Chapter-3
3.0 Materials and Methods

3.1 Hardware Components

In present work all the calculations were carried out with high frequency computational analysis such as molecular modeling, energy minimizations, design and optimization of lead molecules, protein ligand interaction studies etc., at Hi-end server (Pentium 3.4 MHzs, AMD Athlon 64 bit, Quadcore processor with 4 GB RAM) manufactured by HCL Corporation, India.

3.2 Software Components

Most of the soft wares used were either Windows or Linux plat form based which were well accepted and referred in various publications at high rated research journals. Academic license was obtained for the commercial software used in the present study by requesting the concerned suppliers. The software used in the present study was briefly detailed below.

3.2.1 Gromacs (Groningen Machine for Chemical Simulations): provides extremely high performance compared to all other programs. A lot of algorithmic optimizations have been introduced in the code; gromacs have for instance extracted the calculation of the virial from the innermost loops over pairwise interactions, and gromacs use our own software routines to calculate the inverse square root. The innermost loops are generated automatically in either C or Fortran at compile time, with optimizations adopted to your architecture

3.2.2 PyMOL: It is an open-source, user-sponsored, molecular visualization system, which is well suited to produce high quality 3D images of small molecules and biological macromolecules such as proteins. According to the author, almost a quarter of all published images of 3D protein structures in the scientific literature were made using PyMOL. (http://www.delanoscientific.com/)

3.2.3 Visual molecular dynamics (VMD): VMD is a molecular modelling and visualization computer program which is primarily developed as a tool for viewing and analyzing the results of molecular dynamics simulations, but it also includes tools for working with volumetric data, sequence data, and arbitrary graphics objects. Molecular scenes can be exported to external rendering tools such as POV-Ray, Renderman, Tachyon, VRML, and many others. (http://www.ks.uiuc.edu/Research/vmd/)
3.2.4 **AUDOCK-Tool:** AutoDock is a suite of automated docking tools designed to predict how small molecules, such as substrates or drug candidates, bind to a receptor of known 3D structure. AutoDock actually consists of two main programs: AutoDock performs the docking of the ligand to a set of grids describing the target protein; AutoGrid pre-calculates these grids. In addition to using them for docking, the atomic affinity grids can be visualised. This can help, for example, to guide organic synthetic chemists design better binders. (http://autodock.scripps.edu/)

3.2.5 **Clustal** is a widely used multiple sequence alignment computer program. The latest version is 1.83. There are two main variations:
- **Clustal-W:** Available on line at EMBL server
- **Clustal-X:** This version has a graphical user interface. It is available for Windows, Mac OS and Unix/Linux.

3.2.6 **Hyperchem** : HyperChem is the software for molecular modeling, energy minimization and simulation of lead and drug molecules. It calculates the QSAR properties of small existing molecules in database. It is a windows based commercial software. http://www.hyper.com/.

3.2.7 **Chem Office ultra 7.0:** ChemOffice Ultra is the ultimate chemistry & biology suite designed for chemists, which is a suite of software consisting of ChemDraw, Chem3D and ChemFinder. It has encyclopedia of chemical structures, drugs and biological properties with over 10,000 monographs on single substances or groups of related compounds.

### 3.3 On line Tools

In addition to the above software, various on line computational tools used in the present study were as denoted below.

3.3.1 **National Center for Biotechnology Information (NCBI):** The NCBI is part of the United States National Library of Medicine (NLM), a branch of the NIH, located in Bethesda, Maryland and was founded in 1988. The NCBI houses genome sequencing data in GenBank and an index of biomedical research articles in PubMed Central and PubMed, as well as other information relevant to biotechnology. All these databases are available online through the Entrez search engine. (www.ncbi.nlm.nih.gov/).
3.3.2 **PDB: Protein Data Bank** is a repository for 3-D structural data of proteins and nucleic acids. These data, typically obtained by X-ray crystallography or NMR spectroscopy, are submitted by biologists and biochemists from around the world, are released into the public domain, and can be accessed for free. The mission of the PDB is to maintain a single Protein Data Bank Archive of macromolecular structural data that is freely and publicly available to the global community. (www.rcsb.org/pdb).

3.3.3 **Pfam**: Pfam is a collection of protein motifs and families maintained by the Bioinformatics group at the Sanger Centre. Pfam hidden Markov models (HMMs) and the Prosise generalized profiles were developed based on distinct theoretical backgrounds. (http://www.sanger.ac.uk/Software/Pfam/).

3.3.4 **SWISS-Prot (Expasy)**: This server is manually curetted biological database of protein sequences created in 1986 by the Swiss Institute of Bioinformatics and the European Bioinformatics Institute. Swiss-Prot strives to provide reliable protein sequences associated with a high level of annotation (such as the description of the function of a protein, its domains structure, post-translational modifications, variants, etc.), a minimal level of redundancy and high level of integration with other databases. (http://expasy.org/sprot/). This program is available from European Bioinformatics Institute ftp server. (http://www.ebi.ac.uk/Tools/clustalw/index.html)

3.3.5 **Molinspiration Server**: JME Molecular Editor is a Java applet which allows to draw / edit molecules and reactions (including generation of substructure queries) and to depict molecules directly within an HTML page. Editor can generate Dayligth SMILES or MDL mol file of created structures. Due to many requests, the applet (in form of a jar file) has been released to the public and become a standard for molecular structure input on the web with more than 6500 installations worldwide. As recognition of this generous gesture, Molinspiration provides this space for the JME Home. Molinspiration can also offer help with installation and deployment of the JME. (http://www.molinspiration.com/docu/webme/)

3.3.6 **PDBSUM** is a pictorial database that provides an at-a-glance overview of the contents of each 3D structure deposited in the Protein Data Bank (**PDB**).
This server provides cleft and groves on the surface of protein molecules deposited in protein data bank. Pdbsum can also give ligand binding site with Ligplot graphs in 2 dimensional appearances. (http://www.ebi.ac.uk/pdbsum/ )

**3.3.7 PRODRG** : The PRODRG will convert coordinates of small molecules in PDB format to various formats of GROMACS, GROMOS, WHAT IF, CNS, AUTODOCK 2.4 and AUTODOCK 3.0 etc. The output files can be used for the further analysis in various softwares. In addition coordinates for hydrogen atoms are generated. ( www.davapc1.bioch.dungee.ac.uk. )

**3.3.8 VADAR** (Volume Area Dihedral Angle Reporter) is a comprehensive web server for quantitative protein structure evaluation. It accepts Protein Data Bank (PDB) formatted files or PDB accession numbers as input and calculates, identifies, graphs, reports and/or evaluates a large number (>30) of key structural parameters both for individual residues and for the entire protein. These include excluded volume, accessible surface area, backbone and side chain dihedral angles, secondary structure, hydrogen bonding partners, hydrogen bond energies, steric quality, salvation free energy as well as local and overall fold quality. These derived parameters can be used to rapidly identify both general and residue specific problems within newly determined protein structures. The VADAR web server is freely accessible at http://redpoll.pharmacy.ualberta.ca/vadar.

**3.4.0 Methods**

**3.4.1 Sequence analysis**

Various types sequence analysis was carried out through retrieving the sequences either from NCBI or SWISS-Prot databases. Sequence homology search was conducted through the blast-P program available at NCBI. Homology modeling of target sequence needs a template crystal structure coordinates which were obtained by performing blast-P at NCBI with selection of database as PDB. The coordinates of selected crystal structures of sequence similar structures of target protein were obtained from PDB and used for prediction of 3-D structure of target protein using MODELLER 9v1. In order to identify conserved and variable regions of the sequences and in determining the
most robsi gap arrangement, multiple sequence alignment of all homologous proteins of the target sequence Clustal-W (Chenna et. al., 2003) with appropriate parameters were used as per the specified instructions.

### 3.4.2 Molecular dynamics setup

Molecular dynamics (MD) is a computational method that calculates the time dependent behavior of a molecular system. MD simulations provide detailed information on the fluctuations and conformational changes of proteins and nucleic acids, and they are now routinely used to investigate the structure, dynamics and thermodynamics of biological molecules and their complexes. The basic idea of molecular dynamics (MD) is to study atomic fluctuations in solvated system. The principle behind this is application of classical Newton’s equation. All simulations reported in this thesis were performed with GROMACS molecular simulation package (Berendsen et al., 1995). Before MD, the ionization states of amino acids were set to mimic a neutral PH environment i.e. all Lys and Arg carried net positive charge, and all glutamic acid and Aspartic acid carried a net negative charges. The histidine residues were in doubly protonated condition.

A random generation of 100 models from the starting structure was calculated and subsequently the best model with the low RMS value of superposition using Swiss-pdb viewer (Guex et al., 1999) was subjected for further analysis. The best modeled protein were solvated with water molecules in a truncated octahydron box. The size of the box was set to 0.9 nm distance from the surface of the protein. The Single Point Charge (SPC) water model (Berendsen et al., 1987) and ions (Na+ and Cl −) was built. The box model, first with explicit water and then with ions was added to protein containing truncated octahydron box, this was submitted to 400 steps of energy minimization using the steepest descent algorithm till an energy gradient was reached and it was found to be the most appropriate energy gradient to relax the models and afford well Ramachandran plots. In order to constrain all bond length in protein, the LINCS (Hess et al., 1997) algorithm was used. For water molecule bond length constrain, the SETTLE algorithm was implemented (Miyamoto and Kollman, 1992). The electrostatic and Van Der Waal forces are implemented using particle mesh Ewald potential method (Essmann et al., 1995) and Lennard-Jones potential
method respectively. All full MD simulations were performed at 500ps with no restriction using two fs of integration time, constant temperature and pressure. The temperatures of the proteins and solvent molecules were each coupled separately, using (V-resacle) Berendsen thermostat algorithm (Berendsen et al., 1984). The pressure was coupled using (Parrinello-Rahman) Berendsen algorithm at 1 bar with coupling constant $\tau_p = 1$ ps. Co-ordinates and energy terms (total, kinetic and potential for the whole system and electrostatic, distance dependent, distance-independent reaction force field) were saved for each ps. With the aim of evaluating the system stabilization throughout the molecular dynamics time, the total, kinetic and potential energy was plotted versus time. The stabilization was assessed by graphics visualization using VMD (Humphrey et al., 1996) and Xmgrace.

**Atomic force Fields parameterization**

The empirically derived potential function that describes interactions present between the atoms, in a molecule or separate molecules, is usually called as force field. The force field applied for calculations of potential interaction during MD simulations is GROMOS 96 (van Gunsteren et al., 1996). In GROMOS 96 force field, the non-bonded interactions are a sum of electrostatic and van der Waal contributions.

Non-bonded interaction empirical calculation is defined as follows.

$$U_{nb} = \sum_{\text{atom pairs}} \frac{1}{4\pi\varepsilon_0} \frac{q_i q_j}{r_{ij}} + \sum_{\text{atom pairs}} \left( \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right)$$

Where $q$ is the partial charges present on atoms $i$ and $j$, $r_{ij}$ is the distance between the two atoms to be under consideration, $\varepsilon_0$ denotes electric permittivity of simulated system, and $A$ and $B$ are the Lennard-Jones parameters that purely dependent on chemical nature of interacting atoms.

Bonded interaction

$$U_{bn} = \sum_{\text{bonds}} k_b (r_{ij} - r_0)^2 + \sum_{\text{angles}} k_\theta (\theta_{ijk} - \theta_0)^2 + \sum_{\text{tortions}} k_\phi [1 + \cos (n\phi_{ijkl} - \delta)]$$


\( kb, k\theta \) and \( k\phi \) are the bond stretching, angle-bending and torsional force constants respectively. \( r_o \) is the equilibrium bond length and \( \theta_o \) the equilibrium angle. \( \theta \) is the angle between atoms \( i, j \) and \( k \). \( n \) is the number of minima per full turn of the torsional angle \( \phi \), and \( \delta \) is the location of the first barrier.

**Atomic flexibility analysis**

The inherent flexibility of amino acids during MD simulations is measured using Root Mean Square Fluctuation (RMSF) term. The calculation of RMSF at time \( t \) of atoms in a molecule with respect to static structure is defined as

\[
\text{RMSF}(t) = \left( \frac{1}{N} \sum_{i=1}^{N} \left[ r_i(t) - \langle r_i \rangle \right]^2 \right)^{\frac{1}{2}}
\]

Where \( r_i(t) \) is the position of atom type \( i \) at time \( t \) and \( N \) is the number of atoms. With the use of above equations which is readily available in GROMACS program, the significant movement of ET residues was calculated over stipulated MD simulations period. \( \langle r_i \rangle \) is the average position of atom \( i \) in MD simulations.

**Performance of Molecular dynamic simulation**

After successful completion of molecular dynamic simulation set up and force field parameterization, MD simulations are to be implemented. In present study, Leap-frog version of Verlet algorithm was exploited for generating time averaged structural conformations with respect to forces that act on individual atoms (van Gunsteren and Berendsen, 1988). Leap-frog version of Verlet algorithm uses Newton’s second law of motion.

\[
\mathbf{F}_i = -\frac{\partial \mathbf{U}(\mathbf{r}_1, \ldots, \mathbf{r}_N)}{\partial \mathbf{r}_i}
\]

Newton’s law of motion is thus used in MD simulations to calculate the forces up on successive configurations of the system. The velocities, constant temperatures and pressure were constantly maintained. The MD (Molecular Dynamic) simulation algorithm used in present work is implemented in GROMACS package. The system temperature and pressure were kept constant throughout the MD simulation period.
Gromacs need several steps to set up a file input in the simulation. The steps can be seen in flowchart below (Fig.3.1):

Flowchart above illustrates how to do molecular dynamics simulation of a protein. The steps are divided into:

1. **Conversion of the pdb file**

   At this step pdb is converted to gromos file (gro) with pdb2gmx. Pdbgmx also created topology file (.top)

2. **Generate box**

   At this step, the editconf will determine the type of box and the box size that will be used in the simulation. on Gromacs there are three types of box, namely triclinic, cubic, and octahedron.

3. **Solvate protein**

   The next step is solvate the protein in box. The program genbox will do it. Genbox will generate a box defined by editconf based on the type. Genbox also determined the type of water model that will be used and add number of water molecule for solvate protein the water model commonly used is SPC (Simple Point Charge).

4. **Energy minimization**: The process of adding hydrogen bond or termination may cause atoms in protein too close, so that the collision occurred between the atoms. The collision between atoms can be removed by energy minimization. Gromacs use mdp file for setup parameters. Mdp file specified number of step and cut-off distance. Use grompp to generate input file and mdrun to run energy minimization. The energy minimization may take some time, depending on the CPU [21].

5. **Molecular dynamics simulation**: The process of molecular dynamics simulation is the same as energy minimization. Grompp prepare the input file to run mdrun. Molecular dynamics simulations also need mdp file for setup parameters. Most option of mdrun on molecular dynamics is used in energy minimization except –x to generate trajectory file.
6. Analysis

After the simulation has finished, the last step is to analyze the simulation result with the following program:

- Ngmx to perform trajectory
- G_energy to monitor energy
- G_rms to calculated RMSD (root mean square deviation)

**File Formats in Gromacs:**

In Gromacs, there are several types of file format:

1. trr: a file format that contains data trajectory for simulation. It stores information about the coordinates, velocities, force, and energy.

2. Edr: a file format that stores information about energies during the simulation and energy minimization.

3. Pdb: a form of file format used by Brookhaven protein data bank. This file contains information about position of atoms in structure of molecules and coordinates based on ATOM and HETATM records.

4. Xvg: a form of file format that can be run by Grace. This file is used to perform data in graphs.

5. Xtc: portable format for trajectory. This file shows the trajectory data in Cartesian coordinates.

6. Gro: a file format that provides information about the molecular structure in format gromos87. The information displayed in columns, from left to right.

7. Tpr: a binary file that is used as input file in the simulation. This file cannot be read through the normal editor.

8. Mdp: a file format that allows the user to setup the parameters in simulation or energy minimization.
Analysis of Molecular dynamic results
MD simulations produce bunch of structural conformations at different time scale. Therefore, well planned analysis of bunch of structural conformations can provide vital clues of molecular function exactly. From all simulations generated from starting experimental model the Root Mean Square Fluctuation (RMSF), Root mean square deviation (RMSD), potential, kinetic and total energies were analyzed. The stabilities of intramolecular hydrophobic interactions were evaluated in terms of Lennard-Jones potential. Lennard-Jones potential is a good approximation of Van Der Waal (VDW) stabilization energies.

3.4.3 Prediction of secondary structure of protein
The prediction of protein secondary structure is a major parrot of the general protein folding problem and the method of obtaining some structural information
for any sequence. Secondary structure predication is important in establishing alignments during homology modeling. Secondary structure analysis is carried out through the ProFunc and PDBSUM server (Laskowski et. al., 2005), which gives the clear data of protein, alpha helices, sheets, turns, beta hairpins, beta bluges, gamma turns etc.,

3.4.4 GOR-IV method

This method is based on information theory and was developed by J. Garnier, D. Osguthorpe and B. Robson in 1978. The present version, GOR IV, uses all possible pair frequencies within a window of 17 amino acid residues and is reported by Gamier (Garnier. et al., 1996). After cross validation on a data base of 267 proteins, the version IV of GOR has a mean accuracy of 64.4% for a three state prediction (Q3). The program gives two outputs, one eye-friendly giving the sequence and the predicted secondary structure in rows, H=helix, E=extended or beta strand and C=coil; the second gives the probability values for each secondary structure at each amino acid position. The predicted secondary structure is the one of highest probability compatible with a predicted helix segment of at least four residues and a predicted extended segment of at least two residues.

3.4.5 Determination of active site residues

Determination of active site amino acid residues of given protein was performed with the help of literature survey from wet lab results. Based on high identity with active site residues from the crystal structure, the residues of active site of target protein have been established perfectly. The crystal structures are submitted to PDBSUM server which provides the catalytic sites and from that one can determine the conserved and catalytic residues in the active site of the built protein model by sequence alignment. In our study we have identified active site from the literature and those residues were investigated in homology model with visualization tool like Pymol.

3.4.6 Evaluation of the built 3-D protein model

A protein 3D model derived from homology modeling technique may have some sources of errors. It is important, therefore, to have an assessment of structure’s quality and to be able to identify regions that may need modifications especially
at protein folding and turns. The aim of model evaluation is to determine whether the built model is acceptable and suitable to use for molecular analysis such as docking and dynamics. The accuracy of the comparative built structures were tested using the ENERGY command of the MODELLER program (Sali and Blundell, 1993) and tools like PROCHECK, (Laskowski et al., 1993) and WHAT IF (Vriend, 1990) which clearly judges the accuracy of model.

3.4.7 PROCHECK
The PROCHECK suite of programs provides a detailed check on the stereochemistry of a protein structure. The stereo chemical parameter checks implemented in PROCHECK are derived from high-resolution protein structures, against which the structure is compared on a residue-by-residue basis. The criteria are Ramachandran plot, peptide bond planarity, C-alpha tetrahedral distortion, non bonded interactions, hydrogen bond energies, and closeness off side chain dihedral angles to ideal values.

3.4.8 Design and selection of Ligand molecules
From the existing drug molecule which is showing interaction with target protein, the scaffold or skeleton molecular structure of ligand molecule was first drawn using Hyperchem 7.5 . Over thus designed lead molecule (parent) different modifications were performed by taking in to consideration of the database of substituents and spacers (linkers) containing a collection of current drugs at Molinspiraiton server (http://www.molinspiration.com). A series of lead molecules were thus designed as per molecular substitutions and designated the name and number as per instruction manual of Chem Office Ultra 7.0.v (Muegge, 2003). These designed lead molecules were then analyzed for following Lipinski’s Rule of five. The Lipinski’s Rule of five states that in general orally active drug should has

- Not more than 5 hydrogen bond donors (OH and NH groups)
- Not more than 10 hydrogen bond acceptors (notably N and O)
- A molecular weight under 500g/mol
- A partition coefficient log P less than 5.
Among all the designed leads the molecules of high ranking which follow Lipinski’s rule were selected and further analysed for binding with the protein model using docking tools.

**Lipinski rule of parameters**

a) **LogP (octanol/water partition coefficient)**

LogP is calculated by the methodology developed by Molinspiration as a sum of fragment-based contributions and correction factors. Method is very robust and is able to process practically all organic, and most organometallic molecules.

b) **Octanol-water partition coefficient logP**

LogP is used in QSAR studies and rational drug design as a measure of molecular hydrophobicity. Hydrophobicity affects drug absorption, bioavailability, hydrophobic drug-receptor interactions, metabolism of molecules, as well as their toxicity. LogP has become also a key parameter in studies of the environmental fate of chemicals.

Method for logP prediction developed at Molinspiration (miLogP2.2 – November, 2005) is based on group contributions. These have been obtained by fitting calculated logP with experimental logP for a training set more than twelve thousand, mostly drug-like molecules. In this way hydrophobicity values for 35 small simple "basic" fragments have been obtained, as well as values for 185 larger fragments, characterizing intramolecular hydrogen bonding contribution to logP and charge interactions. Molinspiration methodology for logP calculation is very robust and is able to process practically all organic and most organometallic molecules (Fig.3.2). For 50.5% of molecules logP is predicted with error < 0.25, for 80.2% with error < 0.5 and for 96.5% with error < 1.0. Only for 3.5% of structures logP is predicted with error > 1.0. The statistical parameters listed above rank Molinspiration miLogP as one of the best methods available for logP prediction. MiLogP is used due to its robustness and good prediction quality in the popular ZINC database for virtual screen. The designed compounds have given LogP within limits for the drug like compound characterization. With this parameter we have screened the about 50 molecules, among only 13 molecules have shown within range of Log p in the means of miLogP values.
Fig. 3.2. LogP graph has been constructed between predicted logP and experimental logP for standardization of Molinspiration algorithms for calculated drug molecules.

c) Molecular Polar Surface Area TPSA

It is calculated based on the methodology published by Ertl et al. as a sum of fragment contributions. O and N atoms centered polar fragments are considered. PSA has been shown to be a very good descriptor characterizing drug absorption, including intestinal absorption and bioavailability. It is a very useful parameter for prediction of drug transport properties. Polar surface area is defined as a sum of surfaces of polar atoms (usually oxygen, nitrogen and attached hydrogen atoms) in a molecule. This parameter has been shown to correlate very well with the human intestinal absorption, Caco-2 monolayer permeability, and blood-brain barrier penetration.

The calculation of PSA in a classical way, however, is rather time consuming, because of the necessity to generate a reasonable 3D molecular geometry and determine the surface itself. Additionally, calculations require specialized software to generate the 3D molecular structures and to determine the surface. In today's era of drug development shaped by high-throughput screening and combinatorial chemistry, fast bioavailability screening of virtual libraries consisting of hundreds of thousands, even millions of molecules is required. That is the reason why in our molecular property prediction toolkit so called topological polar surface area - TPSA is implemented. Briefly, the procedure is based on the summation of tabulated surface contributions of polar fragments (atoms
regarding also their environment). These fragment contributions were determined by least squares fitting to the single conformer 3D PSA for 34,810 drugs from the World Drug Index. Topological polar surface area provides results of practically the same quality as the classical 3D PSA, the calculations, however, are two to three orders of magnitude faster.

d) Molecular Volume

Method for calculation of molecule volume developed at Molinspiration is based on group contributions. These have been obtained by fitting sum of fragment contributions to "real" 3D volume for a training set of about twelve thousand, mostly drug-like molecules. 3D molecular geometries for a training set were fully optimized by the semiempirical AM1 method.

Molecular volume determines transport characteristics of molecules, such as intestinal absorption or blood-brain barrier penetration. Volume is therefore often used in QSAR studies to model molecular properties and biological activity. Various methods may be used to calculate molecular volume, including methods requiring generation of 3D molecular geometries, or fragment contribution methods such as Mc Gowan volume approximation. Method for calculation of molecule volume developed at Molinspiration is based on group contributions. These have been obtained by fitting sum of fragment contributions to "real" 3D volume for a training set of about twelve thousand, mostly drug-like molecules. 3D molecular geometries for a training set were fully optimized by the semiempirical AM1 method. Calculated volume is expressed in cubic Angstroms (Å³).

Molinspiration methodology for calculation of molecular volume is very robust and is able to process practically all organic and most organo-metallic molecules. The statistical parameters listed above show that Molinspiration fast 2D-based method for calculation of molecular volume provides identical results with computationally much more demanding 3D-based volume calculation for just a fraction of computing time (Fig. 3.3). In present study, designed lead compound screening is too difficult based on molecular volume as well as molecular weight. All 50 lead molecules are within 500 Da, we thought that another parameter is required to screen these compounds. The graph is showing equal distribution of volume in both 2D and 3D model of chemical compounds runs in a straight line.
3.12 ADME Boxes

The new functionality builds on Pharma Algorithms’ molecular property prediction technology based on dynamically defined molecular fragmentation, adding mechanistic modeling of absorption processes such as different routes of permeability and different rates for multiple ionic forms of a compound. The predictive models have been implemented as automated software applications with a straightforward graphical user interface designed to meet the needs of medicinal chemists who require interactive functionality as well as computational chemists who need to perform virtual screening and high-throughput property filtering of virtual libraries.

**Absorption module** contains a mechanistic predictive model of human intestinal permeability. The predictions take into account the transcellular and paracellular routes of permeability and different rates for different ionized forms of a compound. The module can be used to predict the following absorption-related properties.

In the Absorption module, LogP and pKa values that are used in the predictive algorithm can automatically be calculated by the program from compound structure or entered manually if their reliable experimental values are available. Researchers can also explore the effects of deliberate changes to compound
analogues’ lipophilicity and ionization constants on the intestinal absorption profile. By entering modified LogP or pKa values in the graphical interface of the module, researchers can model the changes in compound absorption and estimate the magnitude of the effects. **Protein Binding module** predicts plasma protein bound fraction and the equilibrium binding constant to blood serum albumin of a compound in blood. The protein binding properties are predicted from automatically calculated physicochemical properties such as lipophilicity, ionization constants, and hydrogen bonding capacity. 

**Volume of Distribution module** contains a predictive model which generates a quantitative estimate of the apparent volume of distribution of a compound. Physicochemical parameters, charge state, lipophilicity and hydrogen bonding capacity are automatically calculated and used as inputs to the predictive model of the volume of distribution.

The **Absolv module** algorithm has been updated to increase the accuracy of predictions. The user interface has also been enhanced: contributions of each atom to the currently selected Abraham parameter are color-mapped onto the structure, with intensity of the color indicating the degree of contribution of each atom or substructure to the selected parameter.

### 3.4.9 Preparation of files for AUTODOCK

The advanced molecular docking program AutoDock 3.0.3, which uses a powerful Lamarckian genetic algorithm (LGA) (Morris, et al., 1998) method for conformational search and docking, was applied for the automated molecular docking simulations. Briefly, the LGA described the relationship between the antagonists and receptors by the translation, orientation, and conformation of the antagonists. These so-called ‘state variables’ were the ligands’ genotype, and the intramolecular energies were the antagonists’ phenotype. The environmental adaptation of the phenotype was reverse transcribed into its genotype and became heritable traits. Each docking cycle or generation, consisted of regimen of fitness evaluation, crossover, mutation, and selection. A Solis and Wets local search was carried out to the energy minimization on a user-specified proportion of the population. The docked structures of the ligands were generated after a
reasonable number of evaluations. The whole docking scheme could be stated as follows. First, the receptor molecules were checked for polar hydrogen and assigned for partial atomic charges, the PDBQS file was created, and the atomic salvation parameters were also assigned for the macromolecules. Meanwhile, all of the torsion angles of the antagonists that would be explored during molecular docking stage were defined. Therefore, it allowed the conformation search for ligands during molecular docking process.

Second, the 3D grid was created by Auto Grid algorithm (Morris, et al., 1998) to evaluate the binding energies between the antagonists and receptors. In this stage, the hHH2R antagonist’s receptor was embedded in the 3D grid and probe atom was placed at each grid point. The affinity and electrostatic potential grid were calculated for varies type of atoms in the ligands. The energetic configuration of a particular ligand was found by trilinear interpolation of affinity values and electrostatic interaction of the eight grid points around each atom of the ligand. Third, a series of the docking parameters were set on. The atom types, generations and run numbers for LGA algorithm were properly assigned according to the requirement of the Amber force field. The number of generations, energy evolutions, and docking runs were set to 370,000, 1,500,000, and 20, respectively. The kind of atomic charges were assigned as Kollman-all-atom for hHH2R receptor and Gasteiger-Marsili (Gasteiger, et al., 2005) for the ligands.

3.4.10 Protein – Lead molecules binding studies using AutoDock Tool

AutoDock 3.0/ADT (Rosenfeld, et al., 2002) was used for the docking interactions of lead molecules on to the hHH2R. Autodock is a suite of automated docking tools, which is designed to predict how the small molecules such as substrates or drug candidates bind to receptor of known 3D structure. It calculates the energy minimum by the use of the simulated annealing technique. In the present thesis AutoDock has been used exclusively along with one of its search method called Lamarckian genetic algorithm (LGA) (Morris, 1998). The macromolecule is rigid and fixed while the ligand is flexible and can both translate and rotate. In order to run AutoDock the pdb file of the protein and ligand will be converted into pdbqs file by assigning partial charges, and the ligand pdbq file has been obtained from
PRODRG2 Server (Schuettelkopf and Van Aalten, 2004). All the atoms in the ligand file were also checked if needed, and then autogrid was set for docking. In the docking matrix the file were saved as grid parameter file (gpf). Into the pdbqs file of the enzyme genetic algorithm parameters and local search parameters were set in such a way that a population size of 150 individuals were chosen. These 150 individuals were calculated at 100 different runs (100 dockings) and saved as docking parameter file (dpf) and the docking program was run. After the completion of the docking the interactions were generated in the form of dock log file (dlg) which shows the interactions of ligand molecules to the protein. The interactions are represented in the form of mean docked energy, lowest docked energy and RMSD. Docking log file (dlg file) shows best interactions among all in the form of histogram.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Energy</th>
<th>Cluster</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>20</th>
<th>25</th>
<th>30</th>
<th>35</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-10.59</td>
<td>44</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>-8.94</td>
<td>66</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>-8.63</td>
<td>95</td>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>-8.42</td>
<td>19</td>
<td></td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>-8.39</td>
<td>33</td>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>-8.24</td>
<td>6</td>
<td></td>
<td>18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>-7.85</td>
<td>48</td>
<td></td>
<td>7</td>
<td>***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>-7.80</td>
<td>7</td>
<td></td>
<td>2</td>
<td>***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>-7.67</td>
<td>32</td>
<td></td>
<td>3</td>
<td>***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>-7.65</td>
<td>4</td>
<td></td>
<td>2</td>
<td>***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>-7.59</td>
<td>9</td>
<td></td>
<td>1</td>
<td>***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>-7.57</td>
<td>39</td>
<td></td>
<td>1</td>
<td>***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>-7.50</td>
<td>6</td>
<td></td>
<td>2</td>
<td>***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>-7.33</td>
<td>34</td>
<td></td>
<td>1</td>
<td>***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>-6.61</td>
<td>13</td>
<td></td>
<td>1</td>
<td>***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Basing on the dlg file the best docking interactions of the ligand were observed by using PMV viewer 1.4.5 (http://autodok.scripps.edu/). Analysis of docking interaction gives better picture of the amino acids involved in the binding of ligand molecules. Based on this we can identify the amino acids involved in the active site.