OPEN ACCESS

International
Journal
Of Modern Engineering Research (IJMER)

Structure Based Drug Design of Inhibitors for *Staphylococcus aureus* Biofilm

Le Anh Vu[1], Nguyen Thuy Huong[2]
[1,2]*Faculty of Chemical Engineering, HoChiMinh City University of Technology, Vietnam*

ABSTRACT:- *In silico drug design acts as a new accession for drug discovery and industry. Structure based computer aided drug design was used in this study to find anti-biofilm agents to repress Staphylococcus aureus biofilm production which is conceded the main virulence factor of this bacterium. The sarA protein was chosen as the target for this process as it stimulates icaADBC operon which is responsible for biofilm formation. The first step was modeling a 3-D structure of the protein which was obtained from Protein Data Bank. Pharmacophore generation was performed using the AutoDock and AutoDock Vina engine from LigandScout program. One hundred fifty one molecules were chosen by pharmacophore based virtual screening using ZincPharmer. Then twenty nine molecules were found suitable as having negative binding free energies with sarA protein in EADock engine from the SwissDock website. These molecules can be tested for in vitro studies as anti-biofilm agents.*

Keywords: *anti-biofilm, bioinformatics, drug design, Pharmacophore, Protein Modeling, S. aureus.*

I. INTRODUCTION

Drug discovery process is a decisive issue in the pharmaceutical industry since it is a very cost and time consuming process [1]. Two different methods are widely used in the pharmaceutical industry nowadays for finding leading molecules: high throughput screening (HTS) and virtual screening (VS). The former process is commonly used in all major pharmaceutical companies. However, the cost in synthesis of each compound, in vitro testing and low hit rate are posing huge problems for pharmaceutical industries. Current efforts within the industry are directed to reduce the timeline and costs [2]. At present, hundreds of thousands to millions of molecules have to be tested within a short period for finding novel molecules; therefore, highly effective screening methods are necessary for today's researchers. As the result of the above problems in finding new drugs by HTS; cost effective, reliable *in silico* screening procedures are in practice [3].

In silico drug design means logical design by which drugs are designed/discovered by using computational methods. It can be applied by either of two strategies of design depending on the knowledge of the target, presence of the primary sequence and 3-D structure. The first approach, structure based drug design (SBDD) is one of the earliest techniques used in drug design. Drug targets are typically key molecules involved in a specific metabolic or cell signaling pathway that is known, or believed, to be related to particular disease symptoms. Drug targets are most often proteins and enzymes in these pathways. SBDD uses the known 3D geometrical shape or structure of proteins to assist in the development of new drug compounds, which is derived from X-ray crystallography or nuclear magnetic resonance (NMR) techniques that can resolve the structure of proteins to a resolution of a few angstroms [1]. The other approach is ligand based drug design which is used when the target is unknown, for example, cell surface receptors make excellent drug targets, but are very difficult to crystallize. So if homology modeling was unreliable or low identity score for the homolog protein was observed, in this case the techniques used for structure-based drug design cannot be used. Pharmacophore models and 3D-QSAR models can be used instead [3].

A biofilm can be defined as a microbially-derived sessile community, typified by cells that are attached to a substratum, interface, or to each other, are embedded in a matrix of extracellular polymeric substance, and exhibit an altered phenotype with regard to growth, gene expression and protein production [4]. Bacterial biofilms are found in many aspects of life, including industry, nature, and in human life. In the human domain, bacterial biofilms dwell in the oral cavity as oral plaques as well as on the skin as part of protective microflora against other more aggressive pathogens [5]. In *Staphylococcus aureus*, biofilm formation is regarded as a major pathomechanism as it renders *S. aureus* highly resistant to conventional antibiotics and host defenses. This can be caused by slow diffusion of these compounds through the extracellular polymeric matrix and slow growth of

the bacteria [6]. *S. aureus* biofilm mode of growth is tightly regulated by complex genetic factors. Host immune responses against persistent biofilm infections are largely ineffective and lead to chronic disease. However, current research has taken biofilm formation into account in terms of elucidating host immunity toward infection, and may lead to the development of efficacious anti-biofilm *S. aureus* therapies.

The aim of this study was to predict inhibitors for *S. aureus* biofilm formation using structure based computer aided drug design strategies.

II. MATERIALS AND METHODS

Structure based drug design strategies were used and as follow:

2.1 Target Finding

Target finding was implemented by survey the researches and articles. Then the information and sequence of the chosen target protein was retrieved from NCBI [7] and Uniprot databases [8]. MEGA [9] and Weblogo [10] programs were used to align the sequence of chosen target protein with other sequences in different *S. aureus* strains. The purpose of this step is to identify the conserved region of the target. Finally, other characteristics of the chosen target including cell localization, toxins and antigens activity were evaluated by Cello [11], BTXpred [12] and VaxiJen [13] programs, respectively.

2.2 Protein Modeling

If the target protein has been crystallized, the protein would be modeled by using 3-D structure from the Protein Data Bank [14]. Otherwise, it would be modeled by using 3-D structure of homolog proteins. RaptorX program [15] and Qmean server [16] were used for this purpose.

2.3 Ligands Search

The first step in ligand determination was the search for anti-biofilm groups from research papers. From these groups, ligands were chosen, depending on their relation to biofilm as initial molecules. The structure and information of each ligand was retrieved by using ZINC Database [17]. Then these molecules were tested by docking with target protein 3-D structure by using SwissDock [18]. The optimum molecules would be used for building pharmacophore model.

2.4 Pharmacophore Virtual Screening

The molecules with the best score selected from the previous section were entered in Ligand Scout program to build a pharmacophore. Then the pharmacophore was used in drugs like molecule screening in the ZincPharmer [19] to obtain more suitable ligands. The achieved ligands were finally tested again by T.E.S.T software [20] and SwissDock [18].

III. RESULTS AND DISCUSSIONS

3.1 Target Finding

Some of the proteins of *S. aureus* were studied to select the most effective and suitable target for biofilm formation. SarA protein was chosen as target for this study for its interactions with other proteins in different pathways [21]. SarA protein of 18 strains of *S. aureus* and 7 strains of Staphylococci were aligned to assign the most conserved region. The WebLogo website was used to observe differences between the 25 sequences of the protein (Fig. 1). The Weblogo result displays small differences at amino acids No. 4, 14, 25, 46, 49, 73, 98, 99, 105, 106, 109, 120, 122 and 124. However, these differences do not affect function of sarA protein, as it's core residues are located in amino acids No. 8, 11, 84, 88, 89 and 90 [22]. Then sarA protein of the *S. aureus* strain MW2 was chosen as the target protein sequence (Fig. 2).

SarA protein was chosen out of many targets as it acts as an *icaADBC* operon regulator [23] and blocking it will stop the most important operon in biofilm synthesis process. SarA is a 124-residues DNA binding protein encoded by the *sarA* locus, which consists of three overlapping transcripts, driven by three distinct promoters, P1, P3 and P2 [24].

DNA binding and profiling studies suggest that sarA protein may regulate target genes by directly binding to target gene promoters or indirectly via downstream effects on regulons (e.g. binding to the *agr* promoter) [25] or by stabilizing mRNA during the log phase [26].

The importance of this protein comes from its multifunctional regulatory activity. First of all, it acts at the initiation step of biofilm production by direct binding to *icaA* promoter enhancing transcription of *icaADBC* operon. Furthermore, sarA influences the regulation of biofilm formation via an *agr*-dependent pathway. It has also been found that sarA enhances the proteolytic enzymes activity, which has an important rule in the regulation of biofilm development [23]. So, blocking this protein will afffect the biofilm development process at many stages.

Results retrieved from CELLO, BTXpred and VaxiJen also revealed that sarA is cytoplasmic protein, non-toxin and non-antigen, respectively.

Figure 1. WebLogo results of comparison between the 25 sarA sequences.

```
        10         20         30         40         50
MAITKINDCF ELLSMVTYAD KLKSLIKKEF SISFEEFAVL TYISENKEKE
        60         70         80         90        100
YYLKDIINHL NYKQPQVVKA VKILSQEDYF DKKRNEHDER TVLILVNAQQ
       110        120
RKKIESLLSR VNKRITEANN EIEL
```

Figure 2. Primary structure of sarA protein in *S. aureus* strain MW2.

3.2 Protein Modeling

The sarA protein of *S. aureus* has been crystallized already, and X-ray diffraction studies have been found for it [22]. So, the next step was to download the 3-D structure from the PDB database. Then the protein was modeled by using a code 2FRH from PDB website. Fig. (3) shows the modeled protein in a PDB format.

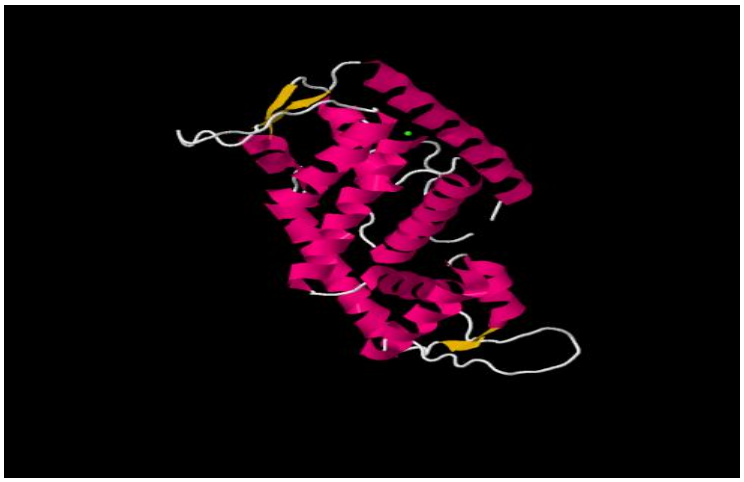

Figure 3. 3-D structure of sar A protein.

Sar A, like its homolog sarR, is a dimeric winged helix structure with each monomer consisting of 5 α-helices, 3 β-strands and several loops (α1α2-β1α3α4-β2β3-α5). The sarA dimer possesses a central helical core and two winged helix motifs. Within each winged helix motif there is a helix-turn-helix motif (α3α4) and a β-hairpin turn wing (β2β3), both of which are putative DNA binding domains [22].

3.3 Ligands Search

Ligands search was carried out by review the literatures and it has been found that 15 molecules from different groups can be used as initial ligands (Acetaminophen, Albendazole, Acetylsalicylic Acid, Diacetyl, Eugenol, Piroxicam, Ibuprofen, Ferric ammonium citrate, Indomethacin, Methyldopa, Pentazocine, Rifampicin, Thymol, Vancomycine, Diclofenac). These molecules were tested by docking with sarA protein 3-D structure by the EADock DSS engine in the SwissDock website. The result was positive (i.e., their minimum binding energy was negative) for only 7 molecules (Fig. 4). Then these molecules were chosen for building pharmacophore model.

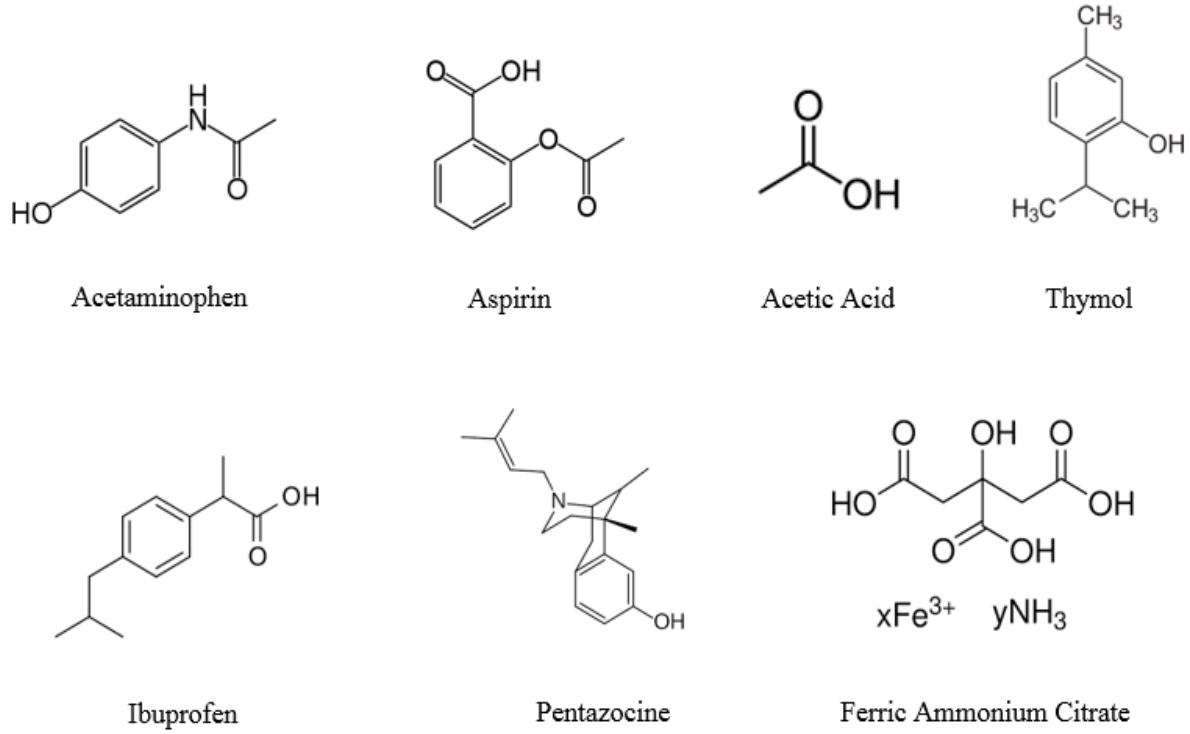

Figure 4. Seven molecules with postive docking results.

Most life science processes involve, at the atomic scale, recognition between two molecules. The prediction of such interactions at the molecular level, by the so-called docking software, is a non-trivial task. Docking programs have a wide range of applications ranging from protein engineering to drug design. SwissDock, a web server dedicated to the docking of small molecules on target proteins. It is based on the EADock DSS engine, combined with setup scripts for curating common problems and for preparing both the target protein and the ligand input files. It also uses calculations performed in the CHARMM force field. An efficient Ajax/HTML interface was designed and implemented, so that workers can easily submit dockings and retrieve the predicted complexes. For automated docking tasks, a programmatic SOAP interface has been set up and template programs can be downloaded in Perl, Python and PHP. The web site also provides an access to a database of manually curated complexes, based on the Ligand Protein Database [27].

3.4 Pharmacophore Virtual Screening

A pharmacophore model is an ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target and to trigger (or block) its biological response [28]. In this study, LigandScout was used for pharmacophore modeling as it uses AutoDock and AutoDock Vina engine for this job (Fig. 5).

The pharmacophore generated by Ligand Scout showed three main features as two hydrogen bond acceptors and negative ionizable. The pharmacophore generated for the chosen group of compounds showed consistency in the above features. Al-Khafaji and Al-Mulla [29] also presented the pharmacophore of sarA inhibitors in *S. epiderpimidis* as having one hydrogen bond acceptor, one hydrogen bond donor, one hydrophobic group and one aromatic ring features. In this way, the pharmacophore has not only been restricted to one active site group but other groups have also been included.

Then the pharmacophore was used in the ZincPharmer to screen more than 215 million different conformations of more than 22 million compounds in Zinc database (Fig. 6).

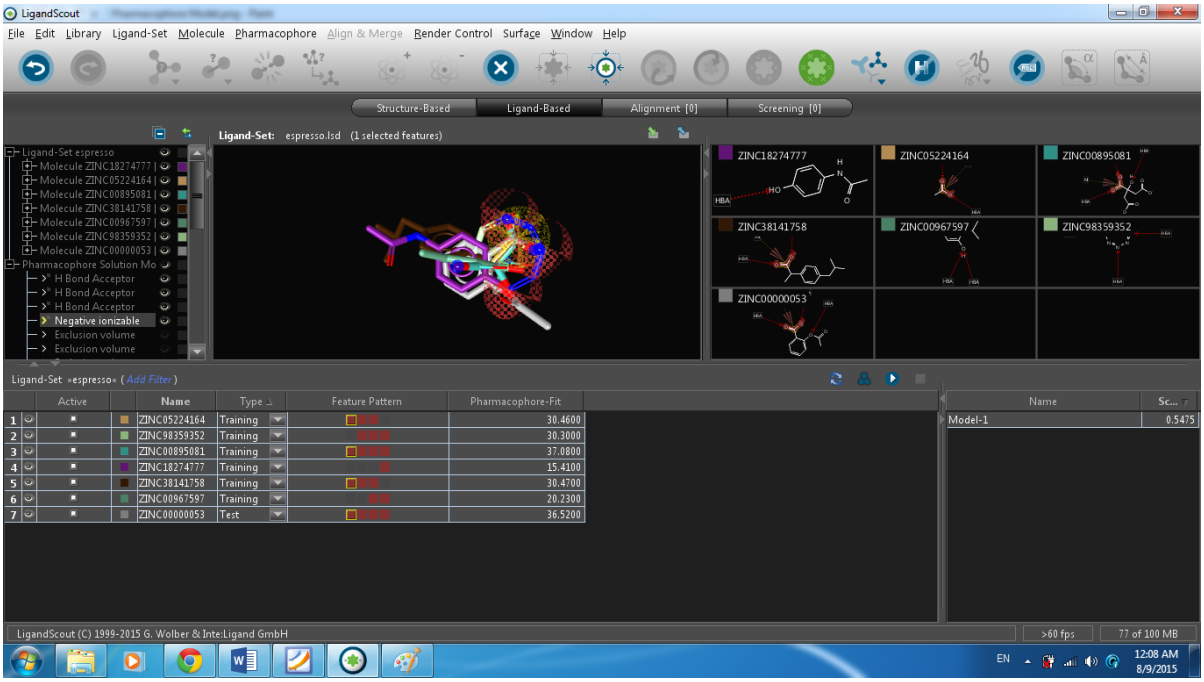

Figure 5. Hypothetical pharmacophore generated in in LigandScout (red colored spheres represent H-bond acceptors while yellow colored sphere represents Negative inonizable).

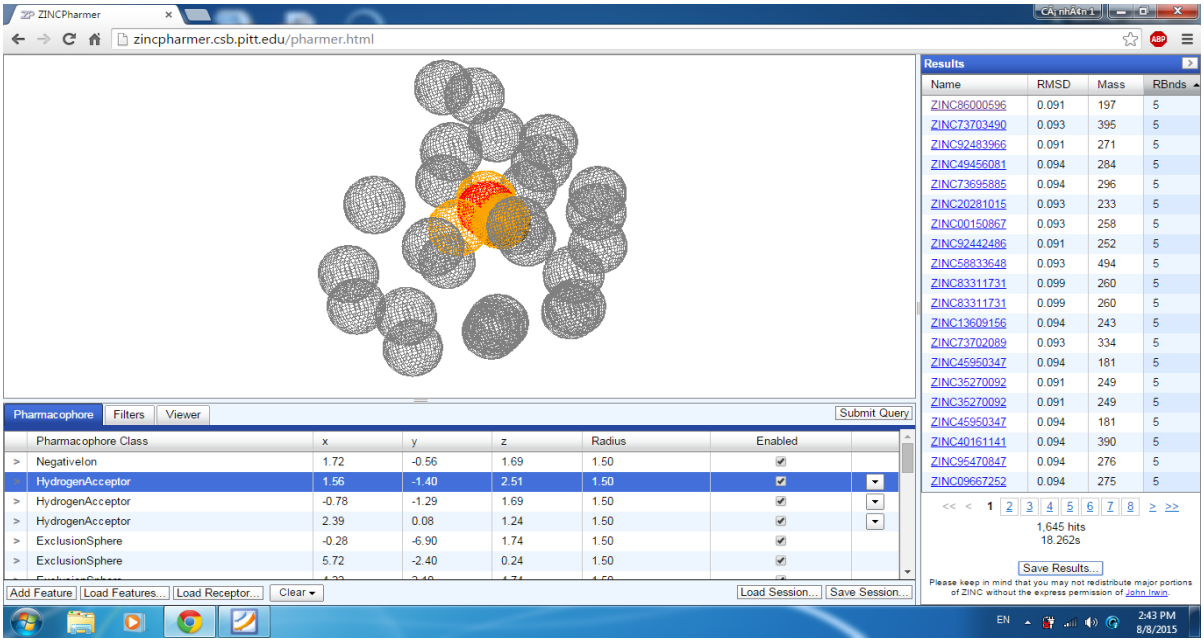

Figure 6. Pharmacophore virtual screening features in Zinc Pharmer.

The goal of virtual screening is to select, relatively rapidly and cheaply, small subset of compounds predicted to have activity against a given biological target out of a large database of compounds. While it is possible to screen large databases using automated high-throughput screening methods, this is expensive and requires a substantial investment in infrastructure and assay development. The idea of virtual screening is to test compounds computationally in order to reduce the number of compounds to be screened experimentally, with the additional advantage that the number of compounds in the final set can easily be adjusted according to the resources available for assaying [30].

One hundred fifty one (151) molecules were obtained from Zinc database after filtering through Lipinski rule of five [31] and through analyzing them with the T.E.S.T program [20] for their mutagenicity. All these molecules were again entered the SwissDock to estimate their binding affinity to the protein. Many criteria from docking results can be used for estimating binding affinity including binding free energy, full fitness, hydrogen bonding and total free energy but binding free energy was used as the main criterion for ranking the best powerful ligands [32]. The final result was 29 molecules having positive docking results after docking through SwissDock (i.e., having negative free binding energy with sarA protein) (Table 1).

Table 1. Molecules with positive docking result.

	Molecules or ZINC ID	Minimum Free Binding Energy (Kcal/Mol)
1	Acetaminophen	-8.26743
2	Acetic Acid	-26.61780
3	Aspirin	-6.43788
4	Ibuprofen	-18.24420
5	Ferric Ammonium Citrate	-4.45522
6	Pentazocine	-5.22241
7	Thymol	-10.30120
8	zinc_86000596	-5.06513
9	zinc_19514206	-6.21830
10	zinc_40141161	-6.81005
11	zinc_45900455	-18.66230
12	zinc_71770235	-3.86902
13	zinc_12496101	-13.97391
14	zinc_06072047	-12.67186
15	zinc_13831151	-6.29054
16	zinc_04212809	-8.17532
17	zinc_00004454	-7.61879
18	zinc_04626657	-7.83685
19	zinc_00056790	-5.39341
20	zinc_00590964	-14.18325
21	zinc_00602086	-3.30871
22	zinc_08766760	-5.85426
23	zinc_00901109	-13.79051
24	zinc_55112784	-38.56271
25	zinc_55128530	-15.23407
26	zinc_65054752	-33.82178
27	zinc_75200618	-20.12643
28	zinc_03074344	-19.99467
29	zinc_70735489	-4.05025

IV. CONCLUSIONS AND RECOMMENDATIONS

Based on the results obtained, the sarA protein can be used as a good target for biofilm suppression and bioassay. Structure based drug design strategy using Ligand Based Virtual Screening was found twenty nine molecules with a negative free binding energy that means high affinity to bind to sarA protein and these molecules can be tested *in vitro* for their biofilm inhibition activity.

ACKNOWLEDGEMENTS

This research was supported by the Division of Biotechnology, Faculty of Chemical Engineering, Ho Chi Minh University of Technology.

REFERENCES

[1]. V. Rao and K. Srinivas. Modern drug discovery process: An *In silico* approach. *Journal of Bioinformatics and Sequence Analysis, 2*, 2011, 89-94.

[2]. H. Böhm, G. Schneider, H. Kubinyi, R. Mannhold and H. Timmerman, *Virtual screening for bioactive molecules (Methods and principles in medicinal chemistry)* (Wiley-VCH, Weinheim, Germany, 2000).

[3]. D. Young, *Computational drug design* (John Wiley & Sons, New Jersey, USA, 2009).

[4]. R. M. Donlan and J. W. Costerton. Biofilms: survival mechanisms of clinically relevant microorganisms. *Clin Microbiol Rev, 15*, 2002, 167-193.

P. Stewart and M. Franklin. Physiological heterogeneity in biofilms. *Nature Reviews Microbiology, 6*, 2008, 199-210.

[5]. N. K. Archer, M. Mazaitis, J. W. Costerton, J. G. Leid, M. E Powers and M. E. Shirtliff. *Staphylococcus aureus* biofilms: Properties, regulation and roles in human disease. *Virulence, 2(5),* 2011, 445-459.

[6]. R. Matthiesen, *Bioinformatics methods in clinical research* (Humana Press, New York, USA, 2010).

[7]. R. Apweiler, A. Bairoch, C. Wu, W. Barker, B. Boeckmann, S. Ferro. UniProt: the Universal Protein knowledgebase. *Nucleic Acids Research, 32*, 2004, 115-119.

[8]. K. Tamura, D. Peterson, N. Peterson, G. Stecher, M. Nei and S. Kumar. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution, 28*, 2011, 2731-2739.

[9]. G. E. Crooks, G. Hon, J.M Chandonia and S. E. Brenner. WebLogo: A Sequence Logo Generator. *Genome Research, 14(6)*, 2004, 1188-1190.

[10]. C. Yu, Y. Chen, C. Lu and J. Hwang. Prediction of protein subcellular localization. *Proteins: Structure, Function and Bioinformatics, 64*, 2006, 643-651.

[11]. S. Saha and G. Raghava. BTXpred: prediction of bacterial toxins. *In Silico Biology, 7*, 2007, 405-412.

[12]. I. Doytchinova and D. Flower. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics, 8*, 2007, 1-7.

[13]. H. M. Berman, J. Westbrook, Z. Feng. The Protein Data Bank. *Nucleic Acids Research, 28(1)*, 2000, 235-242.

[14]. M. Källberg, H. Wang, S. Wang, J. Peng, Z. Wang and H. Lu.Template-based protein structure modeling using the RaptorX web server. *Nature Protocols, 7*, 2012, 1511-1522.

[15]. P. Benkert, M. Künzli and T. Schwede. QMEAN server for protein model quality estimation. *Nucleic Acids Research, 37*, 2009, 510-514.

[16]. J. Irwin and B. Shoichet. ZINC – a free database of commercially available compounds for virtual screening. *Journal of Chemical Information and Modeling, 45*, 2005, 177–182.

[17]. K. Hetal, A. Ratna and B. Pratibha. Docking studies of components of tulsi and mamejavo against plasmodium lactate dehydrogenase. *International Research Journal of Biological Sciences, 2*, 2013, 8-12.

[18]. D. Koes and C. Camacho (2012). ZINCPharmer: pharmacophore search of the ZINC database. *Nucleic Acids Research, 40*, 2012, 409-414.

[19]. I. Sushko, S. Novotarskyi, R. Körner, A. Pandey, A. Cherkasov and J. Li. Applicability domains for classification problems: benchmarking of distance to models for AMES mutagenicity set. *Journal of Chemical Information and Modeling, 50*, 2011, 2094-2111.

[20]. A. Franceschini, D. Szklarczyk, S. Frankild, M. Kuhn, M. Simonovic and A. Roth. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Research, 41*, 2013, 808-815.

[21]. Y. Liu, A. Manna, C. Pan, I. Kriksunov, D. Thiel and A. Cheung. Structural and function analysis of the global regulatory protein SarA from *Staphylococcus aureus. Proceedings of the National Academy of Sciences USA, 103*, 2006, 2392-2397.

[22]. M. P. Trotonda, A. C. Manna, A. L. Cheung, I. Lasa and J. R. Penadés. (2005). SarA positively controls *bap*-dependent biofilm formation in *Staphylococcus aureus*. *Journal of Bacteriology, 187(16)*, 2005, 5790-5798.

[23]. M. Bayer, J. Heinrichs and A. Cheung. The molecular architecture of the *sar* locus in *Staphylococcus aureus*. *Journal of Bacteriology, 178*, 1996, 4563–4570.

[24]. A. Cheung, A. Bayer, G. Zhang, H. Gresham and Y. Xiong. Regulation of virulence determinants *in vitro* and *in vivo* in *Staphylococcus aureus*. *FEMS Microbiology Letters, 1649*, 2004, 1–9.

[25]. C. Roberts, K. Anderson, E. Murphy, S. Projan, W. Mounts and B. Hurlburt. Characterizing the effect of the *Staphylococcus aureus* virulence factor regulator, sarA, on log-phase mRNA half-lives. *Journal of Bacteriology, 188*, 2006, 2593-2603.

[26]. A. Grosdidier, V. Zoete and O. Michielin. SwissDock, a protein-small molecule docking web service based on EADock DSS. *Nucleic Acids Research, 39*, 2011, 270-277.

[27]. S. Yang. Pharmacophore modeling and applications in drug discovery: challenges and recent advances. *Drug Discovery Today, 15*, 2010, 444-450.

[28]. M. Al-Khafaji and F. Al-Mulla. *In silico* design of inhibitors for *Staphylococcus epidermidis* biofilm. *International Journal of Advances in Pharmacy, Biology and Chemistry*, Vol. *3(2)*, 2014, 334-340.

[29]. M. Peach and M. Nicklaus. Combining docking with pharmacophore filtering for improved virtual screening. *Journal of Cheminformatics*, *1*, 2009, 1-6.

[30]. C. A. Lipinski. Lead- and drug-like compounds: the rule-of-five revolution. *Drug Discovery Today: Technologies, 1(4)*, 2004, 337-341.

[31]. T. Romano. Structure-based drug design: docking and scoring. *Current Protein and Peptide Science*, *8*, 2007, 312-328.