Evaluation of novel drug candidates targeting Glycosyl Hydrolase 25 Related Invasion Protein (GHIP) of *Streptococcus pneumonia:* Insights from Molecular docking and MM-GBSA free energy calculations

Subbiah Thamotharan and Arunachalam Jothi *

Department of Bioinformatics, School of Chemical and Biotechnology, SASTRA University, Thanjavur-613 401, INDIA

Abstract

Infectious disease caused by *Streptococcus pneumonia* is one of the major threats among bacterial infections, despite advances in antibiotics based therapy and vaccine programs. It is necessary to identify new drugs to combat pneumococcal infections. Here we target the Glycosyl Hydrolase 25 Related Invasion Protein (GHIP) of *Streptococcus pneumonia*, which is involved in host cell invasion and integration. We have identified twelve lead compounds based on a high throughput virtual screening approach. The ADME analysis of the identified lead compounds also showed good pharmacological properties. The binding free energies were also calculated using MM-GBSA approach for the docked complexes which showed that the identified lead compounds had similar binding affinities.

Keywords   GHIP; *Streptococcus pneumonia;* MM-GBSA; Virtual Screening; Molecular Docking.

1. INTRODUCTION

The most common cause of bacterial pneumonia is *Streptococcus pneumonia,* a gram positive bacterium, which was isolated in 1881 by Louis Pasteur and George Sternberg [1-2]. Apart from pneumonia the above mentioned organism also causes a plethora of infectious diseases such as meningitis, conjunctivitis, bacteraemia, otitis media etc. Infections related to *S. pneumonia* are responsible for the death of one million children every year, which is much more than the mortality rate of malaria, AIDS and measles combined [3-4]. Besides being recognised as a leading child killer, pneumococcal infections are also prevalent in adults [3]. Moreover, Invasive pneumococcal disease (IPD) causes significant mortality in patients with HIV infections [4]. Strategies to cure pneumococcal infections becomes further complicated with the trend of increasing drug resistant and multi drug resistant strains of *S. pneumonia* [5].

To combat pneumococcal infections, several countries have introduced vaccination strategies using multivalent pneumococcal conjugate vaccines such as PCV7, PCV10, PCV13 etc. [3, 5]. However there are reported cases of failures of such vaccines and fatalities due to invasive pneumococcal disease [6]. Due to the mismatch in the distributions of various pneumococcal serotypes in different countries, it is difficult to follow a common strategy. For example, PCV7 is completely unsuitable and PCV10 performs poorly in some countries of West Africa because of the difference in the serotypes prevalent in those countries. Similar difference has been observed in American and European serotypes as well [7-8]. Owing to the variation in the distribution of serotypes, Temime *et al* [3] predict that the effect of vaccination observed today may not sustain for a long term and vaccination alone cannot be a successful strategy to control pneumococcal infections. Moreover, vaccination is not an effective way to control pneumococcal infections in immune compromised situations such as AIDS. All these problems clearly signify the need for the identification of novel drugs to control pneumococcal infections.

The advances in modern computational methods allow rapid screening of thousands of small molecules to identify potential drug candidates for any given disease, which could further be tested experimentally for pharmacological activities. In this study, we use a high throughput structure based virtual screening approach to find potential inhibitors against GHIP, a virulence factor of *S.pneumonia.* GHIP is a Glycosyl Hydrolase 25 Related Invasion Protein which is involved in host cell invasion. Knockout mutation of GHIP gene strongly reduces the *S.pneumonia* virulence [9]. Moreover GHIP protein homologs are absent in *Homo sapiens* which makes *S.pneumonia* GHIP an ideal target for drug discovery and suitable inhibitors for GHIP can be effective drug candidates. The three dimensional structure of *S.pneumonia* GHIP has been solved recently by X ray crystallography at 1.8 Å resolution [9], which was used for our virtual screening. The binding affinities of the lead compounds identified by virtual screening procedure were further analysed by estimating the free energies of the docked ligand receptor complexes.

2. MATERIALS AND METHODS

The detailed procedure of high throughput virtual screening (HTVS) approach is explained as a flowchart in F**ig. 1**. The possible lead compounds were identified through screening the Maybridge HitFinder™ collection (http://www.maybridge.com/). The above mentioned collection comprises of 14,400 premier compounds, which are a non-redundant representation of the drug-like diversity of the Maybridge ligand libraries (~ 56,000 compounds). All these screening compounds were filtered based on Lipinski's rule of five for drug likeness such as – partition coefficient logP being lesser than or equal to 5, Hydrogen bond donors being lesser than or equal to 5, Hydrogen bond acceptors being lesser than or equal to 10 and molecular weight being lesser than 500 Dalton [10].

After filtering, a total of 12,201 compounds were subjected to ligand preparation using the ligprep module of Schrödinger suite. As explained in Fig.1 every step of the screening procedure selected 10% of the compounds which was the input for the subsequent step, yielding 12 lead molecules in the final step. Ligand preparation is a process in which the ligands are subjected to addition of appropriate hydrogen atoms, de salting of metal ions and energy minimization using OPLS 2005 force field.

The *S.pneumonia* GHIP structure (PDB ID 4FF5) used for our study was downloaded from the Protein Data Bank (www.rcsb.org). The structure was optimized and prepared for docking using Protein preparation wizard tool of Schrödinger suite. The process of protein preparation involved addition of missing hydrogen atoms, removal of water molecules and energy minimization using OPLS 2005 force field. The active site of *S.pneumonia* GHIP has been studied in detail [9]. The key residues in the active site were Asp 33, Ser 35, Ser 61, Tyr 98, Tyr 100, Asp 131, Glu 133, Asp 134, Tyr 162, Tyr 186, Asp 189, Ser 210, Asp 220 and Asp 222. It should be noted that there was a discrepancy in residue numbering between the deposited structure in the protein data bank and active site residues mentioned by Niu *et al.* [9]. We have followed the residue numbering based on the deposited structure (PDB ID: 4FF5). The active site was predicted using Sitemap of Schrödinger suite and the predicted active site overlapped exactly with the reported active site residues. The receptor grid was generated based on the predicted active site and the docking studies were performed using the Glide module of Schrödinger suite. The Glide program allows conformational flexibility in ligands while keeping the receptor rigid. The unlikely binding modes were filtered based on a grid based force field evaluation, which considers the rigid body and torsional movements of the ligand. The final models were evaluated using XP Glide scoring function and model energy score (*Emodel*) that combines Glide score, the non-bonded interaction energy and the excess internal energy of the generated ligand conformation is used to choose the best ligand conformation [11].

The binding free energies of the docked complexes were calculated using Prime MM-GBSA (Molecular Mechanics with Generalized Born Surface Area) module of Schrödinger suite [12-13]. The output post-viewer files generated by Glide XP docking protocol were provided as the input to Prime/MM-GBSA for calculating the binding free energies. In order to keep the binding site flexible, the residues within 4 Å from the bound ligand were treated as flexible, while calculating the free energies. The OPLS-2005 force field and GB/SA continuum solvent model were used to estimate energies of the docked complexes. The Prime/MM-GBSA module estimates binding free energies ($\Delta G_{binding}$) by using the following formula

$$\Delta G_{binding} = E_{R:L} - (E_R + E_L) + \Delta G_{solv} + \Delta G_{SA}$$

Where $E_{R:L}$ is the energy of the complex, $E_R + E_L$ is sum of energies of the protein and ligand molecule in unbound state, ΔG_{solv} is the difference in the GBSA solvation energy of the complex and sum total of solvation energies of unbound protein and ligand and ΔG_{SA} is the difference in surface area energies of the complex and sum total of surface area energies of the unbound protein and ligand [14]. The absorption, distribution, metabolism and excretion (ADME) properties were calculated using QikProp module of Schrödinger suite (QikProp, version 3.5. Schrödinger, LLC, New York, NY, 2012).

3. RESULTS AND DISCUSSION

The HTVS docking approach used by us identified a total of 12 lead compounds from a collection of 14,400 compounds which in turn is a representation of approximately 56,000 compounds. Though the identification of the top 12 compounds were based on glide XP score, the potential ligands were subjected to other evaluations such as the MM-GBSA based binding free energy and ADME properties. We have also identified key residues and the corresponding interactions made by them with the ligands. All our evaluations show that the identified 12 compounds are potential candidates for inhibition of *S.pneumonia* GHIP virulence factor. The details of our analysis are given below.

3.1. Overall description of the docked complexes

The Maybridge ligand Ids and the corresponding IUPAC names of the potential inhibitors identified by our screening procedure are given in Table 1. The chemical structures of the 12 compounds are shown in Fig. 2. The glide XP score of the obtained lead compounds were in the range of -8.2 kcal/mol to -5.2 kcal/mol and the glide Emodel scores were in the range of -33.7 kcal/mol to – 56.8 kcal/mol. The details of the XP score and Emodel energy values are depicted in Table 2. The docking pose of the top two compounds (according to glide XP scores) is shown in Fig. 3. The top scoring compound was found to be **PD00612**. However, as mentioned earlier the results were not interpreted in terms of docking scores alone and several identified compounds had interesting pharmacological features. For example, the compound **RJC01223** also known as Clomipramine is an antidepressant drug approved by the FDI and is available in the market. Moreover, Clompiramine was found to be an active inhibitor in 90 different bioassays and is suggested to be a potential drug candidate against *Salmonella typhi*, *Plasmodium falciparum*, Marburg Virus, Human colon cancer and prostate cancer (https://pubchem.ncbi.nlm.nih.gov/summary/summary.cgi?cid=68539). Similarly, the compound KM08436 was found to be active against aldehyde dehydrogenase 1 (https://pubchem.ncbi.nlm.nih.gov/summary/summary.cgi?sid=103011302&viewopt=PubChem).

All the identified 12 compounds showed favourable interactions with the receptor protein *S.pneumonia* GHIP virulence factor, which is shown in Fig. 4. The predominant interaction between the ligands and the receptor was found to be hydrogen bonding interaction. The complete details of the hydrogen bonding geometry of protein-ligand interactions are listed in Table 3. The key hydrogen bond forming amino acid residues of the GHIP virulence factor were Glu133, Q207, Asp131, Tyr162, Tyr186, Arg58 and Asp222 which made six, six, four, two, two and two

hydrogen bonding interactions respectively among the 12 identified compounds. Interestingly, **Asp 131 and Glu 133** are important conserved active site residues among various *Streptococcus* GHIP homologs. Apart from hydrogen bonding interactions, the binding of three identified ligands were also stabilized by π-π stacking interactions. The residues Tyr 98, Trp 182 and Tyr 186 made π-π interactions with the compound **S01517**. Similarly, Tyr 186 was also found to make π-π stacking interaction with the compound **SEW02675**. The compound **HTS04925** was found to make a similar stacking interaction with Tyr 162.

Table 1. Potential inhibitors and their corresponding IUPAC names.

S.No.	Compund ID	IUPAC Name
1	PD00612	1-methyl-4-[(4-methylpiperazino)(2-thienyl)methyl]piperazine
2	SEW02675	N-[2-[2-hydroxy-3-[[4-(trifluoromethyl)phenyl]methylamino]propoxy] phenyl] acetamide
3	BTB12226	1-Carbazol-9-yl-3-dimethylamino-propan-2-ol
4	GK00487	3-[(2,2,2-trifluoroacetyl)amino]thiophene-2-carboxamide
5	RJC01223	3-(3-Chloro-10,11-dihydro-5H-dibenzo[b,f]azepin-5-yl)-N,N-dimethylpropan-1- amine
6	KM08436	N1-(4-chlorophenyl)-2-({5-[3-(dimethylamino)prop-1-ynyl]-3-pyridyl}carbonyl)hydrazine-1-carboxamide
7	S01517	N-(2-furylmethylidene)-(4-{[(2-furylmethylidene)amino]methyl}cyclohexyl)methanamine
8	SCR00967	N-(5-cyclopropyl-2-methylpyrazol-3-yl)-2-(1,3-dioxoisoindol-2-yl)acetamide
9	JFD02837	2-[[6-[(1-hydroxy-3,3-dimethylbutan-2-yl)iminomethyl]pyridin-2-yl]methylideneamino]-3,3-dimethylbutan-1-ol
10	SPB08437	ethyl 1-[[3-(furan-2-carbonylamino)phenyl]methyl]piperidine-4-carboxylate
11	MWP01096	1-(3,4-dihydroisoquinolin-1-ylamino)-3-propylurea
12	HTS04925	2-[4-oxo-2-[4-[5-(trifluoromethyl)pyridin-2-yl]oxyphenyl]-1,3-thiazol-5-yl]acetic acid

Table 2. Glide XP score and Emodel energy values for the identified inhibitors.

S.No.	Compound ID	XP glide score (kcal/mol)	Glide Emodel energy (kcal/mol)
1	PD00612	-8.05	-33.72
2	SEW02675	-6.40	-56.81
3	BTB12226	-6.34	-39.54
4	GK00487	-7.01	-55.22
5	RJC01223	-5.90	-36.26
6	KM08436	-5.78	-44.77
7	S01517	-5.73	-52.38
8	SCR00967	-5.52	-48.56
9	JFD02837	-5.51	-43.85
10	SPB08437	-5.33	-55.75
11	MWP01096	-5.25	-39.31
12	HTS04925	-5.20	-47.24
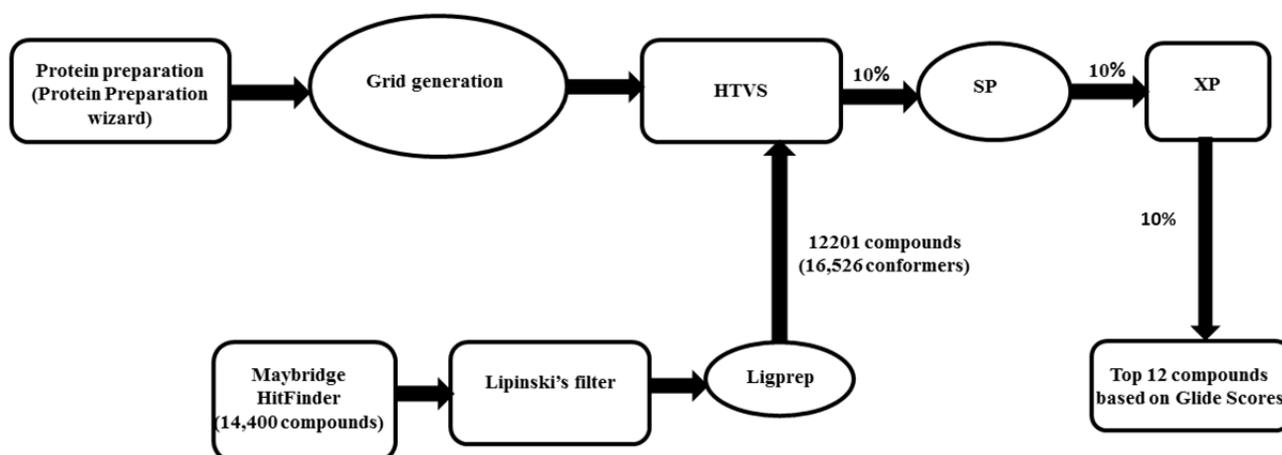
Fig. 1. Flowchart depicting the scheme of HTVS procedure employed, where SP and XP refer to single precision and extra precision respectively.

Table 3. Hydrogen-bonding interactions in the protein-ligand complexes where 'd' and 'a' represent distance in Å and angle in degrees respectively.

S. No.	Ligand	Hydrogen bonding interaction (Protein…ligand)	d(H…A)	d(D…A)	a(D–H…A)
1	PD00612	131 Asn OD1…H–N 186 Tyr H–O…H–N	2.29 2.10	3.03 3.07	129.1 158.2
2	SEW02675	98 Tyr O–H…O 162 Tyr H–O…H–N	1.95 1.76	2.72 2.74	137.4 162.7
3	BTB12226	207 Gln NE2–H…O 222 Asp OD1…H–O	1.85 2.34	2.78 3.01	152.9 127.7
4	KM08436	131 Asp OD1…H–N 133 Glu OE1…H–N 133 Glu OE1…H–N	2.41 2.22 1.82	3.10 3.03 2.80	124.8 136.1 158.1
5	GK00487	131 Asn OD1…H–N 186 Tyr H–O… H–N 207 Gln NE2–H…O 207 Gln NE2–H …O	1.55 2.37 2.02 1.91	2.53 3.32 2.91 2.77	162.4 156.8 145.2 140.6
6	RJC01223	131 Asn OD1…H–N	2.47	3.19	128.0
7	S01517	133 Glu OE1…H–N 222 Asp OD1…H–O	1.52 1.76	2.56 2.72	170.0 153.8
8	SCR00967	38 Gln NE2–H…O 207 Gln NE2–H…N	1.98 2.10	2.90 3.08	151.1 164.8
9	JFD02837	162 Tyr H–O…O–H 210 Ser O…H–O	2.04 1.99	2.96 2.87	163.4 155.0
10	SPB08437	37 Trp NE1–H…O 58 Arg NH1–H…O 133 Glu OE1…H–N 207 Gln NE2–H…O	2.00 2.20 1.92 1.93	2.98 3.03 2.81 2.78	162.5 139.3 144.2 140.8
11	MWP01096	133 Glu OE1…H–N 133 Glu OE2…H–N 207 Gln NE2–H…O	1.60 1.75 2.02	2.64 2.76 3.01	173.7 163.2 164.3
12	HTS04925	37 Trp NE1–H…N 58 Arg NH1–H…N 61 Ser OG–H…O	2.03 2.16 2.03	2.99 3.04 2.87	157.6 145.4 148.5

Table 4. Prime MM-GBSA free energy values for the potential ligands.

S.No.	Compound ID	dG_Bind(kcal/mol)	dG_Bind(NS) (kcal/mol)	Lig_Strain_Energy (kcal/mol)	Rec_Strain_Energy (kcal/mol)
1	PD00612	-55.00	-64.14	5.26	3.87
2	SEW02675	-54.95	-74.91	10.93	9.03
3	BTB12226	-46.20	-53.90	2.35	5.35
4	GK00487	-65.94	-75.94	4.37	5.62
5	RJC01223	-28.13	-36.99	1.21	7.65
6	KM08436	-50.98	-58.11	2.61	4.51
7	S01517	-64.26	-74.59	4.88	5.45
8	SCR00967	-60.96	-70.37	3.31	6.10
9	JFD02837	-58.68	-73.19	10.02	4.48
10	SPB08437	-52.82	-66.12	7.43	5.86
11	MWP01096	-57.78	-71.16	5.64	7.74
12	HTS04925	-48.62	-60.59	4.09	7.87

Table 5. Calculated ADME values for all the 12 identified ligands.

Compound ID	QplogPo/w	QPlogHERG	QPPCaco (nm/sec)	QPPMDCK (nm/sec)	%QP
PD00612	-0.1	-6.8	116.1	89.1	63.1
SEW02675	3.4	-7.1	334.7	731.5	92.3
BTB12226	3.2	-6.0	1192.5	662	100
KM08436	2.4	-6.2	39.6	70.6	69.9
GK00487	1.3	-3.6	536.7	2033.2	83.6
RJC01223	4.6	-6.1	2202.3	3097.3	100
S01517	4.5	-6.1	6352	3649.2	100
SCR00967	2.4	-5.7	403.5	185.5	88.1
JFD02837	3.1	-5.1	1030.4	510.9	100
SPB08437	2.6	-5.2	599.7	314.9	91.9
MWP01096	1.7	-3.8	708	583.1	88.2
HTS04925	3.1	-3.9	35.9	104.9	73.2
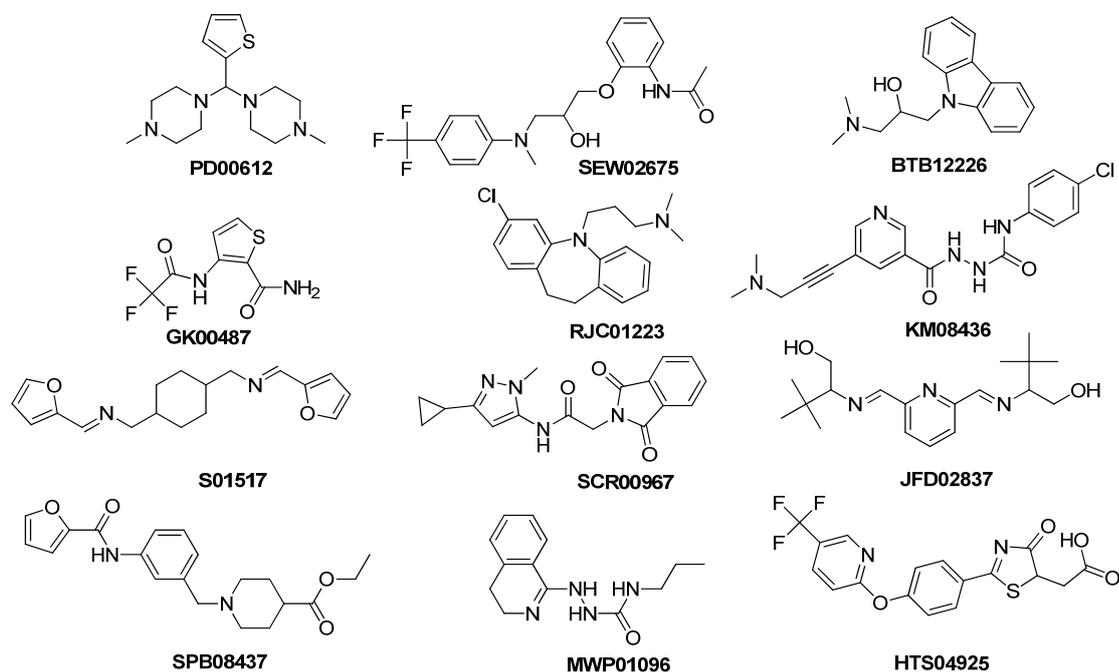
Fig. 2. Chemical structures of the identified potential lead compounds along with Maybridge HitFinder database ID's.
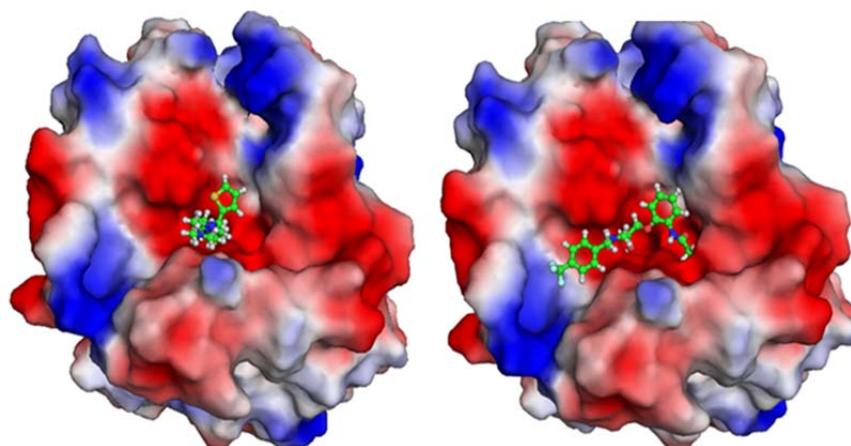
Fig. 3. The docked poses of compounds PD00612 (left) and SEW02675 (right), which were identified to be top 2 compounds based on glide XP score. The ligands are shown on ball and stick representation and the receptor is in charged surface representation.

Fig. 4. Representation of the interactions between the ligand molecules and the receptor. The hydrogen bond interactions are shown in pink colored arrows and π-π interactions are shown in green colored arrows. (A) PD00612 (B) SEW02675 (C) BTB12226 (D) KM08436 (E) GK00487 (F) RJC01223 (G) S01517 (H) SCR00967 (I) JFD02837 (J) SPB08437 (K) MWP01096 (L) HTS04925.

3.2. Free energy calculation of the docked complexes using Prime MM-GBSA

The binding free energies of the protein-ligand complexes were estimated to further consolidate the ability of the identified ligands to bind with the target protein. Along with the binding site residues, the water molecules present in the binding site also play a crucial role in protein-ligand recognition. This recognition could either be by displacement upon ligand binding or by forming water bridges and thereby stabilizing the complex [15]. However, it is difficult to rigorously treat explicit binding-site waters, which requires to completely sample ensembles of water molecules and to consider the free energy cost of replacing waters. The MM-GBSA approach is a computationally efficient method, which employs molecular mechanics, generalized Born model and solvent accessibility method to elicit free energy [16].

The observed binding free energy values of all the 12 complexes are given in Table 4, along with the ligand and receptor strain energy. It could be seen that, apart from the compound **GK00487** (binding free energy: -28.13 kcal/mol), all the other complexes had nearly equal binding free energies, suggesting equal binding affinities. The ligand and receptor strain energies were also found to be lower and equivalent except for the compounds **SEW02675** and **JFD02837**. However, their corresponding observed binding free energy of these compounds were found to be -54.9 and -58.6 kcal/mol, respectively, suggesting compensation by other favourable energy terms. It should also be noted that the compound **KM08436** which was ranked fourth according to glide XP score was predicted to have the highest binding affinity among the 12 compounds with a binding free energy value of -65.9 kcal/mol.

3.3. ADME Screening

The drug like abilities of the identified inhibitors were further emphasized by analyzing their ADME properties which are presented in Table 5. Further description and discussion of the obtained results are given below.

3.3.1. Comparison of logP (o/w)

Partition coefficient (log P) is used to predict the hydrophobicity and hydrophilicity of a drug in the body, where (o/w) represents octanol/water [17]. For an ideal drug, the logP value should be within a range of -2.0 to 6.5. If a compound has high logP value, it refers to high hydrophobicity and if it is less, the compound is highly water soluble. A compound should have both hydrophobic and hydrophilic properties in equal proportion in order to reach the target site. Here, all the obtained ligands are within a range of -0.1 and 4.6 and thus have equal probability of reaching the target site.

3.3.2. Comparison of log HERG

HERG refers to Human Ether-à-go-go-Related Gene which codes for potassium ion channel. This is best known for its contribution to the electrical activity of the heart. The log HERG value gives the predicted IC_{50} value for blockage of HERG K+ channels. There is a risk of sudden death when this channel's ability to conduct electrical current across the cell membrane is inhibited or compromised [18]. Generally, it is a concern if the value is lesser than -5. The values obtained for the identified 12 compounds were in the range of -3.6 to -7.1. Additionally, it should also be noted that the compound **RJC01223** which is an already FDI approved drug and available in the market has a value of -6.1.

3.3.3. Comparison of Caco and Oral Absorption Capability

Caco-2 cells are human epithelial colorectal adeno carcinoma cells. Pharmaceutical industries use Caco-2 monolayers as an in vitro model of the human small intestinal mucosa to predict the absorption of orally administered drugs. They are a model for the gut-blood barrier. These predictions are only for non-active transport. It is considered that values of Caco-2 cell permeability below 25 nm/sec are poor and values above 100 are better [19]. It could be noted from the Table 5 that all the lead molecules have a good oral absorption rate and in particular, the compounds **BTB12226**, **RJC01223**, **S01517** and **JFD02837** have 100% oral absorption rate.

3.3.4. Comparison of MDCK

MDCK stands for Madin-Darby Canine Kidney Cells. It helps to gain a greater understanding of the mechanism of drug efflux and highlights early potential issues with drug permeability. MDCK cells are considered to be a good mimic for the blood-brain barrier [20]. A value lesser than 25 nm/sec is considered to be poor and values greater than 500 is considered to be very good. It could be noted that all the 12 compounds obtained by our screening procedure had admissible values.

Finally, it has to be stated that all the identified 12 compounds fits well within the Lipinski's rule of five, which states that an orally active drug has no more than one violation of the rules such as not having more than 5 hydrogen bond donors, not having more than 10 hydrogen bond acceptors, having a molecular mass less than 500 Daltons and octanol-water partition coefficient (log P) being not greater than 5 [10]. This substantiates that the obtained lead compounds by our screening study have potential pharmacological properties and could be used for further experimental phases in drug discovery.

4. CONCLUSION

In this study, a high throughput structure based virtual screening approach was used to find potential inhibitors against GHIP a virulence factor of *S.pneumonia*, which is a Glycosyl Hydrolase 25 Related Invasion Protein which is involved in host cell invasion. Twelve potential lead compounds were identified by the screening procedure. The favourable ADME properties confirm the drug likeness of the identified compounds. Estimation of binding free energies of the protein-ligand complexes using Prime MM-GBSA calculation showed that the identified compounds have similar binding affinity towards the receptor. The conserved Asp 131 and Glu 133 residues located at the active site made consistent hydrogen bonding interactions with the ligands. One of the lead compounds (**RJC01223** - Clompiramine) identified in the present study was also found to be a potential drug candidate against *Salmonella typhi*, *Plasmodium falciparum*, Marburg Virus, Human colon cancer and prostate cancer.

REFERENCES

[1] Sternberg, G., A fatal form of septicaemia in the rabbit, produced by the subcutaneous injection of human saliva. *Annual Reports of the National Board of Health* **1881,** *3*, 87-108.

[2] Pasteur, L.; Chamberland, M.; Roux, Sur une maladie nouvelle, provoquke par la salive d' un enfant mart de la rage. *Compt Rend Acad d Sci* **1881,** *92*, 159-165.

[3] Verma, R.; Khanna, P., Pneumococcal conjugate vaccine: A newer vaccine available in India. *Human Vaccines & Immunotherapeutics,* **2012,** *8* (9), 1317-1320.

[4] Siemieniuk, R.; Gregson, D.; Gill, M., The persisting burden of invasive pneumococcal disease in HIV patients: an observational cohort study. *BMC Infect Dis,* **2011,** *11* (1), 1-8.

[5] Donkor, E. S.; Dayie, N. T. K. D.; Badoe, E. V., Vaccination against pneumococcus in West Africa: perspectives and prospects. *Int J General Med,* **2013,** *6*, 757-764.

[6] Ladhani, S. N.; Slack, M. P. E.; Andrews, N. J.; Waight, P. A.; Borrow, R.; Miller, E., Invasive Pneumococcal Disease after Routine Pneumococcal Conjugate Vaccination in Children, England and Wales. *Emerging Infectious Diseases,* **2013,** *19* (1), 61-68.

[7] Kries RV, S. A., Schmitt HJ, Rene R Proportion of Invasive Pneumococcal Infections in German Children Preventable by pneumococcal Conjugate Vaccines. *Clinical Infectious Diseases* **2000,** *31*, 482–487.

[8] BG, S.; BM, G., Prevention of pneumococcal disease by vaccination: does serotype replacement matter? *Lancet,* **2000,** *356*, 1210-1211.

[9] Niu, S.; Luo, M.; Tang, J.; Zhou, H.; Zhang, Y.; Min, X.; Cai, X.; Zhang, W.; Xu, W.; Li, D.; Ding, J.; Hu, Y.; Wang, D.; Huang, A.; Yin, Y.; Wang, D., Structural Basis of the Novel <italic>S. pneumoniae</italic> Virulence Factor, GHIP, a Glycosyl Hydrolase 25 Participating in Host-Cell Invasion. *PLoS ONE,* **2013,** *8* (7), e68647.

[10] Lipinski, C. A.; Lombardo F Fau - Dominy, B. W.; Dominy Bw Fau - Feeney, P. J.; Feeney, P. J., Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev,* **2001,** *46*, 3-26.

[11] Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T., Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein−Ligand Complexes. *J Med Chem,* **2006,** *49* (21), 6177-6196.

[12] Jacobson, M. P.; Pincus, D. L.; Rapp, C. S.; Day, T. J. F.; Honig, B.; Shaw, D. E.; Friesner, R. A., A hierarchical approach to all-atom protein loop prediction. *Proteins: Structure, Function, and Bioinformatics,* **2004,** *55* (2), 351-367.

[13] Jacobson, M. P.; Friesner, R. A.; Xiang, Z.; Honig, B., On the Role of the Crystal Environment in Determining Protein Side-chain Conformations. *J Mol Biol,* **2002,** *320* (3), 597-608.

[14] Lyne, P. D.; Lamb Ml Fau - Saeh, J. C.; Saeh, J. C., Accurate prediction of the relative potencies of members of a series of kinase inhibitors using molecular docking and MM-GBSA scoring. *J Med Chem,* **2006,** *49*, 4805-4808.

[15] Sun, H.; Zhao, L.; Peng, S.; Huang, N., Incorporating replacement free energy of binding-site waters in molecular docking. *Proteins: Structure, Function, and Bioinformatics,* **2014,** *82* (9), 1765-1776.

[16] Dill, K. A., Additivity Principles in Biochemistry. *J Biol Chem,* **1997,** *272* (2), 701-704.

[17] Ghose, A.; Crippen, G., Atomic physicochemical parameters for three-dimensional-structure-directed quantitative structure-activity relationships. 2. Modeling dispersive and hydrophobic interactions. *J Chem Inf Comput Sci,* **1987,** *27*, 21-35.

[18] Cavalli, A.; Poluzzi, E.; De Ponti, F.; Recanatini, M., Toward a Pharmacophore for Drugs Inducing the Long QT Syndrome: Insights from a CoMFA Study of HERG K+ Channel Blockers. *J Med Chem,* **2002,** *45* (18), 3844-3853.

[19] Yazdanian, M.; Glynn Sl Fau - Wright, J. L.; Wright Jl Fau - Hawi, A.; Hawi, A., Correlating partitioning and caco-2 cell permeability of structurally diverse small molecular weight compounds. *Pharm Res, 15*, 1490-1494.

[20] Irvine, J. D.; Takahashi L Fau - Lockhart, K.; Lockhart K Fau - Cheong, J.; Cheong J Fau - Tolan, J. W.; Tolan Jw Fau - Selick, H. E.; Selick He Fau - Grove, J. R.; Grove, J. R., MDCK (Madin-Darby canine kidney) cells: A tool for membrane permeability screening. *J Pharm Sci, 88*, 28-33.