CHAPTER-5

EXPERIMENTAL

5. EXPERIMENTAL

The tools, techniques and procedures/methods used for carrying out research work reported in this thesis have been described as follows:

5.1 Molecular Modeling Hardwares and Softwares

5.1.1 Hardwares

- Three-dimensional QSAR (3D-QSAR CoMFA and CoMSIA), docking/calculation of binding energy, pharmacophore generation and virtual screening studies were performed using Silicon Graphics Fuel workstation running on IRIX 6.5 operating system. The system configuration was as follows:

Workstation Name	Silicon Graphics Fuel
Model	Onyx3 Infinite performance Fuel
CPU	1700MHz MIPS R 16000 (IP35) Processor with MIPS R1 6010 FPU
Memory	512 MB
Data cache	32KB
Instruction cache	32KB
Operating system	IRIX 64 Release 6.5
Graphics	V10
Hardware	HUB in module 001c01 Slot 0 Revision 2 speed 200 MHz (enabled)

5.1.2 Softwares

- Three-dimensional QSAR studies were performed using QSAR module in SYBYL (version 7.0) [311] procured from Tripos Inc., USA.
- Docking studies were performed using GLIDE (version 4.0) [102].
- Pharmacophore generation and virtual screening studies were performed using PHASE [38].
- The log P values for the 56 molecules were calculated using ACD/CHEMSKETH FREEWARE (version)10.00 [322].

5.2 Calculation of Biological Activity in Terms of pIC_{50} (Negative Log of IC_{50} in Molar Concentration)

In all of the 3D-QSAR studies, the biological activity in IC_{50} were converted to pIC_{50} using the following equation:

$$pIC_{50} = -\log IC_{50}$$

where, IC_{50} is the concentration (in M) of the inhibitor producing 50% inhibition of enzyme. These values were taken as dependent variables in the 3D-QSAR analysis.

5.3 3D-QSAR Studies

5.3.1 Structure building, optimization, calculation of descriptors and partial Charges

Wherever the X-ray crystallographic data for ligand-target complex was not available, all the compounds were constructed using standard geometry and bond lengths within SYBYL. In case of TACE inhibitors, the co-crystallized ligand (IK-682) was extracted out from the holo structure (pdb code: 2fv5) imported from the protein data bank[323] and all other molecules were built on it. The initial optimization was carried out using TRIPOS force fields employing the Gasteiger Huckel charges. Repeated minimization was performed using steepest descent and conjugated gradient methods till the RMSD (root mean square deviation) value of 0.001 kcal/mol was achieved. Using MULTISEARCH option in SYBYL 7.0, various conformers were obtained out of which the lowest energy conformers were selected for superimposition. Further geometry optimization was performed using MOPAC with AM1 Hamiltonian [324].

A study was performed for determination of structural requirements of influenza neuraminidase Type A inhibitors and binding interaction analysis with the active site of A/H1N1 by 3D-QSAR. In this study, a set of low energy conformations for the template molecule, which was the most active compound in the series was generated by dynamics using simulated annealing technique with Tripos force fields in SYBYL. The molecule was heated to 700 K followed by cooling to 300 K. Time spent for annealing was 1000 fs. Time increment for dynamics computations was 0.5 fs and coupling time for temperature regulation was 2.0 fs. Ten consecutive cycles were calculated. Repeating the cycle many times and collecting the low energy structures resulted in a set of low energy conformations. The lowest energy conformers thus obtained were further minimized using the conjugate gradient method in SYBYL 7.0 using Tripos force fields, atomic charges assigned by the Gasteiger-Huckel method with a distance dependent dielectric

function until a root mean square deviation (rms) of 0.001 kcal/mol Å was achieved. The lowest energy conformer thus obtained was subsequently used for alignments.

5.3.2 General alignment rules

In all the 3DQSAR studies, one or more of the four alignment rules as mentioned below have been employed:

Alignment I : In this alignment, centroids rather than exact superimposition of atoms of the rings were used for RMS fitting to the template.

Alignment II : Here, a combination of both atoms and centroids were used for superimposition and RMS fitting to the template.

Alignment III: In this alignment only atoms were used for RMS fitting on corresponding atoms of the template.

Alignment IV: Each analog was aligned to the template by rotation and translation so as to minimize RMSD between atoms and centroids in template and the corresponding atoms/centroids in the analog using the DATABASE ALIGN command in SYBYL.

Alignment V : Template structure was docked in the active site of TACE enzyme (pdb code: 2fv5) and the resulting conformer was used for the data base alignment as per Alignment IV.

Alignment VI: Docked conformations were aligned on one another and the same were used for the CoMFA model development.

Alignment VII: After docking all the conformations in the active site, all the compounds were aligned on the most active compound as per the procedure adopted in Alignment III.

5.3.3 Segregation of compounds into training and test data sets

For performing CoMFA/CoMSIA 3D-QSAR and pharmacophore studies, compounds were divided into training and test compounds with the following two rules: (a) both training and test set compounds were represented from each class of compounds to ensure structural diversity [if one class had only one compound, it was assigned to the training set] (b) both, training and test sets covered the bioactivities as wide as possible. If there was only one compound with maximum or minimum order of bioactivity in a class, such a compound was assigned to the training set.

5.3.4 CoMFA analysis

CoMFA of the selected series of molecules was carried out on the steric and electrostatic fields using the default values. A 3D cubic lattice, with a 2 Å grid spacing, was generated automatically around these molecules to ensure that the grid extended the molecular dimensions by 4 Å in all directions. A threshold column filtering of 2.0 kcal/mol was set to hasten the analysis and reduce the amount of noise. The steric and electrostatic fields were calculated separately for each molecule using sp^3 carbon atom probe with a charge of +1 (default probe atom in SYBYL) and computed energy cut-off values of 30 kcal/mol Å for both steric and electrostatic interactions with each atom in the molecule using CoMFA standard scaling.

5.3.5 CoMSIA analysis

The standard setting (probe with +1, radius 1 Å and hydrophobicity +1, hydrogen bond donating +1, hydrogen bond accepting +1, attenuation factor of 0.3 and grid spacing 2 Å) was used in CoMSIA to calculate five different fields viz. steric, electrostatic, hydrophobic, hydrogen bond donor and hydrogen bond acceptor.

5.3.6 Partial least square (PLS) analysis

The CoMFA/CoMSIA descriptors were used as independent variables and pIC_{50} values as dependent variables in PLS regression analysis for deducing 3D-QSAR models. Cross-validation was used to check the predictive power of the derived model. The result of analysis corresponds to the regression equation with thousands of coefficients. The predicitive values of models were evaluated using leave-one-out (LOO) cross validation method. The number of components leading to the highest cross-validated r^2 and lowest standard error of prediction (SEP), was set as the optimum number of components (Nc) in PLS analysis. Using these optimum number of components, non-cross-validated r^2 was derived. This analysis was saved and used to predict the biological activity of training and test set of compounds under study. A minimum of 2.0 kcal/mol column filtering (σ) was used as the threshold column filtering value in PLS analysis. To further assess the robustness and statistical confidence of the developed model, the cross-validated results were analyzed by considering the fact that a value of r^2_{cv} above 0.3 indicated that probability of chance correlation was less than 5% [325].

SMT. HANSA MEHTA LIBRARY The M. S. University of Baroda

5.3.7 Evaluation of the predictive ability of CoMFA and CoMSIA models

The predictive ability of each analysis was determined from the test set molecules that were not included in the training set. These molecules were aligned, and their activities were predicted by each PLS analysis. The predictive r^2 (r^2_{pred}) value was defined as:

$$r^2_{pred} = (SD\text{-}PRESS)/SD$$

where, SD is the sum of squared deviations between the biological activities of the test set and mean activity of the training set molecules, and PRESS is the sum of squared deviation between actual and predicted activities of the test set molecules.

5.4 Docking Studies

5.4.1 Selection of crystal structure of TACE

Since crystal structure of porcine TACE (pTACE) is not available in PDB, human TACE (hTACE) pdb code: 2fv5 was used as a part of 3D-QSAR, docking and superimposition and pharmacophore generation studies. The hTACE structure (pdb code: 2fv5, chain B) obtained from Protein Data Bank (PDB, USA) was used for the superimposition of the pharmacophore features because after performing BLAST-2 search [302] with pTACE, it was observed that sequence of hTACE has high degree of homology with that of pTACE (highest identity/BLAST Score) [303]. The sequence of the 839 amino acids of the pTACE was retrieved from the Swiss-Port Database.

5.4.1 Protein preparation and grid generation

5.4.2.1 Preparation of TACE

Protein preparation procedure was imported for attaining accurate docking with GLIDE 4.0. The aim was to remove physically untenable steric clashes often found in crystallographically determined protein sites so that the native ligand can yield favorable van der Waals interaction energies for properly docked structures. It was also important that protonation states and hydrogen bonding patterns be correct. Protein preparation was carried out using 'protein preparation and refinement' option within GLIDE 4.0. Below mentioned steps were followed for protein preparation after importing the protein complex structure (pdb code: 2fv5) from the protein data bank:

- The first step was to prepare the co-crystallized ligand (IK-682) by making sure that multiple bonds are defined correctly and that hydrogens are properly added. "Assign bond orders" option from the Maestro interface was used for this purpose.

- In second step "protein preparation" option from the application main menu was used to run the "pprep" script that is provided with GLIDE 4.0. This procedure neutralizes residues that do not participate in salt bridges and that are at more than a specified distance from the nearest ligand atom. By default, it chooses the value between 10 and 20 Å for this distance that minimizes the total charge of the receptor-ligand complex. The script also sets the tautomeric state for histidine residue, which are assumed to be neutral, by considering potential metal-ligation and hydrogen bonding interactions.
- In the third step "protein refinement" option was selelcted to postprocess the prepared 2fv5 structure in second step. This step was found necessary because the judgement made by the 'Protein preparation" procedure in step 2 might not always be correct. In addition, the original and prepared versions of the protein were compared to search for any discrepancies. In this step the protonation states of amino acid residues were also adjusted.
- The final step which involved addition of hydrogen and minimization of the receptor-ligand complex was done by default setting within GLIDE 4.0. The force constants employed were 3, 1, 0.3, and 0.1 kcal/mol. The procedure stops when the cumulative rms deviation from the initial coordinates for non-hydrogen atoms exceeds a target value, the default for which is 0.3 Å. The last structure having an rms deviation smaller than the target value is then selected further for receptor grid generation and calculation of binding energies.

5.4.2.2 Receptor grid generation

The generated grid files represent physical properties of a volume of the receptor, specifically the active site that will be searched while attempting to dock a ligand. The steps employed for generation of receptor grids for a co-crystallized ligand-receptor complex (2fv5) are mentioned below:

- **Defining the receptor-** The co-crystallized ligand (IK-682) was highlighted so that it will be excluded from the grid generation. The vdW radii of the receptor atoms were not scaled but set to default value of 1.00
- **Defining of the active site-** IK-682 was removed in the first step and the volume for which the grid was calculated was defined correctly. "Site tab" option from the Grid generation menu was used for defining the active site. A purple enclosing box appears that represents the volume of the protein (2fv5). The enclosing box was made small so as to be consistent with the shape and character

of the protein's active site and also with the ligands to be docked. The "size" option was selected in the receptor grid generation box menu. As we had co-crystallized IK-682 in the TACE active site, we highlighted an option which allowed ligands (similar in size to IK682) to be docked. No other specific constraints were applied during receptor grid generation.

- **Receptor grid generation-** After defining receptor and active site, grid generation was started by pressing the start button in the lower left corner of the receptor grid generation panel. Final output files had extension.grd which were imported for the docking calculation.

5.4.3 Docking/calculation of binding energies

GLIDE 4.0 was used for docking and calculating the binding energies. Ligands under study were initially geometry optimized in SYBYL and imported on the workspace of "maestro", a graphical user interface for GLIDE. The grid file generated in the previous step was specified, so that the ligands in the workspace dock in the specified grid. Standard precision mode was used for docking. Finally, the ligands from the workspace to be docked were specified and docking calculations were run by clicking on the "ligand docking" option from the GLIDE applications menu. Ligand docking was performed using OPLSAA force field. Binding energies were calculated and displayed in the output file having extension pv.mae. The constraints to define ligand–receptor interactions were not set. The structure output format was set to poseviewer file so as to view the output of the resulting docking studies from pose-viewer. The H-bonds and van der Waals contacts (good, bad and ugly) to the receptor were visualized using default settings to analyze the binding modes of the ligands to receptor. Docking score was reported as G-Score. More negative G-score (glide score) indicates a better fit in the binding site.

5.5 Pharmacophore Generation

The generation of common pharmacophore hypothesis (CPH) and the alignment based on it was carried out using PHASE [38], version 2.5 (Schrödinger, LLC, New York) installed on an SGI workstation. Following work flowchart (**Figure 5.1**) was adopted to generate the pharmacophore hypothesis.

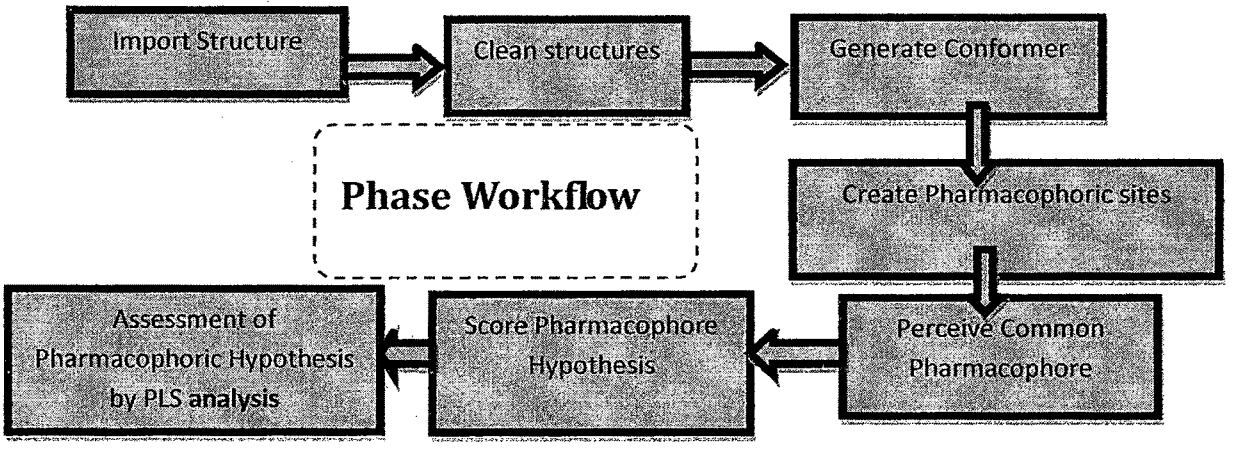

Figure 5.1 Phase workflow adopted for generation of pharmacophore hypothesis

Structures of the compounds were imported from the project table in the "Develop Pharmacophore Model" panel and refined (cleaned) geometrically using Ligprep module available with Schrodinger. Conformations were generated by the mixed MCMM/LMOD method using a maximum of 2,000 steps with a distance-dependent dielectric solvent model and an OPLS-2005 force field. All the conformers were subsequently minimised using truncated Newton conjugate gradient minimisation up to 500 iterations. For each molecule, a set of conformers with a maximum energy difference of 15 kcal/mol relative to the global minimum energy conformer were retained. A redundancy check of 2 Å in the heavy atom positions was applied to remove duplicate conformers. Pharmacophore features- hydrogen bond acceptor (A), hydrogen bond donor (D), hydrophobic group (H), negatively charged group (N), positively charged group (P) and aromatic ring (R) were defined by a set of chemical structure patterns as SMARTS queries and assigned one of the three possible geometries that defined the physical characteristics of the site:

Point : The site is located on a single atom in the SMARTS query.

Vector : The site is located on a single atom in the SMARTS query, and assigned directionality according to one or more vectors originating from the atom.

Group : The site is located at the centre of a group of atoms in the SMARTS query. For aromatic rings, the site is assigned directionality, defined by a vector that is normal to the plane of the ring.

Active and inactive thresholds of pIC_{50} were applied to the dataset of the compounds to yield actives and inactives that were used for pharmacophore generation and subsequent scoring. Common pharmacophoric features were then identified from a set of variants - a set of feature types that define a possible pharmacophore - using a tree-based partitioning algorithm with maximum tree depth of four with the requirement that all actives must match. The final size of the pharmacophore box was fixed at 1 Å to optimise the number of final CPHs. CPHs were examined using a scoring function to yield the best alignment of the active ligands using an overall maximum root mean square deviation (RMSD) value of 1.2 Å with default options for distance tolerance. The quality of alignment was measured by a survival score (S), defined as:

$$S = W_{site} S_{site} + W_{vec} S_{vec} + W_{vol} S_{vol} + W_{sel} S_{sel} + W^{m}_{rew} - W_{E}\Delta E + W_{act}A$$

Where, W's are weights and S's are scores; S_{site} represents alignment score, the RMSD in the site point position; S_{vec} represents vector score, and averages the cosine of the angles formed by corresponding pairs of vector features in aligned structures; S_{vol} represents volume score based on overlap of van der Waals models of non-hydrogen atoms in each pair of structures; and S_{sel} represents selectivity score, and accounts for what fraction of molecules are likely to match the hypothesis regardless of their activity towards the receptor. W_{site}, W_{vec}, W_{vol} have default values of 1.0, while W_{sel} has a default value of 0.0. In hypothesis generation, default values have been used. The reward comes in the form of W^{m}_{rew}, where W_{rew} is user-adjustable (1.0 by default) and *m* is the number of actives that matches the hypothesis minus one. $W_{E}\Delta E$ represents penalty included for high-energy structures by subtracting a multiple of the relative energy from the final score and penalize hypothesis for which the reference ligand activity is lower than the highest activity, by adding a multiple of the reference ligand activity to the score represented by $W_{act}A$, where A is the activity.

5.6 Virtual screening

In order to discover unknown biological activities of existing compounds in corporate/public databases and subsequently to identify new structures for inhibiting TACE, ASINEX database [326] was screened by using following filters:

- **Pharmacophore filter:** All the compounds of the AINEX database were screened by using the common pharmacophore hypothesis (CPH1). Compounds having desired common pharmacophore features were retained and remaining

ones were discarded. In order to decrease the number of hits provided by a 3D database query of CPH1, docking and Lipinski's rule of five were used as filters.

- **Docking:** Virtual screening is used to distinguish potential leads from inactive compounds in a database of chemical samples. One method for accomplishing this is by docking compounds into the structure of a receptor binding site in order to rank-order compounds by the quality of the interactions they form with the receptor. It is generally established that docking can be reasonably successful at generating good poses of a ligand in an active site. In order to decrease the number of hits provided by a 3D database query of CPH1, docking of the hits provided by a 3D database query of CPH1 was carried out by using Glide software. The hTACE structure (pdb code: 2fv5, chain B) obtained from Protein Data Bank (PDB, USA) was used for the docking purpose.
- **Lipinski's rule of five:** Christopher Lipinski's rule-of-five analysis helps to raise awareness about properties and structural features that make molecules more or less drug-like. The guidelines predict that poor absorption or permeation of a orally administered compound are more likely if the compound meets the following criteria:
  - ✓ Molecular mass greater than 500 Da
  - ✓ High lipophilicity (expressed as CLogP greater than 5)
  - ✓ More than 5 hydrogen bond donors
  - ✓ More than 10 hydrogen bond acceptors

The "rule of five" is so called because most of the features start with the number five. This rule helps to understand pharmacokinetic properties of new chemical entities in early preclinical development and could help avoid costly late-stage preclinical and clinical failures of the compounds.