CHAPTER-2

Exploring the Derivatives of SrtA Inhibitors against the Open Lid and Closed Lid Conformation of SrtA Binding Pocket in Bacillus anthracis
INTRODUCTION

In an epoch of the global spread of antibiotic resistance in many human pathogens, concerted efforts must be used toward the recognition and establishment of novel objectives for drug development (Okeke et al., 2011). In the previous chapter, the enzyme plyG efficiency towards the *B. anthracis* is described by utilizing the protein and peptide interactions towards the SrtA. These binding functions between both enzymes are experimentally tested and supports that SrtA is a valid drug target for inhibition of *B. anthracis*. The present scenario is experiencing an urgent demand to carry out fresh approaches for the handling of bacterial infections, because of varying patterns of infectious disease and the emergence of bacterial strain resistant to current antibiotics (Tang et al., 1997). These include all important elements necessary for bacterial survival, even under non-disease conditions as well as virulence determinants whose actions mostly manifested during the infective process (Wilson et al., 2002).

*B. anthracis* utilizes a number of mechanisms to cause disease in human hosts. Especially, these organisms express a broad range of molecules that stick to host cell targets to facilitate a form of different host responses (Njoroge et al., 2009). Proteins attached on the surface of *B. anthracis* have the central role in the contagion, as they boost the bacterial adherence to host cells and tissue to acquire the indispensable foods (Maresso et al., 2008). γ-phage based proteins have much interest towards the SrtA and make the *B. anthracis* to lack of nutrients for its maturation and also termination of cell wall formation (Davison et al., 2005). Cell wall anchoring in Gram positive pathogens occurs by a transpeptidation mechanism requiring surface proteins with C-terminal sorting signals (Navarre et al., 1999). Sortase substrates function as adhesins, internalins, blood clotting factors, immune evasion factors and transporters for nutrients across the microbial cell wall envelope and hence, without them most pathogens cannot sustain an infection (Ton-That et al., 2004).
The *Ba*-SrtA enzyme is a SrtA-type sortase that attaches seven proteins to the cell wall by joining the threonine of the C-terminal LPXTG sorting signal to the amine group of *meso*-diaminopimelic acid (*m*-Dap) within lipid II. *Ba*-SrtA is a potential target for new therapeutics as it is required for *B. anthracis* survival and replication within macrophages, a presumed early step in the development of inhalation anthrax (Weiner *et al.*, 2010). Additionally, it is proven that SrtA is valid small molecule target for suppression of *B. anthracis* (Hecker *et al.*, 2010). *Ba*-SrtA contains several unique active site features that include the presence of an N-terminal extension that contacts the large structurally disordered active site loop. The N-terminal tail in *Ba*-SrtA may be required for later steps in the transpeptidation reaction involving lipid II. An apprehension of the binding site pocket and substrate recognition mechanism of enzymes may be good in the rational growth of SrtA inhibitors (Clancy *et al.*, 2010). The binding pocket with \( \beta 6/\beta 7 \) loop construction in the apostate presumably needs to undergo minimal structural changes to recognize the sorting signal (Ilangovan *et al.*, 2001). To realize the minimal structural alterations in the \( \beta 6/\beta 7 \) loop binding site, the large scale molecular dynamics studies are executed and analyzed the active conformation of SrtA (Weiner *et al.*, 2010). Protein conformation is of paramount importance in understanding Biomolecular interactions. Molecular interactions that involve conformational changes in the interacting molecules are more versatile (Wang *et al.*, 2001). The activity of the compound ultimately plays the role of inhibition with respect to the conformations of proteins and alterations occur in the receptor conformations may vary the actual action of the inhibitory compounds (Christopoulos *et al.*, 2002). In the year of 2009, Suree *et al.*, has reported rhodanine, pyridazinone, and pyrazolethione derivatives as inhibitors of SrtA enzyme from *B. anthracis* (Suree *et al.*, 2009 and Jung *et al.*, 2010). In this present study, 3D QSAR for rhodanine, pyridazinone, and pyrazolethione derivatives is executed for obtaining the bioactive conformation and each conformation are validated by
structural alterations in the β6/β7 loop binding site. The breakthrough of such core structural information along with the strength and selectivity of inhibitor could lead us for the identification of lead molecules and consequently, very useful to design new drug candidates for a particular object.

**MATERIALS AND METHODS**

**Experimental Data Set**

Here, 28 compounds of rhodanine, pyridazinone and pyrazolethione derivatives having experimental activity against *B. anthracis* SrtA, which are reported by Suree *et al.*, 2009 are used for 3D QSAR approach (Suree *et al.*, 2009). The intension of doing the 3D-QSAR is to obtain the bioactive conformation of the ligand molecules (Jiang *et al.*, 2012). The experimental values of all the molecules are available in IC$_{50}$ which are not narrow in range. The value of experimental activity should be in narrow for utilizing the statistical applications and so IC$_{50}$ values are changed over to the pIC$_{50}$ scale (-log IC$_{50}$), in which higher values indicate exponentially greater potency (Schneider *et al.*, 2003). These are the experimental values which are predicted, and coded in nanomolar (NM $\cdot 10^{-9}$ - Describing concentrations one billionth of a molar) and micromolar (µM $10^{-6}$ - Describing concentrations one millionth of a molar) terms of Molar (Santana *et al.*, 2006). There is no tool available as open source software to give the value of pIC$_{50}$ from the IC$_{50}$ values as a pIC$_{50}$ calculator. In this work, the tool for the prediction of pIC$_{50}$ values from the IC$_{50}$ in nanomolar was developed, using the JavaScript programming language, and made available as free and open source software at Sanjeev’s lab web page: www.sanjeevslab.org/tools.html. By using this tool the IC$_{50}$ values are converted into pIC$_{50}$ values and used for the pharmacophore and QSAR study. The whole data set is divided randomly choosing 21 compounds in the training set and seven compounds in test set to maintain the 3:1 ratio.
While splitting the dataset, the uniform distribution is ensured of structurally different compounds with a broad range of pIC$_{50}$ value in both test and training set.

**Ligand Preparation**

The dataset of 28 molecules is minimized by using the ligprep and its conformers are generated by ConfGen along with PHASE. Conformers were generated using a rapid torsion angle search approach followed by minimization of each generated structure using the OPLS-2005 force field, with an implicit GB/SA solvent model. A maximum of 1,000 conformers was generated for each structure using minimum of 100 steps and post process minimization of 50 steps (Shelke et al., 2011). Each minimized conformer was filtered through a relative energy window of 10 kcal mol$^{-1}$ and a minimum atom deviation of 1.00Å. This value (10 kcal mol$^{-1}$) sets an energy threshold relative to the lowest-energy conformer (Reddy et al., 2012). Conformers having higher energy than the threshold are discarded. Distances between all pairs of corresponding heavy atoms must be below 1.00Å for two conformers to be considered identical (Lu et al., 2010). This criterion is applied only after the energy difference threshold, and only if two conformers are within 1 kcal mol$^{-1}$ (Suryanarayanan et al., 2013).

**Generation of Pharmacophoric Sites**

Each compound structure is mapped by a lot of points in 3D space, which coincide with various chemical features that may facilitate binding between the compound and its target receptor. PHASE (PHASE V 3.4) 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) (Tawari et al., 2010). 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 patterns (Ugale et al., 2013). All user-defined
patterns are specified as SMARTS queries and assigned one of three possible geometries, which define physical characteristics of the site like point, vector and group (Kumar Teli et al., 2012). A default setting having acceptor (A), donor (D), hydrophobic (H), negative (N), positive (P), and aromatic ring (R) was used to create pharmacophore sites. Active and inactive thresholds of pIC$_{50}$ 5.9 and 4.3 were applied to the dataset to yield eight actives and five inactives, which is used for pharmacophore modeling and subsequent scoring (Dixon et al., 2006).

Finding Common Pharmacophore and Scoring Hypotheses

Common pharmacophoric features were identified from a set of variants—a set of features that define a possible pharmacophore using a tree-based partitioning algorithm. The terminal size of the box was 1Å, which governs the tolerance on matching; the more closely related pharmacophores having smaller box sizes (Li et al., 2010). Common pharmacophore hypotheses were generated by varying the number of sites ($n_{sites}$) and the number of matching active compounds ($n_{act}$). The common pharmacophore hypotheses were scored by setting the RMSD value below 1.0Å, the vector score value to 0.5 and weighing to include consideration of the alignment of inactive compounds using default parameters (Leone et al., 2008).

Building the 3D-QSAR model

All the common pharmacophore hypotheses generated were used to generate atom-based 3D-QSAR models by correlating the actual and predicted activity for the set of 21 training molecules using Partial Least Square (PLS) analysis (Evans et al., 2007). 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 1.0Å). All models were validated by predicting activity of the set of seven test molecules (Reddy et al., 2012).

Preparation of Protein for Molecular Modeling Calculations
The NMR structure of SrtA (PDB ID: 2KW8) used in the previous chapter is also used in this chapter. But the conformations of SrtA used were minimized by using molecular dynamics study and two different conformations are prepared by using the Protein Preparation Wizard tool implemented in Maestro 9.2. Addition of missing residues, optimization and minimizations are done as described in the previous chapters.

**Molecular Dynamics Simulation**

The SrtA structure *B. anthracis* used in the previous study was also used for this molecular dynamics simulation study. The GROMACS program package (http://www.gromacs.org) adopting the OPLS-AA force field parameters were used for energy minimization and MD simulations of 100ns of timescale (Pronk *et al.*, 2013). For the MD simulation studies, the structure was solvated using the TIP3P water model, and the solvated structure was energy minimized using the steepest descent method, terminating when maximum force is found in smaller than 100 KJ mol⁻¹ nm⁻¹ (Abraham *et al.*, 2011). The structure was energy minimized to eliminate bad atomic contacts and subsequently solvated with water. The total simulation was performed in the NPT ensemble at constant temperature (300K) 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 (Shafreen *et al.*, 2013). In addition to protein simulation studies, the MD simulations were performed for the ligand bound docked structures of SrtA. The initial structure of the receptor and ligands were cleaned using GROMOS96 force field and then the topology files were generated for the receptor and ligands separately using PRODRG tool (Schuttelkopf *et al.*, 2004). The simulation system was created manually by importing the ligand topology into the system pursued along with a dodecahedron box with a margin of 1nm 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 presimulation run of 20ps was applied to relax the system and to remove the geometric restraints which eventually appeared at the initialization of the run. All the simulations were carried out at constant pressure and temperature (NPT) ensemble (Shafreen et al., 2014). The Berendsen coupling was employed to maintain a constant temperature of 300K and constant semi-isotropic pressure of 1 bar with coupling time of 2.0fs 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 (Selvaraj et al., 2014).

**Molecular Docking Simulation**

Docking studies were performed through Glide docking and re-docked through Quantum Polarized Ligand Docking (QPLD) for searching the favorable interactions between ligands (SrtA inhibitors) and a receptor molecule (SrtA with both closed lid and open lid). Grid generation was performed with predicted druggability sites and position of grid box is set as an XYZ axis with the measurement of 57.91, 83.36, and 54.61 respectively (Tripathi et al., 2012). To soften the potential for non-polar part of receptor, Van-der Waal radii of receptor atoms was scaled by 1.00Å with a partial charge cutoff 0.25 (Sengupta et al., 2007). Ligands are set to be rigid conformations, as of all the conformations from 3D QSAR bioactive conformations are used for these docking studies (Agrafiotis et al., 2007). The Glide docking approach was implemented XP (Extra precision) docking were involved for both the conformations of protein (Friesner et al., 2004). For the specificity of ligand partial charges, QM/MM based QPLD docking was also applied for both the protein conformations with all 28 compounds. Ligand polarization was done using the QM/MM based docking approach with existing grid and XP pose viewer file through QPLD (Naik et al., 2004). 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 (Sgrignani et al., 2004).
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 (Cho et al., 2005). Through this QPLD docking, correctness of charge occur in the protein ligand complex and, it will enhance the accuracy of results.

**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) model (Jansen et al., 2004). The LIA model is an empirical method fitted to a set of known binding free energies (Alam et al., 2009). Liaison runs molecular mechanics (MM) simulations of the ligand-receptor complex, and for the free ligand and free receptor using the surface generalized Born (SGB) continuum solvation model (Costanzi et al., 2007). The simulation data and empirical binding affinities are analyzed to generate the Liaison parameters, which are subsequently used to predict binding energies for other ligands with the same receptor (Singh et al., 2005). The empirical function used by Liaison for the prediction of binding affinities is as follows.

\[
\Delta G = \alpha (\langle U_{vdw}^b \rangle - \langle U_{vdw}^f \rangle) + \beta (\langle U_{elec}^b \rangle - \langle U_{elec}^f \rangle) + \gamma (\langle U_{cav}^b \rangle - \langle U_{cav}^f \rangle)
\]

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 (Wildman et al., 1999). Ensemble averages of van der Waals, electrostatic, and cavity (solvent exposed ligand surface area) energies were computed for the docking conformation with open lid and closed lid conformations using an implicit solvation model. The computed energies for each inhibitor complex and corresponding pIC_{50} (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 affinity (Hudock et al., 2008).

**Applicability Domain**

The applicability domain (AD) prediction was performed by using the SIMCA-P 13.0 demo version (Umetrics, 2002). The AD is the Physio-chemical information based on a training set of the model and it’s also applicable to make predictions for new hit compounds (Mitra et al., 2012). By this approach, one can directly analyze the properties of distance matrix, multivariate descriptor to investigate the applicability domain of the training set (Suryanarayanan et al., 2012). These approaches define the AD by calculating distances of a query compound from a defined point within the descriptor space of the training data. 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 for avoiding the errors in the activity prediction applicability domain concept were used.

**RESULTS AND DISCUSSION**

**Determination of Pharmacophore and 3D-QSAR models**

A Pharmacophore modeling and 3D-QSAR studies were performed successfully on a diverse set of rhodanine, pyridazinone and pyrazolethione to understand the effect of spatial arrangement of structural features on SrtA inhibition in *B. anthracis*. Eight compounds with highest activity were selected for common pharmacophore hypothesis generation. Using a tree-based partition algorithm requiring all eight active compounds should match; two probable common pharmacophore hypotheses were generated from the list of variants, based on five pharmacophoric features. On applying the scoring function for five featured pharmacophore hypotheses using default values, two best common pharmacophore
hypotheses AAAHR, and AAADR were selected for 3D-QSAR model building. Training set compounds were aligned on these common pharmacophore hypotheses and are analyzed in PHASE with three PLS factors. 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 is listed in Table 2.1.

<table>
<thead>
<tr>
<th>Statistics</th>
<th>AAAHR</th>
<th>AAADH</th>
</tr>
</thead>
<tbody>
<tr>
<td>SD</td>
<td>0.34</td>
<td>0.41</td>
</tr>
<tr>
<td>R²</td>
<td>0.87</td>
<td>0.82</td>
</tr>
<tr>
<td>F</td>
<td>40.3</td>
<td>25.4</td>
</tr>
<tr>
<td>P</td>
<td>6.079e-08</td>
<td>2.555e-06</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.2637</td>
<td>0.4111</td>
</tr>
<tr>
<td>Q²</td>
<td>0.74</td>
<td>0.6806</td>
</tr>
<tr>
<td>Pearson R</td>
<td>0.88</td>
<td>0.96</td>
</tr>
</tbody>
</table>

Table 2.1: Quantitative Structure Activity Relationship (QSAR) results for the two best Common Pharmacophore Hypotheses

The statistical parameters $r^2$, $q^2$, SD, RMSE and F were used to appraise 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 are 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 $q^2$ value 0.74.
Figure 2.1: Common Pharmacophoric hypothesis of active SrtA inhibitors of *Bacillus anthracis* showing the hypothesis of AAAHR (Three Acceptor, one Hydrophilic and one Aromatic).

Figure 2.2: Scatter plot for the predicted and actual pIC$_{50}$ 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 $R = 0.90$. 
<table>
<thead>
<tr>
<th>Compound</th>
<th>X</th>
<th>R1</th>
<th>R2</th>
<th>R3</th>
<th>R4 Actual IC 50</th>
<th>R4 Predicted IC 50</th>
<th>Residual</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rhodanine, 1-3</td>
<td>−Ph</td>
<td>4-NO₂</td>
<td>3-Bromo</td>
<td>2-OH, 5-NO₂</td>
<td>4.699</td>
<td>4.52</td>
<td>0.179</td>
</tr>
<tr>
<td>Rhodanine, 1-7</td>
<td>−Me</td>
<td>3-Bromo</td>
<td>2-OH, 5-NO₂</td>
<td></td>
<td>4.886</td>
<td>5</td>
<td>-0.114</td>
</tr>
<tr>
<td>Rhodanine, 1-9</td>
<td>Pr</td>
<td></td>
<td></td>
<td></td>
<td>4.276</td>
<td>4.29</td>
<td>0.014</td>
</tr>
<tr>
<td>Rhodanine, 1-12</td>
<td>−Et</td>
<td></td>
<td>2-NO₂</td>
<td></td>
<td>4.131</td>
<td>4.31</td>
<td>-0.179</td>
</tr>
<tr>
<td>Rhodanine, 1-13</td>
<td>−Allyl</td>
<td></td>
<td></td>
<td></td>
<td>4.569</td>
<td>4.52</td>
<td>0.049</td>
</tr>
<tr>
<td>Pyridazine, 2-1*</td>
<td>−SH</td>
<td>−OE1</td>
<td>−Ph</td>
<td>−H</td>
<td>5.854</td>
<td>5.88</td>
<td>-0.026</td>
</tr>
<tr>
<td>Pyridazine, 2-5</td>
<td>−OMe</td>
<td>−SH</td>
<td>−Ph</td>
<td>−H</td>
<td>5.75</td>
<td>5.5</td>
<td>0.245</td>
</tr>
<tr>
<td>Pyridazine, 2-6</td>
<td>−OH</td>
<td>OCH₃Ph</td>
<td>Ph</td>
<td>−H</td>
<td>4.301</td>
<td>3.9</td>
<td>0.401</td>
</tr>
<tr>
<td>Pyridazine, 2-7</td>
<td>−OH</td>
<td>−OMe</td>
<td>Ph</td>
<td>−H</td>
<td>4.301</td>
<td>4.24</td>
<td>0.061</td>
</tr>
<tr>
<td>Pyridazine, 2-8</td>
<td>−OH</td>
<td>−SEt</td>
<td>−Ph</td>
<td>−H</td>
<td>4.301</td>
<td>4.59</td>
<td>-0.289</td>
</tr>
<tr>
<td>Pyridazine, 2-9*</td>
<td>−SH</td>
<td>−SEt</td>
<td>Ph</td>
<td>−H</td>
<td>6.523</td>
<td>6.11</td>
<td>0.413</td>
</tr>
<tr>
<td>Pyridazine, 2-10</td>
<td>−SH</td>
<td>−OEt</td>
<td>−Ph</td>
<td>−H</td>
<td>5.495</td>
<td>5.87</td>
<td>-0.375</td>
</tr>
<tr>
<td>Pyridazine, 2-11</td>
<td>−SH</td>
<td>−OEt</td>
<td>−Ph</td>
<td>4-NO₂</td>
<td>5.174</td>
<td>5.41</td>
<td>-0.236</td>
</tr>
<tr>
<td>Pyridazine, 2-13</td>
<td>−SH</td>
<td>−OEt</td>
<td>−Ph</td>
<td>3-F</td>
<td>5.745</td>
<td>5.7</td>
<td>0.045</td>
</tr>
<tr>
<td>Pyridazine, 2-14*</td>
<td>−SH</td>
<td>−OEt</td>
<td>−Ph</td>
<td>3-Me</td>
<td>5.77</td>
<td>5.68</td>
<td>0.09</td>
</tr>
<tr>
<td>Pyridazine, 2-15</td>
<td>−SH</td>
<td>−OEt</td>
<td>−Ph</td>
<td>3,5-Cl₂</td>
<td>4.854</td>
<td>5.5</td>
<td>-0.646</td>
</tr>
<tr>
<td>Pyridazine, 2-16</td>
<td>−SH</td>
<td>−OEt</td>
<td>Ph</td>
<td>Cyclohexyl</td>
<td>5.854</td>
<td>5.96</td>
<td>-0.106</td>
</tr>
<tr>
<td>Pyridazine, 2-17*</td>
<td>−SH</td>
<td>−OEt</td>
<td>Ph</td>
<td>−H</td>
<td>5.921</td>
<td>5.94</td>
<td>0.019</td>
</tr>
<tr>
<td>Pyridazine, 2-18</td>
<td>−OEt</td>
<td>−SEt</td>
<td>−Ph</td>
<td>−H</td>
<td>5.921</td>
<td>5.66</td>
<td>0.261</td>
</tr>
<tr>
<td>Pyridazine, 2-19*</td>
<td>−OEt</td>
<td>SEt</td>
<td>Ph</td>
<td>3-F</td>
<td>6.046</td>
<td>6.16</td>
<td>-0.114</td>
</tr>
<tr>
<td>Pyridazine, 2-20</td>
<td>−OEt</td>
<td>−SEt</td>
<td>−Ph</td>
<td>3-Me</td>
<td>6.398</td>
<td>6.25</td>
<td>0.148</td>
</tr>
<tr>
<td>Pyridazine, 2-21</td>
<td>−OEt</td>
<td>−SEt</td>
<td>−Ph</td>
<td>3,5-Cl₂</td>
<td>5.284</td>
<td>5.48</td>
<td>-0.196</td>
</tr>
<tr>
<td>Pyridazine, 2-35*</td>
<td>−OEt</td>
<td>−Cl</td>
<td>−Ph</td>
<td>−H</td>
<td>6.523</td>
<td>5.26</td>
<td>1.263</td>
</tr>
<tr>
<td>Pyridazine, 2-36</td>
<td>−OEt</td>
<td>−Cl</td>
<td>−Ph</td>
<td>4-NO₂</td>
<td>3.607</td>
<td>4.04</td>
<td>0.433</td>
</tr>
<tr>
<td>Pyridazine, 2-47</td>
<td>−Cl</td>
<td>−Cl</td>
<td>−Ph</td>
<td>3,5-Cl₂</td>
<td>4.854</td>
<td>4.71</td>
<td>0.144</td>
</tr>
<tr>
<td>Pyrazolethione, 3-9*</td>
<td>S</td>
<td>4-N=N-Ph</td>
<td>H</td>
<td></td>
<td>5.854</td>
<td>5.85</td>
<td>0.004</td>
</tr>
<tr>
<td>Pyrazolethione, 3-17*</td>
<td>S</td>
<td>2,4,6-Br₃</td>
<td>H</td>
<td></td>
<td>6.523</td>
<td>6.72</td>
<td>-0.197</td>
</tr>
<tr>
<td>Pyrazolethione, 3-12</td>
<td>S</td>
<td>2,4,6-Br₃</td>
<td>H</td>
<td></td>
<td>5.854</td>
<td>5.9</td>
<td>-0.046</td>
</tr>
</tbody>
</table>

Table 2.2: Structures and actual versus predicted pIC₅₀ of compounds.

The AAAHR hypothesis has highest $q^2$ value in comparing with other pharmacophore hypotheses, suggesting AAAHR hypothesis is the best model among the other. Additionally, the AAAHR pharmacophore hypothesis has a low RMSE value of the pharmacophore hypotheses, which also supports this hypothesis. Finally, based on $r^2$, $q^2$, SD, and RMSE, the best model was AAAHR hypothesis. Common pharmacophore model alignment of active compounds is shown in figure 2.1. The plots of actual versus predicted pIC₅₀ are shown in the figure 2.2 for training set and test set compounds. The values of each compound predicted
and experimental activity and chemical composition, residual difference is provided in the table 2.2.

**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.

![Diagram](image)

**Figure 2.3:** The significant favorable and unfavorable hydrophobic interactions are shown, when the QSAR model is applied to the (a) most active compound and the (b) most inactive compound.
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 contours generated is shown in figure 2.3. In these representations, blue cubes indicate favorable regions, while red cubes indicate unfavorable regions for the alteration, to enhance the activity. 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 2.3a and 2.3b shows cubes generated from hydrophobic property using QSAR model, these cubes illustrates the significant favorable regions and unfavorable hydrophobic/non-polar interactions that arise when the QSAR model is applied to the most active compound pyridazinone-2-9 and most inactive compound pyridazinone-2-36. Blue cubes were seen on the most active compound pyridazinone-2-9 at the aromatic ring position. Figure 2.3a and 2.3b compares the most significant favorable and unfavorable hydrogen bond acceptors or electron withdrawing features which arise, when the QSAR model is applied to the most active compound pyridazinone-2-9 and to the most inactive compound pyridazinone-2-36. Electron withdrawing group (N) and benzene ring structure were associated with blue-colored cubes in the case of most active compound pyridazinone-2-9 are observed near atoms shows high electron withdrawing feature, while in the case of most inactive compound pyridazinone-2-36, red-colored cubes were visible indicating activity difference in these two molecules. In figure 2.3, blue cubes are representing the sterically favored spatial regions to enhance the activity, while red cubes represent the sterically unfavoured regions.

Validation of Applicability Domain
Chapter 2: Exploring closed and open lid conformations of SrA

The validation of AD results represents that the prediction of test set compounds is 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) is represented in figure 2.4 and from this, it is concluded that these compounds are better for the test set of compounds from this QSAR model.

![Applicability Domain](image)

**Figure 2.4:** The DmodX values of eight test set compounds for applicability domain prediction.

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 the design of better 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. The successful statistical data based ligand conformational changes are carried for further study.

**Molecular Dynamics Simulations**

The stability, flexibility and intramolecular conformations of the SrtA structure is analyzed with this 100ns large scale molecular dynamics simulation. The simulation event is involved in this study is to understand the biological happening in the SrtA structure and alterations occurs in each interval of timescale. From the system, the temperature, total energy, mass density, and volume are relatively stable throughout the simulation period, suggesting that the simulations were carried out satisfyingly and system has been well equilibrated.

![RMSD plot of SrtA for 100ns of timescale.](image)

**Figure 2.5:** Analysis of SrtA structure in dynamic state with the time scale of 100ns (a)

RMSD plot of SrtA for 100ns of timescale.
**Figure 2.6:** RMSF plot showing residue wise changes in the SrtA structure.

**Figure 2.7:** Intramolecular conformational changes of SrtA occur in molecular dynamics simulation of 100ns of timescale (β6/β7 loop active site zoomed with several conformations).
RMSD from the NMR structure of C\(^\alpha\) atoms vs timescale (ns) in figure 2.5 shows, initially the protein is more flexible up to 40ns and so the deviations are inconsistent and especially from 2ns to 6ns the proteins have more deviations and loop structures in the protein structure are responsible for these initial deviations. From the 40\(^{th}\) ns the protein structure was consistent enough as shown in the RMSD figure. 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 2.6. The fluctuations of each amino acid are clearly shown and it clearly shows that the same loop was much fluctuated. RMSF profiles indicate that the residues with higher fluctuation values are those in the long loop Gly53-Asp64 (\(\beta6/\beta7\) Loop). In between the 0-100ns there was much conformational changes occurs in the native protein structure. The MD simulation of the SrtA structure showed that the loop structure in the N-terminal region (53-64), present nearby active site is widely opened at initial position (0ns). It was the just near to the active site regions and this loop region initially looks like tail for the SrtA structure and each ns this particular loop has more conformational changes.

The structural changes were carefully investigated for each 5ns time interval and noticed that the N-terminal tail loop has the biological function of the lid of the active site. The tail loop keeps on fluctuating at the dynamic state till the 30ns and after that tail loop structure slowly changes its position and function as a lid. The lid closes the active site region and after that till 100ns the lid was remain closed. Figure 2.6 and 2.7 explain the conformational changes occur in the SrtA structure (figure 2.7 showing the overall morphed conformations and figure 2.8 showing the separate conformations of SrtA at each 5ns of interval).
Figure 2.8: Observation of SrtA active site loop closure and open at different timescale (Observed with 5ns interval).

Trajectory analysis suggests that the loop structure in the N-terminal is keep on 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. Figure 2.9 and 2.10 indicates the opening and closing of lid in different timescale of the dynamic environment and this lid is called as a universal feature of SrtA 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 not an easy task of designing the suitable compounds for closed binding pockets of SrtA.

**Figure 2.9:** 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 2.10:** Visualization of open lid loop with closed lid loop structure.
Chapter 2: Exploring closed and open lid conformations of SrtA

Figure 2.11: Hydrogen bond analysis of SrtA structure with water contact shows that pair above 0.35 nm (A) and pair within 0.35 nm (B) are showing high contacts in 0-40ns and after that the hydrogen bonds has been decreased.

The overall simulation shows good contacts with the system and protein throughout the simulation. In the hydrogen bond analysis between the SrtA and water model shows that, there is a vast difference in hydrogen bonds. Figure 2.11 shows the Hydrogen bond interactions analysis of SrtA structure with water contact shows that pair above 0.35 nm (A) and pair within 0.35 nm (B). The average of hydrogen bonds 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 and so it is required to understand the interaction difference happen due to this biological function. The average of above 0.35nm and pair within 0.35nm is comparatively high for 0-40ns, when compared with 40-100ns. The number of hydrogen bonds between the protein and the surrounding water molecules were more in the beginning of the timescale (open lid) and comparatively less in later stages of the time scale (closed lid). This is due to closing of active site protects the hydrophobic regions in the 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 lid and open lid structure of SrtA enzyme through molecular modeling techniques. The SrtA loop function has been clearly explored with the closed active site and open active site with reference to the biological mechanism described in figure 2.12.

![Diagram of SrtA inhibitors against Open and Closed lid conformations](image)

**Figure 2.12:** Mechanism of SrtA inhibitors against the open lid and closed lid conformations of SrtA in *Bacillus anthracis*.

The bioactive conformers obtained from the 3D QSAR are validated for both open lid and closed lid conformations, to check the active conformation of SrtA. For obtaining the values of activity based on the conformations of proteins, and used QPLD and LSBD calculations with both closed and open lid conformations. Docking procedures 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. Initially, docking was performed through the XP and for the correctness of interactions QPLD docking was performed. Scoring of docked poses is still regarded as one of the major challenges in the field of molecular docking. The purpose of the scoring procedure is the identification of the correct binding pose by its lowest energy value, and the ranking of protein-ligand complexes according to their binding affinities.

Scoring functions can be divided into empirical scoring functions, scoring functions derived from force fields, and knowledge-based scoring functions. Scoring functions derived from force fields handle the ligand binding prediction with the use of potential energies (non-bonded interaction terms) and sometimes in combination with solvation and entropy contributions. Knowledge-based scoring functions are based on atom pair potentials derived from structural databases. Forces and potentials are collected from known protein-ligand complexes to get a score for their binding affinities. Empirical scoring functions derive from training sets of protein-ligand complexes with determined affinity data. One general aspect in the finding of an accurate empirical scoring function is the assumption that each occurrence of an individual interaction is considered as equivalent.

Here, based on the scoring functions and QM/MM generated partial charges are playing vital role for predictions of theoretical scoring with respect to both close and open lid. Initial docking with XP charges and followed by QPLD docking generates most accurate interactions and scorings according to Das et al, 2009. The QM-Polarized Ligand Docking (QPLD) protocol is an improved docking method, which incorporates quantum mechanical and molecular mechanical (QM/MM) calculations. This method applies the Glide algorithm to generate the best candidate poses for ligand docking. The partial charges on the atoms of the ligand are then replaced with charges derived from QM calculations on the ligand in the field of the receptor for each ligand-receptor complex. The charges are calculated from the
electrostatic potential energy surface of the ligand, which is generated from a single-point calculation using the B3LYP density function for the QM region. Glide then re-docks each of the ligands with updated atom charges, and returns the most energetically favorable pose.

So that, bioactive conformations obtained from 3D QSAR are made rigid conformations and both closed lid and open lid are interacted. The use of ligand charge polarization using QM/MM is more accurate method to substantially improve the interactions between protein-ligand complex structures. QM/MM approach rectifies erroneous electronic charges in protein-ligand molecular docking, through \textit{ab initio} methods to calculate ligand charges within the protein atmosphere using QPLD approach. The correctness in the electronic charge in protein ligand docking implies to increase the accuracy of docking interactions. Decent changes were observed in between XP docking and QM docking with every atom in terms of partial charge and these charges play a vital role in hydrogen bond formation. Protein-ligand interactions are based on physical theory of positive ($+^{\text{ve}}$) and negative ($-^{\text{ve}}$) charges of protein and ligand. The unlike charges of $+^{\text{ve}}$ and $-^{\text{ve}}$ interacts with each other by the mediation of hydrogen bond. The neutral charges seen in XP docking are filled by QM/MM charge calculation and due to allocation of partial charges, there is a difference in H-bond interaction and distance between interacting H-bond. Best conformations of SrtA inhibitors with both protein conformations are taken into account for the energy calculation.

\textbf{Energy Calculation Accounts for Activity Prediction}

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 prediction of activity of a ligand against both open and closed lid conformation of SrtA is equally important in drug design of \textit{B. anthracis}. The experimental activity reported by Suree \textit{et al.}, 2009 is considered in pIC$_{50}$ and liaison calculation is analyzed. 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. Each step is carried out as two tasks, a simulation task and an analysis task.

Figure 2.13: Comparison of predicted and experimental activity of SrtA against the closed and open lid conformation of SrtA from Bacillus anthracis.
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 estimates binding affinities using $<U_{el}>$, $<U_{vdw}>$ and $<U_{cav}>$ 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 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 lid and closed lid. The correlation analysis shows in the figure 2.13, predicts that, the open lid protein conformations show $r^2$ of 0.730 and closed lid conformations 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 calculations.

**Catalytic Triad Involvement in Binding Mechanism**

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. 92% of the active compounds show the interactions with this catalytic triad and compound Pyridazinone-2-17 is additionally showing the hydrophobic interactions with Ala58. As discussed earlier that the lid occupied the binding pocket in closed position and protects the hydrophobic surface of the active site, 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 2.13 shows the ligand aromatic ring structure involved in the hydrophobic interactions with Met56, Val110 and Trp171.
**Figure 2.13:** Active compounds showing the aromatic ring structure showing hydrophobic interactions with MET56, VAL110 and TRP171.

The green colored bonds (Hydrophobic bonds) indicate the hold of aromatic ring structure inside the hydrophobic binding pocket. One of the amino acid MET56 is contributing in the function of the closed lid loop and also its actively participated in the hydrophobic interactions. These three amino acids are core important for targeting this SrtA structure and the aromatic ring structure is core important for the inhibition of SrtA activity.

**Ligand Stability- Molecular Dynamics Simulation**

Based on activity, docking and binding energy, five best ligands were chosen for dynamic behavior, which were analyzed in trajectories.
Figure 2.14: RMSD graph (nm) for understanding the active compound interactions against the closed lid SrtA structure.

A molecular dynamics simulation 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 2.14 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. The figure explained that experimentally validated compounds were stable at interactions and bind properly inside the closed binding pocket of SrtA. The figure 2.14 shows that initially all the compounds is slightly moved from its original position and after that the binding of ligand 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 having role in holding ligand and also interaction with aromatic ring. The average mean values of all RMSD are calculated for pyridazinone-2-19, pyridazinone-2-9, pyrazolethione-3-12, pyridazinone-2-5 and Pyridazinone-2-20 and values 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 the values and graph of the RMSD, 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 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.

Conclusion

Current study explained 3D pharmacophoric features and 3D-QSAR with respect to selected SrtA inhibitors, which are experimentally validated against the *B. anthracis*. Several Pharmacophoric hypotheses of SrtA inhibitors were developed using PHASE, and alignment based on these pharmacophore was used as input for the development of atom-based 3D-
QSAR model. A five-point pharmacophore with three hydrogen bond acceptors (A), one hydrophobic (H), and one aromatic ring (R) as pharmacophoric features were associated with a 3D-QSAR model giving good statistical significance and predictive ability. Visualization of the 3D-QSAR model, providing details of the relationship between structure and activity among these molecules, and thus provides explicit indications for the design of better 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 and for understanding the function of N-terminal tail loop, large scale dynamics were carried out on SrtA. From the Molecular dynamics studies, the N-terminal loop is playing a role of ‘lid’ as a protector of the active site hydrophobic environment. The SrtA flexible structure has been replaced by a dynamic stable model, with reduced internal motions and conformational changes which plays an essential role in their function of active site. The overall dynamic deviation suggested that the open lid loop makes more conformational changes and deviation happen drastically up to 40ns. When the protein reaches above 40ns, the lid loop start closing at the active site and protein establish very stable conformation up to 100ns of timescale. Interestingly, the results of open lid and closed lid of active site demonstrates that SrtA tail loop will be closed after the 40ns of timescale and it protects the active site to bind with the inhibitors. The experimentally proven SrtA inhibitors are docked by means of QPLD and both conformations of open lid and closed lid are analyzed for activity prediction. 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 40$^{\text{th}}$ ns of molecular dynamics simulation is more active and suitable for molecular modeling calculation. The interaction between ligand and protein are looking 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 important role in the inhibition and so this particular aromatic ring structure is required for the interaction with the catalytic triad. 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 studies and docking studies performed here could be very valuable for the development of new and potent leads for SrtA inhibition.