum Paul[1]
1. Department of Biotechnology, the Oxford College of Engineering, Bangalore-560100
*[1]Corresponding Author: *anukailash.09@gmail.com*

Abstract :- Salmonella typhi succeeds in causing diseases like typhoid and paratyphoid in humans and animals. The risk of the diseases is based on food and water contamination that will increases the risk of morbidity and mortality. The proteins such as SipC a *Salmonella* effector known to nucleate actin helps in infection entry in to host cell and SopDEffector protein promote bacterial survival in host tissues and alter host cell physiology.There are several bacterial membrane proteins from outer, inner and secretary layer proteins may shows pathogenicity. The secretary protein is localized with bacterial membranes and causes several types of bacterial infections and component proteins and these proteins highly conserved in different pathological conditions. The SP1 component protein is highly involved in different signaling pathways which are potentially used as drug targets. In the present study identify and characterized signaling molecules interplay between bacteria and host cells; molecular interaction, sequence assembly and structural organization with secreted proteins need to understand. 3D structure of the two proteins SipC and SopD was generated using Phre2 and to design new lead molecules based on Pharmacodynamics and QSAR analysis to control the disease progression and pathogenicity.

Keywords:*Salmonella Typhi, SipC, SopD, QSAR analysis, novel drug molecules of typhoid.*
_____****_____

I.   INTRODUCTION

Typhoid infection is a systemic disease caused by *salmonella typhi*, it is gram negative bacteria which infects humans and animals. 17 million people worldwide affecting Typhoid fever every year with approximate deaths of 600,000 [1].Typhoid fever in the developed world is due to sanitation and chlorination of drinking water. Due to inhabitable condition the problems such as poor sanitation, poverty, and waris plagued in developing world; thereforetyphoid fever is able to causes 5% or more deaths in the developing world [1-2].The most common source of typhoid is Polluted water.Through food and water contaminated by the feces and urine of patients and carriersTyphoid fever is transmitted [2].

SipC is such a *Salmonella* effector known to nucleate actin, translocatetype III effectorsandbundle F-actin. Amino acids 221–260 and 381–409 of full-length SipC were indispensable for its actin binding and bundling activities [3].Effector proteins SopD promote bacterial survival in host tissues and function such as alters host cell physiology. Contributes to replication in macrophages. Plays a role, cooperatively with SopB, in membrane fission and macropinosome formation during invasion[4].

In present study, we described secretary proteins such as SipC (*Salmonella* invasion protein) and SopD(*Salmonella* outer protein) involved in pathogenesis, using bioinformatics methods to predict structures followed by active site prediction and new lead compounds were used for interaction studies, the molecule combrestatinA_3 and CCRIS5440 has strong interaction hence these compounds are used as potential inhibitors.

II.   MATERIALS AND METHODS

Data Collection
Salmonella typhimurium proteins from UniProt database around 2,221 were found, by using advance search against *salmonella* infection screened out to 50 reviewed proteins, among 50 proteins with the help of literature evidence and pathway analysis SipC and SopD are mainly observed in disease progression and pathogenicity, since no structural information from pdb database for the abovetwo proteins,SipC and SopD proteins were selected for, sequence assembly structural analysis and further molecular docking .

Tertiary structure prediction
Salmonella invasion protein (SipC) and*Salmonella* outer protein(SopD)have been taken for this work. SipC and SopD protein sequences were collected from UniProt database. Ab intio method carried out using phyre2 server, for predicting the 3D structures of SipC and SopD proteins. Model built using phyre2 server, for the SipC protein template pdb id c4kxfF_.8CHAIN F Confidence 96.65%, identity 14% and resolution 3.20A^0 similarly for the SopD protein template pdb id c3cxbA_.1A CHAIN Confidence 99.85%, identity 23% and resolution 2.60 A^0 proteins from RCSB PDB database.

_____

Structural validation and Active site Prediction
Generated protein models have beenanalyzed usingRampage (Ramachandran plot).The stereo chemical quality of protein models are analyzed residue by residue geometry using rampage.To determine the compatibility of 3D structure alpha, beta, loop, polar and nonpolar amino acid arrangements the overall quality of all amino acids is placed inside the allowed region. Structural pockets, active sites cavities and ligand binding sites are predicted by using CASTp server. Based on size and shape of 3D structure of protein and ligand fit in active site to serve chemical modification and conformational change in protein structure [5]. The more number of pockets of protein complex represent more active site and ligand binding amino acids is present in overall protein complex.

Ligand identification
Using literature search shows Catechin, Ciprofloxacin and Oxadiazole compounds are strongly inhibit *salmonella typhi* bacterial growth[6-8]. Using Virtual screening methods to find the similar compounds with different molecular descriptor studies from Pubchem compound database shows, the Catechin has 102 similar compounds, using drug likeliness property of rule of 5 shows 22 compounds, using pharmacological actions against *salmonella* inhibitor shows only 3 compounds and these compounds are used as potential lead compounds. Another compound such as Ciprofloxacin has 90 compounds of which only 15 compounds is used for drug likeliness property. Out of these compounds 4 compounds is potentially used as anti-bacterial agents. Oxadiazole is another anti-bacterial drug with similar compounds shows 26 compounds of which 8 compounds alone accepted for drug likeliness property.Only 2 compounds alone shows antibacterial drug and is potentially used as drug candidates. The molecular optimization to study molecular descriptors using Hyperchem 7.5 professional. The pre-optimization of compounds using force field calculation to predict energy minimization based on single point and geometry optimization. The QSAR properties of these compounds to calculate the molecular descriptors against different force fields. The QSAR property of force fields include Surface area (SAP), surface area in grid (SAG), molecular volume (MV), hydration energy (HE), polarizability (MP), refractivity (MR), LogP and molecular weight (MW). The surface area increases the grid also increase that show increase refractivity of compound relatively high.

Molecular Docking
Using Auto Dock 4.2 molecular docking software helps to dock protein with ligand molecule. Two different algorithms are used such as Auto grid to calculate the active site amino acids present in flexible and rigid molecules. We fixed the grid around flexible molecule of X, Y and Z-axis at 80, 80, 80 to calculate the force field energy present in between active site pocket.

III. RESULTS

Salmonella SP1 component proteins as targeted proteins such as SipC and SopD were selected from UniProt database using sequence similarity result shows that SipC has 409 residues of 25% identity and 30% coverage, were as SopD has 317 residues of 34% identity and 12% coverage.The overall results were predicted in table.1

Table: 1. Sequence similarity search of SipC and SopD protein sequences.

Protein	Template	Max. score	e values	Query coverage	Identity	Accession
SipC	Structure of the transforming growth factor-Beta Neutralizing Antibody Gc-1008	28.9	6.5	30%	25%	3Eo0 B
	Structure of the Fab fragment of Gc-1008	28.9	6.8	30%	25%	3Eo1 B
	Crystal structure of the Tqfb2 in complex with Gc2008	28.9	9.4	30%	25%	4KXZ H
SopD	Anqstrom crystal structure of BontAI IN COMPLEX WITH NTNHA	32.3	0.60	12%	34%	3V0A B
	Solution structure of the C2 Domain from human Pi3-kinase P110 Subunit Alpha	28.9	4.0	27%	28%	2ENQ A
	Crystal structure of Sifa and Skip	29.6	4.0	31%	25%	3CXB A

Primary structure prediction

Primary protein sequences were used to understand the physico-chemical properties. The instability index predicts the stability of protein. The aliphatic index (AI) of a protein of relative volume occupied by aliphatic side chain. The positive factor of all protein peptide chains shows positive values that shows increase in thermostability of globular protein. The GRAVY value of proteins shows hydropathy values of amino acids. The physico-chemical properties of SipC and SopDproteins were predicted using PROTPARAM. The overall results were predicted in table:2.

_____

Table: 2. Physico-chemical property of primary protein sequences using Protparam

Protein	Mol.Wt	PI	Total No.of atoms	Ext. Coefficient	Instability Index	GRAVY
SipC	42983.3	8.84	6052	4470	31.31 (Stable)	-0.308
SopD	36141.2	5.71	5039	30285	45.78 (unstable)	-0.361

Secondary structure prediction
The secondary structure of protein was predicted using SOPMA. In whole protein sequenceto understand alpha-helix, beta sheets and random coils arranges, the overall results wererepresented in table: 3.

Table: 3. Secondary structure prediction of primary amino acid sequence using SOPMA.

| Protein Name	Alpha helix(Hh)	Extended strand(Ee)	Beta Turn(Tt)	Random coil(Cc)
SipC	247 is 60.39%	48 is 11.74%	29 is 7.09%	85 is 20.78%
SopD	156 is 49.21%	50 is 15.77%	21 is 6.62%	90 is 28.39%

Tertiary structure prediction

The tertiary structure for the candidate protein was modeled using Phyre2, whichare used to build the 3D structures of target and template sequences. Phyre2 is based on Ab intio method, which gives the result based on CATH and SCOP of secondary structure classification.Whereas Phyre2 uses a profile-profile alignment algorithm along with Phyre2 uses the alignment of hidden Markov models via HHsearchto significantly improve accuracy of alignment and detection rate.

Table: 4. 3D structure of protein structures were predicted using Phyre2.

Target	Template	Identity	Confidence	Query coverage
SipC	C4pkiG	14%	98.3%	2-109
	D1io0a	19%	97.8%	24-166
	C1pvA	10%	97.8	24-166
	C4kxfF	14%	96.6%	192
SopD	C3cxbA	23%	99.9%	27-114
	C4i15B	17%	48.8%	126-203
	C1siyA	22%	40.7%	239-262

Spdb viewer were used to remodel the structures obtained from phyre2.For SipC protein template i.d C4kxfF since because aa GLY191 and THR54 were outside the allowed region, made inserted into the allowed region, were as SopD protein template i.d C3cxbA GLY69 and ARG45were inserted in to the allowed region, since tomake protein structure stable and complex.
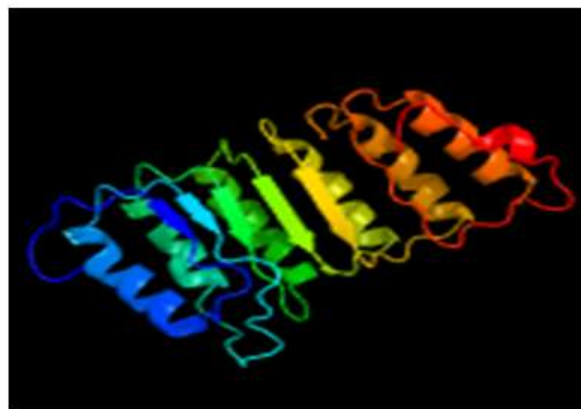
SipC(c4kxfF_.8)

SopD(c3cxbA_.1)

Figure: 1. 3D structure built using Phyre2 technique

_____

_____

RAMPAGE helps to identify stereochemical action of residue-to residue geometry. The overall results shows SipC has 92.1% of quality, 174 residues in favored region,11 residues in allowed region and 4 residues in outlier region, whereasSopD has 93.0% of quality,80 residues in favored region,6 residues in allowed region and no residues in the outlier region.
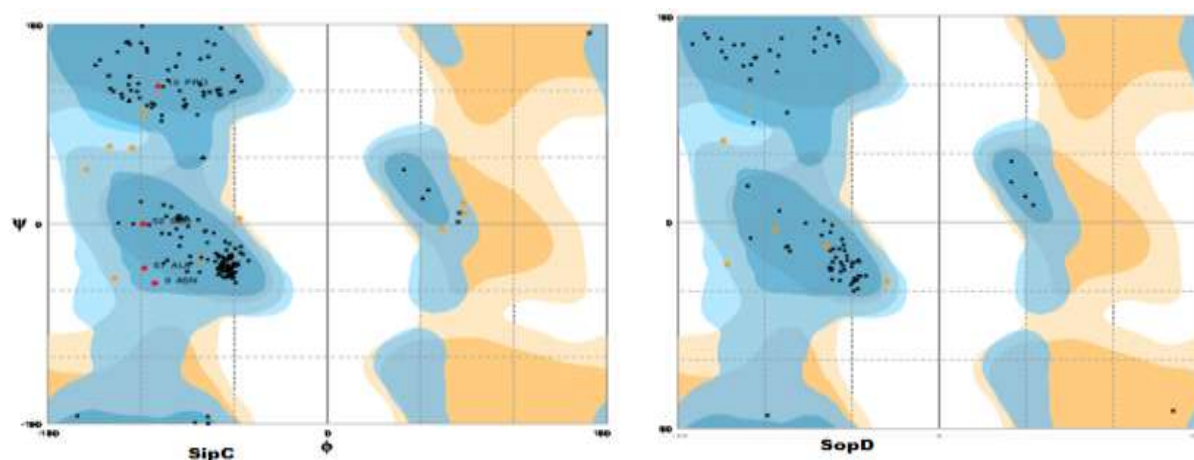

Figure: 2. Validation for SipC and SopD using Rampage

Active site prediction

Structural pockets, active sites cavities and ligand binding sites are predicted by using CASTp
server. The SipC has 28 pockets contains more number of active site and ligand binding amino acids occupied in large surface area (1872.3 $Å^2$) and highest surface volume (2826.8$Å^3$). The SopD has only 20 pockets with SA of 154.3 $Å^2$ and SV161 $Å^3$ in structure prediction and homology modeling; the overall results were predicted in table: 5.
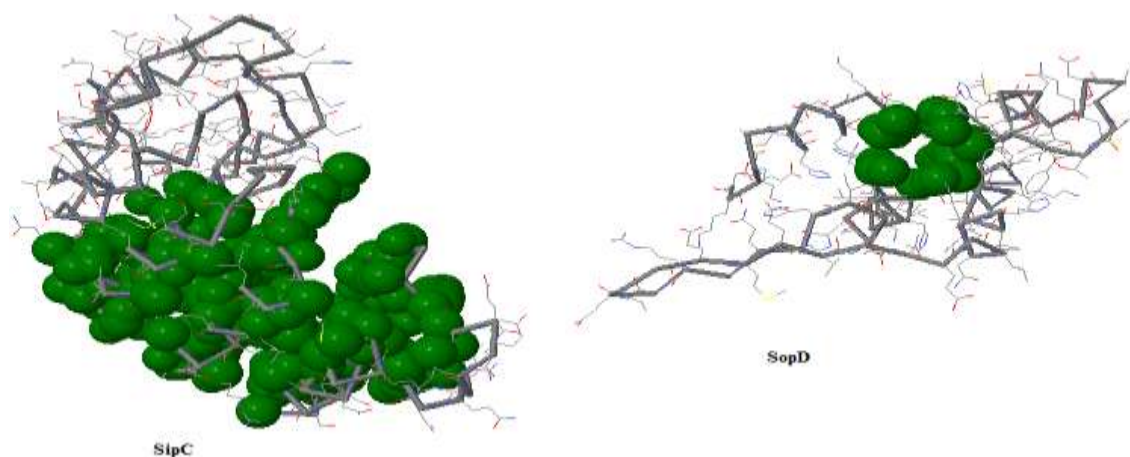

Figure: 3. Actve site pockets for SipC and SopD proteins using CastP.

Table: 5. Active site and ligand binding site prediction for SipC and SopD using CASTp server.

Name of the protein	No.of Pockets	Solvent accessible surface area	Molecular surface volume	Active Site amino acids
SipC	28	1872.3	2826.8	PHE73,MET83,ASP86,LEU87,ALA91,ASN93,VAL94,THR95,ASN99,VAL101,LYS115,PHE116,SER124,VAL127,ALA128,LEU130,MET139,GLN142,SER152.
SopD	20	154.3	161	TRP37,VAL40,HIS43,PHE44,ALA52,LEU56,PHE99,VAL114.

Ligand Identification

The Pharmacological property of lead compounds was selected from Pubchem compound database. Molinspiration was used to predict Pharmacophore analysis to test molecular descriptors of drug likeliness property and the results were shown in table: 6. the

_____

combretastatin and ciloxan compounds have same drug like properties but changes in conformational groups of OH at C12 position.
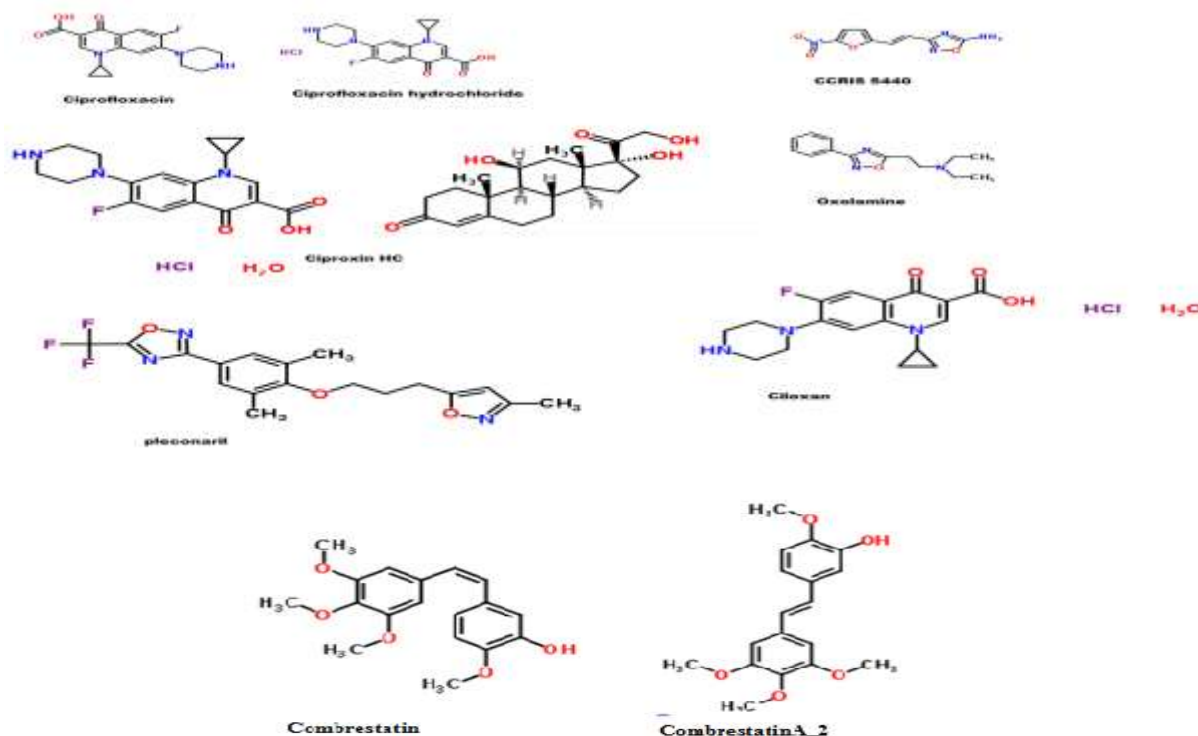

Figure: 4. Lead molecules from pubchem database.

Table: 6. to understand Pharmacophore properties of lead compounds of Lipinski rule of 5 using molinspiration

Ligand	miLogp	TPSA	natoms	MW	nON	nOHNH	nrotb	volume
Combretastatin	1.9	77.392	24.0	334.368	6	2	7	307.3
Combrestatin_A2	1.9	77.392	24.0	334.368	6	2	7	307.3
Combrestatin_A3	1.9	77.392	24.0	334.368	6	2	7	307.3
Ciproxin HC	1.617	94.826	26.0	362.466	5	3	2	343.063
Ciloxan	-0.701	74.569	24.0	331.347	6	2	3	285.46
Ciprofloxacin hydrochloride	-0.701	74.569	24.0	331.347	6	2	3	285.46
Ciprofloxacin	1.617	94.826	26.0	362.466	5	3	2	343.063
CCRIS 5440	0.942	123.911	16.0	222.16	8	2	3	172.315
Oxolamine	3.007	42.162	18.0	245.326	4	0	6	241.82
Pleconaril	4.771	74.19	27.0	381.354	6	0	7	317.898

Pharmacophore analyses of selected lead molecules were screened using QSAR properties. The best confirmations of compound are selected based on aromatic rings (R), hydrogen donors (D) and one positive ionsable group (P). There are seven physico chemical properties of lead compounds such as highest surface area (SAP) and high surface grid (SAG) increases the refractivity, it the refractivity increases the hydration energy also increases. Using Energy minimization of compounds using Semi-empirical calculation shows the electrostatic interaction of HOMO and LUMO properties were represented in table: 7.

Table: 7: Semi-empirical calculation to predict electrostatic energy surface and QSAR analysis of lead compounds.

Ligands	HOMO	LUMO	SAP (A^2)	SAG (A^2)	MV (A^3)	HE (Kcal/mol)	cLogP	MR A^3	MP A^3	MW amu
Combretastatin	+15.689	-8.370	683.62	646.59	1071.52	-19.69	-2.91	97.29	34.93	334.37
Combrestatin_A2	+4.846	-8.771	677.96	639.21	1063.57	-19.41	-2.91	97.29	34.93	334.37
Combrestatin_A3	+4.144	-8.762	683.62	646.59	1071.52	-19.69	-2.91	97.29	34.93	334.37

Ciloxan	+6.356	-3.518	328.36	512.87	858.39	-6.85	-1.85	90.21	32.88	331.35
Ciprofloxacin	+9.219	-12.081	298.67	518.22	872.01	-5.27	-1.85	90.21	32.88	331.35
Ciprofloxacin hydrochloride	+9.593	-2.628	328.36	512.87	858.39	-5.22	-1.85	90.21	32.88	331.35
Ciproxin HC	+9.219	-12.081	298.67	518.22	872.01	-6.85	-1.85	90.21	32.88	331.35
CCRIS 5440	+3.711	-2.506	365.50	423.30	637.88	-25.17	-0.72	53.96	19.66	222.16
Oxolamine	+6.625	-4.851	556.49	539.25	866.75	-7.27	2.49	76.53	28.20	245.32
Pleconaril	+6.554	-4.190	647.43	649.41	1066.50	-	3.02	96.33	35.15	381.35

Molecular Docking

The docking of bioactive reference compounds combretastatin, Ciprofloxacin and Oxolamine along with their similar chemical structures were used for molecular docking against SipC and SopD protein structures.

SipC protein docking with lead molecules

The SipC protein is strong interaction with combretastatinA_3 compound by forming 6 hydrogen bonds of docking score-30.7455 kcal/mol within active site amino acids such as Lys45, Lys75,His17, Ser18, phe116. Other reference compound such asCombrestatin is formed 3 hydrogen bonds within active site amino acid such as lys207,lle78 and Ser181, the predicted ligand such as Ciproxin HC is bound with SipC active site amino acid by forming 3 hydrogen bonds within active site amino acid Lle187, Lys75, His17, Ser18, Phe116. Representation shown in figure 5.

Table 8: Molecular Docking of SopD protein with selected lead molecules using Auto Dock 4.2.

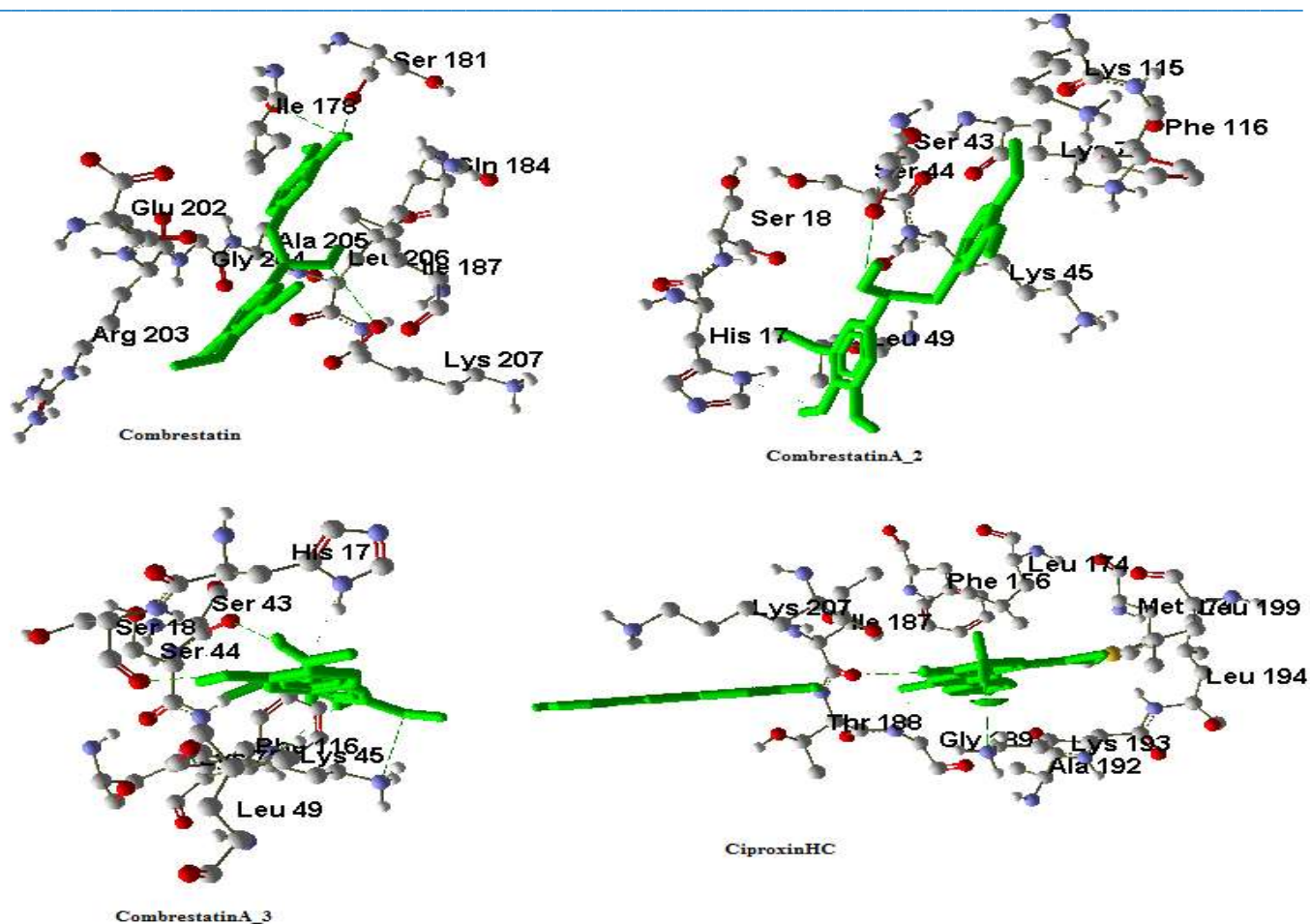| Ligand	No. H Bonds	Docking Score	Amino acid
Combrestatin	3	-15.1686	Ser181,lle178,Lys207
CombrestatinA_2	3	-19.1359	His17,Ser43,Lys75
CombrestatinA_3	6	-30.7455	Lys45,Lys75,His17,Ser18,Phe116
Ciproxin HC	3	-28.0933	Lle187,Gly189,Lys193

Figure: 5. Molecular docking poses of SipC protein with selected lead molecules.

SopD protein docking with lead molecules

The SopD protein is strong interaction with CCRIS 5440 compound by forming 7 hydrogen bonds of docking score -19.902 kcal/mol within active site amino acids such as Lys49, Ala46,Glu47, phe44. Other reference compound such asCombrestatinA_3 is formed 3 hydrogen bonds within active site amino acid such as Arg86 and Ser58, the predicted lead molecule such as Combrestatin is bound with SopD active site amino acid by forming 2 hydrogen bonds within active site amino acid Lle32 and His33.

Table 9: Molecular Docking of SopD protein with selected lead molecules using Auto Dock 4.2.

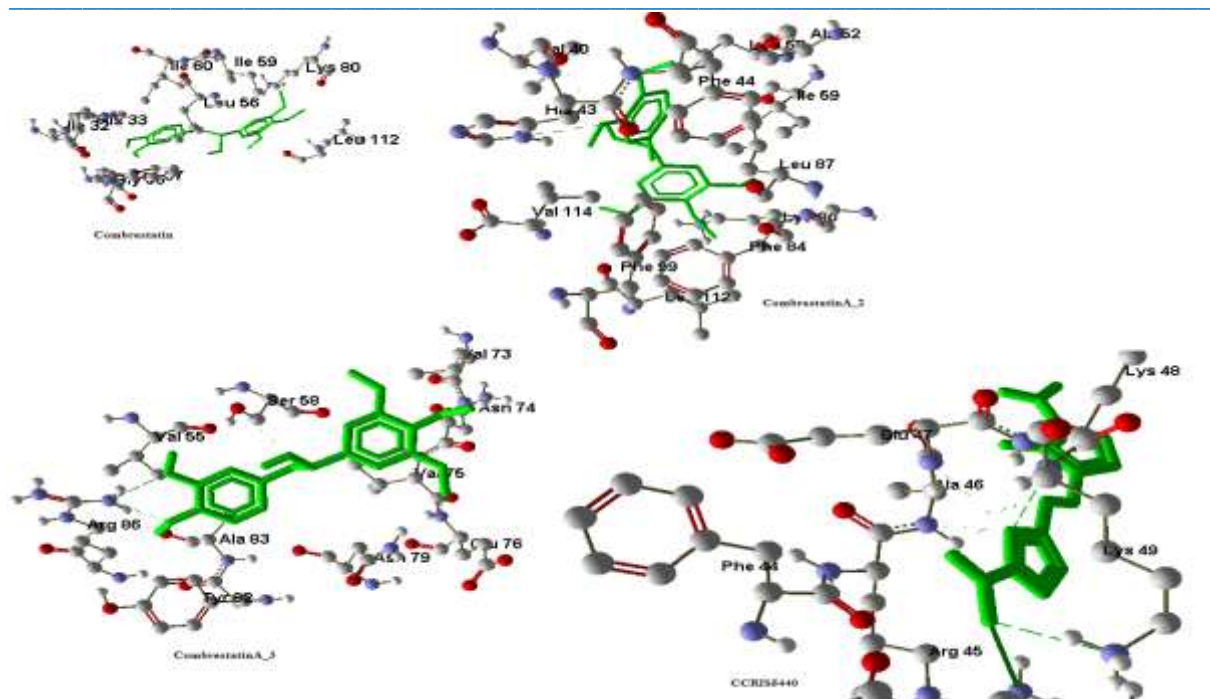| Ligand	No. H Bonds	Docking Score	Amino acid
Combrestatin	2	-5.38874	Lle32,His33
CombrestatinA-2	2	-13.7574	His43,Val40
CombrestatinA-3	3	-15.7901	Ser58,Arg86
CCRIS 5440	7	-19.902	Lys49,Phe44,Ala46,Glu47,Lys48

Figure: 6. Molecular docking poses of SopDprotein with selected lead molecules.

IV.   CONCLUSION

In this study, to understand the physicochemical properties of target proteins along with, secondary and tertiary structures of SipC and SopD proteins were predicted, thenproperties of chemical structures along with disease target proteins using modeling, Pharmacophore analysis, QSAR and molecular docking. The reference compounds such as Combrestatin, Ciprofloxacin and Oxolamine is more effective leadcompounds against *salmonella typhi* infection. But using molecular docking study against SipC and SopD protein structures with reference compounds, Combrestatin and CCRIS 5440 is more effective drug with the selected target protein with strong interaction within active site amino acids by forming 2-7 hydrogen bonds. Only with SipC protein alone is forming 6 hydrogen bonds within target amino acids. The other similar compounds such as CCRIS 5440 of Oxolamine determinant is more effective against all membrane proteins and is more predictive target drug against target proteins, the results of this drug interaction within active site amino acids by forming 2-6 hydrogen bonds with all the protein structures with strong interaction energy. A QSAR property showscombrestatin, ciprofloxacin, CCRIS5440 and Ciproxin HC compounds is good *salmonella typhi* membrane protein inhibitors against typhoid fever. Furthermore ADMET, bioavailability and toxicity can be analyzed.

V.   CONFLICT TO INTERESTS

The authors declare that there is no conflict of interest regarding the publication of this paper.

VI.   ACKNOWLEDGEMENTS

REFERENCES

[1]   J. Parkhill,et,.al "Complete genome sequence of amultiple drug resistant *Salmonella enteric serovarTyphi* CT18" NATURE ,VOL 413: 848-852(2001).

[2]   Groisman EA (1997) "How *Salmonella* became a pathogen". Trends Microbiol 5:343–349

[3]   Sebenzile K. Myeni and Daoguo Zhou1 "The C Terminus of SipC Binds and Bundles F-actin to Promote *Salmonella* Invasion" ,April 30, 2010, THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 285, NO. 18, pp. 13357–13363.

[4] Jiang X., Rossanese O.W., Brown N.F., Kujat-Choy S., Galan J.E., Finlay B.B., Brumell J.H. "The related effector proteins SopD and SopD2 from *Salmonella entericaserovar Typhimurium* contribute to virulence during systemic infection of mice".Mol. Microbiol. 54:1186-1198(2004).

[5] Chowdhury MR1, Bhuiyan MI2, Saha A2, Mosleh IM2, Mondol S2, Ahmed CM3. (2014). "Identification and analysis of potential targets in *Streptococcus*sanguinis using computer aided protein data analysis". AdvApplBioinform Chem;7:45-54.

[6] Muto S1, Fujita K, Yamazaki Y, Kamataki T. (2001). "Inhibition by green tea catechins of metabolic activation of procarcinogens by human cytochrome P450". Mutat Res; 479(1-2):197-206.

[7] T. Elavarasan, D. Bhakiarajand and M. Gopalakrishnan (2014)."Antimicrobial screening and molecular docking studies of some novel triazoloquinazolinonederivatives".DerPharmaChemica, 2014, 6(2):391-400.

[8] TomášGichner, GerassimosVoutsinas, Alexandra Patrineli, Andreas Kappas, Michael J. Plewa (1994)."Antimutagenicity of three isomers of aminobenzoic acid in *Salmonella Typhimurium*". Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis 09/1994; 309(2):201-10.

[9] ChristopherWilliams ,Edouard E. Galyov , and Stefan Bagby, (2004)."Solution Structure, Backbone Dynamics, and Interaction with Cdc42 of *Salmonella* Guanine Nucleotide Exchange Factor SopE2", Biochemistry, 2004, 43 (38), pp 11998–12008.

[10] 10. Manuela Raffatellu, R. Paul Wilson, Daniela Chessa, Helene Andrews-Polymenis, Quynh T. Tran, Sara Lawhon, SangeetaKhare, L. Garry Adams and Andreas J. Bäumler, (2005). "SipA, SopA, SopB, SopD, and SopE2 Contribute to *Salmonella enterica Serotype Typhimurium*Invasion of Epithelial Cells", Infect. Immun. January 2005 vol. 73 no. 1 146-154.

[11] Boguski MS1, McCormick F. (1993). "Proteins regulating Ras and its relatives".Nature.1993 Dec 16; 366(6456):643-54.

[12] Doughari, J. H, Elmahmood, A. M. and Nggada, H. P. "Retrospective study on the antibiotic resistant pattern of *Salmonella typhi*from some clinical samples".African Journal of Microbiology Research pp. 033-036, August, 2007