Advanced Techniques in Biology & Medicine

ISSN: 2379-1764

Balaji et al., Adv Tech Biol Med 2014, 2:1
DOI: 10.4172/2379-1764.1000115

Research article Open Access

Molecular Docking Studies of Wide Spectrum Targets in *Staphylococcus aureus* - An Aim towards Finding Potent Inhibitors

Balaji SR*, Gupta KK, Anusha P and Raveena P

NTHRYS Biotech labs, Hyderabad, Andhra Pradesh, India

Abstract

Methicillin-resistant *staphylococcus aureus* (MRSA) is a bacterium that is evolving towards adaptive changes to certain antibiotics like methicillin, oxacillin, penicillin, and amoxicillin. It is a communicable and most rapidly spreading disease worldwide. It is reported that MRSA is becoming common among children in intensive care units. This disease comes under Emerging Infectious Disease (EID). In this paper, MRSA proteome screening is done and Drug/vaccine targets are proposed based on its essentiality to the pathogen and non-homology with human proteome. Targets validation is done so that its targeting must not affect human proteome and vital pathways. Those targets which has no structure, structure prediction and validation is done and important epitopes and ligands are proposed on suitable targets.

Keywords: MRSA; EPA; Infection; Targets; Epitope; Drug; Pathways

Introduction

MRSA spreads infection through blood and skin. The former infection is of severe kind. But skin infections are the most common one. Skin infection appears as pustules or boils which generally form at areas of the body covered by hair like back of neck, groin, buttock, armpit, beard area of men. The degree of symptoms depends on the stage of infection. The persons who meet MRSA patients are at Risk of Acquiring MRSA Infections. In hospitals, patients has more risk of getting MRSA through catheters inserted into the skin and those patients who has undergo medical procedures like surgery. According to drug bank only three antibiotics namely Arbekacin, Meticillin and Linezolid, are approved for the treatment of MRSA. Drug targets are limited and there is an urgent need for the discovery of novel drug targets. Apart from drug target, we also need Good ligand formulation that would act as a potent inhibitors to the novel targets without affecting human proteome (Figure 1).

Any contaminated surfaces is the source of many kind of infections, MRSA is one of them. Mishandled Hospital procedures can leave immune compromised patients vulnerable to MRSA. Environmental hygienic conditions are the source of good physical and mental health. It is reported that U.S. Environmental Protection Agency (EPA) labeled Cleaners, disinfectants and sanitizers must be used in order to spread all kinds of infections.

Material and Methodology

Staphylococcus aureus proteome screening

Prokaryotic sequence homology analysis tool (PSAT) [1] is used for finding conserved patches in all strains of Staphylococcus aureus and Streptococcus, as both are heavily involved in skin infections. Yersinia Pestis CO 92 is taken as reference strain. Following parameters is considered for finding conserved genes in these selected pathogens: BLAST alignment score thresholds for finding gene homologs: e-value < 10; bit score > 20; % identity > 10

DEG BLAST

All the conserved genes are checked for their key role to pathogens. Database of essential genes (DEG) [2] is used to screen all the conserved genes for their vitality to the selected pathogens. This tool gives hit if query gene is showing any significant similarity with pathogen's growth, reproduction and survival.

NCBI Homo sapiens (human) protein BLAST

Those conserved genes which are giving hits in DEG blast were further checked in NCBI Homo sapiens (human) Protein BLAST [3]. If we are proposing target on pathogen proteome then those conserved genes must be checked if there is any similarity with human proteome because drug/vaccine targeting must not affect any of the human protein.

Annotation and pathways analysis

The conserved genes which are showing no significant similarity checked in NCBI Homo sapiens (human) Protein BLAST was checked for annotation and their involvement in any crucial pathway. Annotation was done by using CELLO [4] and PSORTB [5]. Pathway analysis was done in KEGG genes database [6].

Target proposal and structure prediction

Based on the pathway study, its essentiality and non-homolog to human proteome property of Targets, they are proposed for drug targets and epitope design. Those targets whose have no protein structure was modeled by Protein Homology/analogy Recognition Engine (PHYRE) [7].

Structure validation

These Target proteins were optimized from KOBAMIN [8] and Galaxy WEB server [9] and validated in Rampage [10] and Erratplot

***Corresponding author:** Balaji SR, Director & Research Head, NTHRYS Biotech labs, Hyderabad, Andhra Pradesh, India, Tel: 040-276 212 48; E-mail: balajisrao@nthrys.com

Received January 11, 2014; **Accepted** February 13, 2014; **Published** February 22, 2014

Citation: Balaji SR, Gupta KK, Anusha P, Raveena P (2014) Molecular Docking Studies of Wide Spectrum Targets in *Staphylococcus aureus* - An Aim towards Finding Potent Inhibitors. Adv Tech Biol Med 2: 115. doi: 10.4172/2379-1764.1000115

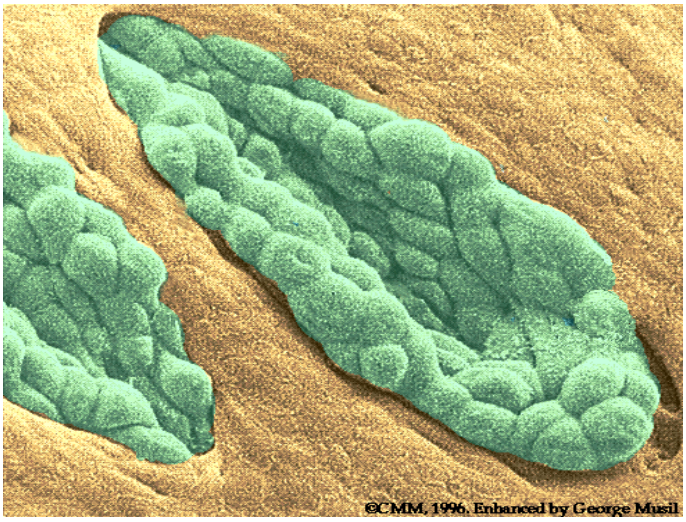


Figure 1 : MRSA.

[11], the former check the stereo chemical properties of modeled structures and latter one analyze the non-bonded interactions.

Epitope design

Transmembrane region of proposed target was predicted with TMHMM [12]. After that, Prediction of B cell epitopes with BCpred (cutoff 0.8, 20-mer epitopes) [13]. Generally B cell epitope sequences are surface-exposed of corresponding proteins. T cell epitopes from propred [14], propred 1[15], MHCpred [16] and T epitope designer [17]. Those T cell epitopes were considered that is part of B cell epitopes and lies at Transmembrane region according to TMHMM. Predict those epitope's antigenicity from vaxijen [18]. Finally, epitopes that bound more than 13 MHC molecules in ProPred and ProPred-I with less than 100 nM IC50 for DRB1*0101 in MHCPred v2.0 and that bound >=80% of HLA molecules in T-epitope designer were selected. Proteins that were antigenic according to Vaxijen (threshold=0.4, ACC output) score above than 0.5 in VaxiJen were selected.

Virtual screening of ligands against target proteins

All the target proteins were screened with 12 million drug-like ZINC12 + 6507 DrugBank drugs from FINDSITE-COMB [19]. Less characterized ZINC DATABASE molecules were selected for drug-likeness analysis from FAF- Drugs2 [20]. FAF results were cross checked with Osiris property explorer [21]. We modified most of the ZINC ID ligands to satify druglikeness and ADME properties. We proposed those novel ligands for each target which is following all the drug-likeness properties.

Molecular docking studies of ADMET following ligands with target proteins

ADMET following ligands were docked with their corresponding target in particular coordinate predicted by Pocket-Finder [22] using Molegro Virtual Docker [23]. All the docking parameters and images were finalized for each ligand-protein docking.

Results

Six targets are finalized for further study (Table 1). Protein accession number, locus tags are the unique identifier of *insilico* discovered targets. All are essential for pathogens as inferred from DEG Blast results. kdpA, opp-1B and icac structures are not available. Therefore, they must be modeled for further molecular docking studies. Pathways and Cellular localization results are shown in Table 2.

All the target proteins are suitable for epitope design as all are predicted to be localized in Membrane from CELLO and PSORTB. They can also be considered for potent drug target. Function of Target Protein is very important for Target validation. Functional annotation is done in Table 3. Epitope design on the target proteins are given in Table 4. The final epitopes following all the criteria from target proteins are given in Table 5. The structural information of refined modeled target proteins from erratplot and rampage are shown in Tables 6 and 7. The coordinates of best pockets for each target proteins are given in Table 8.

Locus Tag	Protein ID	Target name	Structure/ Related (yes/no)	DEG hits (yes/no)
SAR2165	YP_041527.1	kdpA	no	yes
SAR2553	YP_041904.1	Opp-1B	no	yes
SAR2750	YP_042088.1	icac	no	yes
SAR0118	YP_039582.1	sirA	Yes (Related: pdbid: 3MWF)	yes
SAR2537	YP_041888.1	opuCB	Yes (Related: pdbid: 3D31 C)	yes
SAS0639	YP_042767.1	Hypothetical protein	Yes (Related: pdbid: 3MLV L)	yes

Table1: Final Targets based on subtractive proteome screening.

Target name	CELLO	PSORTB	Pathway
kdpA	Membrane	Cytoplasmic membrane	Two-component system
Opp-1B	Membrane	Cytoplasmic membrane	ABC transporters
icac	Membrane	Cytoplasmic membrane	No pathway
sirA	Periplasmic	Cytoplasmic membrane	ABC transporters
opuCB	Inner membrane	Cytoplasmic membrane	ABC transporters
Hypothetical protein	Membrane	Cytoplasmic membrane	No pathway

Table 2: Cellular localization and pathways of target proteins.

Target name	Function	Molecular Function
kdpA	One of the components of the high-affinity ATP-driven potassium transport (or KDP) system, which catalyzes the hydrolysis of ATP coupled with the exchange of hydrogen and potassium ions	ATP binding, potassium-transporting ATPase activity
Opp-1B	Oligopeptide transporter putative membrane permease domain	transporter activity
icac	Presumably involved in the export of the biofilm adhesin polysaccharide poly-beta-1,6-N-acetyl-D-glucosamine (PNAG, also referred to as PIA) across the cell membrane	transferase activity, transferring acyl groups other than amino-acyl groups
sirA	Iron-regulated ABC transporter siderophore-binding protein SirA	Iron-regulated ABC transporter siderophore-binding protein SirA
opuCB	Probable glycine betaine/carnitine/choline ABC transporter opuCB	transporter activity
Hypothetical protein	Uncharacterized protein conserved in bacteria [Function unknown]	Function unknown

Table 3: Function annotation of target proteins.

Target name	TMHMM	Bcpred Position-Bcell epitope (confidance value)	Propred 1 (Sorted in descending score)	Propred	MHCpred
kdpA	Outside: 1-3 86-126 192-244 303-324 373-375 438-483 549-558 TMhelix: 4-26 63-85 127-149 169-191 245-267 280-302 325-347 354-372 376-398 415-437 484-506 526-548	462-AAANNGSGFEGLKDDTTFWN (0.97) 335- FTVITTAFTTGSVNNMHDSL (0.92) 77-LLIVQQWLFLNPNHNLNQSI(0.91) 434 AFMIPGASESITNPSFHGIS (0.89)	LIVQQWLFL NNGSGFEGL MIPGASESI AAANNGSGF LLIVQQWLF IPGASESIT FTVITTAFT TVITTAFTT FMIPGASES ITTAFTTGS VITTAFTTG ANNGSGFEG IVQQWLFLN	FMIPGASES (10/51)	FMIPGASES - 72% MHC alleles are Binding.
Opp-1B	outside 32-106 162-175 257-275 TMhelix: 9-31 107-126 139-161 176-193 234-256 276-298	80- NFGTSYITGDPVAERIGPAF (0.99) 41- AQGTPNVTPELIAETNEKYG (0.91)	GTPNNTPEL TPNVTPELI TSYITGDPV GTSYITGDP AQGTPNVTP FGTSYITGD NFGTSYITG QGTPNVTPE	YITGDPV (8/51)	YITGDPV -100% MHC alleles are Binding.
Icac	outside 30-43 102-115 168-186 234-242 292-305 Tmhelix: 7-29 44-66 79-101 116-138 145-167 187-204 211-233 243-262 269-291 306-328	165- YFTNNTAFHDTVLHYYPLSE (0.7)	NNTAFHDTV FTNNTAFHD TNNTAFHDT	YFTNNTAFH (14/51) VLHYYPLSE (8/51)	FTNNTAFH 100% MHC alleles are Binding.

sirA	Outside: Whole protein	20-GCSGNSNKQSSDSKDKETTS (0.99) 70-LGVKPVGAVESWTQKPKFEY (0.99) 149-KDTTKLMGKALGKEKEAEDL (0.97) 269-LVKKTESEWTSSKEWKNLDA (0.97) 43-AMGTTEIKGKPKRVVTLYQG (0.96) 91-KNDLKDTKIVGQEPAPNLEE (0.94) 177-AAFQKDAKAKYKDAWPLKAS (0.84)	KNDLKDTKI TTKLMGKAL FQKDAKAKY LGVKPVGAV AAFQKDAKA KPVGGAVESW NDLKDTKIV TTEIKGKPK KDTTKLMGK KKTESEWTS LVKKTESEW NDLKDTKIV CSGGNSNKQS KTESEWTSS GCSGGNSNKQ SGNSNNQSS DTTKLMGKA MGTTEIKGKP AFQKDAKAK GVKPGAVE DLKDTKIVG VKPAVES AMGTTEIKG GTTEIKGKP QKDAKAKYK VKKTESEWT GNSNKQSSD TKLMGKALG	LGVKPVGAV (4/51) VKPVGAVES (10/51) FQKDAKAKY (8/51)	LGVKPVGAV 68% MHC alleles are Binding. VKPVGAVES 72% MHC alleles are Binding. FQKDAKAKY 80% MHC alleles are Binding.
opuCB	TMHelix 114 135 115 134 144 170 145 168 174 189 175 188 223 248 225 235 238 243 268 292 269 290	155- YVGAGGLGDFIFNGLNLYDP (0.9)	YVGAGGLGD VGAGGLGDF AGGLGDFIF GAGGLGDFI	YVGAGGLGD (8/51)	YVGAGGLGD 72% MHC alleles are Binding.
Hypothetical protein	TMHelix 64 102 65 102 113 131 115 129 150 181 151 180 194 222 195 202 205 221	No B cell epitope predicted.	No common T cell epitope	No common T cell epitope	No epitope binding affinity prediction.

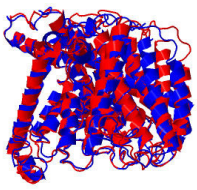
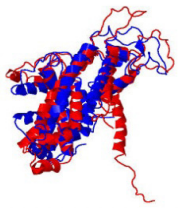
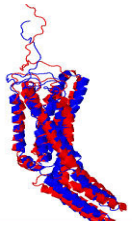
Table 4: Epitope designing on target proteins.

Target name	Final Predicted epitopes	Vaxijen score (>0.5)
kdpA	FMIPGASES	0.4000
Opp-1B	YITGDPV	1.2802
icac	FTNNTAFH	0.3954
sirA	LGVKPVGAV VKPVGAVES FQKDAKAKY	0.0443 0.0286 1.9244
opuCB	YVGAGGLGD	1.5458
Hypothetical protein	No predicted epitope	No predicted epitope

Note: red color indicates non promising peptides whereas green color indicates promising one.
Table 5: Final Epitopes on target proteins.

Target Name/Protein ID	Ramachandran parameters		Errat plot quality factor
kdpA/ YP_041527.1	**RMSD**	2.290	78.46
	Clash score	20.9	
	Poorrotamers	3.5	
	Rama favored	90.6	
Opp-1B/ YP_041904.1	**RMSD**	8.941	82.7
	Clash score	15.7	
	Poor rotamers	0.4	
	Rama favored	94.2	
Icac/ YP_042088.1	**RMSD**	5.597	93.59
	Clash score	17.4	
	Poor rotamers	2.2	
	Rama favored	95.7	
sirA/ YP_039582.1	**RMSD**	4.377	86.00
	Clash score	8.6	
	Poor rotamers	1.0	
	Rama favored	98.2	
opuCB/ YP_041888.1	**RMSD**	5.944	94.44
	Clash score	10.9	
	Poor rotamers	1.2	
	Rama favored	96.2	
Hypothetical protein/ YP_042767.1	**RMSD**	0.749	77.03
	Clash score	17.0	
	Poor rotamers	1.0	
	Rama favored	91.2	

Table 6: Structural information of target proteins.

Target name	Refined modeled target structures
kdpA/ YP_041527.1	 Number of residues in favoured region (~98.0% expected): 504 (90.6%) Number of residues in allowed region (~2.0% expected) : 36 (6.5%) Number of residues in outlier region : 16 (2.9%)
Opp-1B/ YP_041904.1	 Number of residues in favoured region (~98.0% expected): 284 (93.7%) Number of residues in allowed region (~2.0% expected) : 14 (4.6%) Number of residues in outlier region : 5 (1.7%)
Icac/ YP_042088.1	 Number of residues in favoured region (~98.0% expected) : 334 (96.0%) Number of residues in allowed region (~2.0% expected) : 12 (3.4%) Number of residues in outlier region : 2 (0.6%)

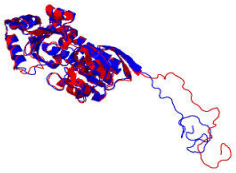
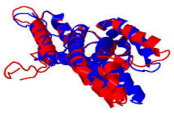
sirA/ YP_039582.1	 Number of residues in favoured region (~98.0% expected) : 322 (98.2%) Number of residues in allowed region (~2.0% expected) : 6 (1.8%) Number of residues in outlier region : 0 (0.0%)
opuCB/ YP_041888.1	 Number of residues in favoured region (~98.0% expected) : 201 (96.2%) Number of residues in allowed region (~2.0% expected) : 7 (3.3%) Number of residues in outlier region : 1 (0.5%)
Hypothetical protein/ YP_042767.1	 Number of residues in favoured region (~98.0% expected) : 506 (91.0%) Number of residues in allowed region (~2.0% expected) : 37 (6.7%) Number of residues in outlier region : 13 (2.3%)

Table 7: Modeled proteins.

Target name	Best pocket with coodinates
Icac/ YP_042088.1	1st site coordinates predicted by Pocket-Finder Min Coords: (-16, -47, 24) Max Coords: (-4, -33, 37)
	Best pocket residues predicted by FINDSITE comb TEMPLATE 1L9HA RES 288 1.00 -0.21 HIS RES 293 1.00 -0.93 ASP RES 318 1.00 -0.51 LEU RES 319 1.00 -0.37 GLY RES 322 1.00 -0.43 ILE DNAE 1.994
Opp-1B/ YP_041904.1	1st site coordinates predicted by Pocket-Finder Min Coords: (-39, 12, -6) Max Coords: (-21, 29, 15)
	Best pocket residues predicted by FINDSITE comb TEMPLATE 1W07A00 TEMPLATE 2FONA00 TEMPLATE 1IS2A00 RES 101 0.67 0.00 ASN RES 105 1.00 0.00 LEU RES 168 1.00 0.00 THR RES 170 1.00 0.00 PRO RES 171 1.00 0.00 GLU RES 172 1.00 0.00 SER RES 173 1.00 0.00 TYR RES 175 0.67 0.00 LEU RES 176 1.00 0.00 PRO RES 252 1.00 0.00 PHE RES 254 1.00 0.00 TRP RES 255 1.00 0.00 PRO RES 256 1.00 0.00 GLY DNAE 2.696

opuCB/ YP_041888.1	1st site coordinates predicted by Pocket-Finder Min Coords: (-35, -74, 64) Max Coords: (-17, -62, 79)
	Best pocket residues predicted by FINDSITE comb TEMPLATE 3PUVG03 TEMPLATE 3PUXG03 TEMPLATE 3PUWG01 TEMPLATE 3QBIB02 TEMPLATE 1QBZA00 TEMPLATE 1IW6A00 TEMPLATE 1CWQB08 RES 32 0.43 0.00 ILE RES 35 0.43 0.00 VAL RES 36 0.57 0.00 PRO RES 128 0.57 0.00 LEU RES 129 0.71 0.00 PRO RES 132 0.71 0.00 LEU DNAE 2.093
kdpA/ YP_041527.1	1st site coordinates predicted by Pocket-Finder Min Coords: (-7, -16, -15) Max Coords: (9, -3, 5)
	Best pocket residues predicted by FINDSITE comb TEMPLATE 1M75B00 TEMPLATE 1F0YA00 TEMPLATE 1F12A00 TEMPLATE 3HDHA00 RES 8 1.00 0.00 THR RES 9 1.00 0.00 MET RES 12 1.00 0.00 MET RES 16 1.00 0.00 VAL RES 20 0.75 0.00 TYR RES 46 1.00 0.00 ILE DNAE 1.749
sirA/ YP_039582.1	1st site coordinates predicted by Pocket-Finder Min Coords: (-5, -4, -16) Max Coords: (17, 11, -1)
	Best pocket residues predicted by FINDSITE comb TEMPLATE 3MWFA00 TEMPLATE 3NU1B00 TEMPLATE 3NU1A00 TEMPLATE 2J6IA00 RES 81 1.00 0.00 TRP RES 82 0.50 0.00 THR RES 84 0.50 0.00 LYS RES 104 0.50 0.00 PRO RES 124 0.50 0.00 VAL RES 125 0.50 0.00 ARG RES 145 0.50 0.00 VAL RES 201 1.00 0.00 ARG RES 206 0.75 0.00 ARG RES 208 0.75 0.00 TYR RES 238 0.50 0.00 ILE RES 258 0.50 0.00 VAL RES 260 0.50 0.00 SER RES 262 1.00 0.00 PRO RES 263 1.00 0.00 ASN RES 299 0.50 0.00 ASP RES 300 0.50 0.00 GLU RES 301 0.50 0.00 ILE RES 304 0.50 0.00 ASN RES 305 0.50 0.00 LEU DNAE -0.912
Hypothetical protein/ YP_042767.1	1st site coordinates predicted by Pocket-Finder Min Coords: (-12, 12, -19) Max Coords: (4, 31, 2)
	Best pocket residues predicted by FINDSITE comb TEMPLATE 2C2FA RES 6 1.00 -0.20 TRP RES 18 1.00 -0.20 VAL RES 19 1.00 -0.20 GLY RES 21 1.00 -0.20 ILE RES 22 1.00 -0.29 LYS DNAE 0.634

DNAE stands for DNA binding pockets (-ve value is most favorable)

Table 8: Best pockets and its coordinates.

References

1. Fong C, Rohmer L, Radey M, Wasnick M, Brittnacher MJ (2008) PSAT: a web tool to compare genomic neighborhoods of multiple prokaryotic genomes. BMC Bioinformatics 9: 170.
2. Zhang R, Lin Y (2009) DEG 5.0, a database of essential genes in both prokaryotes and eukaryotes. Nucleic Acids Res 37: D455-D458.
3. NCBI Homo sapiens (human) Protein BLAST.
4. Yu CS, Lin CJ, Hwang JK (2004) Predicting subcellular localization of proteins for Gram-negative bacteria by support vector machines based on n-peptide compositions. Protein Sci 13: 1402-1406.
5. Yu NY, Wagner JR, Laird MR, Melli G, Rey S, et al. (2010) PSORTb 3.0: Improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics 26: 1608-1615.
6. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M (2012) KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res 40: D109-D114.
7. Kelley LA, Sternberg MJ (2009) Protein structure prediction on the Web: a case study using the Phyre server. Nat Protoc 4: 363-371.
8. Rodrigues JP, Levitt M, Chopra G (2012) KoBaMIN: a knowledge-based minimization web server for protein structure refinement. Nucleic Acids Res 40: W323-W328.
9. Ko J, Park H, Heo L, Seok C (2012) GalaxyWEB server for protein structure prediction and refinement. Nucleic Acids Res 40: W294-W297.
10. Rampage.
11. Colovos C, Yeates TO (1993) Verification of protein structures: patterns of nonbonded atomic interactions. Protein Sci 2: 1511-1519.
12. TMHMM v 2.0.
13. Chen J, Liu H, Yang J, Chou KC (2007) Prediction of linear B-cell epitopes using amino acid pair antigenicity scale. Amino Acids 33: 423-428.
14. Singh H, Raghava GP (2003) ProPred1: prediction of promiscuous MHC Class-I binding sites. Bioinformatics 19: 1009-1014.
15. Singh H, Raghava GP (2001) ProPred: prediction of HLA-DR binding sites. Bioinformatics 17: 1236-1237.
16. Guan P, Doytchinova IA, Zygouri C, Flower DR (2003) MHCPred: bringing a quantitative dimension to the online prediction of MHC binding. Appl Bioinformatics 2: 63-66.
17. Kangueane P, Sakharkar MK (2005) T-Epitope Designer: A HLA-peptide binding prediction server. Bioinformation 1: 21-24.
18. Doytchinova IA, Flower DR (2007) VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC Bioinformatics 8: 4.
19. Zhou H, Skolnick J (2013) FINDSITE(comb): a threading/structure-based, proteomic-scale virtual ligand screening approach. J Chem Inf Model 53: 230-240.
20. Lagorce D, Sperandio O, Galons H, Miteva MA, Villoutreix BO (2008) FAF-Drugs2: free ADME/tox filtering tool to assist drug discovery and chemical biology projects. BMC Bioinformatics 9: 396.
21. Osiris property explorer.
22. Hendlich M, Rippmann F, Barnickel G (1997) LIGSITE: automatic and efficient detection of potential small molecule-binding sites in proteins. J Mol Graph Model 15: 359-363, 389.
23. Thomsen R, Christensen MH (2006) MolDock: a new technique for high-accuracy molecular docking. J Med Chem 49: 3315-3321.