

A novel approach of interstrain docking in accelerating the process of lead identification in Pneumonia

Balasankar Karavadi*, M.Xavier Suresh

Department of Bioinformatics, Sathyabama University, Chennai-INDIA 600119

*Corresponding author:Email: balasankar.sathyabamauniv@gmail.com, xaviersuresh@gmail.com

ABSTRACT

Streptococcus pneumoniae is a human respiratory pathogen that causes serious mucosal and invasive diseases. HUNGARY19A-6, D39, TIGR4, G54, CGSP14, TCH8431-19A strains of *Streptococcus pneumoniae* are extremely virulent strains that cause pneumococcal pathogenesis which contains genes synthesizing proteins. In this work we have identified some proteins from the above strain and performed the homology modeling studies of the virulent protein which is synthesized by the genes and validated the nature of the receptor as a future drug target for the above strains of *Streptococcus pneumoniae* using modeller 9v7.

We have also identified specific ligands for the above mentioned proteins by using the virtual structure based ligand screening approach. Protein-ligand complexes have been analyzed by docking studies using Discovery Studio and the best ligand for each protein is given in below table. Protein-ligand interactions have also been visualized along with the validation of pharmacokinetic descriptors. The best docked analog with individual protein from each strain was used for further interstrain multiple docking, from inter strain docking we analyzed that few analogs had high scores with other strains of *Streptococcus pneumoniae*.

Keywords: *Streptococcus pneumoniae*, Docking, HUNGARY19A, D39, TIGR4, G54, CGSP14 and TCH8431.

INTRODUCTION

Pneumonia and *Streptococcus pneumoniae*: Pneumonia is a lung inflammation which is often caused by infection with bacteria, viruses or other pathogenic organisms, is one among the most significant microbes which cause bacterial disease in humans (Avery et al., 1944; Klein, 1999). *S. pneumoniae* is the most common cause of bacterial meningitis in adults, children, and dogs. The transition from commensal bacterium to opportunistic pathogen often occurs after another respiratory tract infection, e.g., pneumococcal pneumonia has been a leading secondary infection and cause of death during influenza pandemics.

Bacterial pathogens exhibit a number of virulence factors that enable them to invade and colonize the tissues of host. A number of these virulence factors are displayed on the cell surface and include adhesins that mediate attachment to host cells and toxins. Pneumolysin is an important toxin produced by almost all *S. pneumoniae* strains, extensively studied for its ability to cause damage to human tissue (Shak et al., 2013). *S. pneumoniae* has a polysaccharide capsule that acts as a virulence factor for the organism; more than 90 different serotypes are known and these types differ in virulence, prevalence and extent of drug resistance. Serotypes of *S. pneumoniae* are categorized on the structures of their exopolysaccharide capsules. The efficient degradation of host glycoproteins is integral to pneumococcal virulence (Joel et al., 2007; Tettelin et al., 2001). As part of its life cycle, pneumococcus exists as a commensal bacterium that inhabits and colonizes the nasopharynx of 50% healthy adults and children (Cardoso and Lopes, 2007; Del Toro et al., 2006). It is estimated that more than 1 million people die each year from pneumococcal infections worldwide, especially in developing countries (Orihuela, 2004; Orihuela, 2003). General vaccination with the 7-valent pneumococcal conjugate vaccine was recommended in Germany during July 2006 for children greater than 2 years (Blue, 2003; LeMessurier, 2006). In the United States and elsewhere, resistance to a range of antibiotics is increasing among clinical isolates of *S. Pneumoniae* (Nunes, 2005).

Strains of *Streptococcus pneumoniae*: In this work we have chosen the six strains namely HUNGARY19A-6, D39, TIGR4, G54, CGSP14, TCH8431-19A of *Streptococcus pneumoniae* which are extremely virulent. The genome of *S. pneumoniae* HUNGARY19A-6 virulent strain is of single chromosome with 2245615 base pairs having 39.6% of GC content. The genome of *S. pneumoniae* D39 virulent strain is of single circular chromosome with 2 million base pairs having 39.7% of GC content which codes 1914 proteins. Genome of TIGR4 contains a sequence length of 2,347,766 base pairs which includes 2106 proteins and 2302 genes with a GC content of 39.7%. The strain of G54 has length of 2,078,953 base pairs, genes of 2186, protein coding 2114, and GC content of 39%. The genome of *S. pneumoniae* CGSP14 virulent strain is of a chromosome with 2 million base pairs having 39.5% of GC content which codes 2206

proteins and 70 structural RNAs. The genome of *S. pneumoniae* TCH8431/19A virulent strain is of single chromosome with 2088772 base pairs having 39.8% of GC content.

HUNGARY19A-6, D39, TIGR4, G54, CGSP14, TCH8431-19A strains of *Streptococcus pneumoniae* are extremely virulent strains that causes pneumococcal pathogenesis which contains genes synthesizing proteins. In this work we performed the homology modeling studies of the virulent proteins which are synthesized by the genes and validated the nature of the receptor as a future drug target for the above strains of *Streptococcus pneumoniae* using modeller 9v7. The ribosomal protein L4, penicillin-binding protein 3 which are coded by rpLD and protein coding gene respectively in strain Hungary 19A-6, glycoproteins which are coded by amiD, bgaC, pyrDA, ribE in D39, hypothetical proteins coded by various coding genes in TIGR4, capsule polysaccharide biosynthesis proteins CpsB, CpsC, and CpsD of G54, polysaccharide polymerase protein and capsular polysaccharide biosynthesis protein coded by wzy and wzd respectively in CGSP14 and transposase protein and Phage encoded protein coded by protein encoding genes in TCH8431 were modeled and validated.

We have also identified specific ligands for the above mentioned proteins by using the virtual structure based ligand screening approach. Protein-ligand complexes have been analyzed by docking studies using Discovery Studio and the best ligand for each protein. Protein-ligand interactions have also been visualized along with the validation of pharmacokinetic descriptors. The best docked analog with individual protein from each strain was used for further inter strain docking, from inter strain docking we analyzed that few analogs had high scores with other strains of *Streptococcus pneumoniae*. The high score analogs were subjected to molecular dynamic study to find the stability of the protein-ligand complexes. Based on our insilico analysis we conclude from the highest dock scored and stability of protein ligand complex that the receptors and ligands can be considered as the best drug targets and drug candidates, respectively.

MATERIALS AND METHODS

Homology modeling: Homology modeling is a technique of comparative modeling which uses its similar structure as template for predicting 3D structure of a macromolecule (Mondal et al., 2010; Eswar et al., 2006). Homology modeling technique is highly utilized for generating a protein model in an *insilico* environment (Bates and Sternberg, 1999). MODELLER 9V7 is used for homology or comparative modeling of protein three-dimensional structures. MODELLER implements comparative protein structure modeling by satisfaction of spatial restraints, and can perform many additional tasks, including de novo modeling of loops in protein structures, optimization of various models of protein structure with respect to a flexibly defined objective function, multiple alignment of protein sequences or structures, clustering, searching of sequence databases and comparison of protein structures.

Molecular docking: Molecular docking is an important tool in the field of structural and molecular biology which is assisted for a computer based drug design (Roger and Hopfinger, 1994). The ultimate goal of a protein-ligand docking is to predict the binding mode of a ligand molecule (Hou, 1999). Most of the successful methods in docking search for a high dimensional space, so that the scoring function can be used in an efficient manner for ranking the exact candidate with the best pose and in certain cases docking can also be used for screening a library of compounds which is also known as virtual screening (Connolly, 1983).

RESULTS AND DISCUSSION

From the docking studies of Hungary 19A analog (6R,7S)-7-[[2-(2R)-2-hydroxy-2-phenylacetyl]amino]-8-oxo-3-[[1-(sulfomethyl)tetrazol-5-yl]sulfanylmethyl]-5-thia-1-azabicyclo[4.2.0]oct-2-ene-2-carboxylic acid is the best ligand for Ribosomal protein (B1I8J9) with the Dock score of 41.85 with 1 Hydrogen bond and analog 2-[4-[(Z)-4-[4-(4,5-dihydro-1H-imidazol-2-yl)-2-methoxyphenoxy]but-2-en-1-yl]-3-methoxyphenyl]-4,5-dihydro-1H-imidazole is the best ligand for Penicillin-binding protein 3 (B1I40) with the Dock score of 63.949 with 1 Hydrogen bond are the best ligands of the strain.

Docking studies of D39 results in analogs Dichloro (phosphono) methylphosphonic acid, 1,2,6,7,8,8a-hexahydronaphthalen-1-yl]-3,5-dihydroxyheptanoic acid, 2-(2,4-difluorophenyl)-1,3-bis(1,2,4-triazol-1-yl) propan-2-ol; 2-(2,4-difluorophenyl)-1,3-bis(1,2,4-triazol-1-yl) propan-2-ol were the three ligands interacting with maximum amino acid residues of the proteins, amiD-Oligopeptide Permease Protein; bgaC-Beta galactosidase 3; pyrDA-Dihydroorotate dehydrogenase 1; ribE-Riboflavin Synthase; alpha subunit of proteins respectively.

Based on docking studies of TIGR4 analogs Carbonic acid 4-(2-amino-purin-9-yl)-2-hydromethyl-butylester, 1-[2-(2,4-dichlorophenyl)-2-[(2,4-dichlorophenyl) methoxy]ethyl]-3-[(E)-3-phenylprop-2-enyl]imidazol-3-ium,7,8-dimethyl-10-(3,4,5-trihydroxypentyl)benzo[g] pteridine-2,4-dione were the three ligands interacting analogs with maximum dock score with the proteins SP_0372, SP_0192 and SP_0311 respectively.

For strain G54 analogs 5-methyl-N-[4-(trifluoromethyl)phenyl]-1,2-oxazole-4-carboxamide,N, 3-bis (2-chloroethyl)-2-oxo-1,3,2{5}-oxazaphosphinan-2-amine and (4-amino-1-hydroxy-1-phosphonobutyl)phosphonic acid are the best ligand for Tyrosine protein phosphatase wzh (SPG_0315), Capsular polysaccharide biosynthesis protein (SPG_0316) and Tyrosine- kinase (SPG_0317) with the Dock score of 39.76, 49.584 and 47.46 with 2 and 4 Hydrogen bond respectively.

For strain CGSP14 analogs 6R,7R)-7-[[2(Z)-2-(2-amino-1,3-thiazol-4-yl)-2-methoxyiminoacetyl]amino]-3-[(2-methyl-6-oxido-5-oxo-1,2,4-triazin-3-yl)sulfanylmethyl]-8-oxo-5-thia-1-azabicyclo[4.2.0]oct-2-ene-2-carboxylate and (4S,4aR,5S,5aR,6R,12aR)-4(dimethylamino)-1,5,10,11,12a-pentahydroxy-6-methyl-3,12-dioxo-4a,5,5a,6-tetrahydro-4H-tetracene-2-carboxamide are the two analogs with maximum dock score when it binds with Capsular polysaccharide biosynthesis protein (B2ILP4) and Polysaccharide polymerase(B2ILP9) respectively.

For strain TCH8431, analogs 5-O-ethyl 3-O-methyl-4-(2,3-dichlorophenyl)-2,6-dimethyl-1,4-dihydropyridine-3,5-dicarboxylate, 2-[(3Z)-6-fluoro-2-methyl-3[(4methylsulfinylphenyl) methylidene] inden-1-yl] acetate with Phage encoded protein (D6ZPG9) and transposase protein (D6ZS38) had Dock score of 40.7 and 69.655 with 1 Hydrogen bond and 2 hydrogen bonds respectively. ADMET descriptors were also analyzed for the drug candidates.

The best docked analog with individual protein from each strain was used for further intra strain docking, the riboflavin 7,8-dimethyl-10-(3,4,5-trihydroxypentyl) benzo[g]pteridine-2,4-dione has best dock score with D6ZS38 protein of TCH8431-A strain, B5E791 protein of G54 strain. IMIPRAMINE analog 3-(5,6-dihydrobenzo[b] [1] benza zepin-11-yl)-N,N-dimethylpropan-1-amine has best dock score with D6ZPG9 protein of TCH8431 strain and B1IB40 protein of HUNGARY 19-A strain.

The cefazolin analog ((7R)-3-[(5-methyl-1,3,4-thiadiazol-2-yl)sulfanylmethyl]-8-oxo-7-[[2-(tetrazol-1yl) acetyl] amino]-5-thia-1-azabicyclo [4.2.0] oct-2-ene-2-carboxylic acid has best dock score with Q04LD9 protein of D39 strain, B2ILP9 protein of CGSP14 strain.

The fluconazole analog 2-(2,4-difluorophenyl)-1,3-bis(1,2,4-triazol-1-yl) propanol has best dock score with Q97SI9 protein of TIGR-4 strain, B2ILP4 protein of CGSP14 strain.

From intra strain multiple docking we analyzed that few analogs had high scores with other strains of *Streptococcus pneumoniae* as shown in Table 1.

Table.1. Best scoring analogs for each strain after multiple docking of interstrain ligands

Strains	Target protein	Ligands	Best analogues	Best dock score
TCH8431-A	D6ZS38 D6ZPG9	Riboflavin	7,8-dimethyl-10-(3,4,5-trihydroxypentyl) benzo[g]pteridine-2,4-dione	64.699
		Imipramine	3-(5,6-dihydrobenzo[b] [1] benza zepin-11-yl)-N,N-dimethylpropan-1-amine.	73.898
HUNGARY 19-A	B1IB40 B1I8J9	Cefonicid	(6R,7R)-7-[[2(R)-2-hydroxy-2-phenylacetyl]amino]-8-oxo-3-[[1-(sulfomethyl)tetrazol-5-yl]sulfanylmethyl]-5-thia-1-azabicyclo [4.2.0]oct-2-ene-2-carboxylic acid.	75.45
		Cefazolin	(7R)-3-[(5-methyl-1,3,4-thiadiazol-2-yl)sulfanylmethyl]-8-oxo-7-[[2-(tetrazol-1yl) acetyl] amino]-5-thia-1-azabicyclo [4.2.0] oct-2-ene-2-carboxylic acid	31.668
D39	Q04N11 (SPD_0065) Q04LD9 (SPD_0665)	Cefonicid	(6R,7R)-7-[[2(R)-2-hydroxy-2-phenylacetyl]amino]-8-oxo-3-[[1-(sulfomethyl)tetrazol-5-yl]sulfanylmethyl]-5-thia-1-azabicyclo [4.2.0]oct-2-ene-2-carboxylic acid.	75.45
		Cefazolin	(7R)-3-[(5-methyl-1,3,4-thiadiazol-2-yl)sulfanylmethyl]-8-oxo-7-[[2-(tetrazol-1yl) acetyl] amino]-5-thia-1-azabicyclo [4.2.0] oct-2-ene-2-carboxylic acid	31.668

Table.1.Best scoring analogs for each strain after multiple docking of interstrain ligands continuation				
Strains	Target protein	Ligands	Best analogues	Best dock score
TIGR-4	Q97SI9 (SP_0372)	Fluconazole	2-(2,4-difluorophenyl)-1,3-bis(1,2,4-triazol-1-yl)propanol	39.68
G54	B5E790 (SPG_315) B5E791 (SPG_316)	Biotin	5-[(3aS,4S,6aR)-2-oxo-1,3,3a,4,6,6a-hexahydrothieno [3,4-d]imidazol-4-yl] pentanoic acid	48.207
		Riboflavin	7,8-dimethyl-10-(3,4,5-trihydroxypentyl) benzo[g]pteridine-2,4-dione	64.628
CGSP14	B2ILP4 B2ILP9	Fluconazole	2-(2,4-difluorophenyl)-1,3-bis(1,2,4-triazol-1-yl)propanol	47.119
		Cefazolin	(7R)-3-[(5-methyl-1,3,4-thiadiazol-2-yl)sulfanylmethyl]-8-oxo-7-[[2-(tetrazol-1-yl) acetyl] amino]-5-thia-1-azabicyclo [4.2.0] oct-2-ene-2-carboxylic acid	30.668

CONCLUSION

Based on our insilco analysis we conclude on the basis of highest dock score and the stability of protein ligand complexes of different strains of *Streptococcus pneumoniae* and the receptors may be considered as drug targets and the best ligands with maximum dock score may be considered as suitable leads for the pipeline of drug discovery.

ACKNOWLEDGEMENT

The bioinformatics computational facilities available at Department of Bioinformatics, Sathyabama University, and the support by the management of Sathyabama University (Dr. Marie Johnson & Dr. Mariazeena Johnson, Directors) are greatly acknowledged.

REFERENCES

- Avery. OT, MacLeod. CM, and McCarty M, Studies on the chemical nature of the substance inducing transformation of pneumococcal types Journal of Experimental Medicine, 79, 1944, 137-158.
- Bates.P.A and Sternberg M.J.E, Model building by comparison at CASP3: using expert knowledge and computer automation. Proteins, 3, 1999, 47-54.
- Blue. CE and Mitchell. TJ, Contribution of a response regulator to the virulence of *Streptococcus pneumoniae* is strain dependent, Infect. Immun, 71, 2003, 4405-4413.
- Cardoso. TC, Lopes. LM and Carneiro. AH , A case-control study on risk factors for early-onset respiratory tract infection in patients admitted in ICU, BMC Pulm. Med., 7, 2007, 12-17.
- Connolly.M.L, Solvent accessible surfaces of proteins and nucleic acids. Science, 221, 1983, 709-713.
- Del Toro.M.D, Rodriguez-Bano.J. Martinez-Matinez.L, Pascual.A, Perez-Canoa.R, Perea.E.J, Muniain.M.A, Epidemiology, clinical features and prognosis of infections due to *Stenotrophomonas maltophilia*. Enferm Infecc Clin. Microbiol, 24, 2006, 4-9.
- Eswar.N, Comparative Protein Structure Modeling With MODELLER, Current Protocols in Bioinformatics, John Wiley & Sons, 15, 2006, 5.6.1-5.6.30.
- Joel.A. Lanie., Winkler Genome Sequence of Avery's Virulent Serotype 2 Strain D39 of *Streptococcus pneumoniae* and Comparison with That of Unencapsulated Laboratory Strain R6, J Bacteriol, 189(1), 2007, 38-51
- Klein. DL, Pneumococcal disease and the role of conjugate vaccines. Microb. Drug Resist, 5, 1999, 147-157.
- LeMessurier.K.S, Ogunniyi.A.D, Paton.J.C, Differential expression of key pneumococcal virulence genes in vivo. Microbiology, 152, 2006, 305-11.
- Lindahl.E.R, Molecular dynamics simulations. Methods Mol. Biol, Vol.No: 443, 2008, 3-23.

Mondal.U.K, Sen.A and Bothra.A K, Homology modeling of the Cytolethaldistending toxin B gene of *Helicobacter hepaticus* ATCC 51449, *International Journal of Interrogative Biology*,10(1), 2010, 35-40.

Nunes S R, Trends in Drug Resistance, Serotypes, and Molecular Types of *Streptococcus pneumoniae* Colonizing Preschool-Age Children Attending Day Care Centers in Lisbon, Portugal: a Summary of 4 Years of Annual Surveillance, *J Clin Microbiol*, 43(3), 2005, 285-1293.

Orihuela. CJ, Gao KP Francis, J.Yu.E.I, Tuomanen. Tissue-specific contribution of pneumococcal virulence factors to pathogenesis. *J. Infect. Dis*, 190, 2004, 1661-1669.

Orihuela. CJ, Gao. M, McGee, J.Yu, Francis.KP, and Tuomanes.E, Organ-specific models of *Streptococcus pneumoniae* disease. *Scand. J.Infect*, 35, 2003, 647-652.

Price.S.L, Toward more accurate model intermolecular potentials for organic molecules. *Rev. Comput. Chem*, 14, 2000, 225-289.

Roger.D, and Hopfinger A.J, Application of genetic function approximation to quantitative structure–activity relationships and quantitative structure–property relationships. *J. Chem. Inf. Comput. Sci*, 34, 1994, 854-866.

Shak. JR, Ludewick. HP, Howery. KE, Sakai. F, Yi. H, Harvey. RM et.al.,Novel role for the *Streptococcus pneumoniae* toxin pneumolysin in the assembly of biofilms. *MBio*, 4(5), 2013, 655-13.

Tettelin. H, Nelson. KE, Paulsen. IT, Eisen. JA, Read. TD, Peterson. S, Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*. *Science*, 293 (5529), 2001, 498-506.