Title

Structure based virtual screening identifies small molecule effectors for the sialoglycan binding protein Hsa.

Permalink

Journal

Authors

Agarwal, Rupesh
Bensing, Barbara
Mi, Dehui
et al.

Publication Date

DOI

Copyright Information

Peer reviewed

Structure based virtual screening identifies small molecule effectors for the sialoglycan binding protein Hsa

Rupesh Agarwal[1,2],

Barbara A. Bensing[3],

Dehui Mi[4],

Paige N. Vinson[5],

Jerome Baudry[6],

Tina M. Iverson[7],

Jeremy C. Smith[1,8]

[1]UT/ORNL Center for Molecular Biophysics, Oak Ridge National Laboratory, Tennessee 37831-6309

[2]Graduate School of Genome Science and Technology, University of Tennessee, Knoxville, Tennessee 37996

[3]Division of Infectious Diseases, Veterans Affairs Medical Center, Department of Medicine, University of California, San Francisco, and the Northern California Institute for Research and Education, San Francisco, California 94121

[4]Vanderbilt Institute of Chemical Biology, High Throughput Screening Facility, Vanderbilt University, Nashville, TN 37232

[5]Past affiliation: Department of Biochemistry and Vanderbilt Institute of Chemical Biology, Vanderbilt University, Nashville, TN 37232; Present affiliation: Department of Pharmacology and Vanderbilt Center for Neuroscience Drug Discovery, Vanderbilt University, Nashville, TN 37232

[6]Department of Biological Sciences, The University of Alabama in Huntsville, Huntsville, Alabama 35899

[7]Departments of Pharmacology and Biochemistry, Center for Structural Biology and Vanderbilt Institute of Chemical Biology, Vanderbilt University, Nashville, TN 37232

[8]Department of Biochemistry and Cellular and Molecular Biology, University of Tennessee, Knoxville, Tennessee 37996

Abstract

Infective endocarditis (IE) is a cardiovascular disease often caused by bacteria of the *viridans* group of streptococci, which includes *Streptococcus gordonii* and *Streptococcus sanguinis*.

Previous research has found that a serine-rich repeat (SRR) proteins on the *S. gordonii* bacterial surface play a critical role in pathogenesis by facilitating bacterial attachment to sialylated glycans displayed on human platelets. Despite its important role in disease progression, there are currently no anti-adhesive drugs available on the market. Here, we performed structure-based virtual screening using an ensemble docking approach followed by consensus scoring to identify novel small molecule effectors against the sialoglycan binding domain of the SRR adhesin protein Hsa from the *S. gordonii* strain DL1. The screening successfully predicted nine compounds which were able to displace the native ligand (sialyl-T antigen) in an *in vitro* assay and bind competitively to Hsa. Furthermore, hierarchical clustering based on the MACCS fingerprint showed that eight of these small molecules do not share a common scaffold with the native ligand. This study indicates that SRR family of adhesin proteins can be inhibited by diverse small molecules and thus prevent the interaction of the protein with the sialoglycans. This opens new avenues for discovering potential drugs against infective endocarditis.

Introduction

Infective endocarditis (IE) (or bacterial endocarditis (BE)) is a life-threatening infection of cardiac valves and the interior surface of the heart (endocardium)[1]. Oral streptococci account for ~17–45% of all cases of IE[2, 3]. If untreated, infection destroys the valves and results in heart failure[4–6]. Moreover, bacteria may also form clots (emboli) that enter the blood stream and produce strokes. IE affects 10,000–20,000 patients in the US every year and is associated with an in-hospital mortality rate of ~20% and a five year mortality rate of ~40–70 %[1, 7, 8]. Treatment for endocarditis currently requires prolonged antimicrobial therapy, often combined with surgery. The rise in antibiotic resistance[9] has limited our pharmacological options[10, 11], and resistant organisms have increased the mortality rate[12]. Although medical therapy alone often resolves infection, 47% or more of the patients eventually require valve replacement due to the damage incurred[5, 6, 13]. Given the associated morbidity and rising mortality rate, there is an urgent need to develop novel therapies against IE.

Previous studies have reported that the binding of bacteria to host platelets contributes to the colonization of damaged aortic valves[4–6]. A cell wall-anchored serine-rich repeat (SRR) protein mediates the adherence of *S. gordonii* and *S. sanguinis* to sialoglycans displayed on the human platelet[14] glycoprotein GPIb[15, 16]. SRR proteins have been demonstrated to be virulence factors for endocarditis[15, 16], and disrupting the interaction between the SRR protein and sialoglycans on host platelets may therefore reduce virulence. *Streptococcus gordonii* is one of the well-studied species that cause IE and is a normal component of the human oral microbiota[7, 8, 17]. Platelet binding by *S. gordonii* strains M99 and DL1 are facilitated by the homologous SRR proteins GspB and Hsa, respectively[18]. Although these adhesins have high sequence identity, their ligand binding regions (BRs) differ significantly and have different sialoglycan selectivity[15, 19, 20]. GspB binds with narrow selectivity to sialyl-T antigen (sTa) whereas Hsa binds promiscuously to a range of glycans[15, 19, 20]. It is unclear whether an inhibitor could impact infection by the time cases are identified. Sub-acute infective endocarditis disease develops over a period of weeks and by the time disease is identified vegetations that are a mixture of host and bacterial material are well

established. Anti-adhesive therapies have been explored for the treatment of a wide range of other bacterial infections[21–23], but have not yet been pursued for IE. Anti-adhesives can, in principle, complement traditional antibiotics and improve their efficacy, potentially eliminating the need for surgical intervention. Moreover, an inhibitor might also reduce "re-seeding" (bacteremia is a hallmark of IE), or could be used as a prophylactic in some situations. Additionally, anti-adhesive agents are not bactericidal and hence the propagation and spread of resistant strains is much less likely to occur than as a result of exposure to bactericidal agents, such as antibiotics.

The crystal structures of the BRs (Hsa$_{BR}$ (PDB 6EFC)[24], GspB$_{BR}$ (PDB 6EFA)[24], 10712$_{BR}$ (PDB 6EFF)[24], SK150$_{BR}$ (PDB 6EFB)[24], SrpA$_{BR}$ (PDB 5EQ2)[25], and SK678$_{BR}$ (PDB 6EFI)[24]) from a number of *S. gordonii* and *S. sanguinis* SRR proteins have been solved[24–26]. These all have two domains which are associated with sialoglycan binding: the Siglec (Sialic acid-binding immunoglobulin-like lectin) domain and the Unique domain (for which function is not known completely). Furthermore, recent studies have identified that the three loops (CD, EF and FG) adjacent to the sialoglycan binding site are critical for the affinity and selectivity between ligands[24]. Additionally, it has been reported that a partially conserved "YTRY" motif in the binding site is necessary for formation of hydrogen bond interactions with the sialic acid of the native ligand[24] and a crystal structure of Hsa$_{BR}$ bound to sTa (PDB 6EFD[24]) has been resolved showing these interactions. Importantly, there are also human sialoglycan-binding proteins[27], that contain a sialoglycan binding site but the site differs significantly in both geometry and in the location of hydrogen-bonding donors and acceptors from that found in the streptococcal Siglec-like proteins[28]. Moreover, the mode of interaction with sialoglycans is distinct between human Siglecs and bacterial sialoglycan binding adhesin proteins[28].

The above structural information can be leveraged by structure-based approaches for the identification of molecular effectors targeting SRR adhesin proteins. Here, we targeted one of the SRR proteins, Hsa (Hsa$_{BR}$), using the *in-silico* virtual screening of ~105,000 small molecules, with the goal of discovering small molecules that bind to Hsa$_{BR}$ and that disrupt its interactions with sialoglycan ligands. Hsa$_{BR}$ is a good starting structure for a chemical biology approach that would could later be expanded toward inhibitors of the entire family since it well-characterized and is known to bind promiscuously to many glycans[24]. Moreover, we have determined crystal structures of the Hsa sialoglycan-binding domain both in an unliganded state and in complex with di- tri- and tetrasaccharide sialoglycan ligands[24].

Here, instead of using only the crystal structure in our computational approaches, we used molecular dynamics (MD) simulations to describe the flexibility of the binding pocket and generate an ensemble of protein conformations, which has been shown to yield large and diverse sets of molecular effectors to control protein functions [29, 30]. Following subsequent high throughput ensemble docking, we prioritized the compounds using consensus scoring, which has previously shown to reduce the number of false positives and increase the success rate[31]. To further improve our predictions, we cross screened the compounds against the BRs from five Hsa homologues and identified compounds which bound to Hsa$_{BR}$ with relatively higher docking scores compared to other BRs. From our virtual screening

predictions, we were able to achieve a high success rate of ~20%, finding that 9 out of 50 compounds that were suggested for experimental validation were indeed able to displace the native ligand from the Hsa$_{BR}$ binding pocket. Moreover, we were also able to identify scaffolds distant from the native ligand that bind to Hsa$_{BR}$. To our knowledge, these are the first small molecules described to inhibit binding of SRR family of adhesin protein to its native sialoglycan

Methods

System preparation and molecular dynamics simulation

Crystal structures of the sialoglycan binding proteins Hsa$_{BR}$ (PDB 6EFC)[24], GspB$_{BR}$ (PDB 6EFA)[24], 10712$_{BR}$ (PDB 6EFF)[24], SK150$_{BR}$ (PDB 6EFB)[24], SrpA$_{BR}$ (PDB 5EQ2)[25], and SK678$_{BR}$ (PDB 6EFI)[24] were used in this study. Molecular dynamics (MD) simulations was performed on all these proteins using the Amber14 ff14SB force-field parameters[32, 33]. Each of these proteins were solvated in a TIP3P[34] octahedral box with a 10 Å buffer of water around the protein in each direction. First, the protein structure was held fixed with a force constant of 500 kcal mol^{-1} Å^{-2} while the system was minimized with 500 steps of steepest descent followed by 500 steps with the conjugate gradient method. In the second minimization step, the restraints on the protein were removed and 1000 steps of steepest descent minimization were performed followed by 1500 steps of conjugate gradient. The system was heated to 300 K while holding the protein fixed with a force constant of 10 kcal mol^{-1} Å^{-2} for 1000 steps. Then, the restraints were removed, and 1000 MD steps were performed. The SHAKE algorithm[35] was used to constrain all bonds involving hydrogen in the simulations. 200 ns production MD was performed at 300 K using the NPT ensemble and a 2 fs time step with nonbonded cutoff of 10 Å. The temperature was fixed with the Langevin dynamics thermostat[36] and the pressure was fixed with a Monte Carlo barostat[37]. Similar MD simulation protocol was used on all the adhesin. This procedure yielded a total of 20,000 snapshots for subsequent analyses. Three independent runs were performed for each simulation.

In silico screening

Ensemble docking is an *in-silico* structure-based chemical biology method using an 'ensemble' of protein target conformations to discover novel protein effectors[29]. The workflow used is shown in Fig. 1. The ensemble was constructed by clustering snapshots from molecular dynamics (MD) simulation trajectories by root mean square deviation (RMSD) of the binding pocket residues and loops (Table S1) surrounding the binding pocket with the hierarchical agglomerate clustering algorithm using Cpptraj module[38].

The Vanderbilt small molecule collection ("The Discovery Collection") containing ~105,000 compounds was docked to an ensemble of 5 conformations (4 representative structures obtained from clustering from MD and 1 crystal structure) with a cubic box with edges of ~30 Å. This small molecule library has been used in multiple high-throughput screens resulting in hits that have moved to hit-to-lead stages of early drug discovery programs[39–41]. The docking box was centered on the Cα atom of conserved residue THR 339 (Hsa numbering). VinaMPI[42], a parallel version of AutodockVina[43], was used

to perform the *in silico* screening. A similar docking procedure was performed for all the adhesin proteins. The docked poses were then ranked by the AutodockVina scoring function[44]. The compounds were not only screened for Hsa_{BR} but also cross screened with 5 adhesin proteins ($GspB_{BR}$, 10712_{BR}, $SK150_{BR}$, $SrpA_{BR}$, $SK678_{BR}$). The cross screening was performed to remove the promiscuous compounds or "frequent hitters" (i.e., compounds which are always scored high for all the target) and thus reducing the number of false positives[45]. However, it must be noted that the goal of the cross screening was not to get selectivity towards Hsa_{BR}.

From this ranked list of compounds, we tested compounds which were within the top 1% (~1050 compounds) for Hsa_{BR} but not within the top 1% of the other 5 BRs ($GspB_{BR}$, 10712_{BR}, $SK150_{BR}$, $SrpA_{BR}$, $SK678_{BR}$) and narrowed the list down to 250 compounds.

We note that we only experimentally tested binding to Hsa_{BR} and not the selectivity of the predicted binders. Next, the resulting ~250 compounds were refined and rescored using two MOE scoring functions[46]. The non-polar hydrogens (not included in Vina docking protocol) were added before performing the "induced fit" docking protocol in MOE[46]. The docking poses were ranked using "GBVI-WSA dG" and "Affinity DG" scoring functions[46]. Using consensus scoring, the top 50 compounds were then suggested for experimental validation. A flowchart of the screening methodology used is shown in Fig. 1.

Cheminformatics

All the physicochemical properties and fingerprints of small molecules were calculated using combination of MOE[46], ChemBioServer[47] and RDkit[48]. MACCS fingerprints were calculated for each compound and similarity between them were compared with the Tanimoto coefficient, followed by hierarchical clustering to cluster the similarity matrix.

Experimental assays

Protein expression and purification

GST-tagged Hsa_{BR} was expressed and purified as described in ref[20]. GST-Hsa_{BR} was expressed under the control of the pGEX-3X vector in *E. coli* BL21 (*DE3*) in a Terrific Broth medium with 50 μg/ml kanamycin at 37 °C. When the OD_{600} reached 1.0, expression was induced with 1 mM IPTG at 24 °C for 5 hrs. Cells were harvested by centrifugation at $5,000 \times g$ for 15 min, optionally washed with 0.1 M Tris-HCl, pH 7.5, and stored at –20 °C before purification. The frozen cells were resuspended in homogenization buffer (20–50 mM Tris-HCl, pH 7.5, 150–200 mM NaCl, 1mM EDTA, 1 mM PMSF, 2 μg/ml Leupeptin, 2 μg/ml Pepstatin) then disrupted by sonication. The lysate was clarified by centrifugation at $38500 \times g$ for 35–60 min and passed through a 0.45 μm filter. Benchtop purification was performed at 4 °C using Glutathione Sepharose 4B beads, with pure GST-Hsa were eluted with 30 mM GSH in 50 mM Tris-HCl, pH 8.0.

AlphaScreen high-throughput screening assay

We used the AlphaScreen modification of an ELISA as the primary target-based proximity assay to monitor ligand displacement. AlphaPlate (Cat # PE 6005351, Lot # 8220–16081)

with 384-well was used for the screening. In the experimental setup, biotinylated sialyl T antigen (sTa) was coupled to a streptavidin donor bead and GST-tagged Hsa was coupled to an anti-GST conjugated acceptor bead in PBS (phosphate buffered saline). The reaction was excited at 680 nm to stimulate singlet oxygen-mediated energy transfer to the acceptor bead, which can be detected at 615 nm.

The dose-dependent signal reflects the number of bead-coupled adhesins bound to bead-coupled glycans. To determine the optimal ratio of Hsa-GST to biotinylated sTa for occupying binding sites on the beads and, therefore, maximal signal production, the Hsa-GST concentration was titrated in a 10 point-3 fold dilution starting from 1000 nM and the biotinylated sTa concentration was titrated in a 9 point-3 fold dilution starting from 100 nM and the resulting Alpha signal measured. The maximal signal, representing the "hooking point" where either the donor or acceptor beads are saturated, was found to be 3 nM for Hsa-GST and 3 nM for biotinylated sTa. The final chosen concentrations used in the screen was 1 nM of Hsa-GST and 2 nM of biotinylated sTa, slightly below the hook point, to avoid potential excess Hsa that may sequester inhibitors and interfere with signal disruption. DMSO was used as the negative control and unbiotinylated sTa was used as a positive control at a concentration of 30 uM which was determined to provide maximal disruption of the Alpha signal.

We applied this assay to the evaluation of the test compounds that were predicted as binding to Hsa_{BR} using virtual screening. This initial screen was performed with all test compounds in duplicate at a final concentration of 10 μM and DMSO was used as the negative control and unbiotinylated sTa was used as a positive control.

Z' factor calculation

The Z' factor[49] is an indicator of high throughput screening assay performance and was calculated as follows:

$$Z' = 1 - 3(\delta_p + \delta_n)/\left|(\mu_n - \mu_p)\right|$$

The standard deviations and means of the positive (p) and negative (n) controls are denoted by δ_p, μ_p and δ_n, μ_n respectively. DMSO and untagged sTa are the positive and negative control respectively.

Effector identification analyses

The alpha value of each test compound was measured and was filtered using 1-fold, 2-fold or 3- fold of either standard deviation (SD) from the mean of the negative control group, or absolute deviation from the median (MAD) of negative control group. We determined which tested points lay outside the mean of the vehicle control (there were 9 replicates of the negative control (DMSO), 4 replicates of the positive control (untagged sTa)). We used a threshold of both 3 SD and 3 MAD from the negative control group. This was followed by taking the union of the two. Then, it was further filtered by only keeping those molecules which hit twice in the confirmation duplicates. Finally, we calculated the Percentage Control from the control group to identify compounds that disrupted the Hsa-sTa interaction. This

serves as an initial hit identifier but should be followed-up to confirm true actives and rule out false positives.

Percentage control (PC) calculated as follows:

$$\text{Percentage control (PC)} = 1 - \frac{\alpha_{NC} - \alpha_{Com}}{\alpha_{NC} - \alpha_{PC}}$$

where, α is the average alpha value for negative control (α_{NC}), positive control (α_{PC}) and compounds tested (α_{Com}). PC is a measure of the alpha signal of the 10 μM test compound in percentage of the controls.

Results and Discussion

I. Protein dynamics and conformations

We used MD simulations to capture the internal dynamics of the proteins and find binding site conformations not seen in the crystal structure[50]. We calculated the root mean square fluctuation (RMSF) to identify the flexible regions (Fig 2a). Although the overall structure of the Siglec domain is rigid, we observed that the loops (CD, EF and FG) close to the binding pocket are flexible for all the adhesin proteins (Figs 2a, S1). In the case of Hsa$_{BR}$, we observed that the CD and EF loops constitute the most flexible region of the protein. Moreover, critical binding pocket residues other than in these loops were identified from the crystal structure of Hsa$_{BR}$ and the native ligand (sTa) (Table S1).

To capture new conformations of the binding pocket that deviate from the initial crystal structure, the root mean square deviation (RMSD) of the loop residues and other critical residues (previously known to bind to the native ligand) (Table S1) were used to cluster the MD trajectories. The clustering resulted in four different clusters. The structure closest to the centroid of each cluster was used for docking. The "ensemble" of structures obtained from clustering and crystal structure were superimposed to observe the deviation of structures (**as shown in** Fig 2b). We observe that all the structures had similar secondary and backbone structures and the RMSD$_{Calpha}$ of the Siglec domain was calculated to be within ~1.5 Å. However, as seen in the superimposed structures (Fig 2b), the loop regions (especially CD and EF loop) have different orientations in the "ensemble" when compared to the crystal structure. Similarly, we observed that the side chains in the binding pocket residues orient differently between the structures, which can be critical for rigid body docking.

II. Physicochemical properties of small molecule database

The five structures obtained from MD simulations and the existing crystal structure were screened against the Vanderbilt small molecule collection ("The Discovery Collection") containing ~105K compounds. Although, this database has been used in several early drug discovery programs[39–41], it has not been characterized yet. Therefore, before performing the virtual screening, and although the goal of the present work is only to identify a molecular effector, we calculated some useful physicochemical properties. Firstly, we calculated the molecular weight (MW) of the compounds (Fig 3a), which is known to be critical for

safety and tolerability reasons[51]. The Vanderbilt database has compounds with MW less than 500 Da that are considered to improve druglikeness[52, 53] and also has low MW compounds (<300 Da) that are considered better initial precursor because they serve as effective chemical starting points for lead optimization[54]. The polar surface area and the number of rotatable bonds have been found to better discriminate between compounds that are orally active[55]. It has been predicted that compounds with 10 or fewer rotatable bonds and those having a polar surface area of less than 140 Å^2 have a good oral bioavailability[55]. In our database, we observed that most compounds had a mean polar surface area of ~150 Å^2 and less than 10 rotatable bonds (Figs 3b, c). Lipophilicity (SLogP) is another factor which is known to influence drug potency, pharmacokinetics, and toxicity[52, 56]. Compounds with SLog P values between −0.4 to +5.6 range are known to be more "druglike"[53, 57]. Here, we found that most of the compounds fall within this range (Fig 3d). Although the above is a set of physicochemical properties that are considered to be important for different aspects of druggability, there have been numerous FDA approved drugs which violate one or more of these rules[58, 59].

III.   Docking results and poses

After our virtual screening, we first ranked all the top poses for each compound based on the AutodockVina scoring function[44]. Subsequently, we selected those compounds (~250) that were in the top 1% for Hsa_{BR} but did not rank within the top 1% for any other adhesin protein ($GspB_{BR}$, 10712_{BR}, $SK150_{BR}$, $SrpA_{BR}$, $SK678_{BR}$). This was followed by implementing consensus scoring in which the poses (obtained from AutodockVina) were energy-minimized and then rescored using two MOE scoring functions[46] (as mentioned in the Methods section). In the end, compounds that ranked within the top 50 for all the three scoring functions were suggested for experimental validation.

Next, we examined the number of electrostatic intermolecular interaction (hydrogen bonds (HBs) and Pi (π-π/ π-H/π-cation)) which are important for protein-ligand binding[60–62]. The importance of hydrogen bonding in drug design is well recognized and hydrogen-bonding capabilities deeply influence the transport and ADME (Adsorption, Distribution, Metabolism and Excretion) properties of a molecule as well as its specific interaction with biological receptors[63]. Many QSAR studies have been reported in which hydrogen-bonding interactions play a key role in modeling a particular target activity[64, 65]. Therefore, we calculated the number of compounds which form an interaction with the residues known to bind (or in close proximity) sTa to get information about which residues were targeted the most and are easily accessible to interact with a small molecule (Fig S2). Thr 339, Tyr 337 and Lys 335, which form HBs with the native ligand (sTa) in the crystal structure (PDB 6EFD)[24] as shown in Fig. 4a, are also some of the residues which form interactions (HBs or Pi) with majority of the compounds. Moreover, we found that majority of the compounds form 3 or more HBs or Pi interactions with the binding site residues, whereas 10 compounds make more than 5 HBs or Pi interactions (Fig. 4b). Furthermore, out of these 50 compounds, 25 compounds (50%) were predicted to bind to two of the "ensemble" structures generated from MD simulations with higher score than to the crystal structure. This further illustrates the usefulness of using ensemble docking.

IV. Experimental validation

Alpha assay screening was performed for the top 50 compounds predicted to displace sTa (the highest affinity native ligand) from Hsa_{BR}. The Z' factor value of the DMSO (negative control) versus untagged sTa (positive control) was 0.32, which denotes that there is a separation between the high and low signals of the assay in that 3x the sum of the standard deviations of the high and low signals of the assay divided by the difference of the mean of these two experimental groups is 0.68 (the error is relatively small compared to the separation of the mean of the two groups). After filtering the small molecules using the experimental data based on the percentage control (PC), nine small molecules were retained. These nine compounds showed a statistically significant decrease in the signal when the two replicates were averaged (Fig. 5). These compounds have a PC three standard deviations outside the mean of the negative control (DMSO).

The IC_{50} of the untagged sTa (positive control) was calculated as 8.67 μM (Fig. S3a). At 10μM concentration, the PC was 39% (Fig. S3b). At the same concentration, the PC of the 9 small molecules ranges from 23% to 70% and, out of these, two compounds have PC values of less than 39% and one has a PC of 41% (Table S2).

V. Computational and binding pose analyses of experimentally validated effectors

The nine small molecules were screened for 25 known toxic and carcinogenic fragments, such as anthracene, quinone, hydroquinone, butenone--Michael acceptor, chloroethane--Michael acceptor[47]. Of the 9 experimentally validated compounds (C1-C9) (Table 1), only Compound 1 (C1) was identified as potentially toxic, containing a benzo-dioxane and a catechol group. Moreover, to test the similarity between these effectors and the native ligand, fingerprint-based hierarchical clustering was performed. We found four clusters (**as shown in** Fig. S4), which showed that the compounds identified from the screen are diverse among themselves and are not similar to the native ligand. Additionally, we also tested the compounds for Lipinski's rule[53], to evaluate druglike-ness of the compounds. C4 was the only molecule with one violation (with 11 hydrogen bond acceptors), whereas all the other compounds satisfied all the 4 rules.

Following the above cheminformatics analyses of the experimentally validated effectors, we examined the computational models of the best poses and the interactions of the nine small molecules shown to have inhibition experimentally (Fig. S5 **and** Table 1). Interestingly, the inhibitor binding site is adjacent to the sTa binding site (Fig. 6) but both the sites does partially overlap. This site can be further explored for more selective inhibitors in future studies.

Next, we calculated interaction map of each of the nine effector and looked at the hydrogen bonds between the ligand and the side chain and backbone atoms of the binding site residues. In the models C1, C2, C4, and C5 form backbone HBs with Asp 255 and compounds C2–4 form backbone HBs with Val 367 while other compounds (C1, C5–9) form side chain HBs with Val 367. Interestingly, C3 forms four backbone HBs with Gly 362, Phe 366 and Val 367, and five sidechain HBs with Tyr 337, Thr 339, Ser 253, Asn 361 as shown in the interaction map (Fig. S5 **and** Table 1). Other residues that form HBs with most

of the compounds are Thr 339, Val 285 and Asn 361 (Fig. S5 **and** Table 1). The interaction maps of all the nine small molecule effectors are shown in Fig. S5. The orientation of the best docked pose of the nine validated effectors and crystal structure pose of the native ligand in the binding pocket are very different and bind in adjacent sites (Fig. 6). However, the residues that form HBs with the effectors, also form HBs with native ligand or are in close proximity of it. Hence, it is likely that these nine compounds are able to displace the native ligand in part because they form HBs with these critical residues.

Conclusions

The SRR protein Hsa has been considered an attractive molecular target for drug discovery due to its role in infective endocarditis (IE). It is noteworthy that there is no vaccine or anti-adhesive drug approved against IE. Hence, it is important to use chemical biology approaches to identify and describe how small molecules could inhibit the adhering function of the protein, which may, down the road, open the door to drug discovery. Here, we performed structure-based virtual screening to identify competitive small molecule effectors for Hsa_{BR}. We combined three different SBDD strategies; ensemble docking, cross screening, and consensus scoring in one pipeline. For the ensemble docking, we generated an ensemble of receptor conformations from MD simulation, and then cross screened against five homologs ($GspB_{BR}$, 10712_{BR}, $SK150_{BR}$, $SrpA_{BR}$, $SK678_{BR}$). In the last step, three scoring functions (AutodockVina[44] and MOE[46]) were used to rank and prioritize the list of compounds. The Vanderbilt database was used for the small molecules since it covers a wide distribution of different physicochemical properties.

The goal of combining these strategies was to improve the hit rate and reduce the number of false positives. Indeed, we were able to achieve a hit rate of ~20% and identified nine compounds that could displace the native ligand in the experimental assay. The binding poses of all the nine compounds identified from docking show that they are in close proximity with residues known to form HBs with the native ligand (sialyl-T antigen). These compounds may be used as a starting point for medicinal chemistry optimization. Further studies need to be conducted to characterize the binding affinities and poses of these identified compounds, and similar analyses for other sialoglycan-binding SRR proteins are ongoing.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

References

1. Hoen B; Duval X, Clinical practice. Infective endocarditis. N Engl J Med 2013, 368 (15), 1425–33. [PubMed: 23574121]

2. Murdoch DR; Corey GR; Hoen B; Miro JM; Fowler VG Jr.; Bayer AS; Karchmer AW; Olaison L; Pappas PA; Moreillon P; Chambers ST; Chu VH; Falco V; Holland DJ; Jones P; Klein JL; Raymond NJ; Read KM; Tripodi MF; Utili R; Wang A; Woods CW; Cabell CH; International Collaboration on Endocarditis-Prospective Cohort Study, I., Clinical presentation, etiology, and outcome of infective endocarditis in the 21st century: the International Collaboration on Endocarditis-Prospective Cohort Study. Arch Intern Med 2009, 169 (5), 463–73. [PubMed: 19273776]

3. Thornhill MH; Dayer MJ; Forde JM; Corey GR; Chu VH; Couper DJ; Lockhart PB, Impact of the NICE guideline recommending cessation of antibiotic prophylaxis for prevention of infective endocarditis: before and after study. BMJ 2011, 342, d2392.

4. Sullam PM; Sande MA, Role of platelets in endocarditis: clues from von Willebrand disease. J Lab Clin Med 1992, 120 (4), 507–9. [PubMed: 1402325]

5. Nilson B; Olaison L; Rasmussen M, Clinical presentation of infective endocarditis caused by different groups of non-beta haemolytic streptococci. Eur J Clin Microbiol Infect Dis 2016, 35 (2), 215–8. [PubMed: 26610338]

6. Walls G; McBride S; Raymond N; Read K; Coomarasamy C; Morris AJ; Chambers S; Holland D; Murdoch DR, Infective endocarditis in New Zealand: data from the International Collaboration on Endocarditis Prospective Cohort Study. N Z Med J 2014, 127 (1391), 38–51.

7. Dadon Z; Cohen A; Szterenlicht YM; Assous MV; Barzilay Y; Raveh-Brawer D; Yinnon AM; Munter G, Spondylodiskitis and endocarditis due to Streptococcus gordonii. Ann Clin Microbiol Antimicrob 2017, 16 (1), 68. [PubMed: 28978355]

8. Khalid N; Shlofmitz E; Ahmad SA, Aortic Valve Endocarditis. 2020.

9. Carratala J; Alcaide F; Fernandez-Sevilla A; Corbella X; Linares J; Gudiol F, Bacteremia due to viridans streptococci that are highly resistant to penicillin: increase among neutropenic patients with cancer. Clin Infect Dis 1995, 20 (5), 1169–73. [PubMed: 7619995]

10. Fowler VG Jr.; Boucher HW; Corey GR; Abrutyn E; Karchmer AW; Rupp ME; Levine DP; Chambers HF; Tally FP; Vigliani GA; Cabell CH; Link AS; DeMeyer I; Filler SG; Zervos M; Cook P; Parsonnet J; Bernstein JM; Price CS; Forrest GN; Fatkenheuer G; Gareca M; Rehm SJ; Brodt HR; Tice A; Cosgrove SE; Endocarditis S a.; Bacteremia Study, G., Daptomycin versus standard therapy for bacteremia and endocarditis caused by Staphylococcus aureus. N Engl J Med 2006, 355 (7), 653–65. [PubMed: 16914701]

11. Gould IM; Cauda R; Esposito S; Gudiol F; Mazzei T; Garau J, Management of serious meticillin-resistant Staphylococcus aureus infections: what are the limits? Int J Antimicrob Agents 2011, 37 (3), 202–9. [PubMed: 21300528]

12. Boucher H; Miller LG; Razonable RR, Serious infections caused by methicillin-resistant Staphylococcus aureus. Clin Infect Dis 2010, 51 *Suppl* 2 (S2), S183–97. [PubMed: 20731576]

13. Tornos P; Almirante B; Olona M; Permanyer G; Gonzalez T; Carballo J; Pahissa A; Soler-Soler J, Clinical outcome and long-term prognosis of late prosthetic valve endocarditis: a 20-year experience. Clin Infect Dis 1997, 24 (3), 381–6. [PubMed: 9114189]

14. Lizcano A; Sanchez CJ; Orihuela CJ, A role for glycosylated serine-rich repeat proteins in gram-positive bacterial pathogenesis. Mol Oral Microbiol 2012, 27 (4), 257–69. [PubMed: 22759311]

15. Bensing BA; Lopez JA; Sullam PM, The Streptococcus gordonii surface proteins GspB and Hsa mediate binding to sialylated carbohydrate epitopes on the platelet membrane glycoprotein Ibalpha. Infect Immun 2004, 72 (11), 6528–37. [PubMed: 15501784]

16. Plummer C; Wu H; Kerrigan SW; Meade G; Cox D; Ian Douglas CW, A serine-rich glycoprotein of Streptococcus sanguis mediates adhesion to platelets via GPIb. Br J Haematol 2005, 129 (1), 101–9. [PubMed: 15801962]

17. Takamatsu D; Bensing BA; Cheng H; Jarvis GA; Siboo IR; Lopez JA; Griffiss JM; Sullam PM, Binding of the Streptococcus gordonii surface glycoproteins GspB and Hsa to specific

carbohydrate structures on platelet membrane glycoprotein Ibalpha. Mol Microbiol 2005, 58 (2), 380–92. [PubMed: 16194227]

18. Takamatsu D; Bensing BA; Prakobphol A; Fisher SJ; Sullam PM, Binding of the streptococcal surface glycoproteins GspB and Hsa to human salivary proteins. Infect Immun 2006, 74 (3), 1933–40. [PubMed: 16495569]

19. Bensing BA; Khedri Z; Deng L; Yu H; Prakobphol A; Fisher SJ; Chen X; Iverson TM; Varki A; Sullam PM, Novel aspects of sialoglycan recognition by the Siglec-like domains of streptococcal SRR glycoproteins. Glycobiology 2016, 26 (11), 1222–1234. [PubMed: 27037304]

20. Deng L; Bensing BA; Thamadilok S; Yu H; Lau K; Chen X; Ruhl S; Sullam PM; Varki A, Oral streptococci utilize a Siglec-like domain of serine-rich repeat adhesins to preferentially target platelet sialoglycans in human blood. PLoS Pathog 2014, 10 (12), e1004540.

21. Cozens D; Read RC, Anti-adhesion methods as novel therapeutics for bacterial infections. Expert Rev Anti Infect Ther 2012, 10 (12), 1457–68. [PubMed: 23253323]

22. Ribic R; Mestrovic T; Neuberg M; Kozina G, Effective anti-adhesives of uropathogenic Escherichia coli. Acta Pharm 2018, 68 (1), 1–18. [PubMed: 29453908]

23. Ofek I; Hasty DL; Sharon N, Anti-adhesion therapy of bacterial diseases: prospects and problems. FEMS Immunol Med Microbiol 2003, 38 (3), 181–91. [PubMed: 14522453]

24. Bensing BA; Loukachevitch LV; Agarwal R; Yamakawa I; Luong K; Hadadianpour A; Yu H; Fialkowski KP; Castro MA; Wawrzak Z; Chen X; Baudry J; Smith JC; Sullam PM; Iverson TM, Selectivity and engineering of the sialoglycan-binding spectrum in Siglec-like adhesins. bioRxiv 2019, 796912.

25. Bensing BA; Loukachevitch LV; McCulloch KM; Yu H; Vann KR; Wawrzak Z; Anderson S; Chen X; Sullam PM; Iverson TM, Structural Basis for Sialoglycan Binding by the Streptococcus sanguinis SrpA Adhesin. The Journal of biological chemistry 2016, 291 (14), 7230–40. [PubMed: 26833566]

26. Pyburn TM; Bensing BA; Xiong YQ; Melancon BJ; Tomasiak TM; Ward NJ; Yankovskaya V; Oliver KM; Cecchini G; Sulikowski GA; Tyska MJ; Sullam PM; Iverson TM, A structural model for binding of the serine-rich repeat adhesin GspB to host carbohydrate receptors. PLoS Pathog 2011, 7 (7), e1002112.

27. Duan S; Paulson JC, Siglecs as Immune Cell Checkpoints in Disease. Annu Rev Immunol 2020, 38, 365–395. [PubMed: 31986070]

28. Loukachevitch LV; Bensing BA; Yu H; Zeng J; Chen X; Sullam PM; Iverson TM, Structures of the Streptococcus sanguinis SrpA Binding Region with Human Sialoglycans Suggest Features of the Physiological Ligand. Biochemistry 2016, 55 (42), 5927–5937. [PubMed: 27685666]

29. Amaro RE; Baudry J; Chodera J; Demir O; McCammon JA; Miao Y; Smith JC, Ensemble Docking in Drug Discovery. Biophys J 2018, 114 (10), 2271–2278. [PubMed: 29606412]

30. Evangelista W; Weir RL; Ellingson SR; Harris JB; Kapoor K; Smith JC; Baudry J, Ensemble-based docking: From hit discovery to metabolism and toxicity predictions. Bioorganic & medicinal chemistry 2016, 24 (20), 4928–4935. [PubMed: 27543390]

31. Charifson PS; Corkery JJ; Murcko MA; Walters WP, Consensus scoring: A method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. Journal of medicinal chemistry 1999, 42 (25), 5100–9. [PubMed: 10602695]

32. Case DA; Cheatham TE 3rd; Darden T; Gohlke H; Luo R; Merz KM Jr.; Onufriev A; Simmerling C; Wang B; Woods RJ, The Amber biomolecular simulation programs. Journal of computational chemistry 2005, 26 (16), 1668–88. [PubMed: 16200636]

33. Maier JA; Martinez C; Kasavajhala K; Wickstrom L; Hauser KE; Simmerling C, ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. J Chem Theory Comput 2015, 11 (8), 3696–713. [PubMed: 26574453]

34. Price DJ; Brooks CL 3rd, A modified TIP3P water potential for simulation with Ewald summation. J Chem Phys 2004, 121 (20), 10096–103.

35. Miyamoto S; Kollman PA, Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. Journal of computational chemistry 1992, 13 (8), 952–962.

36. Zwanzig R, Nonlinear generalized Langevin equations. Journal of Statistical Physics 1973, 9 (3), 215–220.

37. Åqvist J; Wennerström P; Nervall M; Bjelic S; Brandsdal BO, Molecular dynamics simulations of water and biomolecules with a Monte Carlo constant pressure algorithm. Chemical physics letters 2004, 384 (4–6), 288–294.

38. Roe DR; Cheatham TE 3rd, PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. J Chem Theory Comput 2013, 9 (7), 3084–95. [PubMed: 26583988]

39. Kharade SV; Kurata H; Bender AM; Blobaum AL; Figueroa EE; Duran A; Kramer M; Days E; Vinson P; Flores D; Satlin LM; Meiler J; Weaver CD; Lindsley CW; Hopkins CR; Denton JS, Discovery, Characterization, and Effects on Renal Fluid and Electrolyte Excretion of the Kir4.1 Potassium Channel Pore Blocker, VU0134992. Mol Pharmacol 2018, 94 (2), 926–937. [PubMed: 29895592]

40. Bewley BR; Spearing PK; Weiner RL; Luscombe VB; Zhan X; Chang S; Cho HP; Rodriguez AL; Niswender CM; Conn PJ; Bridges TM; Engers DW; Lindsley CW, Discovery of a novel, CNS penetrant M4 PAM chemotype based on a 6-fluoro-4-(piperidin-1-yl)quinoline-3-carbonitrile core. Bioorganic & medicinal chemistry letters 2017, 27 (18), 4274–4279. [PubMed: 28866269]

41. Kozek KA; Du Y; Sharma S; Prael FJ 3rd; Spitznagel BD; Kharade SV; Denton JS; Hopkins CR; Weaver CD, Discovery and Characterization of VU0529331, a Synthetic Small-Molecule Activator of Homomeric G Protein-Gated, Inwardly Rectifying, Potassium (GIRK) Channels. ACS Chem Neurosci 2019, 10 (1), 358–370. [PubMed: 30136838]

42. Ellingson SR; Smith JC; Baudry J, VinaMPI: facilitating multiple receptor high-throughput virtual docking on high-performance computers. Journal of computational chemistry 2013, 34 (25), 2212–21. [PubMed: 23813626]

43. Trott O; Olson AJ, AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. Journal of computational chemistry 2010, 31 (2), 455–61. [PubMed: 19499576]

44. Morris GM; Huey R; Lindstrom W; Sanner MF; Belew RK; Goodsell DS; Olson AJ, AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. Journal of computational chemistry 2009, 30 (16), 2785–91. [PubMed: 19399780]

45. McGovern S, Promiscuous Ligands. Comprehensive Medicinal Chemistry II 2007, 737–52.

46. Molecular Operating Environment (MOE) 2016; Chemical Computing Group Inc: 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2013.08.

47. Athanasiadis E; Cournia Z; Spyrou G, ChemBioServer: a web-based pipeline for filtering, clustering and visualization of chemical compounds used in drug discovery. Bioinformatics 2012, 28 (22), 3002–3. [PubMed: 22962344]

48. rdkit.org RDKit.

49. Zhang JH; Chung TD; Oldenburg KR, A Simple Statistical Parameter for Use in Evaluation and Validation of High Throughput Screening Assays. J Biomol Screen 1999, 4 (2), 67–73. [PubMed: 10838414]

50. Karplus M; Kuriyan J, Molecular dynamics and protein function. Proceedings of the National Academy of Sciences of the United States of America 2005, 102 (19), 6679–85. [PubMed: 15870208]

51. Hughes JP; Rees S; Kalindjian SB; Philpott KL, Principles of early drug discovery. Br J Pharmacol 2011, 162 (6), 1239–49. [PubMed: 21091654]

52. Leeson PD; Springthorpe B, The influence of drug-like concepts on decision-making in medicinal chemistry. Nature reviews. Drug discovery 2007, 6 (11), 881–90. [PubMed: 17971784]

53. Lipinski CA, Lead- and drug-like compounds: the rule-of-five revolution. Drug Discov Today Technol 2004, 1 (4), 337–41. [PubMed: 24981612]

54. Giordanetto F; Jin C; Willmore L; Feher M; Shaw DE, Fragment Hits: What do They Look Like and How do They Bind? Journal of medicinal chemistry 2019, 62 (7), 3381–3394. [PubMed: 30875465]

55. Veber DF; Johnson SR; Cheng HY; Smith BR; Ward KW; Kopple KD, Molecular properties that influence the oral bioavailability of drug candidates. J Med Chem 2002, 45 (12), 2615–23. [PubMed: 12036371]

56. Edwards MP; Price DA, Role of Physicochemical Properties and Ligand Lipophilicity Efficiency in Addressing Drug Safety Risks. Annu Rep Med Chem 2010, 45, 381–391.

57. Oprea TI, Property distribution of drug-related chemical databases. J Comput Aided Mol Des 2000, 14 (3), 251–64. [PubMed: 10756480]

58. Choy YB; Prausnitz MR, The rule of five for non-oral routes of drug delivery: ophthalmic, inhalation and transdermal. Pharm Res 2011, 28 (5), 943–8. [PubMed: 20967491]

59. Khanna V; Ranganathan S, Physiochemical property space distribution among human metabolites, drugs and toxins. BMC Bioinformatics 2009, 10 *Suppl* 15 (Suppl 15), S10.

60. Salentin S; Haupt VJ; Daminelli S; Schroeder M, Polypharmacology rescored: protein-ligand interaction profiles for remote binding site similarity assessment. Progress in biophysics and molecular biology 2014, 116 (2–3), 174–86. [PubMed: 24923864]

61. Sawada T; Fedorov DG; Kitaura K, Role of the key mutation in the selective binding of avian and human influenza hemagglutinin to sialosides revealed by quantum-mechanical calculations. Journal of the American Chemical Society 2010, 132 (47), 16862–72.

62. Imai YN; Inoue Y; Nakanishi I; Kitaura K, Cl-pi interactions in protein-ligand complexes. Protein science : a publication of the Protein Society 2008, 17 (7), 1129–37. [PubMed: 18434503]

63. Chen D; Oezguen N; Urvil P; Ferguson C; Dann SM; Savidge TC, Regulation of protein-ligand binding affinity by hydrogen bond pairing. Sci Adv 2016, 2 (3), e1501240.

64. Gancia E; Montana JG; Manallack DT, Theoretical hydrogen bonding parameters for drug design. J Mol Graph Model 2001, 19 (3–4), 349–62. [PubMed: 11449575]

65. Fujita T; Nishioka T; Nakajima M, Hydrogen-bonding parameter and its significance in quantitative structure--activity studies. Journal of medicinal chemistry 1977, 20 (8), 1071–81. [PubMed: 894678]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript
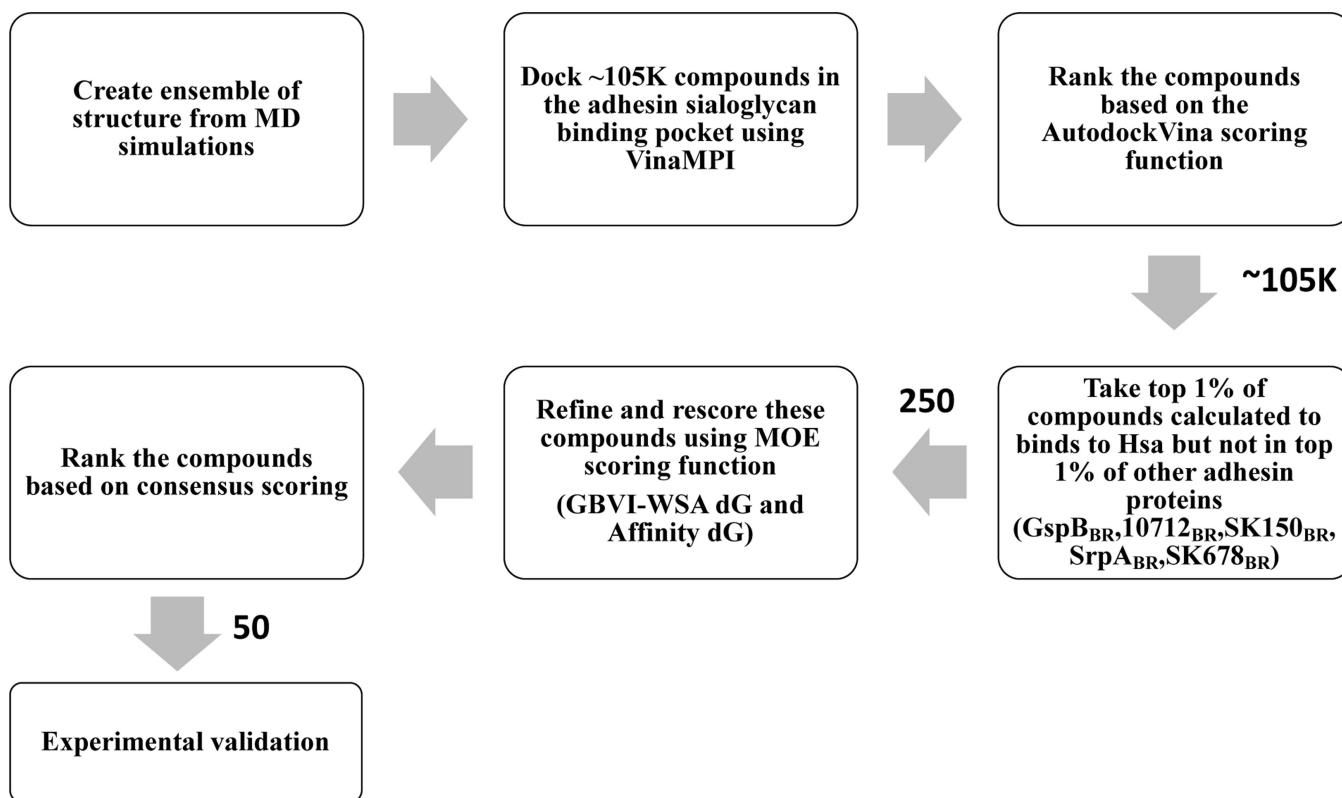
Figure 1:
Structure based virtual screening strategy workflow showing number of compounds which
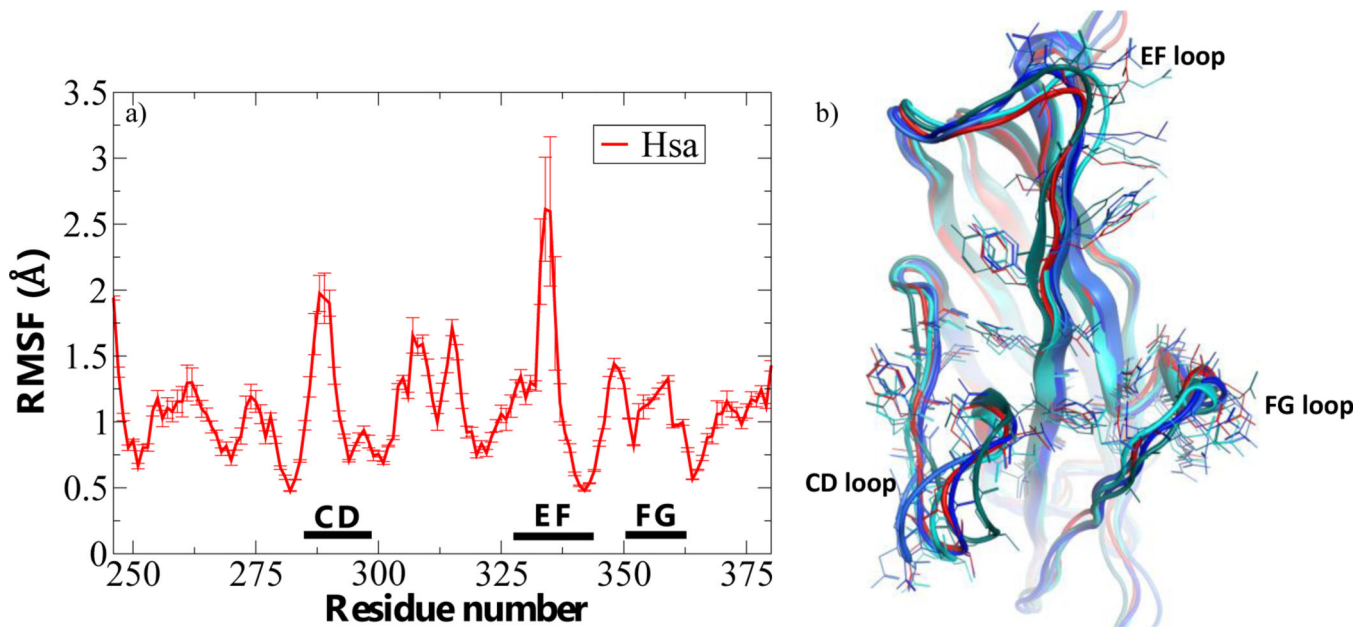were passed on to the next step.

Figure 2:
a) Root mean square fluctuation of Hsa$_{BR}$ from MD simulation showing DC, EF, FG loop regions; b) Superimposed structures (in ribbon) of Hsa$_{BR}$ obtained from different clusters showing residues (in stick) used to during the clustering: crystal structure (in red) and ensemble structure (in shade of blue).
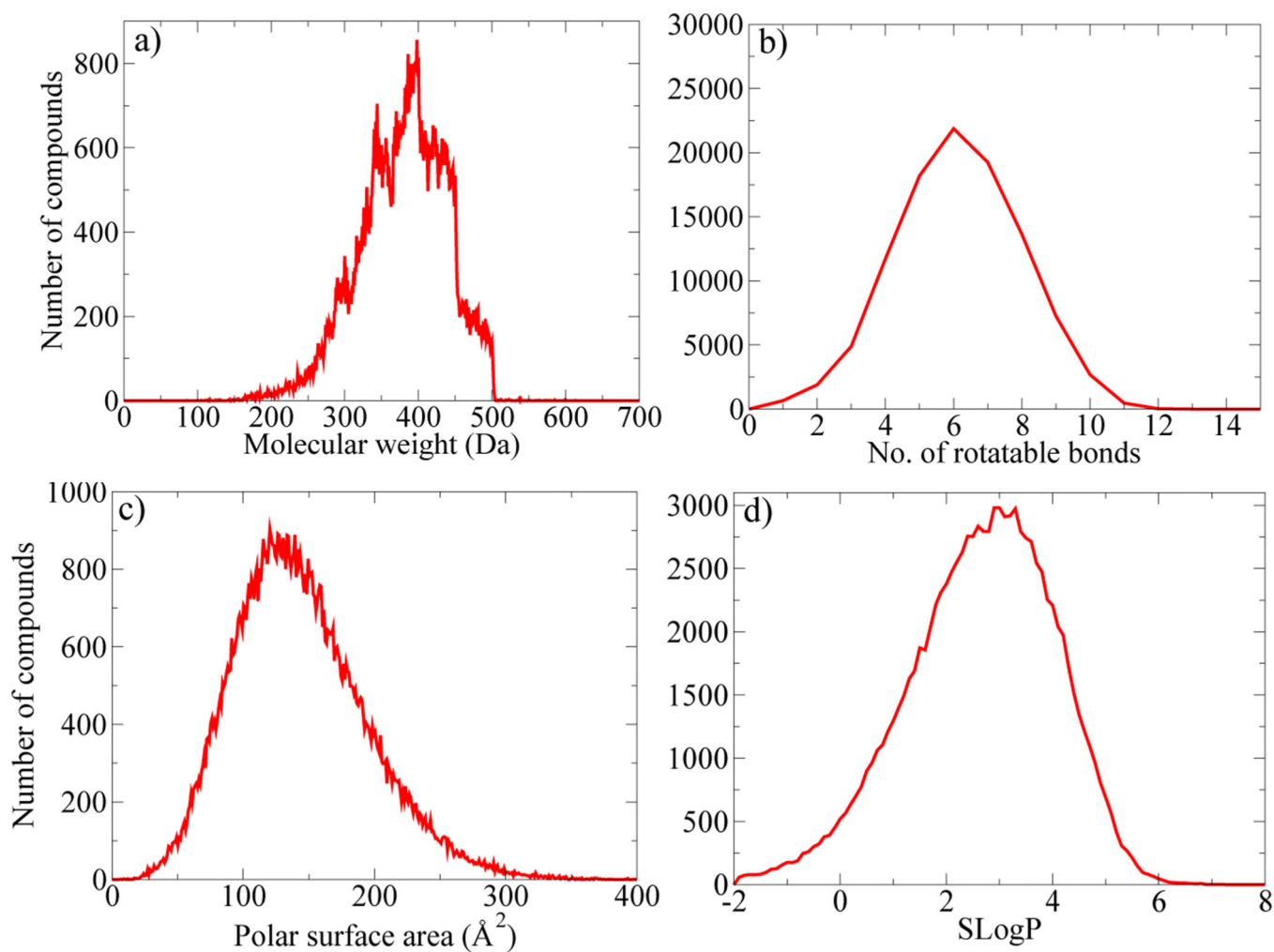
Figure 3:
Density profile of physicochemical properties of small molecule database: molecular weight
(a), number of rotatable bonds (b), polar surface area (c), and Log of the octanol/water
partition coefficient: SLogP (d).
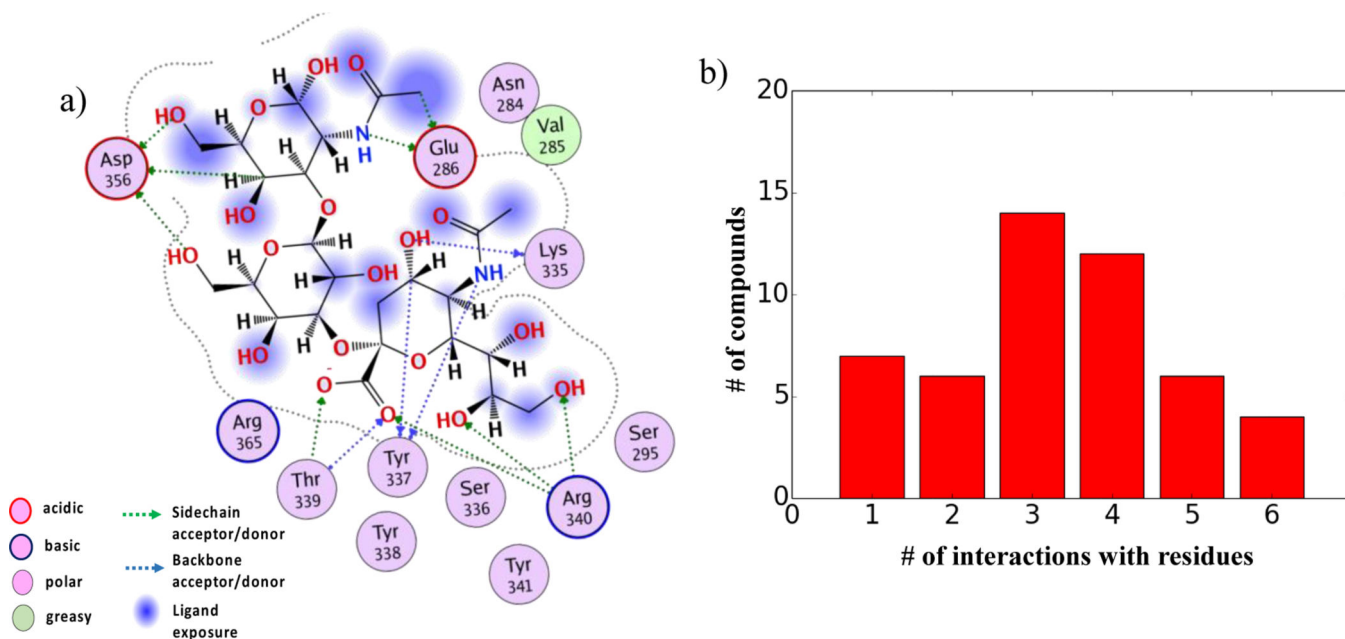
Figure 4:
a) Interaction map of native ligand (sTa) from crystal structure (PDB 6EFD)[24]; b) Bar plot showing number of compounds (within the top 50 compounds) and the number of interactions (Hbond or Pi (π-π/π-H/π-cation) made with protein residues
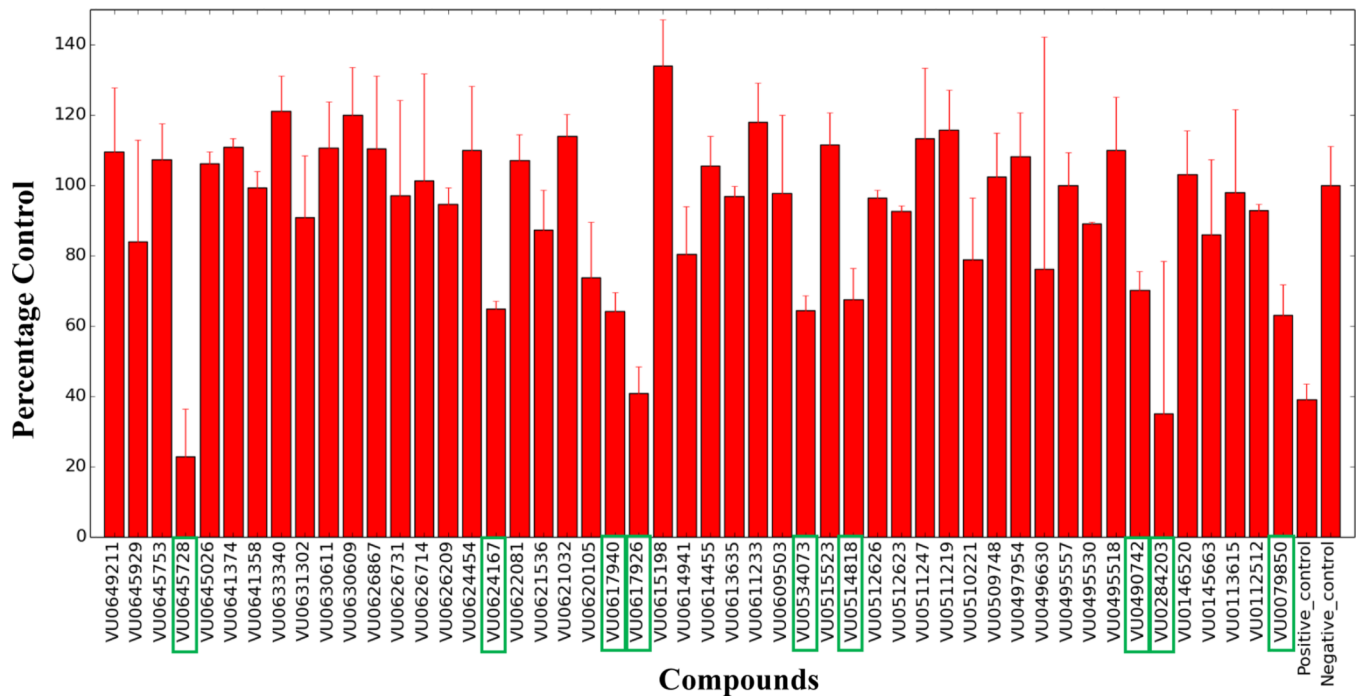
Figure 5:
Alpha Screen assay. Experimentally validated effectors are marked by green boxes on the x-axes. Error bar represent the standard deviation.
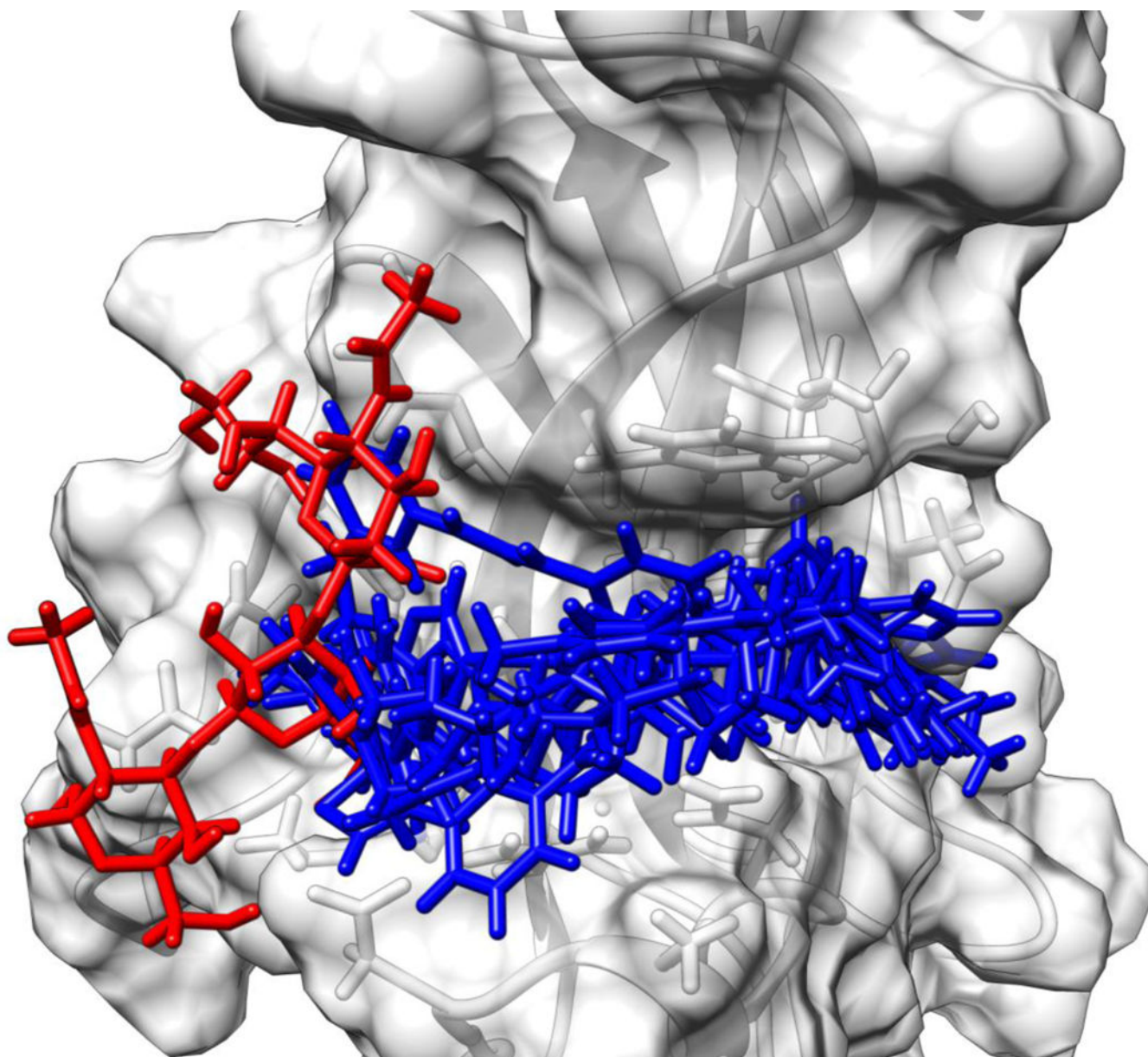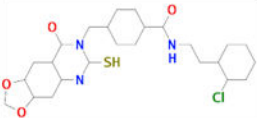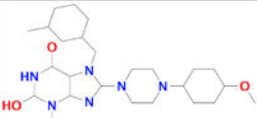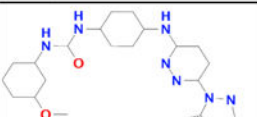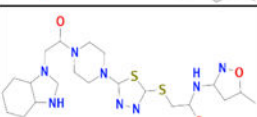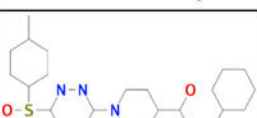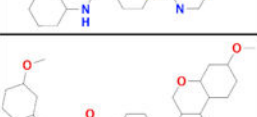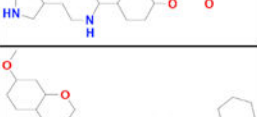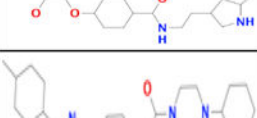
Figure 6:
Crystal structure pose of native ligand (sTa) (in red) and best docked pose of 9 validated compounds (in blue) in the binding pocket of Hsa$_{BR}$

Table 1:

Structure of the nine hits and interaction details in the binding pocket of the Hsa$_{BR}$

Compound number	Structure	ID	Protein residues forming Backbone HBs	Protein residues forming Sidechain HBs
C1		VU0079850	255	285,337,367
C2		VU0284203	255,367	361,367
C3		VU0490742	362,366,367	253,337, 339
C4		VU0645728	255,367	255,337,341,356
C5		VU0514818	255, 362	285,339,340,356 ,361,367
C6		VU0534073	-	255,285,339,361 ,365,367
C7		VU0617926	-	285,337,339,340 ,356,365,367
C8		VU0617940	362	285,337,339,367
C9		VU0624167	285	339,340,361,365 ,367