Molecular Insights on exquisitely Selective SrtA inhibitors towards active site loop forming open/close lid conformations in SrtA from *Bacillus anthracis*

Chandrabose Selvaraj¹, Gurudeeban Selvaraj², Satyavani Kaliyamurthi², Dong-Qing Wei^{2,3*}, and Sanjeev Kumar Singh^{1*}

¹Department of Bioinformatics, Computer Aided Drug Design, and Molecular Modeling Lab, Science Block, Alagappa University, Karaikkudi, 630 004, Tamil Nadu, India

²Center of Interdisciplinary Sciences-Computational Life Sciences, College of Food Science and Engineering, Henan University of Technology, Zhengzhou, 450001, China

³Department of Bioinformatics, The State Key Laboratory of Microbial Metabolism, College of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, 200240, China

For Correspondence:

* skysanjeev@gmail.com; dqwei@sjtu.edu.cn

*Denote both the authors are corresponding authors

Key words: *Bacillus anthracis*, cell adhesion, Molecular dynamics, Pharmacophore modeling, Sortase

Abbreviations: *Ba-Bacillus anthracis*, MD- Molecular Dynamics, SrtA – Sortase A

Abstract

The present study clearly explains the dependency of inhibitory activities in SrtA inhibitors is closely related to protein conformational changes of SrtA from *Bacillus anthracis*. Sortase A (SrtA) protein anchors proteins by recognizing a cell wall sorting signal containing the amino acid sequence LPXTG. In order to analyze conformational changes and the role of SrtA enzyme, especially the loop motions which situated proximal to the active site, molecular dynamic simulation was carried out for 100ns. Particular loop is examined for its various conformations from the MD trajectories and the open/close lid conformations are considered for the enzyme activity validations. Experimentally verified SrtA inhibitors activity was analyzed through 3D-QSAR and Molecular docking approaches. Results indicate that, biological activity of SrtA inhibitors is closely related to the closed lid conformation of SrtA from *Bacillus anthracis*. This work may lead to a better understanding of the mechanism of action and aid to design a novel and more potent SrtA inhibitors.

Introduction

Bacillus anthracis, a gram-positive, spore-forming bio-threat bacterium that is the etiological pathogen of the zoonotic disease anthrax. It constantly exposed to different environmental conditions in its host and strongly influencing its physiology. It is a life-threatening infectious disease, widespread in most parts of the globe, and cases of anthrax have been reported from almost every country. Despite the emphasis on its role as an agent of bioterrorism or biological warfare, anthrax is continuing to be, an important global disease of wildlife and livestock¹. Currently, novel approaches are required for the treatment of bacterial infections, due to continuous changes in the patterns of infectious disease and the emergence antibacterial resistant to present antibiotics. Currently, researchers are involved in finding drug targets of great interest in terms of understanding and inhibiting the infection process and these targets may have potential as targets for therapy².

Surface proteins of this bacterium play an important role in host-pathogen interactions and this mechanism have attracted the researchers in the hunt for the drug target to inhibit the pathogenic microbes. Cell wall anchoring occurs through a transpeptidation mechanism, which requires surface proteins with C- terminal sorting signals³. In sortase family of proteins, class A

enzymes have attracted significant interest as a universal drug target of all Gram positive pathogens and the clinically important Gram positive pathogens use the Sortase A (SrtA) gene to display virulence factors. SrtA enzymes are playing a vital role in cleaving the sorting signals of secreted proteins to form isopeptide (amide) bonds between the secreted proteins and peptidoglycan or polypeptides⁴. SrtA is involved in the covalent anchoring of surface protein to the cell wall peptidoglycan, or link proteins together to construct pili which promote the bacterial adhesion to host cell and tissue to acquire the essential nutrients. These enzymes are actively involved as the principal architects of the bacterial surface that makes SrtA as an attractive drug target⁵.

Recent research has come up with evidence, that virulence of gram positive pathogenic microbes lies on SrtA enzymes, and removal of SrtA gene shows lack of pathogenicity and cell wall formations⁶. This evidence has enlightened and attracted the researchers to develop the promising lead considering SrtA as a drug target. SrtA, is main focus of this study, anchors seven proteins to the cell wall by joining the threonine of the LPXTG sorting signal at the C-terminus of the protein to the amine group of meso-diaminopimelic acid (DAP) within lipid II⁷. In this mechanism, the key active site loop (β 7/ β 8 loop) undergoes a disordered-to-ordered transition upon binding the sorting signal, potentially facilitating recognition of lipid II. Hydrophobic residues Val168 and Leu169 located in β 6- β 7 loop which involved hydrophobic interaction with N-terminal LPETG peptide and the catalytic residues Pro168, Asp169, Lys170 and Trp171 are present in the loop region between β 7/ β 8⁸. Our studies focus on structural changes in SrtA active sites loops (β 7/ β 8) and analyze the activity of known SrtA inhibitors against this various loop conformation⁹. Suree *et al.*¹⁰ had performed structure-activity relationship (SAR) analysis, which leads to the identification of several rhodanine, pyridazinone and pyrazolethione compounds with more potential activity against the *B. anthracis* than any other previously mentioned natural or synthetic inhibitors. In this study, we have evaluated the activity of the reported compounds through 3D QSAR model and the active conformations are cross validated with open and closed conformation of β 7/ β 8 loop in SrtA. This study on analyzing the available SrtA inhibitor's against the β 7/ β 8 loop conformation of SrtA will be much more helpful in discovering of novel lead molecules for further development.

Materials and Methods

Ligand preparation and Pharmacophore Generation

Experimentally active 28 compounds of rhodanine, pyridazinone and pyrazolethione derivatives reported by Suree et al.¹⁰ were taken for the study and the ligand conformations were prepared by using LigPrep/Confgen with OPLS AA. Each compound structure is represented by a set of points in 3D space, which coincide with various chemical features that may facilitate non-covalent bonding between the compound and its target receptor. PHASE (PHASE V 4.2) provides a built-in set of six pharmacophore features, hydrogen bond acceptor (A), hydrogen bond donor (D), hydrophobic group (H), negatively ionizable (N), positively ionizable (P) and aromatic ring (R). The rules that apply to map the positions of pharmacophoric sites are known as feature definitions, and are represented internally by a set of SMARTS patterns¹¹. Each pharmacophore feature is defined by a set of chemical structure patterns. All user-defined patterns are specified as SMARTS queries and assigned one of three possible geometries, which define the physical characteristics of the site like point, vector and group. Active and inactive thresholds of pIC₅₀ 5.9 and 4.3 were applied to the dataset to yield eight actives and five inactives, which are used for pharmacophore modeling and subsequent scoring.

Building the 3D-QSAR model

All the generated common pharmacophore hypotheses were used to generate atom-based 3D-QSAR models by correlating the actual and predicted activity for the set of 21 training molecules using PLS analysis. The PLS regression was carried out using PHASE with a maximum of N/3 PLS factors (N=number of ligands in the training set, and a grid spacing of 1.0 Å) and all the generated models were validated by predicting activity of 7 test set molecules^{12,13}.

External validation

In order to understand the predictive abilities of established model was evaluated to confirm the robustness by external validations¹⁴. The calculated parameter is r^2 , K values and r_{cv}^2 . The r^2 value calculated by the following formula,

$$R = \frac{\sum (y_i - \bar{y}_o) (\hat{y}_i - \bar{y}_p)}{\sqrt{\sum (y_i - \bar{y}_o)^2 (\hat{y}_i - \bar{y}_p)^2}}$$

Where, y_i and $\hat{y}_i$ are the observed and predicted activity, $\bar{y}_o$ and $\bar{y}_p$ are the average values of the observed and predicted pIC50 values of the test set molecules. For the ideal QSAR model, the r^2 value should be close to 1. And the regression of y against $\hat{y}$ through the origin: $\hat{y}_i^{reg} = k y_i$ should be characterized by k close to 1. Slope k is calculated as follow:

$$k = \frac{\sum(y_i \hat{y}_i)}{\sum \hat{y}_i^2}$$

The cross validated co-efficient, r_{cv}^2 , was calculated using the following equation:

$$r_{cv}^2 = 1 - \frac{\sum(Y_{\text{predicted}} - Y_{\text{observed}})^2}{\sum(Y_{\text{observed}} - Y_{\text{mean}})^2}$$

Where, $Y_{\text{predicted}}$, Y_{observed} and Y_{mean} the predicted, observed, and mean values of the target property (pIC50), respectively¹⁵.

Applicability Domain

The applicability domain (AD) prediction was performed by using the SIMCA-P 13.0 demo version (Umetrics. 2002). The AD is the physico-chemical information based on a training set of the model and it's also applicable to make predictions for new hit compounds. By this approach, one can directly analyze the properties of distance matrix, multivariate descriptor to investigate the applicability domain of the training set. These approaches define the AD by calculating distances of a query compound from a defined point within the descriptor space of the training data¹⁶. This measured distance between defined point and the dataset, then compares with a pre-defined threshold. These calculations are done by feature selection and principal component analysis (PCA). The reliable prediction of the source compounds is likely to be similar to the data set and the activity prediction errors were avoided by using activity prediction applicability domain concept¹⁷.

Molecular Dynamics Simulation

Here, MD simulation was performed for SrtA structure and also for ligand bound SrtA structures. For the Apo structure of SrtA, the NMR structure of *B. anthracis* SrtA (PDB ID: 2KW8 (<http://www.rcsb.org>)) was subjected to MD simulation for 100ns of timescale, in order to understand the nature of intra-molecular conformational changes of protein structure. MD simulation is performed with GROMACS program package (<http://www.gromacs.org>) adopting

the GROMOS96 force field parameters¹⁸. The structure was solvated using the SPC water model, and the solvated structure was energy minimized using steepest descent method, and it's terminated when maximum force is found in smaller than $100 \text{ kJ mol}^{-1} \text{ nm}^{-1}$. For equilibration and production runs, the temperature was controlled through Nosé–Hoover thermostat dynamics with a damping coefficient of 2 ps^{-1} . The long range electrostatic interactions were computed by particle-mesh Ewald method and van der Waals (VDW) cut-off were set to 9 Å . The structure was energy minimized to eliminate bad atomic contacts and subsequently solvated with water. The total simulations were carried out in the NPT ensemble at constant temperature (300 K) and pressure (1 bar), with a time step of 2fs. NVT were performed for 1ns, and the minimized structure was equilibrated with a timescale of 100ns^{19, 20}. For the SrtA bound ligand complex, receptor and ligands were cleaned using GROMOS96 force field and then the topology files were generated for the receptor and ligands separately using PRODRG tool²¹. The simulation system was created manually by importing the ligand topology into the system pursued along with a dodecahedron box with a margin of 1 nm and the system was filled with water using the SPC explicit solvation model. The system was applied with energy minimization and the atomic velocities were adjusted according to Maxwell Boltzmann distribution at 300K with a periodic scaling of 0.1ps. A pre-simulation run of 20 ps was applied to relax the system and to remove the geometric restrains which eventually appeared at the initialization of the run²². All the simulations were carried out at constant pressure and temperature (NPT) ensemble. The Berendsen coupling was employed to maintain a constant temperature of 300 K and constant semi-isotropic pressure of 1 bar with coupling time of 2.0 fs and the coordinates were saved. The simulation timescale for ligand bound form is 20ns and the RMSD analysis has been performed for understanding the stability of ligands.

Molecular Docking Simulation

For the interactions, Quantum Polarized Ligand Docking (QPLD), molecular docking, was performed with open and closed active site. Two different conformation of open and closed active site from the MD simulation are subjected to grid generation. To soften the potential for non-polar part of the receptor, Van-der Waal radii of receptor atoms were scaled by 1.00 Å with a partial charge cutoff $0.25^{23,24}$. Ligands are set to be rigid conformations, as of all the conformations from 3D QSAR bioactive conformations were used for these docking studies. The previously reported active site which located around the loop region considered as binding pocket²⁵. For the specificity of ligand partial charges, QM/MM based QPLD docking was also

applied for both the protein conformations with all 28 compounds. For QM calculations, the accuracy level is held in reserve as accurate and background of this calculation was done by using 6-31G*/B3LYP density functional theory^{26,27}. The charges were calculated for a free ligand by Jaguar, with the presence of the hybrid water model. Here, fixed charges of ligands obtained from force field parameter are replaced by QM/MM calculations in protein atmosphere and only ligand treated as the quantum region. However, the protein ligand complex charge correctness occurs through the QPLD docking and it will enhance the accuracy of results²⁸.

Conformation Based Activity Prediction

The Liaison program is an application for estimating the binding affinities between ligands and receptors, using a linear interaction approximation (LIA) model²⁹. The LIA model is an empirical method fitted to a set of known binding free energies. Liaison runs molecular mechanics (MM) simulations of the ligand-receptor complex, and for the free ligand and receptor using the surface generalized Born (SGB) continuum solvation model³⁰. The simulation data and empirical binding affinities were analyzed to generate the Liaison parameters, which are subsequently used to predict binding energies for other ligands with the same receptor. The empirical function used by Liaison for the prediction of binding affinities is as follows.

$$\Delta G = \alpha (<U_{vdw}^b> - <U_{vdw}^f>) + \beta (<U_{elec}^b> - <U_{elec}^f>) + \gamma (<U_{cav}^b> - <U_{cav}^f>)$$

Here, U_{vdw} , U_{elec} , and U_{cav} are the van der Waals, electrostatic and cavity energy terms in the Surface Generalized Born (SGB) continuum solvent model. Default options were specified including a minimization sampling method using a truncated Newton algorithm. Ensemble averages of van der Waals, electrostatic, and cavity (solvent exposed ligand surface area) energies were computed for the docking conformation with open and closed lid conformations using an implicit solvation model. The computed energies for each inhibitor complex and corresponding pIC₅₀ (Experimental Value) were then imported into Strike, where the partial least squares (PLS) and multiple linear regression (MLR) methods were applied, to construct a linear equation representing binding affinity³¹.

Results and Discussion

Determination of Pharmacophore and 3D-QSAR models

In this study, the experimental active 28 compounds of rhodanine, pyridazinone and pyrazolethione derivatives were used to predict theoretical activity of the test compounds, which

are shown in Table 1 and their 2D structure of the scaffoldligand rearrangements are shown in Supplementary figure 1. The biological activity data were reported in the form of IC_{50} values and those values were converted into pIC_{50} using the formula ($pIC_{50} = -\log IC_{50}$) and the data set was divided randomly into 21 training set and 7 test set compounds in order to maintain the 3:1 ratio²⁵. We ensured that the structurally different compounds were distributed uniformly with a wide range of pIC_{50} value in both training and test set. Initial step is tend towards generating the pharmacophore model using a tree-based partition algorithm and two best common pharmacophore hypotheses AAAHR, and AAADR were selected for 3D-QSAR model building by applying the scoring function for five featured Pharmacophore hypotheses with default values. Training set compounds were aligned on these common pharmacophore hypotheses and are analyzed in PHASE with three PLS factors and the predictivity of each hypothesis was re-evaluated by the test set compounds. A summary of the statistical data for the two common pharmacophore hypotheses AAAHR and AAADR was listed in Table 2.

The statistical parameters R^2 , Q^2 , SD, RMSE and F were used to evaluate the good quality QSAR model. The two models show good and consistent R^2 greater than 0.82, SD values lesser than 0.4 and F-test values were moderate. This shows that these two models interpreting structure activity relationship for this series of training set compounds satisfactorily. According to Tropsha, high R^2 is a necessary but not sufficient condition for a QSAR model. Besides the consideration of high R^2 , the best QSAR model should be chosen based on its predictive ability. The AAAHR shows good external predictive ability with Q^2 value 0.74 and this hypothesis has highest Q^2 value when compared with other pharmacophore hypotheses, which suggests that AAAHR hypothesis is the best model among the other generated models³². Additionally, the AAAHR pharmacophore hypothesis has a low RMSE value of the pharmacophore hypotheses, which also supports this hypothesis. Finally, AAAHR hypothesis were considered as the best model based on their R^2 , Q^2 , SD, and RMSE values. The Common pharmacophore model alignment of active compounds was shown in figure S2. The plots of actual versus predicted pIC_{50} in X axis and Y axis in figure 1 for the training set and test set compounds. However, angles and distances of pharmacophore hypothesis AAAHR are given in supplementary material Table S1.

External Validations of AAAHR pharmacophore model

Established 3D-QSAR AAAHR pharmacophore model were evaluated by external validations with 8 tests set compounds and internally using the leave-n-out technique. The

established AAAHR model validated with r^2_{cv} cross validated coefficient produced with the value of 0.56, the high k slope values of regression lines shows the value of 1.002. Non-cross validated correlation coefficient (r^2) value of 0.51. AAAHR pharmacophore model, possibly will predict new derivatives. The summary of external statistical obtained parameter of 3D-QSAR for the AAAHR pharmacophore model was shown in Table S2.

QSAR visualization

One of the major advantages of the PHASE 3D-QSAR technique is to get contour cubes based on favorable and unfavorable regions, which could be visualized in 3D space. The contour cubes obtained from AAAHR shows, how 3D-QSAR methods can identify features which is important for the interaction between ligands and their target protein. Such contour cubes allow identification of those positions that require a particular physicochemical property to enhance the bioactivity of a ligand. A pictorial representation of the generated contours was shown in figure 2. In these representations, blue cubes are indicating the sterically favored spatial regions to enhance the activity, whereas red cubes represent the sterically disfavored regions. These cubes can be generated for different properties such as hydrophobic, hydrogen bond donor, hydrogen bond acceptor (electron withdrawing), positive and negative ionic features, which defines the non-covalent interactions with receptor. Visualization of QSAR model cubes associated with hydrophobic property gives an idea about the topology of the receptor site. Figure 2a and 2b shows cubes generated from hydrophobic property using QSAR model while; these cubes illustrate the significant favorable regions and unfavorable hydrophobic/non-polar interactions that arise when the QSAR model is applied to the most active pyridazinone derivative (2-9) and inactive pyridazinone derivatives (2-36). Blue cubes were seen on the most active compound pyridazinone-2-9 at the aromatic ring position. Although comparison of the most significant favorable and unfavorable hydrogen bond acceptors or electron withdrawing features which arise during QSAR model is applied to the most active pyridazinone-2-9 and to the most inactive pyridazinone-2-36 compound was also shown in figure 2a and 2b. Electron withdrawing group (N) and benzene ring structure were associated with blue-colored cubes in the most active compound of pyridazinone-2-9 are observed near atoms which shows high electron withdrawing feature, whereas, in the case of most inactive compound pyridazinone-2-36, red-colored cubes were visible indicating activity difference in these two molecules.

Validation of Applicability Domain

The validation of AD results represents that the prediction of test set compounds are accurate and reliable. The DmodX values of eight test set compounds are placed below the critical value of 3.4. The residual values of the predicted vs. experimental SD of X-residuals (DmodX) were represented in figure S3, and it is concluded that these compounds are better for the test set of compounds from this QSAR model. Visualization of the 3D-QSAR model, provides the details of the relationship between structure and activity among these molecules, and thus provides explicit indications for designing the best analogues. Analysis of 3D contour maps was shown to understand the essential structural requirement of the compounds to exhibit better inhibition, which is clearly explained in most active and inactive compounds. While, the successful statistical data based ligand conformations are carried for further study.

Open and closed lid in the active site of SrtA

The SrtA structure stability, flexibility and intermolecular conformations were analyzed with this 100ns large scale molecular dynamics simulation. Our previous study on this enzyme for 20ns of simulations has reported about the stability of the protein and so here we tried to extend the simulations for 100ns of time scale. The simulation event involved in this study is to understand the biological happening of the SrtA structure and alterations occurs in each interval of timescale. In the system, the temperature, total energy, mass density, and volume are relatively stable throughout the simulation period, which suggests that the simulations were carried out satisfyingly and system has been well equilibrated. In the figure 3A the RMSD from the NMR structure of C α atoms vs. timescale (ns) shows that initially the protein is more flexible up to 40ns and so the deviations are inconsistent, especially from 2ns to 6ns. The proteins have more deviations due to the presence of loop structures in the protein structure. But after the 40th ns the protein structure was consistent enough as shown in the figure 3. The time evolution of the RMS fluctuation (RMSF) from the averaged structure provides another approach to evaluate the convergence of the dynamical properties of the system as shown in figure 3B. Here, the fluctuations of each amino acid are clearly shown and it's because of the highly fluctuated loop structure. While RMSF profiles indicate that the residues with higher fluctuation values are those in the long loop Gly53-Asp64 residues (Gly53, Ser54, His55, Met56, Asp57, Ala58, Ser59, Lys60, Ile61, Asp62, Gln63 and Pro64). This particular loop structure was already reported that shown much deviation in *Bacillus subtilis*, and holds the interactions with the lead compounds. In similar to *B. subtilis*, the *Enterococcus faecalis*, *Bacillus cereus*, *Bacillus anthracis*, *Staphylococcus aureus* are the additional microbial SrtA, which shows the importance of this

loop towards the drug-protein interactions³³⁻³⁵. There was much conformational changes of native protein structure occur in between the 0-100ns. The MD simulation of the SrtA structure showed that the tail like loop structure in the N - terminal region (53-64) present nearby active site which is widely opened at initial position. But this particular loop undergoes more conformational changes at each nanosecond. Therefore, we carefully investigated the structural changes of SrtA with the time interval of every 5ns until 100ns and noticed that the N-terminal tail loop has the biological function of the lid of the active site. The tail loop continuously fluctuating at the dynamic state till the 30ns and after that it slowly changes its position and act as a lid. The lid closes the active site region and it was remains closed till the 100ns. Meanwhile, the conformational changes occur in the SrtA structure was demonstrate in Figure S4 and 4 shows the overall morphed conformations and fig 4 showing the separate conformations of SrtA at each 5ns of interval. The trajectory analysis suggest that loop structure in N-terminal is continuously fluctuating till the 40ns and shows widely opened active site and after 40ns the protein attain its equilibrium condition and loop moderately closes the active site and function as a lid for SrtA structure. Whereas, figure 5 indicates the opening and closing of lid in different timescale of dynamic environment and this lid is called as universal feature of SrtA enzyme in gram positive pathogens. At 40ns the lid closes the active site and function as protector of the active site hydrophobic environment and so it's difficult task to design the suitable compounds for closed binding pocket of SrtA. The overall simulation studies show good contacts with the system and protein throughout the timescales. In the hydrogen bond analysis between the SrtA and water molecules shows that, there is vast difference in hydrogen bonds. Figure 6 shows the hydrogen bond interactions analysis of SrtA structure with water contact shows that pair above 0.35 nm (A) and pair within 0.35nm (B). The average of hydrogen bond is calculated in two intervals of timescale, i.e. 0-40ns and 40-100ns, as after the 40ns of timescale the loop "lid" has been closed, so it is necessary to understand the interaction difference happen due to this biological function. The average of above 0.35 nm and within 0.35nm was comparatively high for 0-40ns, compared to after 40ns (40-100ns). The number of hydrogen bonds between the protein and the surrounding water molecules are more in beginning of the timescale (open lid) and comparatively less in later stage of the time scale (closed lid). This is due to closing of active site which protects the hydrophobicity regions in active site to interact with water.

SrtA inhibitors against Open and Closed lid conformations

This research works, explains the statistical validation and confirmation of SrtA inhibitors using closed and open lid of SrtA structure through molecular modeling techniques. In *Staphylococcus aureus*, the substrate amino acid Leu plays the crucial role of an anchor to the loop between the beta 6- beta7. This interaction directing the conformational transition of the enzyme from an "open" to a "closed" state subsequent to which the Pro residue facilitates the consummation of binding through predominant engagement of the loop and catalytic motif residues in hydrophobic interactions. Through this, the SrtA loop function in *B. anthracis* has been clearly explored with the closed and open active site conformations taken from the molecular dynamics simulations. In order to check active conformation of SrtA the bioactive conformers of the inhibitors obtained from the 3D QSAR are validated in both open and closed lid conformations. The QPLD and LSBD (ligand and structure based descriptors) calculations of both closed and open lid conformations were used for obtaining the values of activity based on the conformations of the proteins. The docking procedures used this study aim to identify correct poses of ligands in the binding pocket of a protein and to predict the affinity between the ligand and the protein. LSBD calculations predict the activity based on docking pose with respect the experimental activity. The purpose of the scoring procedure is the identification of the correct binding pose with lowest energy value, and the ranking of protein-ligand complexes based on their binding affinities. So that, bioactive conformations obtained from 3D QSAR are made rigid conformations and both closed and open lid are interacted. The analysis of binding patterns from the docking suggests that closed binding pocket has three main amino acids (Met56, Val110 and Trp171), which are mainly involved in hydrophobic interactions. Almost 92% of the active compounds show the interactions with this catalytic triad of SrtA and compound pyridazinone-2-17 is additionally showing the hydrophobic interactions with Ala58. As discussed earlier, the lid closed the binding pocket which protects the hydrophobic surface of the active site and these experimentally validated compounds are well suited for the closed lid binding pocket of SrtA. From the interactions, it is concluded that the aromatic group present in the pyridazinone and pyrazolethione are core important for the activity. The figure 7 shows the ligand aromatic ring structure involved in the hydrophobic interactions with Met56, Val110 and Trp171. The hydrophobic interactions represented in green colored dashed lines in Figure 7 indicate the hold of aromatic ring structure inside the hydrophobic binding pocket. Among the interacting amino acids Met56 is contributing in the function of the closed lid loop and also it's actively participated in the hydrophobic interactions. This result indicates that those three amino acids are

core important for targeting SrtA enzyme and the aromatic ring structure of pyridazinone and pyrazolethione are play a key role in inhibition of SrtA activity. Best conformations of SrtA inhibitors, shown in figure 7, with both protein conformations are taken into account for the Molecular Dynamics Simulation.

Ligand Stability- Molecular Dynamics Simulation

Based on activity, docking and binding energy, five best ligands were chosen for molecular dynamics simulation studies. In the analysis of trajectories shows that all the compounds on this dynamic event having better interaction throughout the timescale of 20ns and ligand does not detach from the protein. The figure S5 shows the RMSD graph of top five active compounds and these compounds are more stable inside the binding pocket due to the engulf of the hydrophobic interaction with the aromatic ring. Moreover, the figure explained that experimentally validated compounds have stable interaction and bind properly inside the closed binding pocket of SrtA. In the initial state all the compounds is slightly moved from its original position and after that the ligand binding position got stable inside the binding pocket and the hydrophobic bonds holds the ligand strongly by not allowing the ligands to detach from the protein structure. Depth analysis of trajectories clearly shows that the catalytic triad of Met56, Val110 and Trp171 are playing role in holding ligand and also interacted with aromatic ring. The average mean values of RMSD for all test compounds (pyridazinone-2-19, pyridazinone-2-9, pyrazolethione-3-12, pyridazinone-2-5 and pyridazinone-2-20) were calculated and those values clearly shows that all are having the average values of less than 0.2nm to 0.4nm. Pyridazinone-2-5 is having 0.4nm average RMSD, which is comparatively higher than the other four compounds. When analyzing RMSD values and graph, all the experimentally verified compounds are stable and have much correlation with each other. Interestingly, all the compounds are having the same binding modes in dynamic state and provides more similar pattern of inhibition. These results suggested that binding of the ligand to the protein showed a deviation from their initial position because of adjustments in their configuration, but remain bound within the catalytic triad of the protein.

Energy Calculation Accounts for Activity Prediction

In general most of the studies measure the accuracy of scoring function by their ability to correctly rank the activity of a congeneric set of ligands. The ligand activity prediction against both open and closed lid conformation of SrtA is equally important in drug design of *B. anthracis*. The experimental activity reported by Sureeet *al.*, 2009 is considered in pIC₅₀ and

liaison calculation is analyzed, while the liaison is a method of predicting ligand-protein binding free energies using a model that has been correlated to known binding energy values. The process involves two steps, a fitting step and a predicting step and each step are carried out as a two tasks, which are the simulation task and an analysis task. The experimental binding affinities given in the property activity (kcal/mol) will be used for the response or dependent variable from which a linear model will be created. The LIA equation outlines in detail and it estimate binding affinities using $\langle U_{ele} \rangle$, $\langle U_{vdw} \rangle$ and $\langle U_{cav} \rangle$ terms from Liaison. Once the LIA model has been created, it must be analyzed to see if it makes intuitive sense and possesses the predictive power that does not arise by chance. Here, the obtained energy levels are analyzed with their reference activity and theoretical activity of both closed lid and open lid ligand binding complex. The obtained activities of experimental and theoretical values are plotted and R^2 is calculated to understand the correlation between the both with respect to open and closed lid. The correlation analysis shows in the figure 8, predicts that, the open lid protein conformations shows r^2 of 0.730 and closed lid conformations shows r^2 of 0.925. This confirms that, the protein conformation obtained from the 40th ns of molecular dynamics simulation is more active and suitable for molecular modeling calculations.

Conclusion

In this current study, 3D pharmacophoric features and QSAR of the selected SrtA inhibitors were explained, which are experimentally validated against the *B. anthracis* SrtA. The pharmacophore with AAAHR pharmacophoric features are associated with a 3D-QSAR model giving the good statistical significance and predictive ability. The 3D contour map analysis was helped to understand the essential structural requirement of the compounds to exhibit the better inhibition, which is clearly explained in the most active and inactive compounds. Meanwhile, the N-terminal tail loop is playing role as the protector lid of the active site hydrophobic regions was through molecular dynamic studies. Through MD simulation, we reported that open lid conformations have a mechanism of closed lid active site at 40ns. Furthermore, the activity prediction of the experimentally proven SrtA inhibitors was analyzed by docking the selected SrtA inhibitors into both conformations of open and closed lid of active site through QPLD

method. Although the open lid protein conformation shows r^2 of 0.730 and closed lid conformation shows r^2 of 0.925. This confirms that, the conformation obtained from the 40th ns of molecular dynamics simulation is more active and suitable for molecular modeling calculation. The interaction between ligand and protein are observed for catalytic triad Met56, Val110 and Trp171 by means of hydrophobic interaction. From these studies, it is understood that the aromatic ring present in the rhodanine, pyridazinone and pyrazolethione derivatives are playing an important role in the inhibition. Consequently, the ligand bound complexes are confirmed and selected compounds have stable interaction by holding the catalytic triad in binding region. Finally, atom-based 3D-QSAR model, molecular dynamics and docking studies performed here could be very valuable for the development of new and potent lead compounds for the inhibition of SrtA.

Reference

1. Goel, A.K. (2015). Anthrax: A disease of biowarfare and public health importance. *World Journal of Clinical Cases*, 3, 20-33. DOI: 10.12998/wjcc.v3.i1.20.
2. Groisman, E.A., & Casadesus, J. (2005). The origin and evolution of human pathogens. *Molecular Microbiology*, 56, 1-7. DOI: 10.1111/j.1365-2958.2005.04564.x.
3. Clancy, K.W., Melvin, J.A., & McCafferty, D.G. (2010). Sortase transpeptidases: insights into mechanism, substrate specificity, and inhibition. *Biopolymers*, 94, 385-96. <https://doi.org/10.1002/bip.21472>.
4. Park, K., Jung, J., Son, J., Kim, S.H., & Chung, B.H. (2013). Anchoring foreign substances on live cell surfaces using Sortase A specific binding peptide. *Chemical Communication (Camb)*, 49, 9585-7. DOI:10.1039/C3CC44753G.
5. Marraffini, L.A., Dedent, A.C., & Schneewind, O. (2006) Sortases and the art of anchoring proteins to the envelopes of gram-positive bacteria. *Microbiology and Molecular Biology Reviews*, 70, 192-221. DOI: 10.1128/MMBR.70.1.192-221.2006
6. Schneewind O., & Missiakas, D.M. (2012). Protein secretion and surface display in Gram-positive bacteria. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367, 1123-39. DOI: 10.1098/rstb.2011.0210.
7. Chan, A.H., Yi, S.W., Terwilliger, A.L., Maresso, A.W., & Jung, M. E. (2015). Structure of the bacillus anthracis sortase A enzyme Bound to its sorting signal. *Journal of Biological Chemistry*, 290, 25461. DOI: 10.1074/jbc.M115.670984.
8. Schneewind, O., & Missiakas, D.M. (2014). Sec-cretion and sortase-mediated anchoring of proteins in Gram-positive bacteria. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1843, 1687-97. DOI: 10.1016/j.bbamcr.2013.11.009.

9. Chan, A.H., Yi, S.W., Terwilliger, A.L., Maresso, A.W., Jung, M.E., & Clubb, R.T. (2015). Structure of the Bacillus anthracis Sortase A Enzyme Bound to Its Sorting Signal: A Flexible Amino-Terminal Appendage Modulates Substrate Access. *Journal of Biological Chemistry*, 290, 25461-74. DOI: 10.1074/jbc.M115.670984.
10. Suree, N., Yi, S.W., Thieu, W., Marohn, M., Damoiseaux, R., Chan, A., Jung, M.E., & Clubb, R.T. (2009). Discovery and structure-activity relationship analysis of Staphylococcus aureus sortase A inhibitors. *Bioorganic & Medicinal Chemistry*, 17, 7174-85. <https://doi.org/10.1016/j.bmc.2009.08.067>.
11. Dixon, S.L., Smondyrev, A.M., & Rao, S.N. PHASE: a novel approach to pharmacophore modeling and 3D database searching. *Chemical Biology and Drug Design*, 67, 370-2. <https://doi.org/10.1111/j.1747-0285.2006.00384.x>.
12. Phase, Schrödinger, LLC, New York, NY, 2018
13. Evans, D.A., Doman, T.N., Thorner, D.A., & Bodkin M.J. (2007). 3D QSAR methods: Phase and Catalyst compared. *Journal of Chemical Information and Modeling*, 47, 1248-57. DOI: 10.1021/ci7000082.
14. Chirico, N., & Gramatica, P. (2011). Real external predictivity of QSAR models: how to evaluate it? Comparison of different validation criteria and proposal of using the concordance correlation coefficient. *Journal of Chemical Information and Modeling*, 51, 2320-35. DOI: 10.1021/ci200211n.
15. Schuurmann, G., Ebert, R.U., Chen, J., Wang, B., & Kuhne, R. (2008). External validation and prediction employing the predictive squared correlation coefficient test set activity mean vs training set activity mean. *Journal of Chemical Information and Modeling*, 48, 2140-5. DOI: 10.1021/ci800253u.
16. Roy, K., & Mitra, I. (2011). On various metrics used for validation of predictive QSAR models with applications in virtual screening and focused library design. *Combinatorial Chemistry & High Throughput Screening*, 14, 450-74. DOI: 10.2174/138620711795767893.
17. Keefer, C.E., Kauffman, G.W., & Gupta, R.R. (2013). Interpretable, probability-based confidence metric for continuous quantitative structure-activity relationship models. *Journal of Chemical Information and Modeling*, 53, 368-83. DOI: 10.1021/ci300554t.
18. Pronk, S., Pall, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., Shirts, M.R., Smith, J.C., Kasson, P.M., van der Spoel, D., Hess, B., & Lindahl, E. (2013). GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*, 29, 845-54. <https://doi.org/10.1093/bioinformatics/btt055>.
19. Selvaraj, C., Singh, P., & Singh, S.K. (2014). Molecular insights on analogs of HIV PR inhibitors toward HTLV-1 PR through QM/MM interactions and molecular dynamics studies: comparative structure analysis of wild and mutant HTLV-1 PR. *Journal of Molecular Recognition*, 27, 696-706. DOI: 10.1002/jmr.2395.
20. Selvaraj, C., Sivakamavalli, J., Vaseeharan, B., Singh, P., Singh, S.K. (2014). Structural elucidation of SrtA enzyme in Enterococcus faecalis: an emphasis on screening of

- potential inhibitors against the biofilm formation. *Molecular Biosystems*, 10, 1775-1789. DOI:10.1039/C3MB70613C.
21. Schüttelkopf, A.W., & van Aalten, D.M. (2004). PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallographica D Structural Biology*, 60, 1355-63. <https://doi.org/10.1107/S0907444904011679>.
 22. Hetenyi, C., Paragi, G., Maran, U., Timar, Z., Karelson, M., & Penke, B. (2006). Combination of a modified scoring function with two-dimensional descriptors for calculation of binding affinities of bulky, flexible ligands to proteins. *Journal of American Chemical Society*, 128, 1233-9. DOI: 10.1021/ja055804z.
 23. Kombo, D.C., Grinevich, V.P., Hauser, T.A., Sidach, S., Bencherif, M. (2013). QM-polarized ligand docking accurately predicts the trend in binding affinity of a series of arylmethylene quinuclidine-like derivatives at the $\alpha_4\beta_2$ and $\alpha_3\beta_4$ nicotinic acetylcholine receptors (nAChRs). *Bioorganic and Medicinal Chemistry Letters*, 23, 4842-7. <https://doi.org/10.1016/j.bmcl.2013.06.094>.
 24. Schrödinger Suite 2018-3 QM-Polarized Ligand Docking protocol; Glide, Schrödinger, LLC, New York, NY, 2018.
 25. Selvaraj, C., & Singh, S.K. (2014). Validation of potential inhibitors for SrtA against *Bacillus anthracis* by combined approach of ligand-based and molecular dynamics simulation. *Journal of Biomolecular Structure and Dynamics*, 32, 1333-49. [ps://doi.org/10.1080/07391102.2013.818577](https://doi.org/10.1080/07391102.2013.818577).
 26. Chung, J.Y., Hah, J.M., & Cho, A.E (2009). Correlation between performance of QM/MM docking and simple classification of binding sites. *Journal of Chemical Information and Modeling*, 49, 2382-7. DOI: 10.1021/ci900231p.
 27. Jaguar, Schrödinger, LLC, New York, NY, 2018.
 28. Zhong, H., Kirschner, K.N., Lee, M., & Bowen, J.P. (2008). Binding free energy calculation for duocarmycin/DNA complex based on the QPLD-derived partial charge model. *Bioorganic Medicinal Chemistry Letters*, 18, 542-5. DOI: 10.1016/j.bmcl.2007.11.090.
 29. Liu, S., Zhou, L.H., Wang, H.Q., & Yao, Z.P. (2010). Superimposing the 27 crystal protein/inhibitor complexes of beta-secretase to calculate the binding affinities by the linear interaction energy method. *Bioorganic Medicinal Chemistry Letters*, 20, 6533-7. <https://doi.org/10.1016/j.bmcl.2010.09.050>.
 30. Alam, M.A., & Naik, P.K. (2009). Applying linear interaction energy method for binding affinity calculations of podophyllotoxin analogues with tubulin using continuum solvent model and prediction of cytotoxic activity. *Journal of Molecular Graphics and Modeling*, 27, 930-43. DOI: 10.1016/j.jmgm.2009.02.003.
 31. Vilar, S., Karpiak, J., & Costanzi, S. (2010). Ligand and structure-based models for the prediction of ligand-receptor affinities and virtual screenings: Development and application to the beta (2)-adrenergic receptor. *Journal of Computational Chemistry*, 31, 707-20. doi: 10.1002/jcc.21346.

32. Tropsha, A., Gramatica, P., & Gombar, V.K. (2003). The importance of being earnest: Validation is the absolute essential for successful application and interpretation of QSPR models. *Molecular Informatics*, 22, 69-77. <https://doi.org/10.1002/qsar.200390007>.
33. Clancy, K.W., Melvin, J.A., & McCafferty, D.G. (2010). Sortase transpeptidases: insights into mechanism, substrate specificity, and inhibition. *Peptide Science*, 94, 385-96. <https://doi.org/10.1002/bip.21472>.
34. Selvaraj, C., Bharathi Priya, R., Lee, J.K., & Singh, S.K. (2015). Mechanistic insights of SrtA-LPXTG blockers targeting the transpeptide mechanism in *Sterptococcus mutants*. *RSC Advances*, 5, 100498-100510. DOI:10.1039/C5RA12869B.
35. Selvaraj, C., Sivakamavalli, J., Vaseeharan, B., Singh, P., & Singh, S.K. (2014). Examine the characterization of biofilm formation and inhibition by targeting SrtA mechanism in *Bacillus Subtilis*: a combined experimental and theoretical study. *Journal of Molecular Modeling*, 20, 2364. <https://doi.org/10.1007/s00894-014-2364-8>.

Legend for Tables

Table 1. Structures and actual versus predicted pIC₅₀ of compounds

Table 2. Quantitative Structure Activity Relationship (QSAR) results for the two best Common Pharmacophore Hypotheses

Table 3. Scoring parameters of active compounds in QSAR towards closed lid structure of SrtA

Compounds	X	R1	R2	R3	R4	Actual	Predicted	Residual
						IC ₅₀	IC ₅₀	
Rhodanine								
1-1		-Ph	4-NO ₂	3-Br, 2-OH, 5-NO ₂		4.69	4.52	0.179
1-2		-Me		3-Br, 2-OH, 5-NO ₂		4.88	5	-0.11
1-3		-Pr			-H	4.27	4.29	-0.01
1-4		-Et			2-NO ₂	4.13	4.31	-0.17
1-5		-Allyl			-H	4.56	4.52	0.04
Pyridazinone								
2-1 *		-SH	-OEt	-Ph	-H	5.85	5.88	-0.02
2-5		-OMe	-SH	-Ph	-H	5.74	5.5	0.24
2-6		-OH	OCH ₂ Ph	-Ph	-H	4.30	3.9	0.40

2-7		-OH	-OMe	-Ph	-H	4.30	4.24	0.06
2-8		-OH	-SEt	-Ph	-H	4.30	4.59	-0.28
2-9*		-SH	-SEt	-Ph	-H	6.52	6.11	0.41
2-10		-SH	-OEt	-Ph	-H	5.49	5.87	-0.37
2-11		-SH	-OEt	-Ph	4-NO ₂	5.17	5.41	-0.23
2-13		-SH	-OEt	-Ph	3-F	5.74	5.7	0.04
2-14*		-SH	-OEt	-Ph	3-Me	5.77	5.68	0.09
2-15		-SH	-OEt	-Ph	3,5-Cl ₂	4.85	5.5	-0.64
2-16		-SH	-OEt	Cyclohexyl		5.85	5.96	-0.10
2-17*		-SH	-OEt	-Ph	-H	5.92	5.94	-0.01
2-18		-OEt	-SH	-Ph	-H	5.92	5.66	0.26
2-19*		-OEt	-SH	-Ph	3-F	6.04	6.16	-0.11
2-20		-OEt	-SH	-Ph	3-Me	6.39	6.25	0.15
2-21		-OEt	-SH	-Ph	3,5-Cl ₂	5.28	5.48	-0.19
2-35*		-OEt	-Cl	-Ph	-H	6.52	5.26	1.26
2-36		-OEt	-Cl	-Ph	4-NO ₂	3.60	4.04	-0.43
2-47		-Cl	-Cl	-Ph	3,5-Cl ₂	4.85	4.71	0.14
Pyrazolethione								
3-9*	S	4-N=N-Ph	H			5.85	5.85	0.004
3-17*	S					6.52	6.72	-0.19
3-12	S	2,4,6-Br ₃	H			5.85	5.9	-0.05

Table 1. Structures and actual versus predicted pIC50 of compounds (* Test Set Compounds)

	AAAHR	AAADH
SD	0.34	0.41
R ²	0.87	0.82
F	40.3	25.4
P	6.079e-08	2.555e-06
RMSE	0.2637	0.4111
Q ²	0.74	0.6806
Pearson R	0.88	0.96

Table 2. Quantitative Structure Activity Relationship (QSAR) results for the two best Common Pharmacophore Hypotheses

	pIC₅₀	IFD score	Glide score	Glide Energy	Binding Energy	Hydrophilic interactions
Pyridazinone-2-19	6.046	489.21	-6.591	-42.62	-38.61	Met56, Val110 and Trp171
Pyridazinone-2-9	6.523	480.67	-6.486	-43.19	-40.51	Met56, Val110 and Trp171
Pyrazolethione-3-12	6.523	477.99	-6.209	-39.57	-38.67	Met56, Val110 and Trp171
Pyridazinone-2-5	5.745	481.21	-5.916	-36.24	-39.24	Met56, Val110 and Trp171
Pyridazinone-2-20	6.398	467.26	-5.858	-40.29	-39.98	Met56 and Val110
Pyridazinone-2-16	5.854	459.38	-5.614	-36.21	-38.21	Met56, Val110 and Trp171
Pyridazinone-2-35	6.523	461.50	-5.512	-38.29	-38.20	Met56, Val110 and Trp171
Pyridazinone-2-17	5.854	446.55	-5.367	-39.18	-36.19	Met56, Ala58, Val110 and Trp171

Table 3. scoring parameters of active compounds in QSAR towards closed lid structure of SrtA

Figure 1: Scatter plot for the predicted and actual pIC50 values for AAAHR hypothesis showing the statistical values of $R^2 = 0.92$, $SD = 0.16$, $F = 84.8$, $N = 40$, $Q^2 = 0.71$, $RMSE = 0.06$, Pearson $R = 0.90$

Figure 2: The significant favorable and unfavorable hydrophobic interactions are shown, when QSAR model is applied to the most active compound (**a**) and the most inactive compound (**b**).

Figure 3. Analysis of SrtA structure in dynamic state with the time scale of 100ns (A) RMSD plot of SrtA for 100ns of timescale. (B) RMSF plot showing residue wise changes in the SrtA structure.

Figure 4: Observation of SrtA active site loop close and open at different timescale (Observed with 5ns interval)

Figure 5: Active site loops on dynamic state (A) showing the active site lid in open state at 0ns (B) Active site lid closed at 40ns of timescale

Figure 6. Hydrogen bond analysis of SrtA structure with water contact shows that pair above 0.35 nm (A) –and pair within 0.35nm (B) - are showing high contacts in 0-40ns and after that the hydrogen bonds has been decreased.

Figure 7: Active compounds showing the aromatic ring structure showing hydrophobic interactions with Met56, Val110 and Trp171

Figure 8. R^2 cross validation of theoretical and experimental activity with respect to open lid (a) and closed lid (b) conformation

Figure 1

Figure 1: Scatter plot for the predicted and actual pIC50 values for AAAHR hypothesis showing the statistical values of $R^2 = 0.92$, $SD = 0.16$, $F = 84.8$, $N = 40$, $Q^2 = 0.71$, $RMSE = 0.06$, Pearson $R = 0.90$

Figure 2

Figure 2: The significant favorable and unfavorable hydrophobic interactions are shown, when QSAR model is applied to the most active compound (a) and the most inactive compound (b).

Figure 3

Figure 3. Analysis of SrtA structure in dynamic state with the time scale of 100ns (A) RMSD plot of SrtA for 100ns of timescale. (B) RMSF plot showing residue wise changes in the SrtA structure.

Figure 4

Figure 4: Observation of SrtA active site loop close and open at different timescale (Observed with 5 ns interval)

Figure 5

Figure 5: Active site loops on dynamic state (A) showing the active site lid in open state at 0ns
(B) Active site lid closed at 40ns of timescale

Figure 6

Figure 6. Hydrogen bond analysis of SrtA structure with water contact shows that pair above 0.35 nm (A) –and pair within 0.35nm (B) - are showing high contacts in 0-40ns and after that the hydrogen bonds has been decreased.

Figure 7

Figure 7: Active compounds showing the aromatic ring structure showing hydrophobic interactions with Met56, Val110 and Trp171

Figure 8

Figure 8. R^2 cross validation of theoretical and experimental activity with respect to open lid (a) and closed lid (b) conformation