CHAPTER TWO

MOLECULAR MODELLING STUDIES

2.1. QSAR Studies (2D and 3D QSAR STUDIES)

2.2. Design of New Chemical Entities (NCEs) Containing Pyrimidinone Pharmacophore

2.3. Molecular Docking Studies

2.4. Structure Based Drug Design

2.5. ADME Prediction of designed compounds
2.1. QSAR STUDIES:

2.1.1. INTRODUCTION:

Quantitative structure activity relationship (QSAR) paradigm is based on the assumption that there is an underlying relationship between the molecular structure and the biological activity which arises from a systematic variation. On this basis, QSAR attempts to establish a correlation between the chemical structure of the compound and biological activity. In QSAR model, the physicochemical properties expressing the structural components of the compounds are transformed into a set of numerical descriptors of properties relevant to biological activity and then it establishes the quantitative relationships between the physicochemical properties and biological activities of the compound.

The fundamental hypothesis of QSAR states that the biological activity of a series of compounds can be quantitatively correlated to the physicochemical and/or structural properties of the compound. There are two main objectives for the development of QSAR:

1. Development of predictive and robust QSAR model with a specified chemical domain for prediction of activity of untested molecules.
2. It acts as an informative tool concerning the mechanisms causing a given biological activity i.e. informative by extracting significant patterns in descriptors related to the measured biological activity. The process of QSAR model development is divided into three key steps:

1. Data Preparation
2. Data Analysis
3. Model Validation

Drug design by QSAR is an ever expanding field. The first model for the QSAR studies was reported in 1964 and is known as the "Hansch model". It is the most popular model used in drug design by QSAR. Other frequently used models include free Wilson model, pattern recognition, some statistical procedures such as factor analysis, discriminant analysis and principle component analysis.
Hansch model:

It is a linear free energy related approach, which states that the biological activity of a series of compounds can be correlated with a combination of lipophilic, electronic and steric parameters. Linear and parabolic dependence of lipophilicity on the biological activity are as shown in the below equation:

\[ \log BA = a \pi + b \sigma + c E_s + d \] \text{ - Linear Equation}

\[ \log BA = a \pi^2 + b \pi + c \sigma + d E_s + e \] \text{ - Eq. 2.1}

Where \( a, b, c, d \) and \( e \) are the constants generated by regression equation.

When a worthwhile correlation between biological activity and physicochemical parameters is obtained, further synthesis of only “active” compounds, as predicted by the QSAR equation is attempted.

**Data requirement:**

1. **Biological activity:**

   For QSAR analysis, a dataset of a series of synthesized molecules tested for its desired biological activity is required. For a QSAR to be valid and reliable, the activity of all of the chemicals covered must be elicited by a common mechanism. The quality of the model is totally dependent on the quality of the experimental data used for building the model. Biological activity can be of two types:

   1) Continuous Response: MIC, IC\textsubscript{50}, ED\textsubscript{50}, % inhibition

   2) Categorical response: Active/Inactive

   In order to have confidence in QSAR analysis, biological data of at least 20 molecules is required:

   1. Preferably tested in the same lab and by the same biological assay method.

   2. With wide range and uniform distribution of the activity data.

   3. Activity should be well-defined in terms of either real number (continuous response, and cannot be e.g. >1000 or <1000) or in a particular class (categorical response).
2. Molecular descriptors:

Molecular Descriptors can be defined as a numerical representation of chemical information encoded within a molecular structure via mathematical procedure.

Type of QSAR is based on the dimensionality of molecular descriptors used:

- **0D QSAR:** These are descriptors derived from molecular formula (e.g. molecular weight, number and type of atoms etc.).

- **1D QSAR:** A substructure list representation of a molecule can be considered as a 1D molecular representation and consists of a list of molecular fragments (e.g. functional groups, rings, bonds, substituents etc.).

- **2D QSAR:** A molecular graph contains topological or two dimensional (2D) information. It describes how the atoms are bonded in a molecule, both the type of bonding and the interaction of particular atoms (e.g. total path count, molecular connectivity indices etc.).

- **3D QSAR:** These are calculated starting from a geometrical or 3D representation of a molecule. These descriptors include molecular surface, molecular volume and other geometrical properties. There are different types of 3D descriptors e.g. electronic, steric, shape etc.

- **4D QSAR:** In addition to the 3D descriptors the 4th dimension is generally in terms of different conformations or any other experimental condition.

3. Selection of training and test Sets:

QSAR models are used increasingly to screen chemical databases and/or virtual chemical libraries for potentially bioactive molecules. These developments emphasize the importance of rigorous model validation to ensure that the models have both the ability to explain the variance in the biological activity (internal validation) and also the acceptable predictive power (external validation).

For model validation the dataset is required to be divided into training set (for building the QSAR model) and test set (for examining its predictive ability). For any QSAR model, it is of crucial importance that the training set selected to calibrate the model exhibits a well balanced distribution and contains representative molecules.

Following are the methods for division of dataset into training and test set:
1. **Manual selection:** This is done by visualizing the variation in the chemical and biological space of the given dataset.

2. **Random selection:** This method creates training and test set by random distribution.

3. **Sphere exclusion method:** This is a rational method for creation of training and test set. It ensures that the points in both the sets are uniformly distributed with respect to chemical and biological space.

4. **Others:**

   Experimental design, Full factorial, Fractional factorial, Principal component analysis, Self organizing maps (SOM).

5. **Variable selection methods:**

   There are hundreds of molecular descriptors available for building a QSAR model. Not all of the molecular descriptors are important in determining the biological activity and hence to find the optimal subset of the descriptors which plays an important role in determining activity, a variable selection method is required. The variable selection method could be divided mainly into two categories:

   **I. Systematic variable selection:** These methods add and/or delete a descriptor in steps one-by-one in a model.

   (a) Stepwise forward
   (b) Stepwise forward-backward
   (c) Stepwise backward

   **II. Stochastic variable selection:** These methods are based on simulation of various physical or biological processes. These methods create model starting from randomly generated model(s) and later modifying these model(s) by using different process operator(s) (e.g. perturbation, crossover etc.) to get better model(s).

   (a) Simulated annealing
   (b) Genetic/ Evolutionary algorithms;
   (c) Modified particle swarm optimization;
   (d) Artificial colony system.
5. Use of statistics in QSAR analysis:

Various statistical tools are available to analyze the data of molecular properties and to correlate various molecular descriptors with their corresponding biological activity. A suitable statistical method coupled with a variable selection method allows analyzing all of this data in order to establish a QSAR model with a subset of descriptors that are most statistically significant in determining the biological activity.

The statistical methods can be broadly divided into: Linear and non-linear methods. In statistics, a correlation is established between dependent variables (biological activity) and independent variables (molecular descriptors). The linear method fits a line between the selected descriptors and activity as compared to non-linear method which fits a curve between the selected descriptors and activity. The statistical method to build QSAR model is decided based on the type of biological activity data. Following are few commonly used statistical methods:

i) Categorical dependent variable:
- Discriminant analysis.
- Logistic regression.
- K-Nearest neighbor classification.
- Decision trees.
- SIMCA.

ii) Continuous dependent variable:
- Multiple regressions.
- Principal component regression.
- Continuum regression.
- Partial least squares (PLS) regression.

6. Evaluation of the model: Various statistical measures available for evaluation of the significance of the model, the most commonly used measures are enlisted below:

- n - Number of molecules (> 20 molecules)
- Number of descriptors in a model (statistically n/5 descriptors in a model)
• df- Degree of freedom (n-k-1) (higher is better)
• $r^2$- Coefficient of determination (> 0.7)
• $q^2$- Cross validated $r^2$ (>0.5)
• pred_ $r^2$. $r^2$ for external test set (>0.5)
• SEE- Standard error of estimate (smaller is better)
• F-test- F-test for statistical significance of the model (higher is better, for same set of descriptors and molecules)
• F_prob.- Alpha error probability (smaller is better)
• Z score – Z score calculated by the randomization test (higher is better)
• best_ran_q^2 - Highest $q^2$ value in the randomization test (as low as compared to $q^2$)
• Alpha- Statistical significance parameter by randomization test (<0.01)

Note: Comment(s) in the parenthesis suggests the minimum recommended values for significant QSAR model.

2.1.2. COMPUTATIONAL DETAILS:

All computational studies were performed using VLife Sciences MDS Version 3.5. The compounds were constructed from the fragments in the VLife Molecular Builder database with standard bond lengths and bond angles. The geometry optimization was carried out using the standard Merck molecular force field (MMFF) with distance dependent-dielectric function and energy gradient of 0.001 kcal/mol Å. The initial conformations were selected and minimized using the Powell method till root-mean-square deviation 0.001 kcal/mol Å was achieved.

2.1.3. BIOLOGICAL DATA:

A data set consisting of 28 substituted benzyl pyrimidine derivatives reported by Gasse et. al. for anti-mycobacterial activity was taken from the literature. The inhibitory activities (ki) were converted into the corresponding pKi [$pKi = -\log(Ki)$]. The compound number 3 is the least potent compound of the series and the compound number 11 is the most potent compound of the series.
Table 2.1: Structure and antimycobacterial activity data $Ki$ (in μmoles)

<table>
<thead>
<tr>
<th>Compound</th>
<th>X</th>
<th>Y</th>
<th>$Ki$ (μM)</th>
<th>p$Ki$ (μM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>-CH$_3$</td>
<td>-CH$_2$CH$_2$CONH$_2$</td>
<td>0.89</td>
<td>-1.59493</td>
</tr>
<tr>
<td>02</td>
<td>-CH$_3$</td>
<td>-CH$_2$CH$_2$COOH</td>
<td>0.55</td>
<td>-1.7403</td>
</tr>
<tr>
<td>03</td>
<td>-CH$_3$</td>
<td>-CH=CHCONH$_2$</td>
<td>1.95</td>
<td>-1.2900</td>
</tr>
<tr>
<td>04</td>
<td>-CH$_3$</td>
<td>-CH=CHCOOH</td>
<td>0.39</td>
<td>-1.5910</td>
</tr>
<tr>
<td>05</td>
<td>-CH$_3$</td>
<td>-CH$_2$COOH</td>
<td>0.13</td>
<td>-1.1139</td>
</tr>
<tr>
<td>06</td>
<td>-CH$_3$</td>
<td>-CH$_2$CONH$_2$</td>
<td>1.12</td>
<td>-2.0492</td>
</tr>
<tr>
<td>07</td>
<td>-CH$_3$</td>
<td>-C === CH$_2$CH$_2$OH</td>
<td>0.70</td>
<td>-1.8450</td>
</tr>
<tr>
<td>08</td>
<td>-CH$_3$</td>
<td>-CH$_2$OH</td>
<td>0.51</td>
<td>-1.7075</td>
</tr>
<tr>
<td>09</td>
<td>-H</td>
<td>-CH$_2$COOH</td>
<td>2.02</td>
<td>-2.3053</td>
</tr>
<tr>
<td>10</td>
<td>-Br</td>
<td>-CH$_2$COOH</td>
<td>0.10</td>
<td>-1.0000</td>
</tr>
<tr>
<td>11</td>
<td>-Cl</td>
<td>-CH$_2$COOH</td>
<td>6.50</td>
<td>-0.8129</td>
</tr>
<tr>
<td>12</td>
<td>-Br</td>
<td>-CH$_2$CONH$_2$</td>
<td>0.39</td>
<td>-1.5910</td>
</tr>
<tr>
<td>13</td>
<td>-Cl</td>
<td>-CH$_2$CONH$_2$</td>
<td>0.39</td>
<td>-1.5910</td>
</tr>
<tr>
<td>14</td>
<td>-CH$_3$</td>
<td>-CH$_2$COOH</td>
<td>0.58</td>
<td>-1.7634</td>
</tr>
<tr>
<td>15</td>
<td>-CH$_3$</td>
<td>-CH$_2$CONH$_2$</td>
<td>0.55</td>
<td>-1.7403</td>
</tr>
<tr>
<td>16</td>
<td>-Br</td>
<td>-CH$_2$COOH</td>
<td>0.38</td>
<td>-1.5797</td>
</tr>
<tr>
<td>17</td>
<td>-Cl</td>
<td>-CH$_2$COOH</td>
<td>0.16</td>
<td>-1.2041</td>
</tr>
<tr>
<td>18</td>
<td>-Br</td>
<td>-CH$_2$CONH$_2$</td>
<td>0.35</td>
<td>-1.5440</td>
</tr>
<tr>
<td>19</td>
<td>-CH$_3$</td>
<td>-C === CH$_2$CH$_2$COOH</td>
<td>0.96</td>
<td>-1.9822</td>
</tr>
<tr>
<td>20</td>
<td>-CH$_3$</td>
<td>-CH$_2$COOH</td>
<td>0.47</td>
<td>-1.6721</td>
</tr>
<tr>
<td>21</td>
<td>-CH$_3$</td>
<td>-CH$_2$CONH$_2$</td>
<td>2.65</td>
<td>-1.4232</td>
</tr>
<tr>
<td>22</td>
<td>-Br</td>
<td>-CH$_2$COOH</td>
<td>0.34</td>
<td>-1.5314</td>
</tr>
<tr>
<td>23</td>
<td>-Cl</td>
<td>-CH$_2$COOH</td>
<td>0.23</td>
<td>-1.3617</td>
</tr>
<tr>
<td>24</td>
<td>-Br</td>
<td>-CH$_2$CONH$_2$</td>
<td>1.95</td>
<td>-1.2900</td>
</tr>
<tr>
<td>25</td>
<td>-Cl</td>
<td>-CH$_2$CONH$_2$</td>
<td>0.26</td>
<td>-1.4149</td>
</tr>
<tr>
<td>26</td>
<td>-Br</td>
<td>-CH$_2$COOCH$_3$</td>
<td>0.52</td>
<td>-1.7160</td>
</tr>
<tr>
<td>27</td>
<td>-Cl</td>
<td>-CH$_2$COOCH$_3$</td>
<td>0.59</td>
<td>-1.7708</td>
</tr>
<tr>
<td>28</td>
<td>-CH$_3$</td>
<td>-Br</td>
<td>0.38</td>
<td>-1.5797</td>
</tr>
</tbody>
</table>
For QSAR model building, the compounds were divided in training and the test set using Sphere exclusion method. The Sphere exclusion is a well known dissimilarity based compound selection. In this algorithm, molecules are selected whose similarities with each of the other selected molecules are not higher than the defined threshold. Therefore, each selected molecule creates a (hyper) sphere around itself, so that any candidate molecules inside the sphere are excluded from the selection. The radius of the sphere is an adjustable parameter, determining the number of compounds selected and the diversity among them. The original method starts with the "most descriptive compound" and in each cycle identifies the compound most similar to the centroid of the already selected compounds.

After this selection, a Uni-Column statistics (Table 2.2) for training set and test set was generated to check the correctness of selection criteria for training and test set molecules. The maximum and minimum value in training and set were compared in a way that:

1. The maximum value of pKi of test set should be less than or equal to maximum value of pKi of training set.
2. The minimum value of pKi of test set should be higher than or equal to minimum value of pKi of Training set.
3. The vale for standard deviation of training set should be more than or equal to the Standard deviation of test set.

Table 2.2: Uni-Column statistics for training and test set

<table>
<thead>
<tr>
<th>Particulars</th>
<th>Column Name</th>
<th>Average (mean)</th>
<th>Max*</th>
<th>Min*</th>
<th>Std Dev</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training Set</td>
<td>pKi</td>
<td>-1.5925</td>
<td>-0.8130</td>
<td>-2.3050</td>
<td>0.3778</td>
<td>-31.8500</td>
</tr>
<tr>
<td>Test Set</td>
<td>pKi</td>
<td>-1.6464</td>
<td>-1.2900</td>
<td>-2.0490</td>
<td>0.2883</td>
<td>-8.2320</td>
</tr>
</tbody>
</table>
2.1.4. 2D QSAR STUDIES:

Various 2D descriptors (220) like topological, physicochemical, alignment independent and atom type count descriptors were calculated. The preprocessing of the independent variables was carried out by removing invariable column resulting in around 150 descriptors. Further selection of descriptors has been carried out by using correlation to obtain most representative descriptors.

The most widely used Multiple linear regression (MLR) analysis was used to correlate biological activities with physicochemical properties and in turn chemical composition of the selected series of compounds. MLR is the standard method for multivariate data analysis. It is also called as Ordinary least squares regression (OLS) method. This method of regression estimates the values of the regression coefficients by applying least squares curve fitting method. For getting reliable results, dataset having typically 5 times as many data points (molecules) as independent variables ( descriptors) is required. The regression equation takes the form as mentioned below:

\[
Y = b_1 x_1 + b_2 x_2 + b_3 x_3 + c,
\]

--- Equation 2.1

Where, \(Y\) is the dependent variable (Biological activity, \(pIC_{50}\)), the \(b_1\) to \(b_3\) are regression coefficients (contribution of respective descriptors i.e. \(x_1\) to \(x_3\)), \(x_1\) to \(x_3\) are independent variables (Descriptors), and \(c\) is a regression constant or intercept.

The various QSAR models were developed by MLR method by optimizing the statistical results by variation of the descriptors in these models. Physicochemical descriptors, Baumann alignment independent descriptors and MMFF Atom type count descriptors were used for the 2D QSAR model generation.

Physicochemical descriptors are based on the physicochemical properties of molecule whereas Alignment independent (AI) descriptors were calculated as discussed by Baumann. For calculation of AI descriptors every atom in the molecule was assigned at least one and at most three attributes. The first attribute is ‘T-attribute’ to thoroughly characterize the topology of the molecule. The second attribute is the atom type. The atom symbol is used here. The third attribute is assigned to atoms taking part in a double or triple bond. After all atoms have been assigned their respective attributes, selective distance count statistics for all combinations of different attributes...
are computed. A selective distance count statistic \( XY2 \) (e.g. 'TOPO2N3') counts all the fragments between start atom with attribute 'X' (e.g. '2' double bonded atom) and end atom with attribute 'Y' (e.g. 'N') separated by the graph distance 3.

The graph distance can be defined as the smallest number of atoms along the path connecting two atoms in molecular structure. The atom type count descriptors are based on MMFF atom types and their count in each molecule. In MMFF, there are 99 atom types and hence 99 descriptors indicating number of times that atom type has occurred in a given molecule are generated.

2.1.4.1. Partial least squares regression method:

PLS is an effective technique for finding the relationship between the properties of a molecule and its structure. In mathematical terms, PLS relates a matrix \( Y \) of dependent variables to a matrix \( X \) of molecular structure descriptors, i.e., a latent variable approach to modeling the covariance structures in these two spaces \(^{123}\). PLS has two objectives: to approximate the \( X \) and \( Y \) data matrices, and to maximize the correlation between them. Whereas the extraction of PLS components is performed stepwise and the importance of a single component is assessed independently, a regression equation relating each \( Y \) variable with the \( X \) matrix is created. PLS decomposes the matrix \( X \) into several latent variables that correlate best with the activity of the molecules \(^{124}\).

Three molecules, viz., 5, 10 and 11 were considered as outlier molecules. Literature Survey revealed that the compounds that have unexpected biological activity and are unable to fit in a QSAR model are known as outliers\(^ {125-126}\).

Three QSAR models were generated (SA-MLR, SA-PLS and SA-PCR) \(^ {110-11, 123-124}\) and the SA-MLR model, found to be the best, (Equation 2.2) was further considered to predict the activity. Results of 2D QSAR study along with the contribution of the descriptors are presented in Table 2.3 and Equation 2.2. The obtained \( r^2 \) was 0.7830 with a \( q^2 \) of 0.6862 and predicted \( r^2 \) of 0.6588. The observed and predicted activities for the training and test set are shown in Table 2.4 and Table 2.5 respectively. The plot of observed versus predicted activities according to the equation 1 is shown in Graph 2.1 for training and test sets respectively.
The best regression equation obtained is represented below:

\[
pKi = 0.2791(\pm 0.1564) \text{SaaCHE-index} + 1.4179(\pm 0.0714) \text{SdsCHE-index} + 0.0969(\pm 0.0470) \log p - 2.2042(\pm 0.1618) \text{SdsCHcount} - 4.0935
\]

\[\text{--- Eqation 2.2}\]

**Table 2.3: Results of 2D QSAR**

<table>
<thead>
<tr>
<th>Statistical Parameter</th>
<th>SA-MLR</th>
<th>SA-PLS</th>
<th>SA-PCR</th>
</tr>
</thead>
<tbody>
<tr>
<td>(r^2)</td>
<td>0.7830</td>
<td>0.7896</td>
<td>0.7072</td>
</tr>
<tr>
<td>(q^2)</td>
<td>0.6862</td>
<td>0.6699</td>
<td>0.5434</td>
</tr>
<tr>
<td>F test</td>
<td>13.5292</td>
<td>14.07140</td>
<td>12.8815</td>
</tr>
<tr>
<td>(\text{pred}_r^2)</td>
<td>0.6588</td>
<td>0.6392</td>
<td>0.6416</td>
</tr>
<tr>
<td>(\text{pred}_r^2\text{se})</td>
<td>0.1690</td>
<td>0.1738</td>
<td>0.1732</td>
</tr>
<tr>
<td>Positively Contributing Descriptors</td>
<td>SaaCHE-index (2.67%)</td>
<td>H-Donor Count (12.83%)</td>
<td>(\chi_4\text{pathCluster}) (19.16%)</td>
</tr>
<tr>
<td></td>
<td>SdsCHE-index (46.81%)</td>
<td>(\chi_3\text{Cluster}) (40.28%)</td>
<td>(\text{SssSNE-index}) (34.47%)</td>
</tr>
<tr>
<td></td>
<td>(\log p) (1.77%)</td>
<td>(\text{XlogP}) (21.56%)</td>
<td>(\text{XlogP}) (18.33%)</td>
</tr>
<tr>
<td>Negatively Contributing Descriptors</td>
<td>SdsCHcount (-48.75%)</td>
<td>3ClusterCount (-25.33%)</td>
<td>(\text{SssO count}) (-28.04%)</td>
</tr>
<tr>
<td>best_ran(r^2)</td>
<td>0.5537</td>
<td>0.4182</td>
<td>0.6302</td>
</tr>
<tr>
<td>best_ran(q^2)</td>
<td>0.0005</td>
<td>0.1096</td>
<td>0.3887</td>
</tr>
<tr>
<td>(Z\text{ score}_\text{ran}r^2)</td>
<td>5.2712</td>
<td>4.4241</td>
<td>4.9052</td>
</tr>
<tr>
<td>(Z\text{ score}_\text{ran}q^2)</td>
<td>2.7152</td>
<td>3.5240</td>
<td>3.2817</td>
</tr>
<tr>
<td>(\alpha\text{-statistical significance parameter by randomization test})</td>
<td>0.0500</td>
<td>0.0500</td>
<td>0.0000</td>
</tr>
</tbody>
</table>
## Table 2.4: Observed and predicted activity of the training set of molecules

<table>
<thead>
<tr>
<th>Sr. No</th>
<th>Molecule No.</th>
<th>Biological Activity pKi (μM)</th>
<th>SA-MLR Predicted Activity</th>
<th>Residual</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>01</td>
<td>-1.9493</td>
<td>-1.9014</td>
<td>0.0479</td>
</tr>
<tr>
<td>02</td>
<td>02</td>
<td>-1.7403</td>
<td>-1.9205</td>
<td>-0.1802</td>
</tr>
<tr>
<td>03</td>
<td>03</td>
<td>-2.2900</td>
<td>-2.2924</td>
<td>-0.0024</td>
</tr>
<tr>
<td>04</td>
<td>04</td>
<td>-1.5910</td>
<td>-1.5917</td>
<td>-0.0007</td>
</tr>
<tr>
<td>05</td>
<td>07</td>
<td>-1.8450</td>
<td>-1.8861</td>
<td>-0.0411</td>
</tr>
<tr>
<td>06</td>
<td>08</td>
<td>-1.7075</td>
<td>-1.6360</td>
<td>0.0715</td>
</tr>
<tr>
<td>07</td>
<td>09</td>
<td>-2.3053</td>
<td>-2.2988</td>
<td>0.0064</td>
</tr>
<tr>
<td>08</td>
<td>13</td>
<td>-1.5910</td>
<td>-1.6481</td>
<td>-0.0571</td>
</tr>
<tr>
<td>09</td>
<td>14</td>
<td>-1.7634</td>
<td>-1.6754</td>
<td>0.0879</td>
</tr>
<tr>
<td>10</td>
<td>15</td>
<td>-1.7403</td>
<td>-1.6855</td>
<td>0.0547</td>
</tr>
<tr>
<td>11</td>
<td>16</td>
<td>-1.5797</td>
<td>-1.5099</td>
<td>0.0697</td>
</tr>
<tr>
<td>12</td>
<td>17</td>
<td>-1.2041</td>
<td>-1.5420</td>
<td>-0.3379</td>
</tr>
<tr>
<td>13</td>
<td>18</td>
<td>-1.5440</td>
<td>-1.5200</td>
<td>0.0239</td>
</tr>
<tr>
<td>14</td>
<td>19</td>
<td>-1.9822</td>
<td>-1.8360</td>
<td>0.1461</td>
</tr>
<tr>
<td>15</td>
<td>21</td>
<td>-1.4232</td>
<td>-1.6046</td>
<td>-0.1814</td>
</tr>
<tr>
<td>16</td>
<td>23</td>
<td>-1.3617</td>
<td>-1.4526</td>
<td>-0.0909</td>
</tr>
<tr>
<td>17</td>
<td>25</td>
<td>-1.4149</td>
<td>-1.4713</td>
<td>-0.0564</td>
</tr>
<tr>
<td>18</td>
<td>27</td>
<td>-1.7708</td>
<td>-1.5343</td>
<td>0.2364</td>
</tr>
<tr>
<td>19</td>
<td>26</td>
<td>-1.7160</td>
<td>-1.5664</td>
<td>0.1495</td>
</tr>
<tr>
<td>20</td>
<td>28</td>
<td>-1.5797</td>
<td>-1.5256</td>
<td>0.0540</td>
</tr>
</tbody>
</table>

*All values are micro molar concentration (Ki) required to inhibit TMPKmt enzyme.

## Table 2.5: Observed and predicted activity of the test set of molecules

<table>
<thead>
<tr>
<th>Sr. No</th>
<th>Molecule No.</th>
<th>Biological Activity pKi (μM)</th>
<th>SA-MLR Predicted Activity</th>
<th>Residual</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>06</td>
<td>-2.0492</td>
<td>-1.7814</td>
<td>0.2677</td>
</tr>
<tr>
<td>2</td>
<td>20</td>
<td>-1.6720</td>
<td>-1.5859</td>
<td>0.0860</td>
</tr>
<tr>
<td>3</td>
<td>12</td>
<td>-1.5910</td>
<td>-1.6159</td>
<td>-0.0249</td>
</tr>
<tr>
<td>4</td>
<td>22</td>
<td>-1.5314</td>
<td>-1.4204</td>
<td>0.1109</td>
</tr>
<tr>
<td>5</td>
<td>24</td>
<td>-1.2900</td>
<td>-1.4391</td>
<td>-0.1491</td>
</tr>
</tbody>
</table>

*Outliers: 5, 10, 11*
2.1.4.2. QSAR model validation

In order to test the stability and predictive ability of the developed QSAR models, the models were validated as follows:

1. The internal validation to check whether training and test set molecules are properly distributed was checked using the values of $r^2 > 0.7$ and $q^2 > 0.5$ (as $r^2$ is an indication of training and $q^2$ is an indication of cross validated $r^2$ i.e. test set molecules). The generated models are found to have value in the required range.

2. The external validation of the QSAR models was done in order to know the predictive power of the models for external test set. This is done by considering the value of pred $r^2$ which should be above 0.5. The generated models are found to have value in the required range.

3. The robustness of the models for training sets was examined by comparing these models to those derived for random datasets (generated by rearranging the activities of the molecules in the training set) using the results generated in randomization test.
Chapter 2  
Molecular Modeling Studies

The best \( \text{ran}_q^2 \), highest \( q^2 \) value in the randomization test is low as compared to \( q^2 \); 
\( \text{best}_\text{ran}_r^2 \), highest \( r^2 \) value in the randomization test is low as compared to \( r^2 \); \( \alpha \)-statistical significance parameter by randomization test (<0.01).

4. Due consideration was also given to the SEE, standard error of estimate while selecting a particular model compared to the others.

2.1.5. CONTRIBUTION OF DESCRIPTORS:

The Contribution of Descriptors for biological activity developed using MLR equation is shown in Graph No.2.2.

1. SaaCHE-index: - This electrotopological state index represents no. of -CH groups connected with two aromatic bonds. This is positively contributing descriptor. Positive contribution of this descriptor reveals that presence of phenyl substitution may increase the activity.

2. SdsCHE index: - This descriptor signifies the alkene atom type in governing the activity. This is positively contributing descriptor. Positive contribution of this descriptor reveals that presence of such carbon may increase the activity.

3. \text{slogp} : - This descriptor signifies log of octanol/water partition coefficient. This is an atomic contribution model that calculates logP from the given structure. Positive contribution of this descriptor reveals that presence of hydrophobic groups may increase the activity.

4. SdsCHcount: - This descriptor evaluates total number of -CH groups connected with one double and one single bond. This is negatively contributing descriptor indicating the decrease in activity due to presence of side chain containing unsaturation.
2.1.6. 3D QSAR STUDIES:

The k-Nearest Neighbor Molecular Field Analysis (kNN-MFA) was used to develop 3D QSAR models that adopts a k-nearest neighbor principle for generating relationships of molecular fields with the experimentally reported activity. Like many 3D QSAR methods, k-nearest neighbor molecular field analysis (kNN-MFA) requires suitable alignment of given set of molecules (Fig 2.1). This is followed by generation of a common rectangular grid around the molecules. The steric and electrostatic interaction energies are computed at the lattice points of the grid using a methyl probe of charge +1. These interaction energy values are considered for relationship generation and are utilized as descriptors to decide nearness between molecules. The optimal training and test sets were generated using the sphere exclusion algorithm. Once the training and test sets were generated, kNN methodology was applied to the descriptors generated over the grid.

The kNN methodology relies on a simple distance learning approach whereby an unknown member is classified according to the majority of its k-nearest neighbors in the training set. The nearness is measured by an appropriate distance metric (e.g., a
molecular similarity measure calculated using field interactions of molecular structures). The standard kNN method is implemented simply as follows.

1. Calculate distances between an unknown object (u) and all the objects in the training set.
2. Select k objects from the training set most similar to object u, according to the calculated distances; and
3. Classify object u with the group to which the majority of the k objects belong. An optimal k value is selected by optimization through the classification of a test set of samples or by leave-one out (LOO) cross-validation. The variables and optimal k values were chosen using Simulated annealing (SA) variable selection methods.

Simulated annealing\textsuperscript{130} (SA) is the simulation of a physical process, 'annealing', which involves heating the system to a high temperature and then gradually cooling it down to a preset temperature (e.g., room temperature). During this process, the system samples possible configurations distributed according to the Boltzmann distribution so that at equilibrium, low energy states are the most populated.

The SA kNN-MFA method employs the kNN classification principle combined with the SA variable selection procedure. For each predefined number of variables (Vn) it seeks to optimize the following using stochastic sampling and simulated annealing as an optimization tool; (i) the number of nearest neighbors (k) used to estimate the activity of each molecule and (ii) the selection of variables from the original pool of all molecular descriptors that are used to calculate similarities between molecules (i.e., distances in Vn-dimensional descriptor space).

The 3-D QSAR models were developed using kNN MFA method using following steps:

1. The trial solution was generated to the underlying optimization problem; i.e., a kNN-MFA model was built based on a random selection of descriptors.
2. The value of the fitness function, q\textsuperscript{2} value for a kNN-MFA model which characterizes the quality of the trial solution to the underlying problem, was calculated and analyzed.
3. The trial solution was perturbed to obtain a new solution; i.e., a fraction of the current trial solution descriptors were changed to other randomly selected descriptors and a new kNNMFA model was built for the new trial solution.

4. The value of the fitness function (q\textsuperscript{2} new) was calculated for the new trial solution.

5. The optimization criterion was applied: if q\textsuperscript{2} curr \leq q\textsuperscript{2} new; the new solution was accepted and used to replace the current trial solution; if q\textsuperscript{2} curr > q\textsuperscript{2} new, the new solution was accepted provided the Metropolis criterion was satisfied; i.e. \textsuperscript{\text{rnd} < e^{(q2\text{curr} - q2\text{new})/T}}\text{ where rnd means a random number uniformly distributed between 0 and 1 and T means a parameter analogous to the temperature in the Boltzmann distribution.}

6. Steps 3-5 were repeated until the termination condition was satisfied. The temperature-lowering scheme and the termination condition used in this work have been adapted from Sun et al. Thus, when a new solution was accepted or when a preset number of successive steps of generating trial solutions (20 steps) did not lead to a better result, the temperature was lowered by 10% (the default initial temperature was 1000 K).

The calculations were terminated, when either the current temperature of simulations reached 10\textsuperscript{-6} K or the ratio between the current temperature and the temperature corresponding to the best solution was found to be 10\textsuperscript{-6}.

**Fig 2.1:** Alignment of substituted benzyl pyrimidine derivatives
The 3D-QSAR studies were performed using partial least squares (PLS) and kNN-MFA. As the predictivity of the SA-kNN method was low, the study was continued with SA-PLS.

Several 3D QSAR models were generated using SA-PLS method and best model was selected based on value of statistical parameters (Table 2.6). Thus generated QSAR model was used to optimize the electronic requirement around pharmacophore as all the data points' generated showed requirement of electronic parameters.

The predictive ability of this SA-PLS model was evaluated by predicting the biological activities of the test set molecules. Residuals values obtained by subtraction of predicted activities from biological activities were found near to zero. Therefore it can be concluded that the resultant QSAR model have very good predictive ability.

The predicted, actual activities and residuals of the training set and test set molecules are given in Table 2.7 and Table 2.8 respectively.

<table>
<thead>
<tr>
<th>Sr. No.</th>
<th>Parameters</th>
<th>SA-KNN MFA</th>
<th>SA-PLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$r^2$</td>
<td>-</td>
<td>0.8317</td>
</tr>
<tr>
<td>2</td>
<td>$q^2$</td>
<td>0.5972</td>
<td>0.7287</td>
</tr>
<tr>
<td>3</td>
<td>$q^2$se</td>
<td>0.2398</td>
<td>0.2215</td>
</tr>
<tr>
<td>4</td>
<td>pred $r^2$</td>
<td>0.8895</td>
<td>0.7580</td>
</tr>
<tr>
<td>5</td>
<td>$R^2$se</td>
<td>0.0995</td>
<td>0.1473</td>
</tr>
<tr>
<td>6</td>
<td>N</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>7</td>
<td>Contributing parameters</td>
<td>E_1367; E_1131; E_97; S_574</td>
<td>E_676; E_672; E_713; E_283; E_833; E_189</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sr. No.</th>
<th>Compound No.</th>
<th>Biological Activity $pKi$ ($\mu M$)</th>
<th>SA-PLS</th>
<th>Predicted Activity</th>
<th>Residual</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>01</td>
<td>-1.9493</td>
<td>-1.9980</td>
<td>-0.0487</td>
<td></td>
</tr>
<tr>
<td>02</td>
<td>02</td>
<td>-1.7403</td>
<td>-1.6466</td>
<td>0.0936</td>
<td></td>
</tr>
<tr>
<td>03</td>
<td>03</td>
<td>-2.2900</td>
<td>-1.8769</td>
<td>0.4130</td>
<td></td>
</tr>
<tr>
<td>04</td>
<td>04</td>
<td>-1.5910</td>
<td>-1.7688</td>
<td>-0.1748</td>
<td></td>
</tr>
</tbody>
</table>
### Table 2.8: Results of SA-PLS for test set of benzyl pyrimidine derivatives.

<table>
<thead>
<tr>
<th>Sr. No.</th>
<th>Compound No.</th>
<th>Biological Activity pKi (μM)</th>
<th>SA-PLS Predicted Activity</th>
<th>Residual</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10</td>
<td>-1.0000</td>
<td>-1.0830</td>
<td>-0.0830</td>
</tr>
<tr>
<td>2</td>
<td>13</td>
<td>-1.5910</td>
<td>-1.6915</td>
<td>-0.1005</td>
</tr>
<tr>
<td>3</td>
<td>14</td>
<td>-1.7634</td>
<td>-1.7378</td>
<td>0.0255</td>
</tr>
<tr>
<td>4</td>
<td>15</td>
<td>-1.7403</td>
<td>-1.9275</td>
<td>-0.1872</td>
</tr>
<tr>
<td>5</td>
<td>20</td>
<td>-1.6720</td>
<td>-1.6258</td>
<td>0.0461</td>
</tr>
<tr>
<td>6</td>
<td>22</td>
<td>-1.5314</td>
<td>-1.5846</td>
<td>-0.0532</td>
</tr>
<tr>
<td>7</td>
<td>23</td>
<td>-1.3617</td>
<td>-1.4033</td>
<td>-0.0416</td>
</tr>
</tbody>
</table>
The plots of observed vs. predicted activity for the optimal cross validated QSAR model are depicted in Graph 2.3.

**Graph 2.3**: Observed Vs predicted activity for training (a) and test Set (b)

3D data points generated which contribute to SA kNN MFA model are shown in Fig 2.2. The range of property values for the generated data points definitely help for the design of new compounds.

**Fig 2.2**: Stereo view of the molecular rectangular field grid
Chapter 2

Molecular Modeling Studies

Results obtained and points generated around the benzyl pyrimidine pharmacophore using the 3D QSAR studies helped in correlating the structure of substituents around pharmacophore with their observed antitubercular activity.

![Chemical Structure](image)

X- Substitution requirements:
3D QSAR studies using the SA-PLS method helped us to find out the importance of electron withdrawing groups at this position. The electrostatic data point generated negative value viz. E_189 (0.1885). It was found that the electron withdrawing groups like -Cl, -Br were essential for potent antitubercular activity and accordingly the substituents were finalized while designing new chemical compounds.

Y- Substitution requirements:
3D QSAR studies here showed mixed contribution of electron withdrawing and electron releasing groups at this position towards antitubercular activity.

2.1.7. OPTIMIZATION OF PHARMACOPHORE:

2.1.7.1. Information obtained from QSAR studies:

The information obtained from 2D and 3D QSAR studies was used to optimize the pyrimidinedione nucleus for selective inhibition of the TMPKmt. Descriptor with positive coefficient, slog corroborates that lipophilic substituents are required for potent antitubercular activity of these compounds. Similarly electrotopological state indices for number of -CH group connected with two aromatic bonds (SaaCHE-index) and Electro-topological state indices for number of -CH group connected with one double and one single bond (SdsCHE-index) also contribute positively for
activity. Descriptor with negative coefficient SdsCHcount defines the total number of –CH group connected with one double and one single bond and indicate that such groups shall be as minimum as possible. Similarly the contribution of electronic points (electropositive and electronegative) generated in the 3D QSAR was considered while designing the new compounds.

2.1.1.1. Information obtained from literature:

The close analysis through literature \(^{94, 92}\) helped us to conclude the following key structure activity relationship points and the whole study indicates that the pharmacophoric requirements for selective and effective binding with the target enzyme TMPK\(_{mt}\) are as follows:

- The sugar moiety though reported earlier that the 3'-OH contribute for binding to -COOH group of Asp9 is not essential for selective inhibition of TMPK\(_{mt}\). Sugar moiety can be replaced with other groups. (Fig 2.3)

- Other important findings of the pharmacophore without sugar moiety i.e. benzyl group containing substituted containing pyrimidine derivatives are as follows: i) The –COOH group of side chain on -Ph of benzyl form hydrogen bond with Arg 95. ii) pyrimidine form pi stacking interaction with Phe 70. iii) Carbonyl ‘O’ of C4 of thymine forms hydrogen bond with -H of -NH group of guanidine group of Arg 74. iv) N1 of thymine froms H-bond with -H of -COOH of Asn 100 and with H of water molecule v) the -COOH group of ligand forms H bond with Arg 95 .vi) Compounds containing -CONH\(_2\) linked through 3-4 carbon chain on benzyl substituted on N1 interact with Glu 166. vii) Van Dael et al. have reported pharmacophore substituted with aryl substituted thiourea on C3 of sugar at N1 of thymidine i.e. aryl thiourea alpha thymidine analogues. N1-H of thiourea of these compounds form H-bond with carbonyl oxygen of Asp 9. This indicates that 3'-OH of sugar can be replaced by any other H-bond donor group like thiourea in this case. This clearly indicates that sugar moiety is not essential for activity.

Taking all above interactions of compounds without sugar moiety on thymine ring into consideration, we came to conclude that only required functional groups can be substituted by substitution of required H-bond donor on aromatic ring.
Therefore, we have designed pharmacophore for selective and effective inhibition of TMPK$_{mt}$ as shown below.

**Fig 2.3:** Replacement of sugar moiety with other functional groups

**Fig 2.4:** Designed pharmacophore
2.2. DESIGN OF NEW CHEMICAL ENTITIES (NCES) CONTAINING PYRIMIDINONE PHARMACOPHORE:

In order to better determine structural characteristic that were able to improve the affinity for and in turn antitubercular activity, different chemical modifications were carried out based on analysis of extensive SAR studies from literature, findings of 3D QSAR studies and regression equation obtained by 2D QSAR. The new chemical entities (NCEs) were designed by varying the nature and position of substituents using Pharmacophore shown in Fig 2.4 and predicted bioactivity (pIC₅₀ calculated) using CombiLib tool of V Life MDS 3.5 software using Lipinski’s screen (the rule of 5) as the filters. The following descriptors comprised the Lipinski’s screen used for design of New Chemical Entities and are enlisted as follows. The cut off values set for each descriptor are mentioned in the parenthesis.

a. Hydrogen bond acceptor (A), (not more than 8)
b. Hydrogen bond donor (D), (not more than 5)
c. Number of rotatable donds (R), (not more than 10)
d. xlog p (4), (not more than 5)
e. Molecular weight (W), (not more than 450)
f. Dynamic polar surface Area (PSAd) (S), (not more than or equal to 140 Å²)

More than hundred molecules were generated by the CombiLib tool of MDS using a particular template (template 1, 2 and 3 respectively) which follows the Lipinski’s rule, but we have selected only the most active molecules based on the predicted activity. The most potent compounds have positive values of pKi, similarly the predicted activity values for NCE’s were found either towards positive or in between the actual activities of most potent and least potent compound mentioned in the series. Compounds qualifying all required parameters set for Lipinski’s Screen/filter are indicated by ADRXWS strings. The columns containing the Lipinski’s screen score and other column containing the strings of alphabets, ADRXWS indicate that all 6 conditions are satisfied by that corresponding compound.

If any compound does not satisfy all 6 criteria, then those numbers of alphabet strings are missing in that column and the screen score also changed accordingly. Lesser the screen score, lesser is the pharmacokinetic compatibility (drug likeliness) for that
designed compound. The results of CombiLib generated along with predicted activities are tabulated in Table 2.9 a, Table 2.9b and Table 2.9 c respectively.

Table 2.9 a: NCEs generated by template 1

**Template 1**

<table>
<thead>
<tr>
<th>Compound Code</th>
<th>12X</th>
<th>13X</th>
<th>Prediction</th>
<th>Extrapolation</th>
<th>Screen Results</th>
<th>Screen Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>TSIib</td>
<td>3,4,5-Trimethoxy benzaldehyde</td>
<td>3-Chloro, 4-fluoro-aniline</td>
<td>0.3098</td>
<td>1.5699</td>
<td>ADRXS</td>
<td>5</td>
</tr>
<tr>
<td>TSIbii</td>
<td>4-Fluoro-benzaldehyde</td>
<td>4-Fluoro-aniline</td>
<td>-0.3045</td>
<td>0.9555</td>
<td>ADRXWS</td>
<td>6</td>
</tr>
<tr>
<td>TSIbiii</td>
<td>4-Fluoro-benzaldehyde</td>
<td>4-Nitro-aniline</td>
<td>-0.5292</td>
<td>0.7307</td>
<td>ADRXWS</td>
<td>6</td>
</tr>
<tr>
<td>TSIbiv</td>
<td>4-Fluoro-benzaldehyde</td>
<td>2-Methoxy-aniline</td>
<td>-0.4356</td>
<td>0.8243</td>
<td>ADRXWS</td>
<td>6</td>
</tr>
<tr>
<td>TSIbii</td>
<td>4-Fluoro-benzaldehyde</td>
<td>3-Chloro, 4-fluoro-aniline</td>
<td>0.5991</td>
<td>1.8592</td>
<td>ADRXWS</td>
<td>6</td>
</tr>
<tr>
<td>TSIbiiii</td>
<td>4-Fluoro-benzaldehyde</td>
<td>-H</td>
<td>0.3045</td>
<td>0.7307</td>
<td>ADRXWS</td>
<td>6</td>
</tr>
<tr>
<td>TSIbii</td>
<td>4-Chloro-benzaldehyde</td>
<td>4-Fluoro-aniline</td>
<td>0.5991</td>
<td>1.8592</td>
<td>ADRXWS</td>
<td>6</td>
</tr>
<tr>
<td>TSIbiii</td>
<td>4-Chloro-benzaldehyde</td>
<td>4-Nitro-aniline</td>
<td>0.3743</td>
<td>1.6344</td>
<td>ADRXWS</td>
<td>6</td>
</tr>
<tr>
<td>TSIbii</td>
<td>4-Chloro-benzaldehyde</td>
<td>2-Methoxy-aniline</td>
<td>0.4679</td>
<td>1.7280</td>
<td>ADRXWS</td>
<td>6</td>
</tr>
<tr>
<td>TSIbiiii</td>
<td>4-Chloro-benzaldehyde</td>
<td>3-Chloro, 4-fluoro-aniline</td>
<td>-1.8449</td>
<td>2.7628</td>
<td>ADRXWS</td>
<td>6</td>
</tr>
<tr>
<td>TSIbiv</td>
<td>4-Chloro-benzaldehyde</td>
<td>Aniline</td>
<td>0.3743</td>
<td>0.730757</td>
<td>ADRXWS</td>
<td>6</td>
</tr>
</tbody>
</table>
Table 2.9 b: NCEs generated by template 2

<table>
<thead>
<tr>
<th>Compound code</th>
<th>12X</th>
<th>13X</th>
<th>Prediction</th>
<th>Extrapolation</th>
<th>Screen Results</th>
<th>Screen Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>TS III</td>
<td>-H</td>
<td>3,4,5-Trimethoxy-benzaldehyde</td>
<td>-2.7454</td>
<td>-0.4368</td>
<td>ADRXWS</td>
<td>6</td>
</tr>
<tr>
<td>TS IV</td>
<td>-H</td>
<td>4-Methoxy-benzaldehyde</td>
<td>-2.2582</td>
<td>0.2083</td>
<td>ADRXWS</td>
<td>6</td>
</tr>
<tr>
<td>TS V</td>
<td>-H</td>
<td>3-Chloro-benzaldehyde</td>
<td>-1.5524</td>
<td>0.0605</td>
<td>ADRXWS</td>
<td>6</td>
</tr>
<tr>
<td>TS II</td>
<td>-H</td>
<td>4-Chloro-benzaldehyde</td>
<td>-1.5524</td>
<td>0.0605</td>
<td>ADRXWS</td>
<td>6</td>
</tr>
<tr>
<td>TS I</td>
<td>-H</td>
<td>4-Fluoro-benzaldehyde</td>
<td>-2.4561</td>
<td>0.2895</td>
<td>ADRXWS</td>
<td>6</td>
</tr>
<tr>
<td>TS VIII</td>
<td>4-Fluoroaniline 3-Chloro, 4-fluoroaniline</td>
<td>-1.2502</td>
<td>0.9081</td>
<td>ADRXWS</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>TS IX</td>
<td>4-Fluoroaniline 3-Chloro-benzaldehyde</td>
<td>-0.5674</td>
<td>0.2253</td>
<td>ADRXWS</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>TS VI</td>
<td>2-Methoxyaniline 3-Chloro-benzaldehyde</td>
<td>-0.5814</td>
<td>0.0941</td>
<td>ADRXWS</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>TS VII</td>
<td>4-Methoxyaniline 3-Chloro-benzaldehyde</td>
<td>-0.6750</td>
<td>0.1849</td>
<td>ADRXWS</td>
<td>6</td>
<td></td>
</tr>
</tbody>
</table>
Table 2.9: NCEs generated by template 3

<table>
<thead>
<tr>
<th>Compound code</th>
<th>11X</th>
<th>Prediction</th>
<th>Extrapolation</th>
<th>Screen Results</th>
<th>Screen Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>TSIIa</td>
<td>4-Fluorobenzaldehyde</td>
<td>1.5020</td>
<td>0.0000</td>
<td>ADRXWS</td>
<td>6</td>
</tr>
<tr>
<td>TSIla</td>
<td>4-Chlorobenzaldehyde</td>
<td>-0.9411</td>
<td>0.1531</td>
<td>ADRXWS</td>
<td>6</td>
</tr>
<tr>
<td>TSVa</td>
<td>4-Methoxybenzaldehyde</td>
<td>1.2692</td>
<td>0.0000</td>
<td>ADRXWS</td>
<td>6</td>
</tr>
<tr>
<td>TSVa</td>
<td>3-Chlorobenzaldehyde</td>
<td>-0.9411</td>
<td>0.1532</td>
<td>ADRXWS</td>
<td>6</td>
</tr>
</tbody>
</table>

2.3. MOLECULAR DOCKING STUDIES OF DESIGNED COMPOUNDS:

2.3.1. Introduction:

The application of computational methods to study the formation of intermolecular complexes has been the subject of intensive research during the last decade. It is widely accepted that drug activity is obtained through the molecular binding of one molecule (the ligand) to the pocket of another, which is commonly a protein. In the conformations of a complex of a protein with a therapeutic drug, the molecule exhibits the geometric and chemical complementarities, both of which are essential for successful drug activity. The computational process of searching for a ligand that is able to fit both geometrically and energetically to the binding site of a protein is called molecular docking. The molecular docking tool, Glide software, version 4.5 (Schrodinger Inc. U.S.A.) was used for ligand docking studies in to the enzyme binding pocket. Glide is one of the most accurate docking programs available for ligand protein, protein-protein binding studies. Glide was found to produce least number of inaccurate poses and 85% of Glides binding models had an RMSD of 1.4 Å or less from native co-crystallized structures. Glide is validated software designing for calculating the accurate binding interaction energies of the 3-D...
structures of a known protein receptor with ligand or another protein molecule. The Glide docking program approximated a complete systematic search of the conformational, orientational and positional space of the docked ligand molecules in to the receptor (protein) binding pocket.

2.3.2. Methodology:

2.3.2.1. Selection and preparation of protein structures:

The crystal structure of TMPKmt was obtained from protein data bank (PDB). The crystal structure used was 1W2H. The typical structure file from the PDB is not suitable for immediate use in molecular modeling calculations. A typical PDB structure file consists only of heavy atoms and may include a co-crystallized ligand, water molecules, metal ions, and cofactors. Some structures are multimeric, and may need to be reduced to a single unit. Because of the limited resolution of X-ray experiments, it can be difficult to distinguish between NH and O, and the placement of these groups must be checked. PDB structures may be missing information on connectivity, which must be assigned, along with bond orders and formal charges. Schrödinger has therefore assembled a set of tools to prepare proteins in a form that is suitable for modeling calculations. All structures were prepared for docking using the ‘protein preparation wizard’ in Maestro wizard 8.5. Water molecules in the crystal structures were deleted and termini were capped by adding ACE and NMA residue. The protein preparation was carried out two steps, preparation and refinement. After ensuring chemical correctness, the hydrogens were added where hydrogen atoms were missing. Side chains that are not close to the binding cavity and do not participate in salt bridges were neutralized. The hydrogen-bonding network was optimized by reorienting hydroxyl groups, water molecules, and amide groups of Asn and Gln, and selecting appropriate states and orientations of the imidazole ring in His residues. Each hydrogen bond donor, His ring, and Asn and Gln terminal amide is considered a separate orientable species. Optimizing the orientation of the various groups is an iterative process, which passes over all the groups whose H-bonds need to be optimized multiple times. The refinement component performs a restrained impact minimization of the co-crystallized complex was performed. This helps in reorientation of side chain hydroxyl group. It uses the PLS all-atom force field (OPLS.AA force field) for this purpose.
2.3.2.2. Receptor grid generation:

For receptors that adopt more than one conformation on binding, it is necessary to prepare grids for each conformation, to ensure that possible actives are not missed. Grid files represent physical properties of a volume of the receptor (specifically the active site) that are searched when attempting to dock a ligand. Also the shape and properties of the receptor are represented on a grid by several different sets of fields that provide progressively more accurate scoring of the ligand poses. Grids were generated by Receptor Grid Generation panel which define receptor structure by excluding any co-crystallized ligand that may be present, determine the position and size of the active site as it will be represented by receptor grids, and set up Glide constraints. Grids were defined by centering them on the ligand in the crystal structure using the default box size.

2.3.2.3. Ligand preparation:

Each designed structure was assigned an appropriate bond order and were converted to mae format and optimized by means of the OPLS2005 force field using a default setting using the ligprep script shipped by Schrodinger. The LigPrep process consists of a series of steps that perform conversions, apply corrections to the structures, generate variations on the structures, eliminate unwanted structures, and optimize the structures. The simplest use of LigPrep produces a single low-energy 3D structure with correct chiralities for each successfully processed input structure and can also produce a number of structures from each input structure with various ionization states, tautomers, stereochemistries, and ring conformations, and eliminate molecules using various criteria including molecular weight or specified numbers and types of functional groups present.

2.3.2.4. Docking and scoring functions:

The ligands were docked with the active site using the ‘Extra precision’ Glide algorithm. Glide uses a hierarchical series of filters to search for possible locations of the ligand in the active-site region of the receptor. The initial filters test the spatial fit of the ligand to the defined active site, and examine the complementarity of ligand-receptor interactions using a grid-based method.
Poses that pass these initial screens enter the final stage of the algorithm, which involves evaluation and minimization of a grid approximation to the OPLS-AA nonbonded ligand-receptor interaction energy. Final scoring is then carried out on the energy-minimized poses. Finally, the minimized poses are re-scored using Schrödinger's proprietary Glide Score scoring function. Glide Score is based on Chem Score, but includes a steric-clash term and adds buried polar terms devised by Schrödinger to penalize electrostatic mismatches.

\[
G\text{-Score} = 0.065\text{vdW} + 0.130\text{Coul} + \text{Lipo} + \text{HBond} + \text{Metal} + \text{BuryP} + \text{RotB} + \text{Site}
\]

Where, \(\text{vdW}\) - Van Der Waal energy;
\(\text{Coul}\) - Coulomb energy;
\(\text{Lipo}\) - Lipophilic contact term;
\(\text{HBond}\) - Hydrogen-bonding term;
\(\text{Metal}\) - Metal-binding term;
\(\text{BuryP}\) - Penalty for buried polar groups;
\(\text{RotB}\) - Penalty for freezing rotatable bonds;
\(\text{Site}\) - Polar interactions at the active site; and

The coefficients of \(\text{vdW}\) and \(\text{Coul}\) are: \(a = 0.065\), \(b = 0.130\), respectively.

The choice of best-docked structure for each ligand is made using a model energy score (Emodel) that combines the energy grid score, the binding affinity predicted by GlideScore, and (for flexible docking) the internal strain energy for the model potential used to direct the conformational search algorithm.

2.3.3. RESULT AND DISCUSSION:

Glide results were examined with an emphasis on both visual and numerical appraisal. The glide results can be viewed through glide pose viewer panel in the Tools menu of main 'Maestro' window. Each ligand pose was identified by index number, title, ligand number, conformation number, and pose number. For each pose, the output window in the form of spread sheet generates the Glide score (G-Score), Energy, and various contact counts like good contacts, bad contacts, ugly contacts associated with the ligand receptor pose combination. The docking results are presented below:
Table 2.10: Docking results for designed NCEs

<table>
<thead>
<tr>
<th>Sr.No.</th>
<th>Compound code</th>
<th>G score</th>
<th>E model</th>
<th>Energy</th>
<th>H Bond</th>
<th>Good Vdw</th>
<th>Bad Vdw</th>
<th>Ugly Vdw</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>TSIa</td>
<td>-8.85</td>
<td>-89.6</td>
<td>-41.6</td>
<td>2</td>
<td>263</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>02</td>
<td>TSIIVa</td>
<td>-8.74</td>
<td>-83.3</td>
<td>-44.0</td>
<td>1</td>
<td>262</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>03</td>
<td>TSVII</td>
<td>-8.62</td>
<td>-81.0</td>
<td>-42.3</td>
<td>1</td>
<td>256</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>04</td>
<td>TSVIII</td>
<td>-8.62</td>
<td>-80.4</td>
<td>-45.5</td>
<td>2</td>
<td>254</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>05</td>
<td>TSVa</td>
<td>-8.32</td>
<td>-78.5</td>
<td>-41.3</td>
<td>1</td>
<td>249</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>06</td>
<td>AZTMP</td>
<td>-8.23</td>
<td>-74.2</td>
<td>-51.7</td>
<td>4</td>
<td>246</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>07</td>
<td>TSIIX</td>
<td>-7.86</td>
<td>-54.3</td>
<td>-41.5</td>
<td>0</td>
<td>236</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>08</td>
<td>TSIib</td>
<td>-7.84</td>
<td>-68.4</td>
<td>-43.7</td>
<td>0</td>
<td>283</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>09</td>
<td>TSIibii</td>
<td>-7.79</td>
<td>-61.6</td>
<td>-45.8</td>
<td>2</td>
<td>254</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>TSI</td>
<td>-7.77</td>
<td>-60.7</td>
<td>-42.2</td>
<td>2</td>
<td>224</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>TSIibiv</td>
<td>-7.75</td>
<td>-64.7</td>
<td>-46.0</td>
<td>0</td>
<td>261</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>TSV</td>
<td>-7.73</td>
<td>-60.7</td>
<td>-42.4</td>
<td>1</td>
<td>223</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>TSIibiii</td>
<td>-7.67</td>
<td>-66.2</td>
<td>-43.8</td>
<td>1</td>
<td>263</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>TSIib</td>
<td>-7.53</td>
<td>-62.4</td>
<td>-44.3</td>
<td>2</td>
<td>270</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>TSIibiv</td>
<td>-7.46</td>
<td>-62.5</td>
<td>-47.1</td>
<td>1</td>
<td>237</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>TSIIB</td>
<td>-7.44</td>
<td>-60.5</td>
<td>-44.1</td>
<td>1</td>
<td>213</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>17</td>
<td>TSIIV</td>
<td>-7.41</td>
<td>-58.7</td>
<td>-42.7</td>
<td>1</td>
<td>234</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>18</td>
<td>TSIbi</td>
<td>-7.33</td>
<td>-64.2</td>
<td>-46.7</td>
<td>1</td>
<td>267</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>19</td>
<td>TSIIBa</td>
<td>-7.37</td>
<td>-53.2</td>
<td>-42.1</td>
<td>1</td>
<td>220</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>20</td>
<td>TSIibiii</td>
<td>-7.34</td>
<td>-60.1</td>
<td>-43.3</td>
<td>1</td>
<td>253</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>21</td>
<td>TSIIB</td>
<td>-6.34</td>
<td>-54.0</td>
<td>-37.0</td>
<td>0</td>
<td>215</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>22</td>
<td>TSIibv</td>
<td>-6.33</td>
<td>-55.3</td>
<td>-41.6</td>
<td>2</td>
<td>243</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>23</td>
<td>TSIIBi</td>
<td>-6.02</td>
<td>-43.6</td>
<td>-33.0</td>
<td>0</td>
<td>244</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>24</td>
<td>TSIIBb</td>
<td>-5.30</td>
<td>-53.5</td>
<td>-45.5</td>
<td>0</td>
<td>214</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>25</td>
<td>TSIIBii</td>
<td>-5.63</td>
<td>-52.4</td>
<td>-40.5</td>
<td>1</td>
<td>234</td>
<td>4</td>
<td>0</td>
</tr>
</tbody>
</table>
Chapter 2  Molecular Modeling Studies

The docking studies helped to find out the binding modes of the designed series of compounds which would be the key for further discovery of potent compounds.

1. **G-score:**

The scoring function of Glide 4.5, docking program is presented in the G-score form. Thus G-score indicates the binding affinity of the designed compound to the receptor/ enzyme. The G-score of the standard compound AZTMP was found to be -8.23. The G-score of six of the designed compounds was found to be more than that of the standard viz., AZTMP.

The G-score of the other designed compounds were also found to be comparable with the standard compound. Beside the G-score, the energy, E-model was also used for the evaluation of the docking results. Those were also found to be comparable with the G-score.

2. **H-bond interactions:**

This is one of the most widely used alternative parameter for the evaluation of the docking results. As H-bond is an influential parameter in the activity of the drug compound, this parameter was also taken into consideration. The number of H-bond interactions in the standard compounds was compared with that of designed compounds. The standard PDB SUM (1W2H) is shown in (Fig 2.8). The numbers of H-bond in the standard compound, AZTMP after carrying out the docking process were found to be 4 (Fig 2.9).

3. **Contacts:**

The contacts are represented in the form of Ven der Waals (vdw) Interaction: –

- Good Contact vdw.
- Bad Contact vdw.
- Ugly Contact vdw.

It is well established and accepted fact that number of good van der Waals interaction decides the binding affinity for any ligand with receptor/ enzyme protein. Therefore we have analyzed the binding modes and abilities, taking in to consideration the
number of vdw contacts and interaction of the standard (AZTMP) and designed compounds with the Thymidine Monophosphate Kinase active binding site. It was found that many of the designed compounds have shown more number of good vdw interactions, less number of bad vdw when compared with the standard AZTMP (Table 2.10). In conclusion, G-score and E-model in addition to number of H-bond interactions, number of good, bad and ugly vdw contacts decide the possible binding affinity and in turn potency of the designed new chemical compounds.

Fig 2.5: PDB SUM (1W2H)

Fig 2.6: Binding modes of AZTMP with TMPKmt enzyme
Fig 2.7: Binding modes of TSIA with TMPKmt enzyme

Fig 2.8: Binding modes of TSIIbi with TMPKmt enzyme
Fig 2.9: Binding modes of TSIhv with TMPKmt enzyme

Fig 2.10: Binding modes of TSIIbi iii with TMPKmt enzyme
2.4. STRUCTURE BASED DRUG DESIGN:

Structure-based drug design (SBDD) methods were used to search for novel inhibitors of herpes simplex virus type 1 (HSV-1) thymidine kinase and *Mycobacterium tuberculosis* thymidylate kinase. The method involved the use of crystal structure complexes to guide database searching for potential inhibitors. The available chemicals directory (ACD) was used as a database of potential screening compounds. When the structure of the target is known (available), usually from X-ray crystallography, the most commonly used virtual screening method is molecular docking. Molecular docking can also be used to test possible hypotheses before conducting costly laboratory experiments. Molecular docking programs try to predict how a drug candidate binds to a protein target without performing a laboratory experiment. X-ray crystallographic techniques have been applied in medicinal chemistry to perform structural elucidation of large complex molecules. The basic assumption in SBDD is that good inhibitors must possess significant structural and chemical complementarity to their target receptor. With the above literature survey, it was decided to screen a data base of designed compounds by carrying out their docking with 1W2H according to the procedure mentioned in section 2.3. Around 100 molecules were docked and search was trimmed to around 17 molecules after considering their requisite interactions.
2.4.1. RESULT AND DISCUSSION:
Glide results were examined with an emphasis on both visual and numerical appraisal. The glide results can be viewed through glide pose viewer panel in the Tools menu of main ‘Maestro’ window. Each ligand pose was identified by index number, title, ligand number, conformation number, and pose number. For each pose, the output window in the form of spread sheet generates the Glide Score (G-Score), Energy, and various contact counts like good contacts, bad contacts, ugly contacts associated with the ligand receptor pose combination. The docking result is presented below:

<table>
<thead>
<tr>
<th>Compound Code</th>
<th>G score</th>
<th>E model</th>
<th>Energy</th>
<th>H Bond</th>
<th>Good Vdw</th>
<th>Bad Vdw</th>
<th>Ugly Vdw</th>
</tr>
</thead>
<tbody>
<tr>
<td>T8</td>
<td>-8.87</td>
<td>-63.6</td>
<td>-48.4</td>
<td>3</td>
<td>333</td>
<td>12</td>
<td>0</td>
</tr>
<tr>
<td>T7</td>
<td>-8.86</td>
<td>-63.6</td>
<td>-48.4</td>
<td>3</td>
<td>361</td>
<td>14</td>
<td>0</td>
</tr>
<tr>
<td>T5</td>
<td>-8.80</td>
<td>-62.8</td>
<td>-48.3</td>
<td>3</td>
<td>331</td>
<td>14</td>
<td>0</td>
</tr>
<tr>
<td>T1</td>
<td>-8.83</td>
<td>-52.8</td>
<td>-42.3</td>
<td>0</td>
<td>343</td>
<td>03</td>
<td>0</td>
</tr>
<tr>
<td>T2</td>
<td>-8.70</td>
<td>-55.4</td>
<td>-40.3</td>
<td>0</td>
<td>330</td>
<td>03</td>
<td>0</td>
</tr>
<tr>
<td>T3</td>
<td>-8.67</td>
<td>-62.8</td>
<td>-49.3</td>
<td>0</td>
<td>301</td>
<td>01</td>
<td>0</td>
</tr>
<tr>
<td>T10</td>
<td>-8.63</td>
<td>-61.8</td>
<td>-40.3</td>
<td>0</td>
<td>320</td>
<td>02</td>
<td>0</td>
</tr>
<tr>
<td>T13</td>
<td>-8.45</td>
<td>-62.3</td>
<td>-41.1</td>
<td>0</td>
<td>300</td>
<td>00</td>
<td>0</td>
</tr>
<tr>
<td>T11</td>
<td>-8.33</td>
<td>-60.8</td>
<td>-50.3</td>
<td>0</td>
<td>335</td>
<td>02</td>
<td>0</td>
</tr>
<tr>
<td>AZTMP</td>
<td>-8.28</td>
<td>-73.2</td>
<td>-51.7</td>
<td>4</td>
<td>253</td>
<td>05</td>
<td>0</td>
</tr>
<tr>
<td>T12</td>
<td>-7.89</td>
<td>-59.7</td>
<td>-47.6</td>
<td>0</td>
<td>319</td>
<td>03</td>
<td>0</td>
</tr>
<tr>
<td>T14</td>
<td>-7.57</td>
<td>-58.7</td>
<td>-41.6</td>
<td>0</td>
<td>288</td>
<td>06</td>
<td>0</td>
</tr>
<tr>
<td>T6</td>
<td>-7.88</td>
<td>-54.0</td>
<td>-38.7</td>
<td>2</td>
<td>268</td>
<td>08</td>
<td>0</td>
</tr>
<tr>
<td>TSC6</td>
<td>-7.11</td>
<td>-50.3</td>
<td>-37.5</td>
<td>1</td>
<td>183</td>
<td>00</td>
<td>0</td>
</tr>
<tr>
<td>TSC2</td>
<td>-7.06</td>
<td>-52.3</td>
<td>-41.6</td>
<td>2</td>
<td>248</td>
<td>02</td>
<td>0</td>
</tr>
<tr>
<td>T9</td>
<td>-6.84</td>
<td>-52.1</td>
<td>-38.0</td>
<td>3</td>
<td>276</td>
<td>05</td>
<td>0</td>
</tr>
<tr>
<td>TSC5</td>
<td>-6.81</td>
<td>-51.4</td>
<td>-38.4</td>
<td>2</td>
<td>221</td>
<td>01</td>
<td>0</td>
</tr>
<tr>
<td>TSC7</td>
<td>-6.68</td>
<td>-48.1</td>
<td>-37.7</td>
<td>3</td>
<td>145</td>
<td>00</td>
<td>0</td>
</tr>
<tr>
<td>TSC4</td>
<td>-6.60</td>
<td>-54.3</td>
<td>-43.2</td>
<td>1</td>
<td>264</td>
<td>06</td>
<td>1</td>
</tr>
<tr>
<td>TSC1</td>
<td>-6.26</td>
<td>-48.3</td>
<td>-38.6</td>
<td>0</td>
<td>223</td>
<td>00</td>
<td>0</td>
</tr>
<tr>
<td>TSC3</td>
<td>-6.25</td>
<td>-48.2</td>
<td>-38.3</td>
<td>1</td>
<td>217</td>
<td>00</td>
<td>0</td>
</tr>
</tbody>
</table>
3. **G-score:**

The scoring function of Glide 4.5, docking program is presented in the G-score form. Thus G-score indicates the binding affinity of the designed compound to the receptor/ enzyme. The G-score of the standard compound AZTMP was found to be -8.283. The G-score of the designed compounds were also found to be more than AZTMP or comparable with the standard compound. Beside the G-score, the energy, E-model was also used for the evaluation of the docking results. Those were also found to be comparable with the G-score.

4. **H-bond Interactions:**

This is one of the most widely used alternative parameter for the evaluation of the docking results. As H-bond is an influential parameter in the activity of the drug compound, this parameter was also taken into consideration. The number of H-bond interactions in the standard compounds were compared with that of designed compounds.

3. **Contacts:**

The contacts are represented in the form of Ven der Waals (vdw) Interaction: –

- Good contact vdw.
- Bad contact vdw.
- Ugly contact vdw.

It is well established and accepted fact that number of good Van Der Waals (vdw) interaction decides the binding affinity for any ligand with receptor/ enzyme protein. Therefore we have analyzed the binding modes and abilities, taking in to consideration the number of vdw contacts and interaction of the standard (AZTMP) and designed compounds with the Thymidine monophosphate kinase active binding site.

It was found that many of the designed compounds have shown more number of good vdw interactions, less number of bad vdw when compared with the standard AZTMP (Table 2.11). In conclusion, G-score and E-model in addition to number of
hydrogen bond interactions, number of good, bad and ugly vdw contacts decide the possible binding affinity and in turn the potency of the designed new chemical compounds.

Fig 2.15: Binding modes of T5 with TMPKmt enzyme

Fig 2.16: Binding modes of T7 with TMPKmt enzyme
2.5. ADSORPTION, DISTRIBUTION, METABOLISM AND EXCRETION (ADME) PREDICTION OF DESIGNED COMPOUNDS USING QIKPROP (3.0) SCHRODINGER LLC (8.0)

2.5.1. Introduction:

Nearly 40% of drug candidates fail in clinical trials due to poor ADME (absorption, distribution, metabolism, and excretion) properties. These late-stage failures contribute significantly to the cost of new drug development. The ability to detect problematic candidates early will dramatically reduce the amount of wasted time and resources, and streamline the overall development process. Accurate ADME properties prediction prior to expensive experimental procedures, such as high-throughput screening (HTS), can eliminate unnecessary testing on compounds that are doomed to fail; it can also focus lead optimization efforts to enhance the desired ADME properties. Finally, incorporating ADME predictions as a part of the development process will result in lead compounds that are more likely to exhibit satisfactory ADME performances during clinical trials.

2.5.2. Methodology:

Prediction of the ADME parameters prior to the experimental studies is one of the most important aspects in the drug discovery and development of the drug molecule. Drug may fail to reach the market phase if those properties are not fulfilled by the drug candidate. Taking into consideration the above mentioned aspects, the ADME profile of the designed NCEs was studied using the QikProp\textsuperscript{141} 3.0 software.

QikProp is a quick, accurate, easy-to-use absorption, distribution, metabolism, and excretion (ADME) prediction program in Maestro, Schrodinger and which was used here to predicate ADME properties. QikProp predicts physically significant descriptors and pharmaceutically relevant properties of organic molecules, either individually or in batches. In addition to predict molecular properties, QikProp provides ranges for comparing a particular molecule's properties with those of 95% of known drugs.
QikProp also flags 30 types of reactive functional groups that may cause false positives in HTS assays. The range of values that cause a molecule to be flagged can be similar or dissimilar to other known drugs.

Forty four physically descriptors and pharmaceutically relevant properties of designed compounds have been analyzed using QikProp, among which significant descriptors are reported here which are required for predicting the drug like properties of molecules. These properties are as follows:

1. Molecular Weight (mol_MW) (150 - 650)
2. Octanol/water partition coefficient (Log P_o/w) (-2 - 6.5)
3. Aqueous Solubility (QPlogS) (-6.5 - 0.5)
4. Apparent MDCK cell permeability (QPMDCK) (<25 poor, >500 great)
5. Brain/blood partition coefficient (QPlogBB) (-3.0 - 1.2)
6. Percent human oral absorption (> 80% is high, < 25% is poor)

2.5.3. RESULT AND DISCUSSION:

QikProp tool predicts both physicochemical significant descriptors and pharmacokinetically relevant properties. It also evaluates the acceptability of analogues based on Lipinski’s rule of 5\textsuperscript{131} which is essential to ensure drug like pharmacokinetic profile while using rational drug design. The ADME values of design compound inhibitors are given in Table 2.12.

<table>
<thead>
<tr>
<th>Compound Code</th>
<th>M.W.</th>
<th>Log P_o/w</th>
<th>Log S</th>
<th>Log BB</th>
<th>PMDCK</th>
<th>% Oral Absorption</th>
</tr>
</thead>
<tbody>
<tr>
<td>AZTMP</td>
<td>251.244</td>
<td>-0.846</td>
<td>-2.096</td>
<td>-2.033</td>
<td>0008.932</td>
<td>046.816</td>
</tr>
<tr>
<td>T1</td>
<td>348.354</td>
<td>3.260</td>
<td>-5.458</td>
<td>-0.370</td>
<td>1516.425</td>
<td>100.000</td>
</tr>
<tr>
<td>T2</td>
<td>360.389</td>
<td>3.076</td>
<td>-5.226</td>
<td>-0.563</td>
<td>0847.777</td>
<td>100.000</td>
</tr>
<tr>
<td>T3</td>
<td>394.834</td>
<td>3.672</td>
<td>-5.755</td>
<td>-0.140</td>
<td>3910.039</td>
<td>100.000</td>
</tr>
<tr>
<td>T5</td>
<td>324.763</td>
<td>4.071</td>
<td>-5.195</td>
<td>-0.361</td>
<td>1119.121</td>
<td>100.000</td>
</tr>
<tr>
<td>T6</td>
<td>335.315</td>
<td>2.994</td>
<td>-4.436</td>
<td>-1.220</td>
<td>0086.764</td>
<td>085.652</td>
</tr>
<tr>
<td>T7</td>
<td>335.315</td>
<td>2.890</td>
<td>-4.607</td>
<td>-1.470</td>
<td>0047.705</td>
<td>080.742</td>
</tr>
<tr>
<td>T8</td>
<td>306.317</td>
<td>2.791</td>
<td>-4.046</td>
<td>-1.066</td>
<td>0126.523</td>
<td>087.178</td>
</tr>
</tbody>
</table>
### Molecular Modeling Studies

<table>
<thead>
<tr>
<th>Compound Code</th>
<th>M.W.</th>
<th>Log $P_{ow}$</th>
<th>Log S</th>
<th>Log BB</th>
<th>PMDCK</th>
<th>% Oral Absorption</th>
</tr>
</thead>
<tbody>
<tr>
<td>T9</td>
<td>320.344</td>
<td>3.695</td>
<td>4.707</td>
<td>-0.600</td>
<td>0455.241</td>
<td>100.000</td>
</tr>
<tr>
<td>T10</td>
<td>316.356</td>
<td>4.283</td>
<td>5.210</td>
<td>-0.709</td>
<td>0455.222</td>
<td>100.000</td>
</tr>
<tr>
<td>T14</td>
<td>267.305</td>
<td>1.840</td>
<td>2.492</td>
<td>0.061</td>
<td>1866.584</td>
<td>100.000</td>
</tr>
<tr>
<td>TSI</td>
<td>247.246</td>
<td>1.440</td>
<td>3.265</td>
<td>-0.545</td>
<td>913.335</td>
<td>82.19</td>
</tr>
<tr>
<td>TSIa</td>
<td>265.692</td>
<td>2.303</td>
<td>4.372</td>
<td>0.122</td>
<td>7369.048</td>
<td>96.30</td>
</tr>
<tr>
<td>TSIbi</td>
<td>352.342</td>
<td>3.270</td>
<td>7.188</td>
<td>-1.324</td>
<td>0531.86</td>
<td>89.03</td>
</tr>
<tr>
<td>TSIbii</td>
<td>359.780</td>
<td>4.702</td>
<td>7.971</td>
<td>0.187</td>
<td>10000.000</td>
<td>100.00</td>
</tr>
<tr>
<td>TSIbiii</td>
<td>307.344</td>
<td>4.026</td>
<td>6.961</td>
<td>-0.064</td>
<td>5814.773</td>
<td>100.00</td>
</tr>
<tr>
<td>TSIbiv</td>
<td>325.335</td>
<td>4.180</td>
<td>7.215</td>
<td>0.038</td>
<td>10000.000</td>
<td>100.00</td>
</tr>
<tr>
<td>TSIb</td>
<td>340.35</td>
<td>3.262</td>
<td>5.458</td>
<td>-0.370</td>
<td>01500.215</td>
<td>95.20</td>
</tr>
<tr>
<td>TSI</td>
<td>263.701</td>
<td>1.646</td>
<td>3.682</td>
<td>-0.504</td>
<td>01255.165</td>
<td>83.40</td>
</tr>
<tr>
<td>TSIa</td>
<td>282.147</td>
<td>2.518</td>
<td>4.737</td>
<td>0.170</td>
<td>10000.000</td>
<td>100.00</td>
</tr>
<tr>
<td>TSIb</td>
<td>368.797</td>
<td>3.579</td>
<td>7.571</td>
<td>-1.282</td>
<td>00730.947</td>
<td>90.84</td>
</tr>
<tr>
<td>TSIbii</td>
<td>341.790</td>
<td>4.514</td>
<td>7.692</td>
<td>0.098</td>
<td>10000.000</td>
<td>100.00</td>
</tr>
<tr>
<td>TSIbiii</td>
<td>376.235</td>
<td>4.966</td>
<td>8.356</td>
<td>0.241</td>
<td>10000.000</td>
<td>100.00</td>
</tr>
<tr>
<td>TSIbiv</td>
<td>323.799</td>
<td>4.290</td>
<td>7.345</td>
<td>-0.011</td>
<td>07991.108</td>
<td>92.10</td>
</tr>
<tr>
<td>TSI</td>
<td>319.334</td>
<td>1.574</td>
<td>4.032</td>
<td>-0.804</td>
<td>00657.373</td>
<td>84.79</td>
</tr>
<tr>
<td>TSIbii</td>
<td>431.868</td>
<td>4.798</td>
<td>7.742</td>
<td>0.072</td>
<td>10000.000</td>
<td>100.00</td>
</tr>
<tr>
<td>TSIb</td>
<td>368.840</td>
<td>3.182</td>
<td>5.011</td>
<td>-0.500</td>
<td>00900.200</td>
<td>92.10</td>
</tr>
<tr>
<td>TSI</td>
<td>259.282</td>
<td>1.222</td>
<td>3.355</td>
<td>-0.741</td>
<td>00510.588</td>
<td>80.91</td>
</tr>
<tr>
<td>TSIb</td>
<td>277.728</td>
<td>2.164</td>
<td>4.248</td>
<td>-0.062</td>
<td>04119.241</td>
<td>95.49</td>
</tr>
<tr>
<td>TSI</td>
<td>263.701</td>
<td>1.646</td>
<td>3.682</td>
<td>-0.504</td>
<td>01255.181</td>
<td>803.40</td>
</tr>
<tr>
<td>TSIb</td>
<td>282.147</td>
<td>2.518</td>
<td>4.737</td>
<td>0.17</td>
<td>10000.000</td>
<td>100.00</td>
</tr>
<tr>
<td>TSI</td>
<td>369.825</td>
<td>3.211</td>
<td>5.181</td>
<td>-0.539</td>
<td>00913.235</td>
<td>95.23</td>
</tr>
<tr>
<td>TSIb</td>
<td>369.825</td>
<td>3.324</td>
<td>5.699</td>
<td>-0.572</td>
<td>01128.228</td>
<td>95.69</td>
</tr>
<tr>
<td>TSI</td>
<td>392.234</td>
<td>3.951</td>
<td>6.369</td>
<td>-0.261</td>
<td>03810.093</td>
<td>100.00</td>
</tr>
<tr>
<td>TSIX</td>
<td>357.789</td>
<td>3.546</td>
<td>5.823</td>
<td>-0.373</td>
<td>01982.961</td>
<td>96.995</td>
</tr>
<tr>
<td>TSC1</td>
<td>226.241</td>
<td>0.433</td>
<td>3.991</td>
<td>-2.017</td>
<td>00045.004</td>
<td>65.934</td>
</tr>
<tr>
<td>TSC2</td>
<td>256.266</td>
<td>0.574</td>
<td>4.324</td>
<td>-2.103</td>
<td>00046.906</td>
<td>67.060</td>
</tr>
<tr>
<td>TSC3</td>
<td>227.225</td>
<td>0.821</td>
<td>3.967</td>
<td>-1.989</td>
<td>00047.089</td>
<td>68.528</td>
</tr>
</tbody>
</table>
Chapter 2  
Molecular Modeling Studies

<table>
<thead>
<tr>
<th>Compound Code</th>
<th>M.W.</th>
<th>Log P&lt;sub&gt;o/w&lt;/sub&gt;</th>
<th>Log S</th>
<th>Log BB</th>
<th>PMDCK</th>
<th>% Oral Absorption</th>
</tr>
</thead>
<tbody>
<tr>
<td>TSC4</td>
<td>257.251</td>
<td>0.893</td>
<td>-4.189</td>
<td>-1.994</td>
<td>00047.09</td>
<td>068.957</td>
</tr>
<tr>
<td>TSC5</td>
<td>220.233</td>
<td>-0.752</td>
<td>-2.869</td>
<td>-1.586</td>
<td>00007.672</td>
<td>045.545</td>
</tr>
<tr>
<td>TSC6</td>
<td>221.218</td>
<td>-0.052</td>
<td>-3.818</td>
<td>-1.581</td>
<td>00063.563</td>
<td>065.576</td>
</tr>
<tr>
<td>TSC7</td>
<td>203.163</td>
<td>-1.055</td>
<td>-2.280</td>
<td>-1.583</td>
<td>019.612</td>
<td>051.247</td>
</tr>
<tr>
<td>T11</td>
<td>307.300</td>
<td>2.701</td>
<td>-4.036</td>
<td>-1.022</td>
<td>123.200</td>
<td>067.178</td>
</tr>
<tr>
<td>T12</td>
<td>367.400</td>
<td>3.170</td>
<td>-5.140</td>
<td>-0.520</td>
<td>913.240</td>
<td>095.012</td>
</tr>
<tr>
<td>T13</td>
<td>281.600</td>
<td>2.408</td>
<td>-4.615</td>
<td>0.180</td>
<td>925.350</td>
<td>068.860</td>
</tr>
</tbody>
</table>

The first three properties are based on Lipinski rule of five, molecular weight (mol_MW) less than 650, partition coefficient between octanol and water (logP o/w) between -2 and 6.5 and solubility (QPlogS) greater than -7. Brain/blood partition coefficient (QPlogBB) parameter indicated about the ability of the drug to pass through the blood brain barrier which is mandatory for inhibition of HIV infection.

Whereas QPPMDCK Predicted apparent MDCK cell permeability in nm/sec. MDCK cells are considered to be a good mimic for the blood-brain barrier. Higher the value of MDCK cell higher the cell permeability. All designed compounds have shown the ADME properties in acceptable range.