Chapter 3

Materials and methods

3.1 DATASETS

The 3D dimensional structures of native and mutant influenza targets used in our analysis were taken from the Brookhaven Protein Data Bank (PDB) (Berman et al., 2000). The PDB structures are considered on the basis of experimental evidences and resolution factors.

The drug molecules were collected from PubChem database. The 3D structures were generated by means of SMILES stings. All the water molecules and the hetero atoms were removed before performing MD simulations.

3.2 IDENTIFICATION OF BINDING SITE RESIDUES

It was a challenging task to extrapolate a mechanism of action from the view of 3D structures. Detailed biochemical information about the enzyme can be used to design substrate or transition state analogues, which can then be bound into the enzyme for structure determination. These can reveal binding site locations and identify residues, which are likely to take part in the receptor–ligand interaction. From this, a catalytic mechanism can be proposed. In order to identify the binding residues in the structure of NA and M2, we submitted the protein-drug complex structures into the ligand contact tool (LCT) program (Lopez et al., 2007) This program calculates contacts between the binding residues of Protein and drug by using default parameters.
3.3 FLEXIBILITY OF BINDING RESIDUE BY NORMAL MODE ANALYSIS

A quantitative measure of the atomic motions in proteins can be obtained from the mean square fluctuations of the atoms relative to their average positions. These can be related to the normalized mean square displacements, \(<R^2>\) (Yuan et al., 2005; Ringe and Petsko, 1986). Therefore, \(<R^2>\) analysis is likely to provide newer insights into protein dynamics, flexibility of amino acids and protein stability (Parthasarathy and Murthy, 2000). It is to be noted that protein flexibility is important for protein function and rational drug design (Carlson and McCammon, 2000). Also, flexibility of certain amino acids in protein is useful for various types of interactions. Moreover, flexibility of amino acids in drug binding pocket is considered to be a significant parameter to understand binding efficiency. In fact, the loss of flexibility in the binding site residues results in the decrease in binding free energies of the protein–drug complex (Rajasekaran et al., 2008; Hinkle and Tobacman, 2003). Hence, we computed normalized mean square displacements, \(<R^2>\), with the aid of the Elnemo program (Suhre and Sanejouand, 2004) in order to understand the flexibility of the binding residues both in the native and the mutant types of NA and M2.

3.4 COMPUTATION OF DOCKING ENERGY

To investigate the free energy of binding between the ligand and the target protein, flexible docking was carried out using GAdock docking engine, a genetic algorithm search technique implemented in ArgusLab (http://www.ArgusLab.com) with the following parameters: population size 250, max generations 50,000 and grid resolution of 0.2 Å. The ArgusLab 4.0.1 docking program has been extensively validated with docking accuracy at ~3 Å, for root mean square deviation (RMSD) value between the predicted and the original crystallographic pose. Flexible ligand docking of ArgusLab
is available by describing the ligand as a torsion tree. Groups of bonded atoms that do not have rotatable bonds are nodes, while torsions are the connections between the nodes. Topology of a torsion tree is a determinative factor influencing efficient docking. In the docking calculations, the scoring method A score from the ArgusLab 4.0.1 suite is employed. A score is based on the decomposition of the total protein–ligand-binding free energy, taking into account the following contributions: the van der Waals interaction between the ligand and the protein, the hydrophobic effect, the hydrogen bonding between the ligand and the protein, the deformation effect and the effects of the translational and rotational entropy loss in the binding process, respectively. The A score function, with the parameters read from the AScore.prm file, was used to calculate the binding energies of the resulting docked structures. This file contains the coefficients for each term in the scoring function. A maximum of 150 poses were allowed to be analyzed. To improve the accuracy of docking study, program of energetic analysis of receptor ligand system (PEARLS) (Han et al., 2006) and PatchDock server (Schneidman et al., 2006) was used to calculate binding energy of protein–ligand complexes of both native and mutant. After performing the docking simulations, we employed LIGPLOT (Wallace et al., 1995) for studying hydrogen bond interactions present in the native and mutant complex. And PyMOL (Warren, 2002) was used to calculate hydrogen bond interaction distances between the drug and the enzyme in both the native and the mutant complexes.

### 3.5 MOLECULAR DYNAMICS SIMULATION

The docked complex both native and mutant types of NA and M2 was used as starting point for MD simulation using GROMACS package 4.5.3 (Hess et al., 2008; Spoel et al., 2005) adopting the GROMOS43a1 force-field parameters. The structures were solvated in cubic 0.9 nm, using periodic boundary conditions and the SPC water model (Meagher et al., 2005). PRODRG server (Schuttelkopf and Van Aalten, 2004) was used to generate ligand topology. This server utilizes GROMOS force field for generating topology file and assigning atom types. Counter ions were added to neutralize the total charge of the system. One thousand steps of steepest descent energy minimization were carried out for the protein–ligand complex. After energy minimization, the system was
equilibrated at constant temperature and pressure. The equilibrated structures were then subjected to MD simulations with the integration time step was set to 2 fs. The non-bonded list was generated, using an atom-based cutoff of 8 Å. The long range electrostatic interactions were handled by the particle-mesh Ewald algorithm (Darden et al., 1999). 0.9 nm cutoff was employed to Lennard-Jones interaction. During the simulations, all bond lengths containing hydrogen atoms were constrained utilizing the Lincs algorithm (Lindahl et al., 2001), the trajectory snapshots were stored for structural analysis at every pico-second. RMSD, salt bridge, solvent accessibility, radius of gyration, radial distribution function, root mean square fluctuation (RMSF), principal component analysis (PCA), H bonds and PMF formed between protein and drug were analyzed through GROMACS utilities g_rms, g_saltbr, g_sas, g_gyrate, g_rdf, g_rmsf, g_covar and g_hbond and g_wham respectively.

### 3.6 VIRTUAL SCREENING

Virtual screening (Shoichet, 2004) is the computational analogue of biological screening. The approach has become increasingly popular in the pharmaceutical research for lead identification. The basic goal of the VS is the reduction of the massive virtual chemical space of small organic molecules, to screen against a specific target protein, to a manageable number of the compound that inhibit a highest chance to lead to a drug candidate (Tondi et al., 1999). The PubChem database, Traditional Chinese medicine database (TCMD) and DugBank database was utilized to screen new lead compounds in our study. Several hits were obtained from the database, which were further screened using molecular docking and bioavailability studies.

### 3.7 ADME AND TOXICITY

Molecular properties such as membrane permeability and bioavailability are always associated with some basic molecular descriptors such as logP (partition coefficient), molecular weight (MW), or counts of hydrogen bond acceptors and donors in a molecule (Ertl et al., 2000). These molecular properties were used in formulating
‘rule of five’ (Lipinski et al., 1997) The rule states that most molecules with good membrane permeability have \( MW \leq 500 \), calculated octanol–water partition coefficient, \( \log P \leq 5 \), hydrogen bond donors \( \leq 5 \), acceptors \( \leq 10 \). Therefore, Lipinski’s Rule of Five was used to test the bioavailability characteristics such as adsorption, distribution, metabolism, elimination (ADME) of the lead compounds. In this study, these molecular properties for all the lead compounds were estimated using MOLINSPIRATION program (Buntrock, 2002).

Additional screening was also carried out by restricting the number of rotatable bonds to a maximum of ten (Oprea, 2000). Successful drug discovery requires high-quality lead structures which may need to be more drug-like than commonly accepted (Proudfoot et al., 2002). Toxicity and poor pharmacokinetics should be eliminated in the early stages of drug discovery. Hence, the hits were further screened using drug-likeness, drug score and toxicity characteristics. These physico-chemical properties were therefore calculated for the filtered set of hits using the programs OSIRIS (http://www.organic-chemistry.org/prog/peo/).

The OSIRIS program calculates the drug-likeness based on a list of about 5,300 distinct substructure fragments created by 3,300 traded drugs as well as 15,000 commercially available chemicals yielding a complete list of all available fragments with associated drug-likeness. The drug score combines drug-likeness, \( c\text{LogP}, \log S, MW \) and toxicity risks as a total value which may be used to judge the compound’s overall potential to qualify for a drug.

3.8 PREDICTION OF RAT ORAL TOXICITY (LD50)

Successful drug discovery requires high quality lead structures which may need to be more drug-like than commonly accepted (Proudfoot, 2002). Toxicity and poor pharmacokinetics should be eliminated in the early stages of drug discovery. Hence in our study, rat oral toxicity is predicted using the CORAL software (http://www.insilico.eu/coral). This software is used to predict the endpoint of the
chemical compounds which is not determined experimentally and it has experimentally predetermined Rat oral toxicity (LD50) data of 689 compounds. Initially these LD50 value of compound is converted into \(-pLD50 (\log_{10} [1/LD50])\) (Toropova et al., 2011). Further these compounds are divided into 6 splits with sub-training, calibration, and test sets to predict the Rat oral toxicity for the virtually screened lead compound. Using these six splits were prepared six various models for the pLD50. Using these models one can calculate the endpoint for a compound and estimate the value of the pLD50 together with the dispersion of this endpoint over these six "measurements".