ISSN: 0975-8585

Research Journal of Pharmaceutical, Biological and Chemical Sciences

Pharmacoinformatics Analysis to Identify Inhibitors Against transposase, an Ally to Develop Antibiotic Resistance in *Streptococcus pneumonia.*

M Xavier Suresh*, and K Balasankar.

Department of Bioinformatics, Faculty of Bio and Chemical Engineering, Sathyabama University, Chennai – 600 119, Tamil Nadu, India.

ABSTRACT

Streptococcus pneumoniae is increasingly drug resistant in recent years and becoming potential threat health hazard. The availability of the complete genomes of the virulent strains paves ways to combat the disease with pharmacoinformatics approaches. Transposase is an enzyme that catalyzes the movement of the transposon to another part of the genome, and helps the organism to develop antibiotic resistance. In the present work, a combined structure and ligand based pharmacophore modeling, virtual structure based ligand screening, molecular docking and molecular dynamics approaches were employed to identify potent inhibitors of *S.pneumoniae* transposase. The phamacophore models were used to screen the chemical compounds dataset. ADME properties and the toxicity of the screened compounds were analyzed followed by molecular docking. Further molecular dynamics simulation studies were carried out on the docked complex and the analysis shows that the ligand binding is largely guided by domain movements that help the molecule bury deeply inside and compounds remain bound to the key residues of the binding site. Visualization of the transposase-ligand interactions reveals that Lys137 involved in key interactions in addition to Asp184, and Gln119. Further investigations into the antipathogenic potential of the identified compounds may open new avenues for the design more potent inhibitors.

Keywords: *Streptococcus pneumoniae*, pharmacoinformatics, molecular docking, molecular dynamics.

Corresponding author

ISSN: 0975-8585

INTRODUCTION

*Streptococcus pneumoniae (*also referred as the pneumococcus) is a gram positive bacterium, mostly forming in diplococcic and a member of the family of lactic acid bacteria. As a human pathogen, *S. pneumoniae* is the most common bacterium to cause meningitis, sepsis, pneumonia, and upper respiratory tract infections, such as acute otitis media (AOM) [1-2], and is estimated to result in over 3 million deaths of children in every year around the world. More than 50% of children acquiring at least one strain during their first year of life, although an individual may harbour multiple strains simultaneously or sequentially. Each carriage episode lasts for days to months, but by age 3 carriage prevalence steadily declines until adulthood [3-4]. Progression from carriage to disease is a relatively uncommon event, but the consequences for the host are significant [5]. A high incidence of pneumococcal infections is combined with a constantly growing antibiotic resistance of this pathogen. In recent years the antibiotic resistance in bacteria is increasing [6].

The whole genome sequencing of various *Streptococcus pneumoniae* strains have increased in recent years. *S. pneumoniae* has 37 genome groups completed till today. There are 318 strains in *S. pneumoniae* which have genome assembly and annotations. These genomes have several numbers of genes coding different proteins. Among those proteins there may be new targets for vaccine and antibiotic development. Serotype 19A has become the most common pneumococcal serotype isolated from chronic or recurrent pneumococcal sinusitis in children and is used in the evolutionary study of vaccines with higher valency. Serotype 19A isolates have high rates of antimicrobial resistance and are frequently isolated along with multiple other organisms [7]. The genome of *S. pneumoniae* TCH8431/19A virulent strain is of single chromosome with 2088772 base pairs. There are about 2275 proteins and 2355 genes coding them. The GC content is 39.8%. The total size of the genome is 2.09 mb. The entire genome summary showing genes, proteins encoded by genes etc., which were further exploited for in silico drug design purposes [8]. Recombination-mediated genetic plasticity and homologous recombination ability of *S. pneumoniae* enables it to acquire drug resistance and evade vaccine pressure and acquire new traits by taking up naked DNA from the environment and incorporating it into its genome [9].

Transposases are enzymes involved in moving a transposon, is a piece of DNA that includes several genes for antibiotic resistance, along with the gene needed to build the transposase itself. Thus blocking the function of transposase may interrupt the bacterial function. Therefore, in this work transposase protein from *S. pneumoniae* was modeled, validated and specific ligands were identified using the virtual structure based ligand screening approach. The high score analogs were subjected to further docking, and molecular dynamics studies which would enhance our understanding whereby some of the ligands can be considered as the best choice for leads/drug candidates, respectively.

MATERIALS AND METHODS

The genome of *Streptococcus pneumoniae* TCH8431*/19A* virulent strain is analyzed and the transposase protein coded by protein encoding genes is identified as a drug target. The structure of the above mentioned virulent protein is obtained using homology modeling technique. Sequence of the target protein was retrieved from UniProtKB/Swiss-Prot. Templates for the target protein was identified using PDB-sum database based on the identity of the protein sequence and the coordinates were retrieved from PDB. The instability index and the aliphatic index of the protein were predicted using PROTPARAM tool [10]. The secondary structure was predicted using the NNPREDICT server [11]. Modeller 9v7 was used for homology modeling of protein three-dimensional structure. The stereochemical quality of the models was assessed by PROCHECK which indicate the amino acids with unusual backbone conformation [12]. SAVS was used to visualize dihedral angles ψ against ϕ of amino acid residues in structure through Ramachandran plot. Similarly the non-bonded interactions between different atom types were calculated by ERRAT [13].

The ligand binding sites from the modeled structure were predicted with the help of binding pocket detection server tools such as pocket finder and Q-site finder (www.modelling.leeds.ac.uk/qsitefinder) [14-15]. Additionally, the binding pockets of the receptor were also determined by using Accelrys Discovery Studio 2 (DS). Ligands for transposase protein (Sequence ID: D6ZS38) were retrieved from DrugPort sharing more identity with related protein sequence for which already a drug exists. The best analog structures for each ligand were retrieved from PubChem [16]. In addition to this dataset a few known antibiotic compounds and their similar compounds from PubChem were also included. Virtual screening [17] (VS) is performed to identify

ISSN: 0975-8585

those structures, molecules or compounds which are most likely to bind to a drug target. Virtual screening automatically evaluates very large libraries of compounds using computer programs and mainly focuses on how to reduce the massive number of compounds into a manageable number that can be synthesized and tested in the laboratory [18].

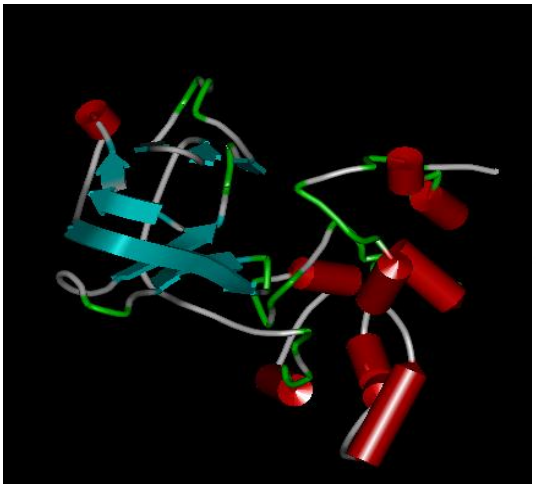
Molecular docking is a method to evaluate the feasible binding geometries of a putative ligand with a target whose target site is known. The binding geometries is often known as binding poses, includes, in principle, both the position of the ligand relative to the receptor and conformational state of the ligand and the receptor. The selected compounds were docked into the binding site of the receptor using Ligand fit protocol implemented in Accelrys Discovery Studio [19]. The ligand fit protocol of receptor-ligand interaction is run with the protein and its corresponding ligands are given as input. The docking run generated 10 poses for each of the analog compounds. The ligscore, Jain, PLP and PMF scoring functions were used to identify the best docked pose. Prior to docking CHARMm forcefield is applied to prepare both ligands and proteins [20]. The ADMET analysis for all the ligand molecules was also studied through Discovery Studio. Protein-ligand interactions have also been visualized along with the validation of pharmacokinetic descriptors. Finally simulation protocol of DS is used for molecular dynamics (MD) simulation to calculate the time dependent behavior of molecular system which provides detailed information on the conformational changes occur in the protein-ligand complexes.

RESULTS AND DISCUSSION

Structure prediction of transposase protein

The tertiary structure of *S.pneumoniae* transposase from the strain TCH8431/19A is not available in PDB. The prediction of the three dimensional structure of the transposase from *S.pneumoniae* is achieved by homology modeling. The sequence of the transposase protein (D6ZS38) was obtained from UniProtKB and homology model building was performed using Modeler9v7 with the template structure (PDB Id: 1TDH_A) which share an identity of 40%. The modeled protein was validated through SAVS. From the result it is found that 86.2% residues are present in the most favored region, 11.6% in additional allowed region, 2.1% in generously allowed region and 0% residues were present in disallowed regions of Ramachandran Plot and the modeled protein and its plot is shown in figure 1.

>tr|D6ZS38|D6ZS38_STRP0 Transposase OS=Streptococcus pneumoniae serotype A19 (strain TCH8431) GN=HMPREF0837_11199 PE=4 SV=1
MVGEILCHVQNILGKQTLKPLSRLKIKSQFKKKRKRFKTSRPFHKVRLMYQDEAGFGRIS
KLGSCWAPIGVGPHVHSHYIREFHYCYGAVDAHTGESFFLRAGGCNTEWMNAFLEELSQA
YPDDYLLLVMDNAIWHKSSTLKIPTNIGFAFIPPYTPEINPIEQVWKEIRKRGFKNKAFQ
TLEDVMNQLQDVIQGLEKEVIKSIVNRRWTRMLCESR



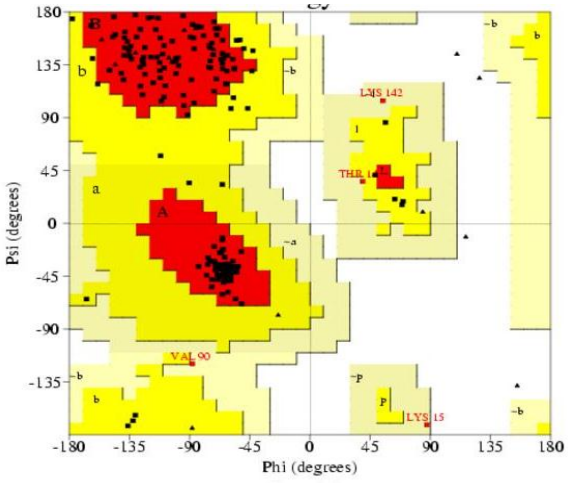


Figure 1: The sequence, structure and corresponding Ramachandran plot of the homology modeled transposase

ISSN: 0975-8585

Pharmacophore modeling

Ligands for transposase protein were retrieved from DrugPort database sharing more than 40% identity with related protein sequence for which already a drug exists. Then the structurally similar analog compounds for each ligand were searched and retrieved from PubChem database. This search resulted in to 51 drugs molecules, and more than 1523 analog compounds whose 2D and 3D structures if available were retrieved from PubChem database.

A common-feature pharmacophore model was derived with the HipHop module of catalyst for the drugs that were validated by pharmacokinetics and toxicity studies. Five kinds of features including hydrogen-bond acceptor (HBA), hydrogen-bond donor (HBD), hydrophobic group (HYD), and positive ionizable (P) and ring aromatic (R) features were selected to initiate the pharmacophore hypotheses generation process. The three most potential analogs were selected for pharmacophore modeling which is one of the most powerful methods to categorize and identify key features from a group of molecules. The pharmacophore, predicted to identify the common functional moieties for these analogs, showed a hydrogen bond donor, hydrogen bond acceptor, and a hydrophobe. The aligned molecule with the pharmacophore for each of the analogs is shown in [Figure 2]. This pharmacophore model will provide a new insight to design novel molecules that can inhibit the function of the target and will be useful in drug discovery strategies.

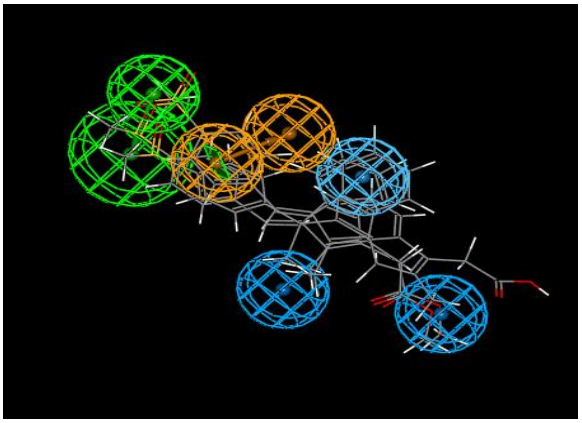


Figure 2: The alignment of the selected compounds with the generated pharmacophore model

Molecular docking and ADMET screening

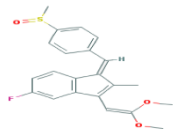
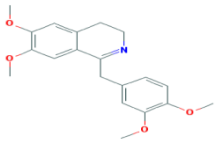
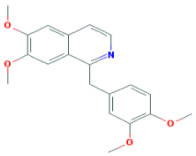
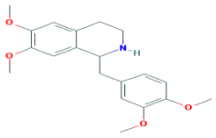
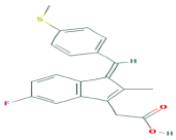
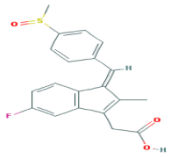
Molecular docking was performed with those analogs using Ligand fit module of the Discovery Studio software. Dock score was calculated for all the analogs. The analog compounds with docking score more than 30.0 were considered to be the best and are given in table 1 and interactions are shown in figure 3.

The best analog compounds based on docking scores are 2-[(3Z)-6-fluoro-2-methyl-3[(4methylsulfinylphenyl)methylidene]inden-1-yl]acetate,2-[(3Z)-6-fluoro-2-methyl-3 [(4Methylsulfanylphenyl) methylidene]inden-1-yl]acetic acid, 2-[(3E)-6-fluoro-2-methyl-3-[(4-methylsulfinylphenyl) methylidene] inden-1-yl]acetic acid, 21-[(3,4-dimethoxyphenyl)methyl]-6,7-dimethoxyiso quinoline, 1-[(3,4-dimethoxy phenyl) methyl]-6,7-dimethoxy-3,4-dihydroiso quinoline, 1-[(3,4-dimethoxyphenyl)methyl]-6,7-dimethoxy-1,2,3,4-tetrahydroisoquinoline, N-[6-[[(1S,3S)-3-acetyl-3,5,12-trihydroxy-10-methoxy-6,11-dioxo-2,4-dihydro-1H-tetracen-1-yl]oxy]-3-hydroxy-2-methyloxan-4-yl]formamide, which are the analogs of sulindac molecule and papaverine molecules respectively.

Based on docking studies 2-[(3Z)-6-fluoro-2-methyl-3[(4methylsulfinylphenyl) methylidene] inden-1-yl] acetate with D6ZS38 of strain TCH8431/19A with the dock score 69.655 with 2 hydrogen bond is selected as the best ligand molecule. The ligand molecule and the lining of the binding site residues surrounding it are shown in figure 3.

ISSN: 0975-8585

Table 1: The dock score for ligands obtained from PubChem for transposase protein

Ligands	Structure of ligands	Site	No of hydrogen bonds	Amino acid	Dock score
CID5372384		2	2	ASP184 LYS137	69.66
CID22176		1	1	TRP66	55.23
CID4680		1	1	ARG81	42.88
CID5418		1	1	CYS86	42.19
CID5352624		1	1	ARG81	36.81
CID1548887		1	1	ASN11	39.68

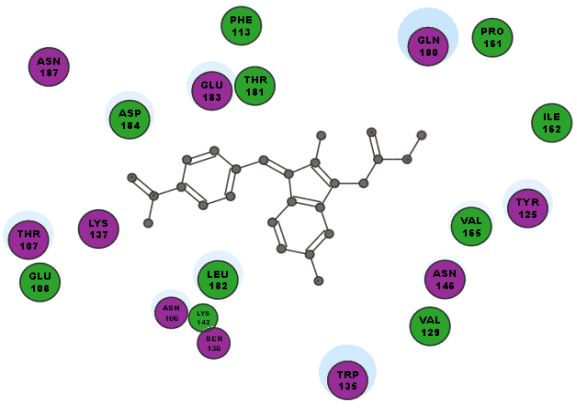


Figure 3: Interaction profiles of transposase-ligand complex

Discovery Studio was used to analyze pharmacokinetics of the ligand molecules. The toxicity of the ligand molecules have also been analyzed using ADMET descriptors of DS. Based on our analysis, it has been found that the analogs which had maximum dock score have proper logP, Absorption and Blood Brain Barrier

ISSN: 0975-8585

values. The plot of polar surface area (PSA) vs logP is shown in (Figure 4). Few of the compounds did not satisfy the ADMET properties are outliers in the plot.

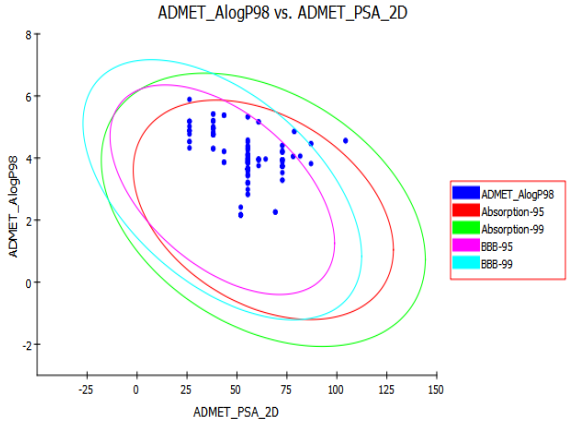


Figure 4: ADMET plot for ligands of TCH8431/19A

Molecular dynamics simulation study of ligand-transposase complex

Further to understand the stability of the receptor-ligand complex, molecular dynamics simulations were performed over a period of 1 ns. Molecular dynamics trajectory of information have been collected and analyzed. Analysis reveals a number of intersting informations which are helpful in designing the inhibitors for the protein, several of them are given below.

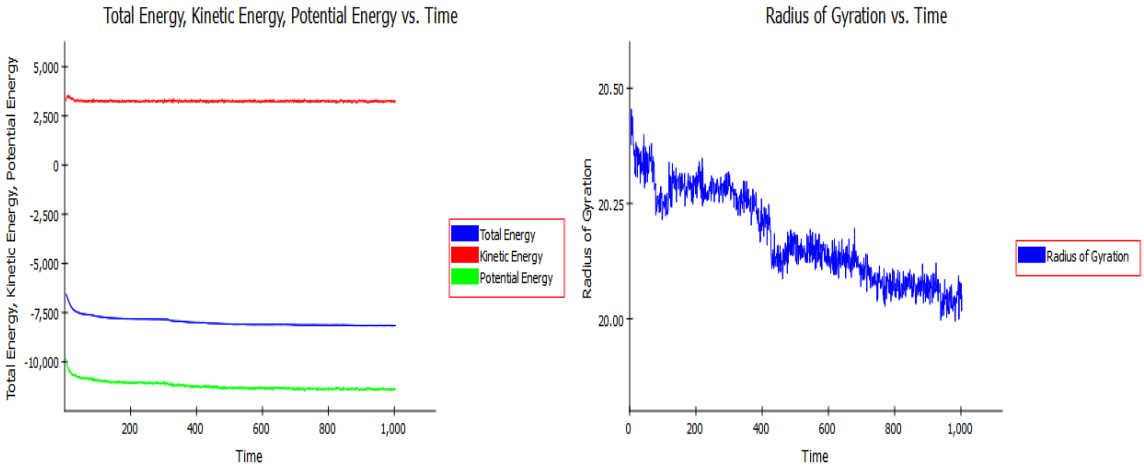


Figure 5: Energy profiles and radius of gyration of transposase-2-[(3Z)-6-fluoro-2-methyl-3[(4methylsulfinylphenyl) methylidene] inden-1-yl] acetate complex

The potential energy, kinetic energy, total energy profiles and radius of gyration of transposase and 2-[(3Z)-6-fluoro-2-methyl-3[(4methylsulfinylphenyl)methylidene]inden-1-yl]acetate complex are shown in figure 5. The energy profiles clearly depict the stability of the complex. Analysis of the molecular dynamics trajectory suggests that that the compound is deeply buried in the binding site cavity and is stabilized by number of intermolecular hydrogen bonding interactions. The interactions with the residue Asp184 is significant and Gln119 and Lys137 are some of the important residues involved in the interactions of the complex at the ligand binding site. The ligand at the binding site is shown in figure 6. Further analysis of the MD trajectory reveals that the flexibility of the walls of the binding site plays a major role in ligand binding. In addition, Ser138 and Trp135 are prone to have interactions at the other tail of the compound. Trp135 is shown to have a hydrophobic interaction with the compound. Moreover it is worth to mention that there exist local domain movements which may contribute significantly to push the ligand deep into the binding site.

ISSN: 0975-8585

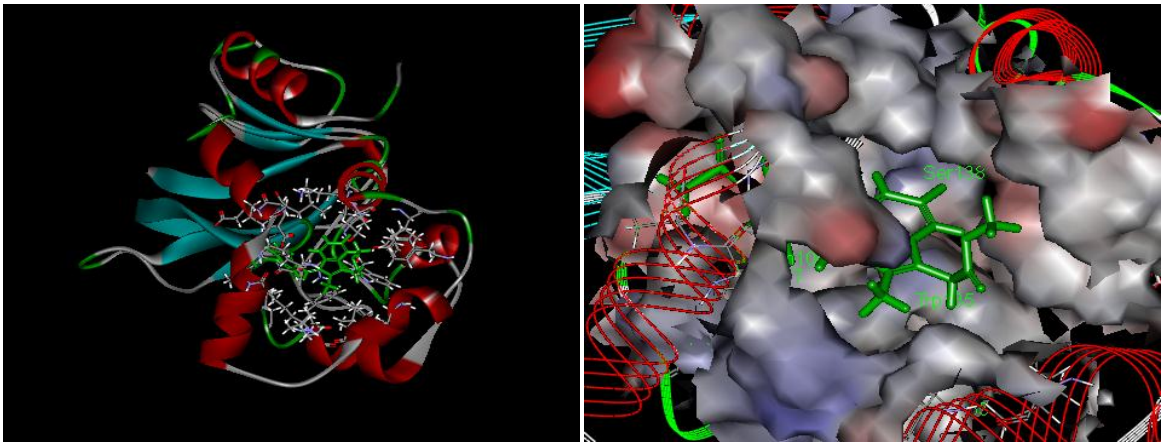


Figure 6: Molecualr graphics view of the docked compound in the binding site of transposase protein

CONCLUSION

In conclusion, the receptor identified and analog compounds identified based on the pharmacoinformatics methodologies such as pharmacophore modeling, molecular docking and molecular dynamics can be considered as the suitable drug targets and drug candidates respectively, which may provide guidance for the rational design of more potent inhibitors against *S.pneumoniae*. The pharmacophore features of the drugs provided suggests the key functional groups that can aid in the design and synthesis of more potential inhibitors for *S.pneumoniae*.

ACKNOWLEDGEMENTS

The authors thank the management of Sathyabama University for the cluster computing facilities installed in the department of Bioinformatics and for all the support to carry out this work.

REFERENCES

[1] Alonso M, Marimon J M, Ercibengoa M, Perez-Yarza E G and Perez-Trallero E PLoS One 2013;8: e54333.
[2] Hathaway LJ, et al. J Bacteriol 2004; 186(12): 3721-9.
[3] Lipsitch M. Emerg Infect Dis, 1999; 5(3): 336-
[4] Lipsitch M, et al. PLoS Med, 2005; 2(1): e15.
[5] Cripps AW, et al. Fifth International Symposium on Pneumococci and Pneumococcal Diseases, Alice Springs, Central Australia, 2-6 April 2006. Vaccine, 2007. 25(13): 2361-5.
[6] Raghunath D. J Biosci 2008; 33(4): 593-603.
[7] Nelson S, Mason CM, Kolls J, Summer WR. Clin Chest Med 1995; 16: 1-12.
[8] Nelson KE, et al. Science 2010;328 (5981): 994-9.
[9] Dockrell DH, Whyte M K and Mitchell T. J Chest 2012; 142: 482-491.
[10] Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A. The Proteomics Protocols Handbook, Humana Press 2005; 571-607
[11] Kneller DG, Cohen FE, Langridge R. J Mol Biol 1990; 214: 171-182.
[12] Laskowski RA, Mac Arthur MW, Moss DS, Thornton JM. J Appl Cryst 1993; 26: 283- 291.
[13] Colovos C, Yeates TO. Protein Sci 1993; 2: 1511-1519.
[14] Weisel M, Proschak E, Schneider G. Chem Cent J 2007;1:7.
[15] Laurie AT, Jackson RM. Bioinformatics 2005; 21:1908–1916.
[16] Wang Y, et al. Nucleic Acids Res 2012; 40(1): D400-412.
[17] Brian K. Stoichet. Nature 2004; 432: 862-865.
[18] Tondi D, Slomczynska U, Costi MP, Watterson DM, Ghelli S, Shoichet BK. Chem Biol 1999; 6: 319–331.
[19] Venkatachalam CM, Jiang X, Oldfield T, Waldman M. J. Mol Graph Model 2003; 21(4):289-307.
[20] Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. J Comp Hem 1983; 4: 187–217.