In silico prediction of binding affinity of drugs against P-glycoprotein

UNDER THE SUPERVISION OF

Dr Pradeep Naik

BY-

Akshi Bassi-101502

Priyanka Tripathi-101505

DEPARTMENT OF BIOTECHNOLOGY AND BIOINFORMATICS

JAYPEE UNVERSITY OF INFORMATION TECHNOLOGY,WAKNAGHAT

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY

WAKNAGHAT, HIMACHAL PRADESH

CERTIFICATE

This is to certify that project report entitled "*In silico* prediction of binding affinity of drugs against P-glycoprotein", submitted by **Akshi Bassi(101502) and Priyanka Tripathi(101505)** in partial fulfilment for the award of degree of Bachelor of Technology in Bioinformatics of Jaypee University of Information Technology, Waknaghat, Solan has been carried out under my supervision.

This work has not been submitted partially or fully to any other University or Institute for the award of this or any other degree or diploma.

Dr. Pradeep Naik   Dated: 24th May 2014
Associate Professor, Biotechnology and Bioinformatics

ACKNOWLEDGEMENT

Akshi Bassi
101502
Dated: 26th May 2014

Priyanka Tripathi
101505
Dated: 26th May 2014

ABSTRACT

P-glycoprotein (P-gp) is a plasma membrane efflux transporter belonging to ATP-binding superfamily, responsible for multidrug resistance in tumor cells. Over-expression of P-gp in cancer cells limits the efficacy of many anticancer drugs. A clear understanding of P-gp substrate binding will be advantageous in early drug discovery process. However, substrate poly-specificity of P-gp is a limiting factor in rational drug design. In this investigation, we report a dynamic trans-membrane model of P-gp that accurately identified the substrate binding residues of known anticancer agents. The study included homology modeling of human P-gp based on the crystal structure of Murine P-gp, molecular docking, molecular dynamics analyses and binding free energy calculations. The model was further utilized to speculate substrate propensity of in-house anticancer compounds. The model demonstrated promising results with one anticancer compound Noscapine and its derivatives (Amino-noscapine and Bromo-noscapine). As per our observations, the molecules could be a potential lead for anticancer agents devoid of P-gp mediated multiple drug resistance.

LIST OF TABLES

LIST OF FIGURES

3.4.4(e)	The ligplot for the same to show the interaction between Human p-gp and Bromo-Noscapine.
3.4.4(f)	The complex of Human p-gp (site3) with Noscapine.
3.4.4(g)	The ligplot for the same to show the interaction between Human p-gp and Noscapine.

LIST OF ABBREVATIONS & TERMS USED

1. **MDR:** Multiple Drug Resistance

2. **P-gp:** P-glycoprotein

3. **ATP:** Adenine tri-phosphate

4. **TM :** Trans-membrane

5. **NBD:** Nucleotide-binding Domain

7. **TMD:** Trans-membrane Domain

8. **NBFs**: Nucleotide-binding Folds

9. **ADP:** Adenine di-phosphate

10. **NMR**: Nuclear Magnetic Resonance

11. **BLAST:** Basic Local Alignment Search Tool

12. **FASTA:**FAST-All

13.**PSI-BLAST: Position specific**

14. **PDB:** Protein Data Bank

15. **MD:** Molecular Dynamics

17.**PME:** Particle Mesh Ewald

19.**NPT:** Normal Pressure Temperature

20.**RMSD:** Root Mean Square Deviation

21. **RMSF**: Root Mean Square Fluctuation

22. **XP:** Xtra Precision

21. **MM-GBSA:** Molecular Mechanics Generalized Born Surface Area

22. MM-PBSA: Molecular Mechanics Poisson Boltzmann

DEDICATION

This project is dedicated to my project guide & mentor Dr. Pradeep Kumar Naik for teaching and guiding us in every part of this project.

CONTENTS

Chapter 1: Introduction

P - glycoprotein (permeability glycoprotein) a member of ATP-binding cassette is(ABC) an important protein of the cell membrane that pumps many foreign substances out of cells. It was first isolated by Ling and Juliano in 1976. It is also known as ABC1 or MDR1(Multi drug resistance). More formally, it is an ATP-dependent efflux pump with broad substrate specificity [1]. It exists in animals, fungi and bacteria and likely evolved as a defense mechanism against harmful substances.

P-glycoprotein (P-gp) is one of the major export pump proteins in the body and plays an important role exporting harmful compounds from the cell and transporting lipids from the inner to the outer leaflet of the cell membrane. It is up-regulated in tumour cells where it protects the cell by pumping out anti-cancer drugs, and can give rise to multidrug resistance. In the brain, P-gp has a neuroprotective role preventing harmful compounds from reaching the brain.

The pump belongs to a group of transporters known as the ATP Binding Cassette (ABC) family which needs ATP to pump compounds out of the cell. The protein is thought to alternate between two different conformations when it functions – the substrate binds to an inward facing conformation, and export proceeds in an ATP-dependent way through conformational changes to an outward facing form.

P-gp is approximately 170 kDa protein, consisting of 1280 amino acid residues. This trans-membrane single polypeptide is structurally composed of two homologous parts; each homolog contains six trans-membrane (TM) segments followed by a consensus nucleotide-binding domain (NBD). The two homologous parts are separated by an intracellular linker region of about 60 amino acid Residues.

It comprised of two pseudosymmetric halves, each containing a nucleotide-binding domain (NBD) and a transmembrane domain (TMD). Currently, the "alternating access" model is the most widely accepted paradigm explaining the mechanics of transport by ABC transporters . According to this model, binding of ATP at the NBDs drives conformational changes in the TMDs and switches the transporter's overall conformation from inward-facing to outward-facing (inward/outward refer to the opening of the drug-binding pocket relative to the cell). This ATP-driven switch results in the vectorial transport of substrates out of the cell. The

hydrolysis of ATP and release of Pi/ADP are essential for resetting the transporter back to the inward-facing conformation.
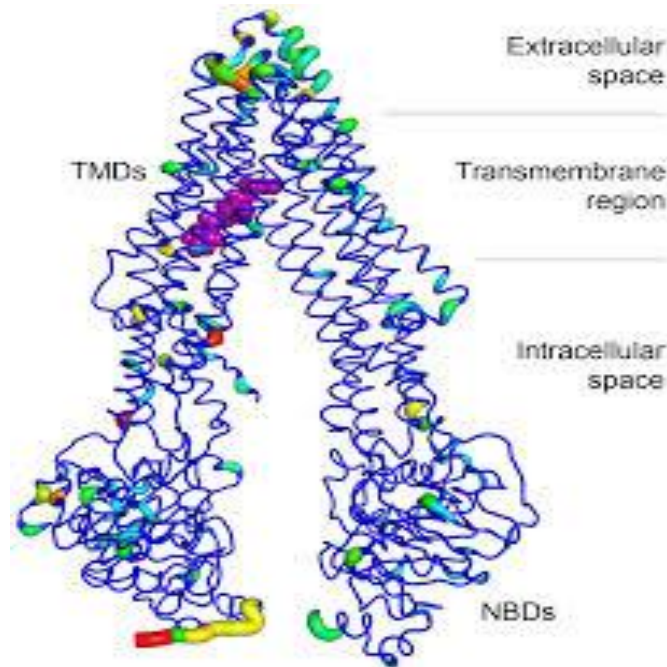
Fig 1.1 Structure of P-glycoprotein

There are two ATP-binding domains of P-gp, located in the cytosol side. ATP-binding domain(s) are also known as nucleotide-binding folds (NBFs). The NBFs are located in the cytoplasm and they transfer the energy to transport the substrates across the membranes.

Each ATP-binding domain contains three regions: Walker A, B, and signature C motifs(fig. 2).Highly conserved Lys residue within the walker A motif of histadine permease[2] is directly involved with the binding of ATP and a highly conserved Asp residue within the walker B motif serves to bind the Mg+ ion. Human P-gp, the MDR1 gene product, requires both Mg(+)-ATP-binding and hydrolysis to function as a drug transporter. It has also been proposed that magnesium may play a role in stabilizing the ATP-binding site. Signature C motifs probably participate to accelerate ATP hydrolysis via chemical transition state interaction and is also suggested to be involved in the transduction of the energy of ATP hydrolysis to the conformational changes in the membrane-integral domains required for translocation of the substrate

Unlike the ATP-binding sites that are restricted to Walker A motifs of ATP-binding domains, many substrate-binding sites have been identified throughout the transmembrane (TM) domain of P-gp. The major drug-binding sites reside in or near TM6 and TM12. In addition to this, TM1, TM4, TM10, and TM11 have drug-binding sites. Amino acids in TM1 are involved in the formation of a binding pocket that plays a role in determining the suitable substrate size for P-gp, whereas Gly residues in TMs 2 and 3 are important in determining substrate specificity. The close proximity of TM2/TM11 and TM5/TM8 indicates that these regions between the two halves must enclose the drug-binding pocket at the cytoplasmic side of P-gp. They may form the "hinges" required for conformational changes during the transport cycle. In addition to the TM domains, intracellular loops and even ATP-binding domains have drug-binding sites
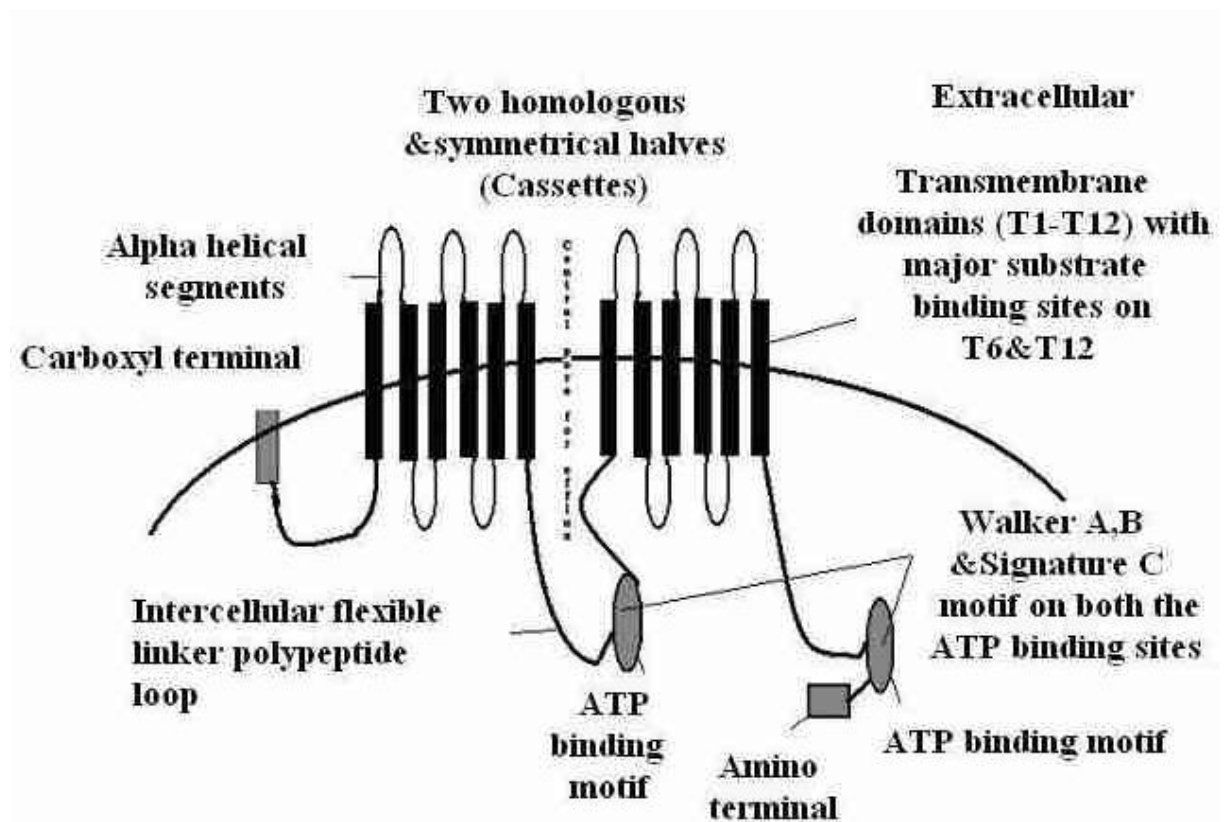
Fig 1.2 Schematic diagram of structur of P-gp

MECHANISM OF ACTION

Substrate enters P-gp either from an opening within the inner leaflet of membrane or from an opening at the cytoplasmic side of the protein. ATP binds at the cytoplasmic side of the protein. Following binding of each, ATP hydrolysis shifts the substrate into a position to be excreted from the cell. Release of the phosphate (from the original ATP molecule) occurs concurrently with substrate excretion. ADP is released, and a new molecule of ATP binds to the secondary ATP-binding site. Hydrolysis and release of ADP and a phosphate molecule resets the protein, so that the process can start again
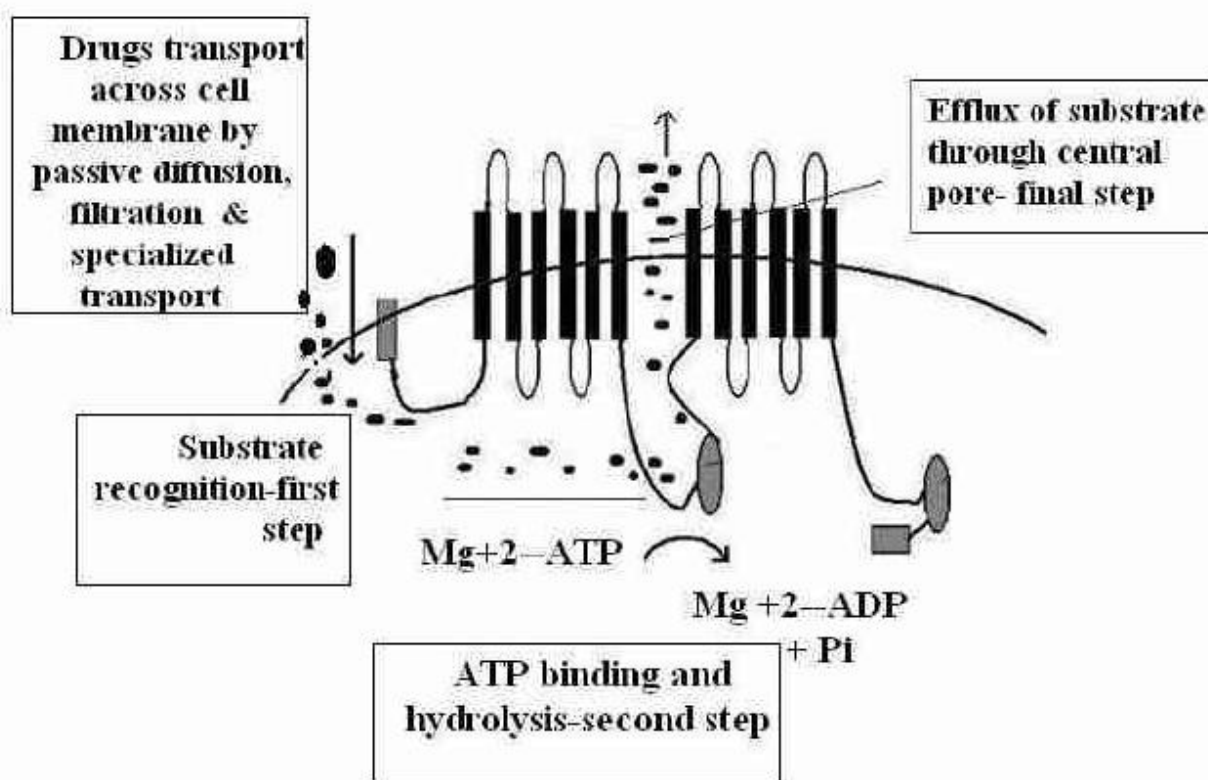
Fig 1.3 Mechanism of drug transportation by P-gp

Fig 1.3 depicts that drugs or substrates can cross into the cell membrane by simple diffusion, filtration, or by specialized transport, and the first step in drug efflux is drug recognition by P-gp followed by ATP-binding and subsequent hydrolysis.

Multi drug resistance developed by p-glycoprotein:

Continuous administration of chemotherapeutic agents results in development of natural and acquired resistance in tumor cells, which is imputable to overexpression of P-gp in the tumour cells. Due to the substrate promiscuity of P-gp, the drug resistance is not only limitedto a single chemotherapeutic agent, instead many of the chemotherapeutic agents are actively effluxed out of the cancer cells resulting in the MDR(multi drug resistance)[3]. for instance, overexpression of P-gp in tumor cells results in reduced intracellular drug concentration of broad spectrum antineoplastic agents including 5,12-anthracyclinediones (e.g.,doxorubicin), vinca alkaloids (e.g., vincristine), podophyllotoxins (e.g., etoposide), and taxanes (e.g., paclitaxel).

Crystal structure of human p-glycoprotein

Due to unavailability of human x-ray crystallography structure of p-glycoprotein ,we have tried to build the structure of human p-gp with the help of p –gp structure of murine (3GGU) and c elegans(4F4C) as template. Out of these crystallographic structures, 3G5U was reported with a resolution of 3.80 A ° and 87 % sequence identity with human P-gp, providing a better template for homology modeling than structure of C. elegans P-gp, with a resolution of 3.4 Å ,which shares less sequence identity (49%) to human P-gp for modelling of human p-gp structure. But there were certain errors in the structure of murine p-gp,these errors were gap between the residues. Those gaps were

gap 1: 184 – 200

gap 2: 217 – 251

gap 3: 251 – 255

gap 4 :265 – 268

Chapter 2: Methodology

2.1 Homology modelling

2.1.1 Introduction

It is also known as comparative modeling of protein, refers to constructing an atomic-resolution model of the target protein from its amino acid sequence and an experimental three-dimensional structure of a related homologous protein (template). Homology modeling relies on the identification of one or more known protein structures likely to resemble the structure of the query sequence, and on the production of an alignment that maps residues in the query sequence to residues in the template sequence. It has been shown that protein structures are more conserved than protein sequences amongst homologues, but sequences falling below a 20% sequence identity can have very different structure.

Evolutionarily related proteins have similar sequences and naturally occurring homologous proteins have similar protein structure. It has been shown that three-dimensional protein structure is evolutionarily more conserved than would be expected on the basis of sequence conservation alone. The sequence alignment and template structure are further used to produce a structural model of the target.

2.1.2 Motivation

The method of homology modeling is based on the observation that protein tertiary structure is better conserved than amino acid sequence Thus; even proteins that have diverged appreciably in sequence but still share detectable similarity will also share common structural properties, particularly the overall fold. Because it is difficult and time-consuming to obtain experimental structures from methods such as X-ray crystallography and NMR for every protein of interest, homology modeling can provide useful structural models for generating hypotheses about a protein's function and directing further experimental work.

2.1.3 Steps in Model Building

The homology modeling procedure can be broken down into four sequential steps: template selection, target-template alignment, model construction, and model assessment. The first two steps are often essentially performed together, as the most common methods of identifying templates rely on the production of sequence alignments; however, these alignments may not

be of sufficient quality because database search techniques give priority to speed over alignment quality. These processes can be performed iteratively to improve the quality of the final model, although quality assessments that are not dependent on the true target structure are still under development.

2.1.4 Template Selection and Sequence Alignment

The critical first step in homology modeling is the identification of the best template structure, if indeed any are available. The simplest method of template identification relies on serial pair wise sequence alignments aided by database search techniques such as FASTA and BLAST. More sensitive methods based on multiple sequence alignment – of which PSI-BLAST is the most common example – iteratively update their position- specific scoring matrix to successively identify more distantly related homologs. . Therefore, choosing the best template from among the candidates is a key step, and can affect the final accuracy of the structure significantly. This choice is guided by several factors, such as the similarity of the query and template sequences, of their functions, and of the predicted query and observed template secondary structure. Perhaps most importantly, the *coverage* of the aligned regions: the fraction of the query sequence structure that can be predicted from the template, and the plausibility of the resulting model.

In our case we were supposed to model human p-glycoprotein. We used MOE for modelling our structure of human p- glycoprotein. The sequence similarity with Murine p-glycoprotein (PDB ID 3G5U) is 87 % which is higher than that of C. Elegans (PDB ID 4F4C) which is 49 % and the resolution is 3.8 Å and 3.4 Å . So Murine p-glycoprotein should be ideally used as a template. But, as 4F4C contained four gap regions , so we could not directly use this as a template.

Initially we used 4F4C as a template to fill the gap regions in 3G5U using homology modeling . And refined the structure of 3G5U. Then this refined structure of 3G5U was utilised to build our model for human p-glycoprotein.

The structure was validated using various validation programs such as an ERRAT score of 62.39 and Ramachandran plot [4] analysis revealed only 2.40% of residues in the disallowed region and 3.0 % of residues in generously allowed regions (Figure 3.1.2 (a) and (b)).

2.2 Structure refinement using MD Simulation

One of the principal tools in the theoretical study of biological molecules is the method of molecular dynamics simulations (MD). This computational method calculates the time dependent behavior of a molecular system. MD simulations have provided detailed information on the fluctuations and conformational changes of proteins and nucleic acids. These methods are now routinely used to investigate the structure, dynamics and thermodynamics of biological molecules and their complexes. They are also used in the determination of structures from x-ray crystallography and from NMR experiments.

MD simulations were run for 1000 picoseconds on the modeled structure to further refine the modeled structure. **Gromacs Pakage** was used with **Gromos96(43a)** force field for all simulations.All simulations were carried in a dodecahedron solvation box with simple point charges of water molecules, using periodic boundary conditions.The Lenard-Jones and electrostatic interaction cut-off were set at 1.0 nm distance. Particle-Mesh-Elwald algorithm was employed to calculate the electrostatic contributions to energy and forces. LINCS algorithm was used to constraint the bond lengths. During the Simulation the modeled protein was first energy minimized using steepest descent algorithm.

To specifically investigate the binding affinity of amino-noscapine (the most potent derivative in our library), bromo-noscapine (the clinical derivative) and noscapine (the lead molecule) onto the Human p-glycoprotein, I have calculated their binding free energy. Towards this end, the complex of Human p-glycoprotein with amino-noscapine, bromo-noscapine and noscapine, obtained after Glide docking was used as an initial conformation for MD simulation. The MD simulation was performed in AMBER 11.0 [5] software suite and the force fields used were AMBER ff99SB [6] for the protein and general AMBER (GAFF) [7] for the ligands. To solvate the system, TIP3P water model was used in an octahedral box with a distance of 15 Å between the wall of the box and the closest atom of the complex. To neutralize the system 31 Na^+ ions were added as counter ions. The molecular systems were first energy minimized with 500 steps of steepest descent energy minimization, followed by 500 steps of conjugate gradient energy minimization so as to remove the bad contacts in the structure; this was done in 3 consecutively rounds. With the force constants of 10 and 2 $kcal^{-1}Å^{-2}$ respectively, positional restraints were applied to the whole system for the first and second round to allow for relaxation of the solvent molecules. In the third round the

whole system was minimized without restraint. Finally a 10 ns MD simulation was carried out following 200 ps of equilibration at 300 K. With a time step of 2 fs a total of 5000 frames were generated. SHAKE algorithm [8] was applied for all the bonds involving hydrogen bonds. The non-bonded cut off distance was 10 Å. The particle mesh Ewald (PME) method [6] was applied to treat long-range electrostatics interactions. The temperature of the system was regulated using the langevin thermostat. All equilibration and subsequent MD stages were carried out in an isothermal isobaric (NPT) ensemble using a Berendsen barometer [8,10,11] with a target pressure of 1 bar, recording trajectories every 2 ps.

2.3 Lead Molecule preparation

Molecular structures of noscapine and its derivatives (Figure 3.3.1(a),(b)and (c)) were built using the molecular builder of Maestro (version 9.2, Schrödinger). All these structures were energy minimized using Macromodel (version 9.9, Schrödinger) and OPLS 2005 force field with PRCG algorithm (1000 steps of minimization and an energy gradient of 0.001). Appropriate bond order for each structure was assigned using Ligprep (version 2.5, Schrödinger). Furthermore, these molecular structures were geometrically optimized using hybrid density functional theory with Becke's three-parameter exchange potential and the Lee-Yang-Parr correlation functional (B3LYP) [12,13] with basis set 3-21G* [14-16]. Jaguar (version 7.7, Schrödinger, LLC) was used for the geometrical optimization of the ligands.

2.4 Glide Docking

Glide is a ligand docking program for predicting protein-ligand binding modes and ranking ligands via high-throughput virtual screening. Glide utilizes two different scoring functions, SP and XP Glide Score, to rank-order compounds. Three modes of sampling ligand conformational and positional degrees of freedom are available to determine the optimal ligand orientation relative to a rigid protein receptor geometry. This unit presents protocols for flexible ligand docking with Glide, optionally including ligand constraints or ligand molecular similarities.

Glide uses a hierarchical series of filters to search for possible locations of the ligand in the active-site region of the receptor. 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.Conformational flexibility is handled in Glide by an extensive conformational search,augmented by a heuristic screen that rapidly eliminates unsuitable conformations, such as conformations that have long-range internal hydrogen bonds.

Molecular docking of noscapinoids (Figure 3.3.1 (a),(b) and (c)) onto was performed using "Extra Precision" (XP) algorithm of Glide docking (version 5.7, Schrödinger) [17,18]. The noscapinoid binding pocket on Human p-gp was defined using a concentric grid box at the centroid of the noscapinoid binding site using the Glide grid-receptor generation program [18]. A bounding box of size 12Å x 12Å x 12Å was defined in order to confine the mass center of the docked ligand. The larger enclosing box of size 12Å x 12Å x 12Å was also chosen so that it occupied all the atoms of the docked poses. All the ligands were then docked into the binding site using Glide XP (extra precision) and evaluated using a Glide XP$_{Score}$ function [14,15]. For the ligand docking stage, a scale factor of 0.4 for van der Waals radii was applied to atoms of protein with absolute partial charges less than or equal to 0.25. Out of the 10,000 poses that were sampled, 1000 were taken through minimization (conjugate gradients) and the 30 structures having the lowest energy conformations were further evaluated for a favorable Glide docking score. The single best conformation for each ligand was considered for further analysis. The residues within 12 Å of the docked ligands were extracted and analyzed for differences in molecular interactions with respect to Human p-glycoprotein.

The human p-glycoprotein was firstly prepared , optimized and finally minimized.For our protein i.e. human p-glycoprotein the sites were identified and 5 sites were generated and they were arranged in the order of their binding scores.The ligand preparation was done then for three ligand molecules. Twelve conformations each were generated for the ligand molecules. Then for each site a grid box was generated. And then first the SP- docking was performed.Using the best scores for each of the ligand molecule from the SP- docking then XP- docking was done.Then the values for glide score and energy values from the project table were exported into an Excel sheet.Then the ligand molecules along with the protein were merged and there structures were saved for each of the five sites generated.

2.5 Calculation of binding free energy

We have used molecular mechanics generalized Born surface area (MM-GBSA) and molecular mechanics Poisson Boltzmann (MM-PBSA) [19, 20] to calculate the binding free energy implemented in AMBER 11.0 [5]. For this calculation a total of 1000 snapshots generated from the last 2 ns of the MD trajectory for each molecular species were considered. The binding free energy was computed as the difference between the energy of the complex with the combination energy of the receptor and ligand for each frame for a total of 1000 frames that were generated. The binding free energy was then calculated for each molecular species as

$$\Delta G_{bind} = G_{complex} - (G_{receptor} + G_{ligand}).$$

The free energy, G for each species was calculated by the following scheme using the MM-PBSA and MM-GBSA methods [16-17].

$$G = E_{gas} + G_{sol} - TS$$

$$E_{gas} = E_{int} + E_{ele} + E_{vdw}$$

$$G_{ele, PB (GB)} = E_{ele} + G_{PB (GB)}$$

$$G_{sol} = G_{sol-np} + G_{PB (GB)}$$

$$G_{sol-np} = \gamma SAS$$

Here, E_{gas} is the gas-phase energy; E_{int} is the internal energy; E_{ele} and E_{vdw} are the coulomb and van der Walls energies, respectively. E_{gas} was calculated using the ff99SB molecular mechanics force field. G_{sol} is the solvation free energy and can be split into polar and non-polar contributions. $G_{PB(GB)}$ is the polar solvation contribution calculated by solving the GB and PB equations. $G_{ele, PB(GB)}$ is the polar interaction contribution. G_{sol-np} is the nonpolar solvation contribution and was estimated via the solvent-accessible surface area (SAS), which was determined using a water probe radius of 1.4 Å. T and S are the temperature and the total solute entropy, respectively.

The Prime MM-GBSA panel can be used to calculate ligand binding energies and ligand strain energies for a set of ligands and a single receptor, using the MM-GBSA technology available with Prime. The ligands and the receptor must be properly prepared beforehand, for example, by using LigPrep and the Protein Preparation Wizard. The ligands must be pre-positioned with respect to the receptor, and the receptor must be prepared as for a Prime refinement calculation. Specify the source of the structures. You can take structures from a

Pose Viewer file (Glide output), or from separated ligand and protein structures. If you choose the latter option, you must ensure that the ligands and the protein are properly prepared and aligned. Choose calculation settings. If you want to evaluate ligand strain energies as well as ligand binding energies, select Calculate ligand strain energies. With this option, an extra calculation on the ligand is performed at its geometry in the complex, and this result is combined with the free ligand calculation to calculate the strain energy. Select a region within a certain distance of the ligand for which the protein   structure will be relaxed in the calculation.The atoms in all residues within the specified distance of the first ligand processed are included in the flexible region. By default, all protein atoms are frozen, and only the ligand structure is relaxed. The larger the flexible region, the longer the calculation takes.Then you may click 'start job' or you can also run Prime MM-GBSA from the command line on Unix.

Chapter 3: RESULTS AND DISCUSSIONS

3.1 Template preparation and homology modeling of Human p-glycoprotein.

We used MOE for modeling our structure of human p- glycoprotein. The sequence similarity with Murine p-glycoprotein (PDB ID 3G5U) is 87 % which is higher than that of C. Elegans (PDB ID 4F4C) which is 49 % and the resolution is 3.8 Å and 3.4 Å . So, Murine p-glycoprotein should be ideally used as a template. But, as 4F4C contained four gap regions , so we could not directly use this as a template.

Initially we used 4F4C as a template to fill the gap regions in 3G5U using homology modeling . And refined the structure of 3G5U. Then this refined structure of 3G5U was utilised to build our model for human p-glycoprotein. And finally the structure of human p-glycoprotein was modelled (Fig 3.1.1).

The structure was validated using various validation programs such as Ramachandran plot [4] analysis revealed only 2.40% of residues in the disallowed region and 3.0 % of residues in generously allowed regions (Figure 3.1.2 (a) and Table 3.1.1(a)) and ERRAT score of 62.39 (Figure 3.1.2 (c)).

Fig 3.1.1 Structure of human p-glycoprotein modeled using Murine p-glycoprotein as template in MOE.
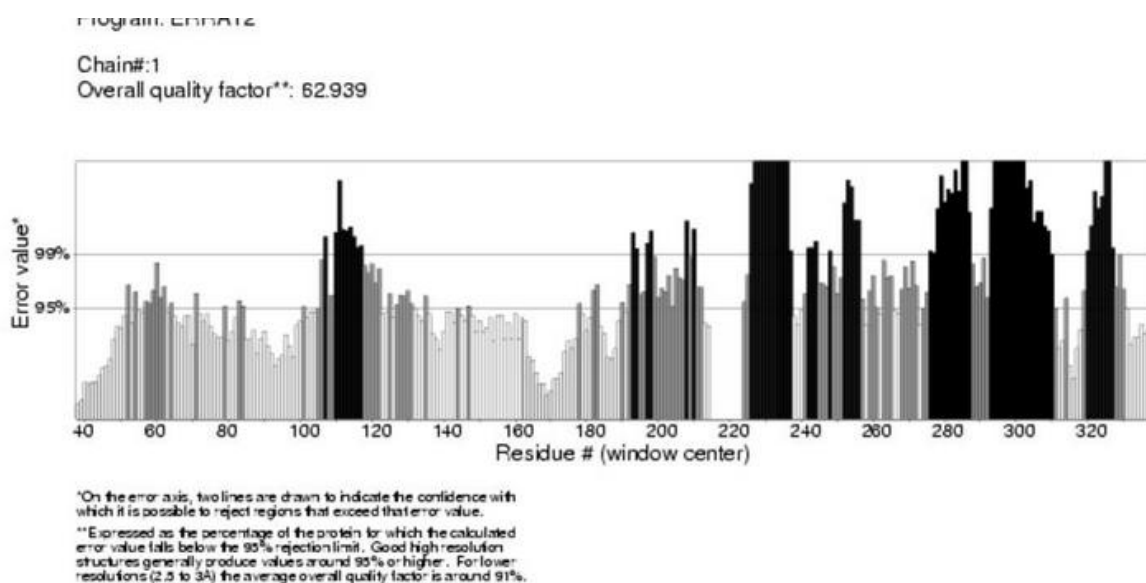
Fig   3.1.2 (a) Ramachandran plot analysis using   PROCHECK for Human p-glycoprotein.

Residues in most favoured regions	73.70%
Residues in additional allowed regions	20.90%
Residues in generously allowed regions	3.00%
Residues in disallowed regions	2.40%

Table 3.1.1 (a) Table containing the percentage of residues present in most favoured , generously allowed, additional allowed and disallowed regions in Ramachandran Plot.

*On the error axis, two lines are drawn to indicate the confidence with which it is possible to reject regions that exceed that error value.

**Expressed as the percentage of the protein for which the calculated error value falls below the 95% rejection limit. Good high resolution structures generally produce values around 95% or higher. For lower resolutions (2.5 to 3A) the average overall quality factor is around 91%.

Fig 3.1.2 (c) Structure validation program Errat shows a score of 62.939 which indicates good quality of the template structure.(a portion of the residue window is represented)

3.2 Structure refinement using MD Simulation

The convergence of the MD trajectories was monitored by plotting root mean square deviation (RMSD) of the backbone C_α atoms with respect to time. Infact the relative fluctuation of the RMSD value is very small after initial equilibration suggesting the stability of the system (Figure 3.2.1(a)). A total of 5000 frames were generated in the MD trajectories, out of which the last 1000 frames were used to generate an average structure (Figure 3.2.1 (b)). Furthermore, the root mean square fluctuations (RMSF) of the residues of the structure were calculated to reveal the flexibility of these residues (Figure 3.2.2).Also the energy for the entire trajectory was calculated which revealed the decrease in energy with the progress of simulation(Figure 3.2.3).

The modelled structure after MD Simulation is shown in (Fig 3.2.4).The modelled structure after performing MD Simulation was validated using various validation programs such as Ramachandran plot [4] analysis revealed only 2.2 % of residues in the disallowed region and 2.1 % of residues in generously allowed regions (Figure 3.2.5 (a) and table 3.2.1 (a)) and ERRAT score of 94.248 (Figure 3.2.5 (b)). We superimposed the human p-gp

structure along with the Murine p-gp in order to see how well are we able to build the human
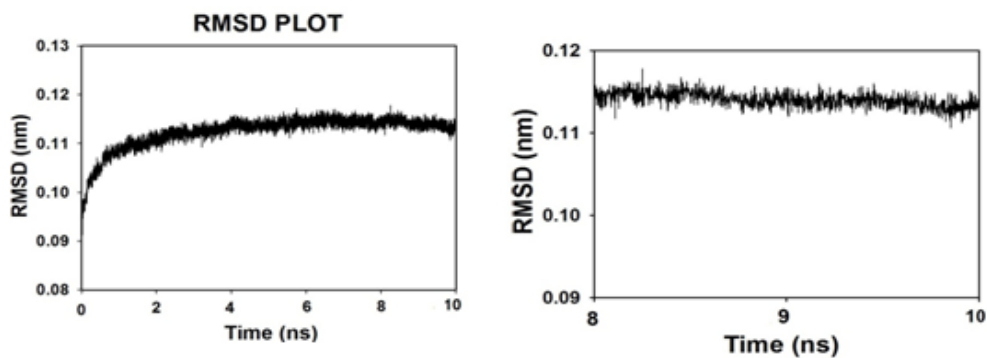p-gp structure (Fig 3.2.6)

Figure RMSD 3.2.1 (a) Time series of the root-mean-square deviations
(RMSD) for the Cα carbon atoms of protein during 10 ns of MD simulation
starting from the initial structure. The relative fluctuation in the rmsd of the
Cα atoms is very small after the initial equilibration (~4 ns) demonstrating
the convergence of the simulation. (b) The relative fluctuation in the rmsd of
the Cα atoms for the last 1000 frames that are used to generate the average
structure of protein for further docking study.
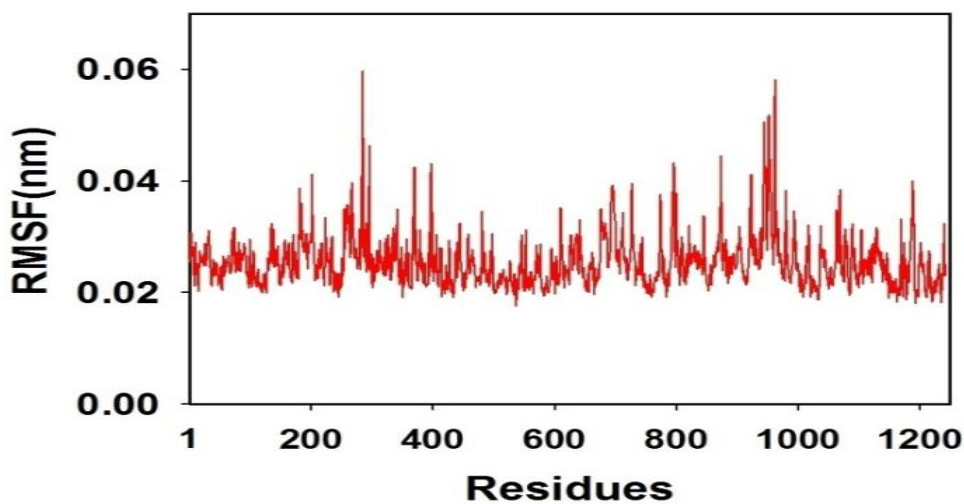
Fig 3.2.2 Root mean square fluctuation (RMSF) of residues (1-1242) of the entire
protein structure. The figure clearly shows a fluctuation of less than 0.06 revealing the
stsbility of the structure.
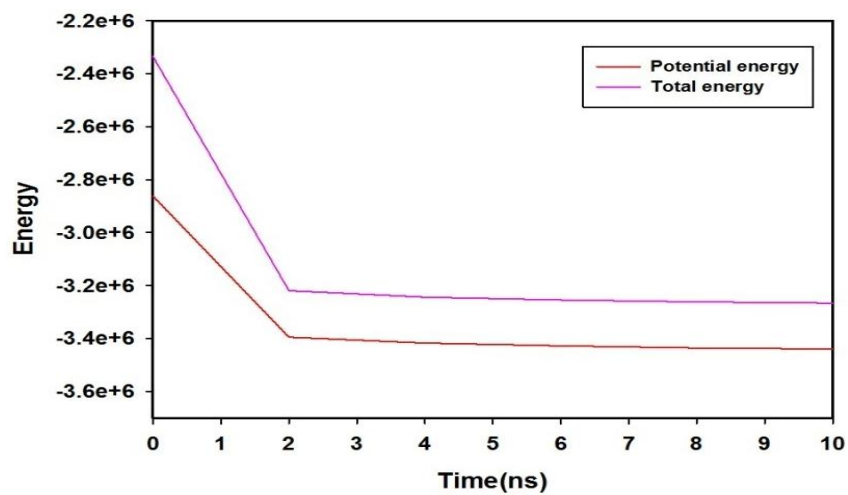
Fig 3.2.3   the energy calculated for the entire simulation showing decrease in energy with the progress of simulation.
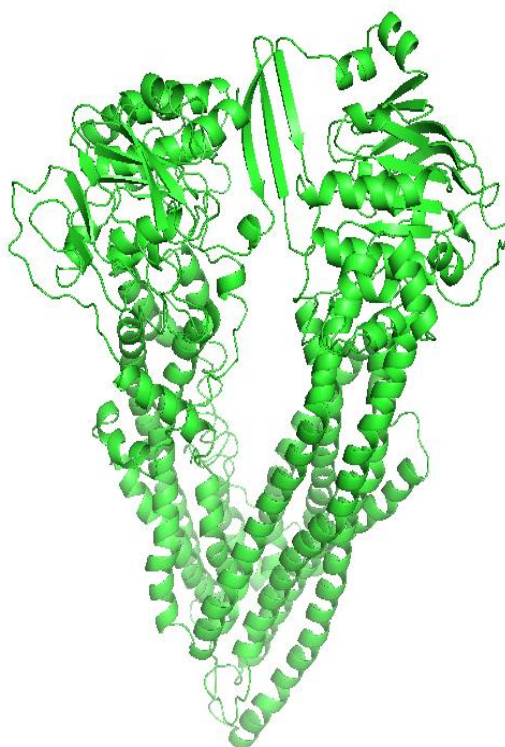

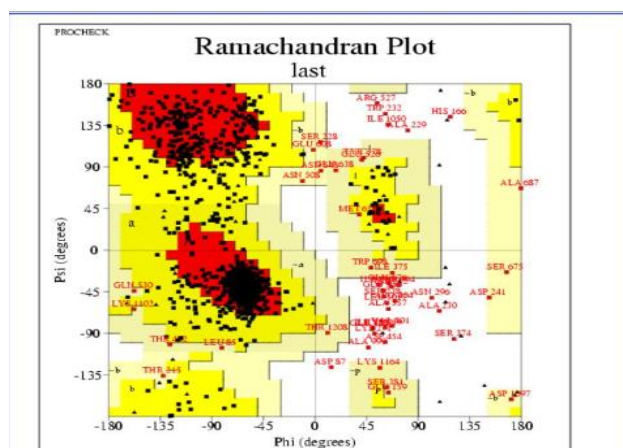Fig 3.2.4 The refined modeled structure of Human p-gp after MD Simulation.

Fig 3.2.5 (a)   Ramachandran plot analysis using PROCHECK for Human p-glycoprotein.

Table 3.2.1(a) Table containing the percentage of residues present in most favoured , generously allowed, additional allowed and disallowed regions in Ramachandran Plot.

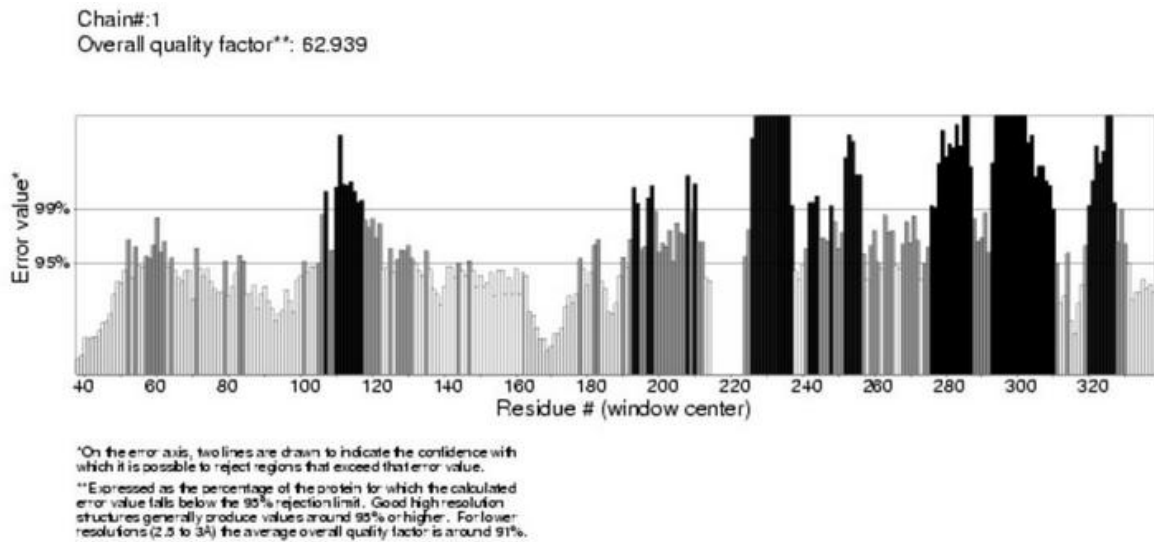|	
Residues in most favoured regions	77.2%
Residues in additional allowed regions	18.5%
Residues in generously allowed regions	2.1%
Residues in disallowed regions	2.2%

Chain#:1
Overall quality factor**: 62.939

'On the error axis, two lines are drawn to indicate the confidence with
which it is possible to reject regions that exceed that error value.
**Expressed as the percentage of the protein for which the calculated
error value falls below the 95% rejection limit. Good high resolution
structures generally produce values around 95% or higher. For lower
resolutions (2.5 to 3A) the average overall quality factor is around 91%.

Fig 3.2.5 (b)   Structure validation program Errat shows a score of 94.28 which indicates good quality of the template structure.(a portion of the residue window is represented).

Table   3.2.1   (b) Comparison between the validation scores for p-glycoprotein among different species.

Structures (PDB Format)	Verify 3-d score (%)	Errat score (%)	RC Plot Core (%)	RC Plot Disallowed (%)
Murine P-gp (3G5U)	0.0	74.254	65.8	0.0
C.Elegans P-gp (4F4C)	57.07	88.335	90.8	0.0
Modelled Human P-gp	42.88	62.939	73.7	2.4
Modelled Human P-gp (MD Simulation)	44.09	94.248	77.2	2.2

Table 3.2.1 (c) The resolution values of Murine p-gp and C.elegans p-gp superimposed with human p-gp.

Structure	RMSD Value
Murine p-gp super imposed with human p-gp	1.2450
C.elegans p-gp super imposed with human p-gp	2.0330

Human p-gp : blue
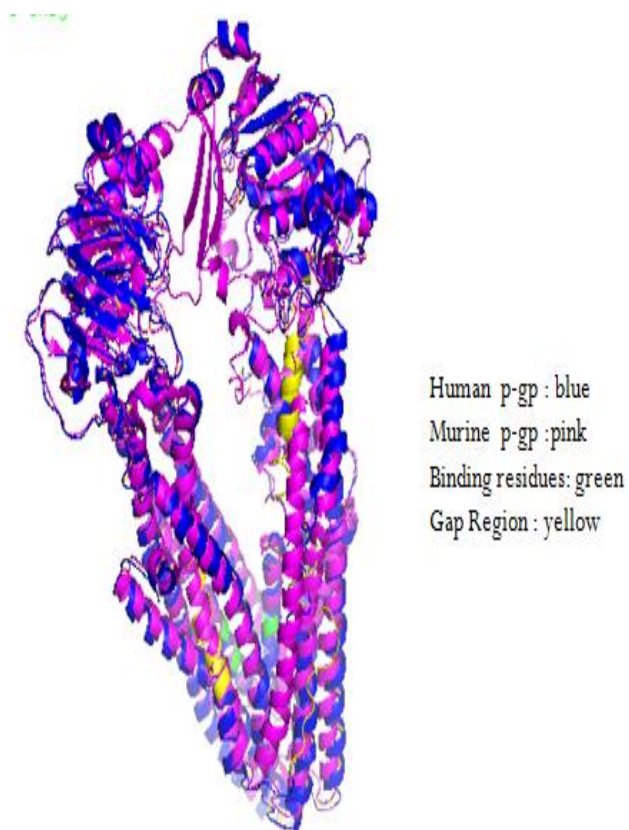Murine p-gp :pink
Binding residues: green
Gap Region : yellow

Fig 3.2.6   Murine P-gp structure superimposed with human P-gp.

3.3 Lead Molecule Preparation

Molecular structures of noscapine and its derivatives (Figure 3.3.1(a),(b)and (c)) were built using the molecular builder of Maestro (version 9.2, Schrödinger). All these structures were energy minimized using Macromodel (version 9.9, Schrödinger) and OPLS 2005 force field with PRCG algorithm (1000 steps of minimization and an energy gradient of 0.001). Appropriate bond order for each structure was assigned using Ligprep (version 2.5, Schrödinger). Furthermore, these molecular structures were geometrically optimized using hybrid density functional theory with Becke's three-parameter exchange potential and the Lee-Yang-Parr correlation functional (B3LYP) [9,10] with basis set 3-21G* (11-13). Jaguar (version 7.7, Schrödinger, LLC) was used for the geometrical optimization of the ligands.
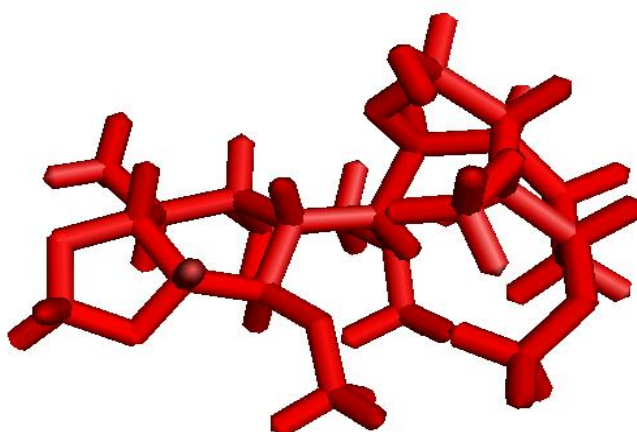
Fig 3.3.1 (a) Optimized structure of Amino Noscapine .
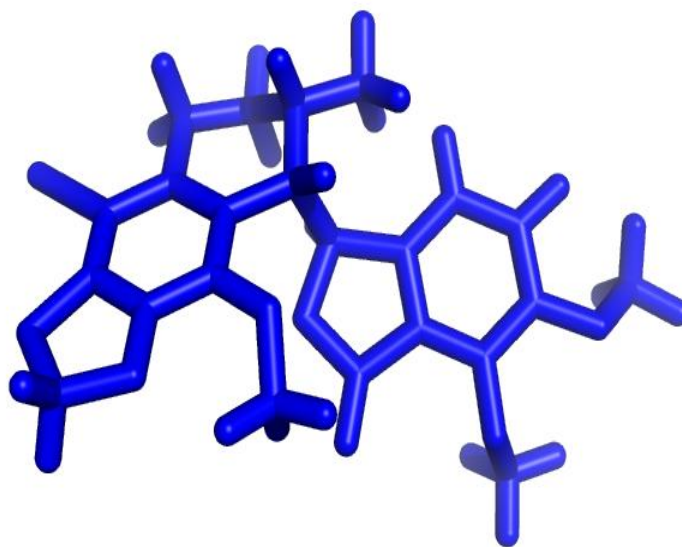
Fig 3.3.1 (b) Optimized structure of Bromo Noscapine.
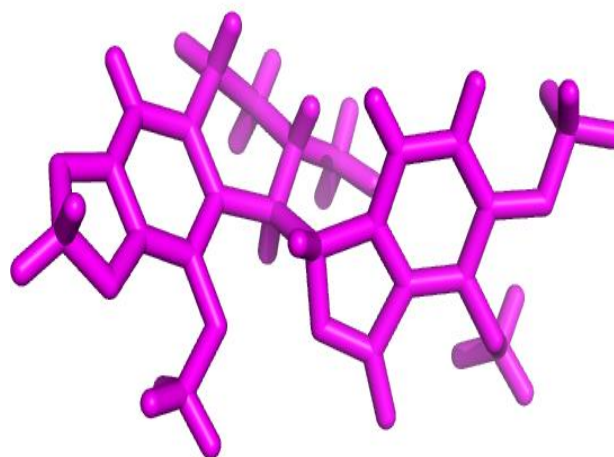
Fig 3.3.1 (c) Optimized structure of Noscapine.

3.4 Molecular Docking of Noscapinoids onto human P-gp

Various sites were generated for the Human p-gp using site map while doing Glide Docking and site 2 was found to be the best site as depicted (Table 3.4.1).

 The electrostatic , van der Walls ,internal and distance energy contribution of residues within the 12 Å diameter of noscapinoid binding site. It clearly shows that binding site residues contribute differently to the electrostatic , van der Walls internal and distance binding energy with amino-noscapine, bromo-noscapine and nosacpine with respect to Human p-gp for the best two sites for the Human p-gp (Fig 3.4.1 (a) ,(b), (c) and (d)), (Fig 3.4.2 (a), (b) , (c) and (d)).

Since site 2 and site 3 for the Human p-gp three ligands have shown interaction with the Human p-gp (Fig 3.4.3 (a),(b),(c),(d),(e),(f) and (g)) and (Fig 3.4.4 (a),(b),(c),(d),(e),(f) and (g)) so we have studied the interaction of these two sites with the ligands using ligplot.

The docking score , glide gscore and other values such as emodel were calculated for all the sites in Glide Docking (Table 3.4.2 (a),(b),(c),(d),(e)).

Table 3.4.1 This represents the site score and other values for the various sites generated for the human p-gp using sitemap.

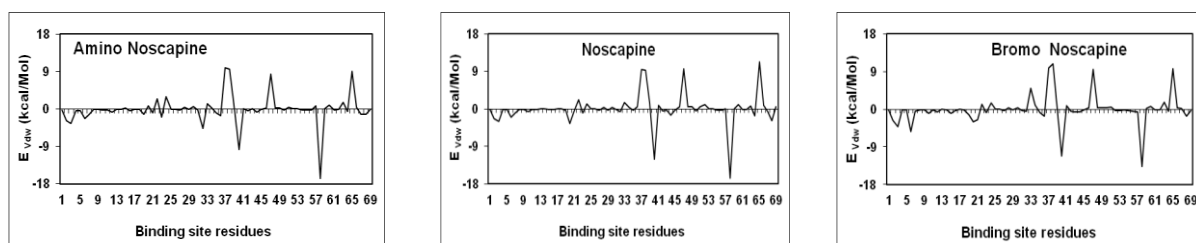| Sites	SiteScore (a.u)	Volume(Å)^3	Exposure (a.u)	Enclosure (a.u)	Contact (a.u)	Phobic (kcal/mol)	Philic (kcal/mol)	Balance (a.u)	don/acc (a.u)
1	0.576	138.23	0.654	0.653	0.772	0.806	0.86	0.936	0.898
2	0.820	160.87	0.5	0.663	0.867	1.053	0.61	1.737	0.710
3	0.660	109.76	0.705	0.620	0.656	0.469	0.53	0.883	1.418
4	0.743	171.16	0.550	0.593	0.741	0.886	0.62	1.439	1.438
5	0.634	122.45	0.629	0.629	0.805	0.611	0.80	0.762	1.261

Fig 3.4.1 (a) Vanderwaals energy plotted for ligand Amino Noscapine, Noscapine and Bromo Noscapine against binding sites residues for site 2 generated for the human p-glycoprotein.
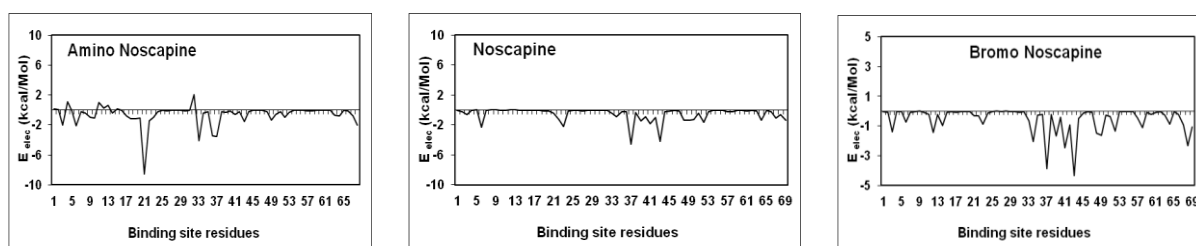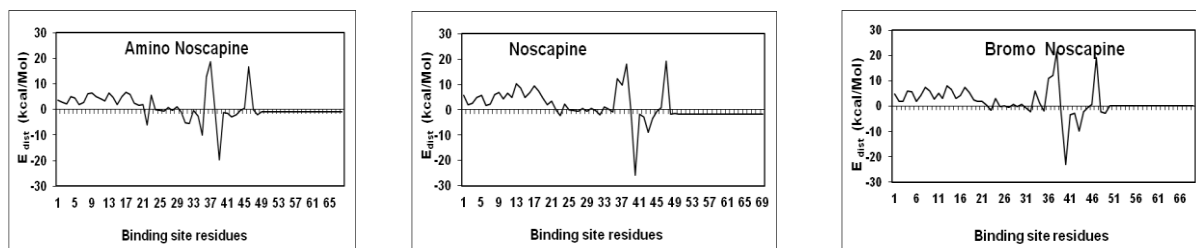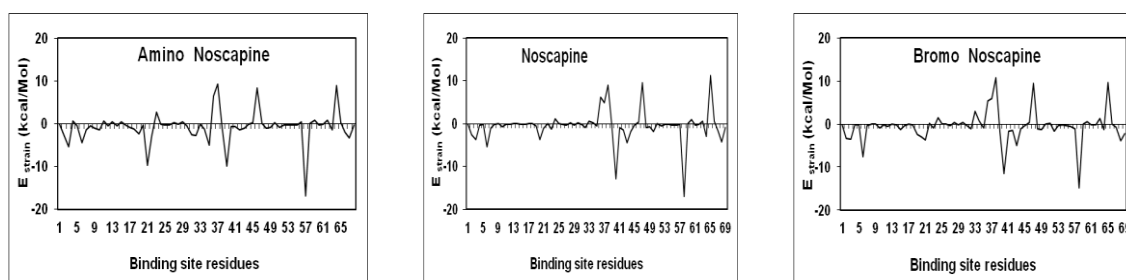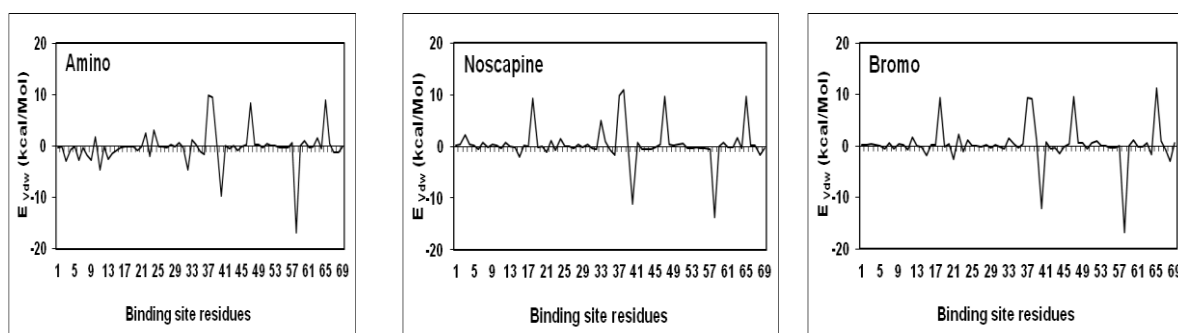

Fig 3.4.1 (b) Electrostatic energy plotted for ligand Amino Noscapine, Noscapine and Bromo Noscapine against binding sites residues for site 2 generated for the human p-glycoprotein.
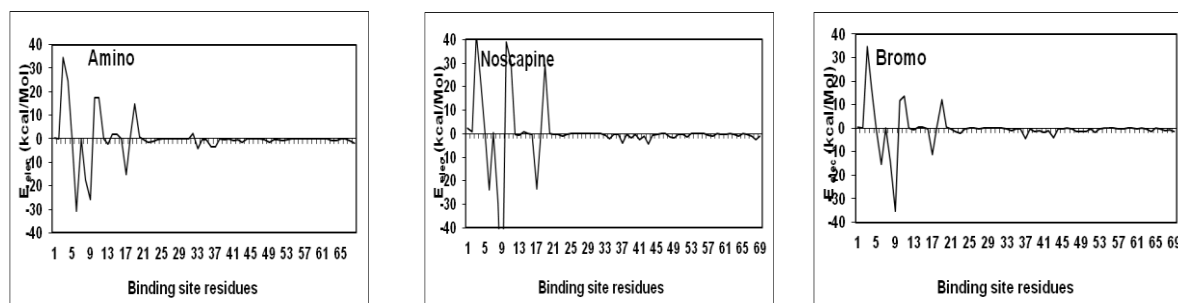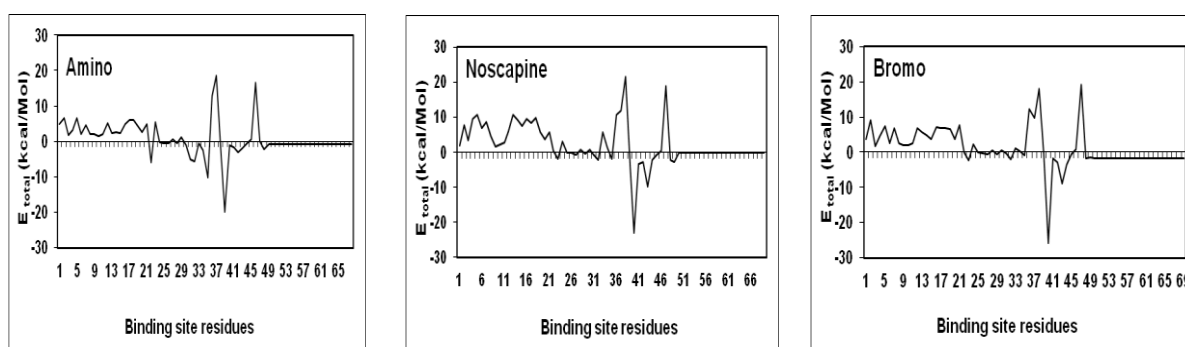
Fig 3.4.1 (c) Distance value plotted for ligand Amino Noscapine, Noscapine and Bromo Noscapine against binding sites residues for site 2 generated for the human p-glycoprotein.

Fig 3.4.1 (d) Internal energy value plotted for ligand Amino Noscapine, Noscapine and Bromo Noscapine against binding sites residues for site 2 generated for the human p-glycoprotein.
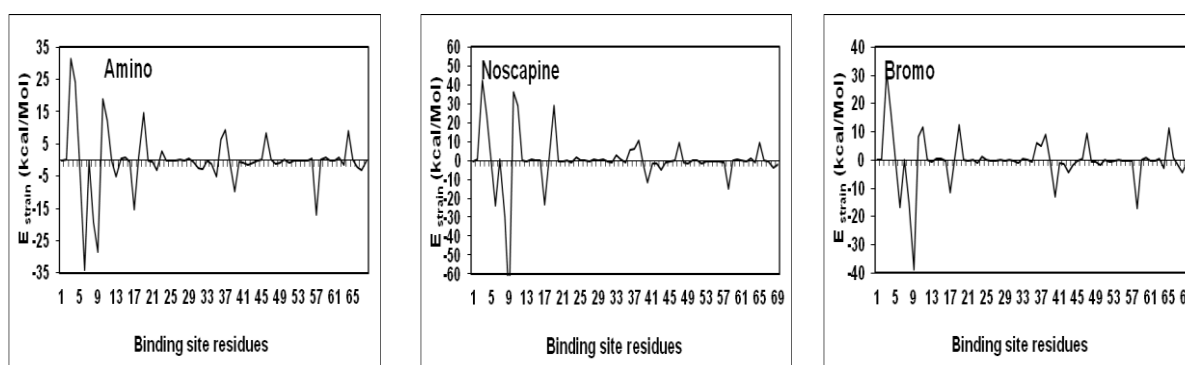
 Fig 3.4.2 (a) Vanderwaals energy plotted for ligand Amino Noscapine, Noscapine and Bromo Noscapine against binding sites residues for site 3 generated for the human p-glycoprotein.

Fig 3.4.2 (b) Electrostatic energy plotted for ligand Amino Noscapine, Noscapine and Bromo Noscapine against binding sites residues for site 3 generated for the human p-glycoprotein.

Fig 3.4.2 (c) Distance value plotted for ligand Amino Noscapine, Noscapine and Bromo Noscapine against binding sites residues for site 3 generated for the human p-glycoprotein.
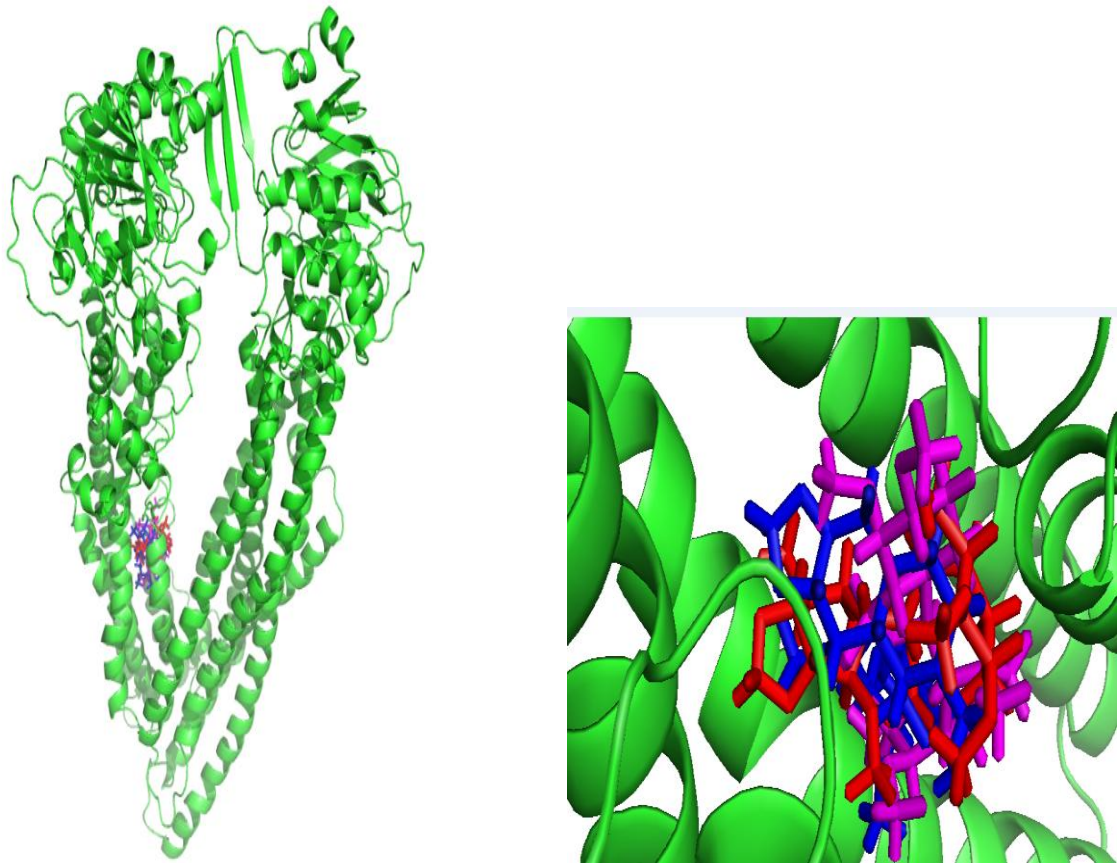

Fig 3.4.2 (d) Internal energy value plotted for ligand Amino Noscapine, Noscapine and Bromo Noscapine against binding sites residues for site 3 generated for the human p-glycoprotein.

Fig 3.4.3 (a) The complex of Human p-gp in green (site2) with Amino-Noscapine in red, Bromo-Noscapine in blue and Noscapine in pink.
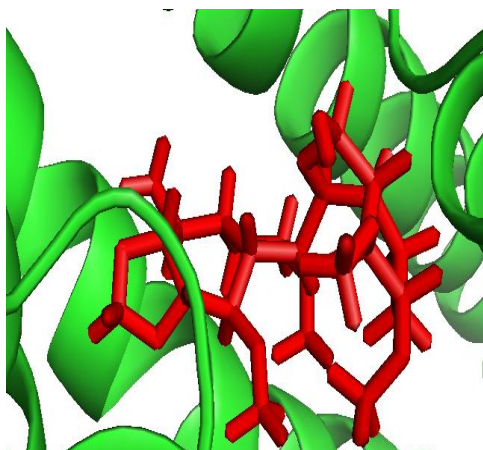
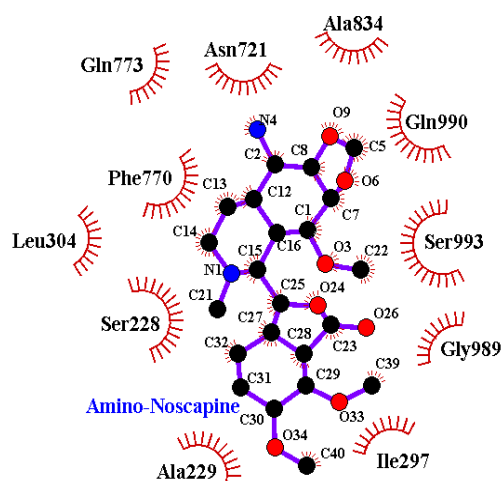Fig 3.4.3 (b) The complex of Human p-gp (site2) with Amino-Noscapine in red.

Fig 3.4.3 (c) the ligplot for the same to show the interaction between Human p-gp and Amino-Noscapine.
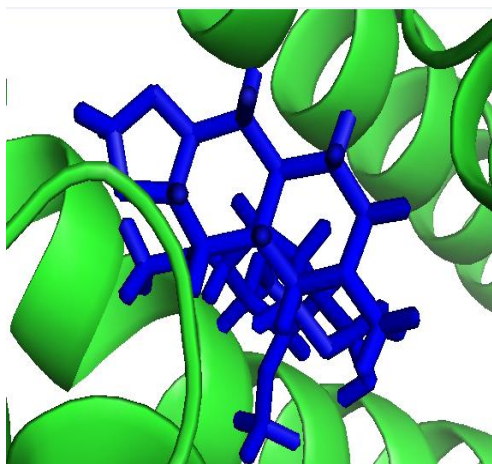
Fig 3.4.3 (d) The complex of Human p-gp (site2) with Bromo-Noscapine in blue.
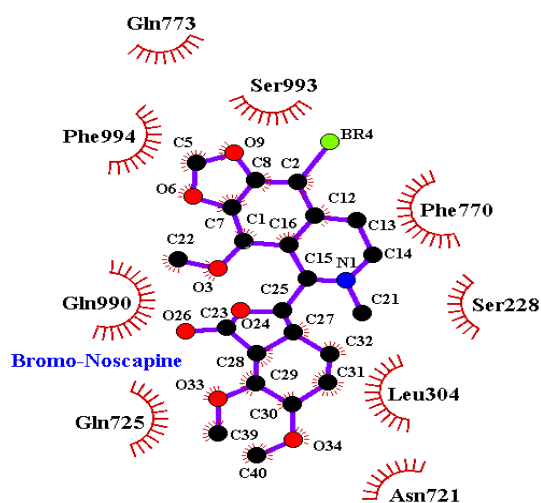

Fig 3.4.3 (e)   The ligplot for the same to show the interaction between Human p-gp and Bromo-Noscapine.
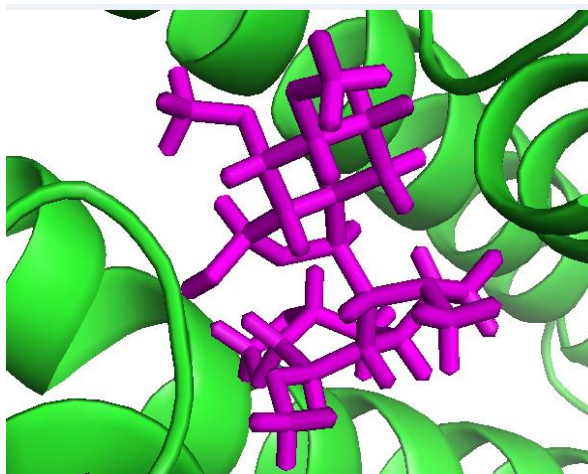
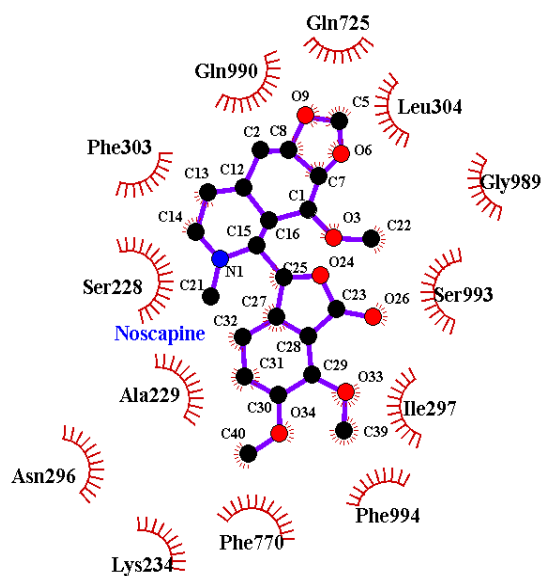Fig 3.4.3 (f) The complex of Human p-gp (site2) with Noscapine in pink.

Fig 3.4.3 (g) The ligplot for the same to show the interaction between Human p-gp and Noscapine.
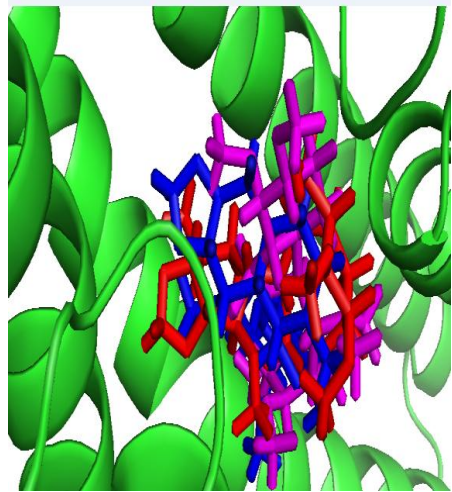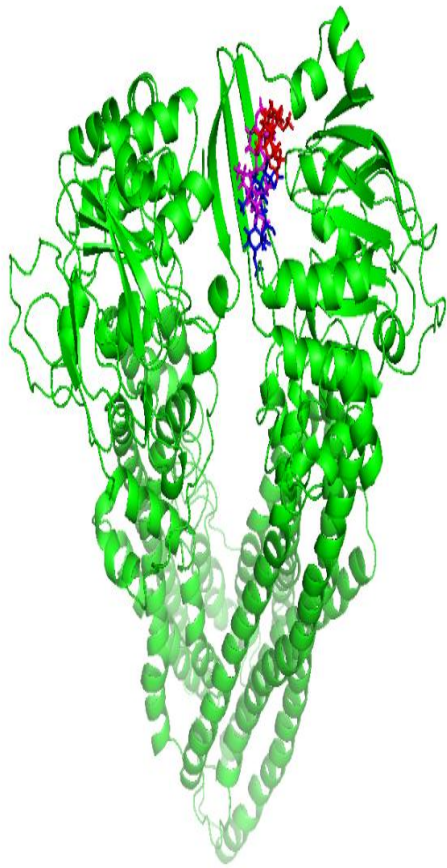
Fig 3.4.4 (a) The complex of Human p-gp (site3) with Amino-Noscapine in red, Bromo-Noscapine in blue and Noscapine in pink.
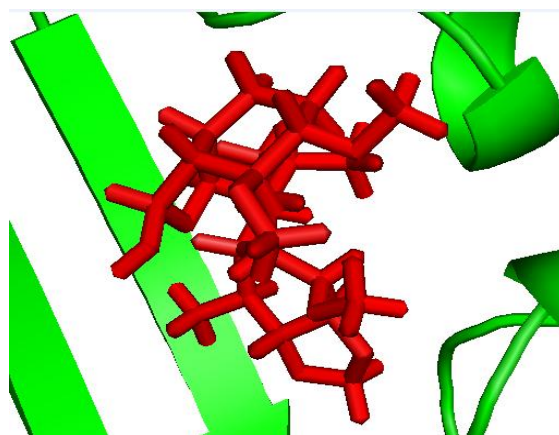
Fig 3.4.4 (b) The complex of Human p-gp (site3) with Amino-Noscapine in red.
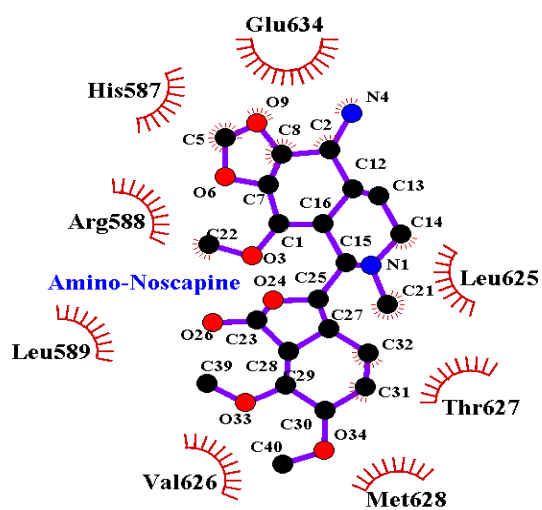
Fig 3.4.4 (c) The ligplot for the same to show the interaction between Human p-gp and Amino-Noscapine.

Fig 3.4.4 (d)   The complex of Human p-gp (site3) with Bromo-Noscapine in blue.

Fig 3.4.4 (e)   The ligplot for the same to show the interaction between Human p-gp and Bromo-Noscapine.
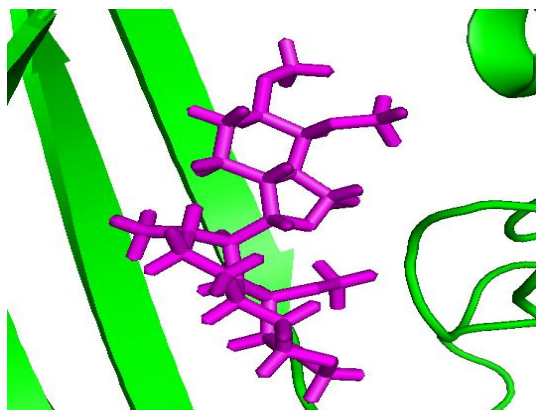
Fig 3.4.3 (f) The complex of Human p-gp (site3) with Noscapine and the ligplot for the same to show the interaction between Human p-gp and Noscapine.
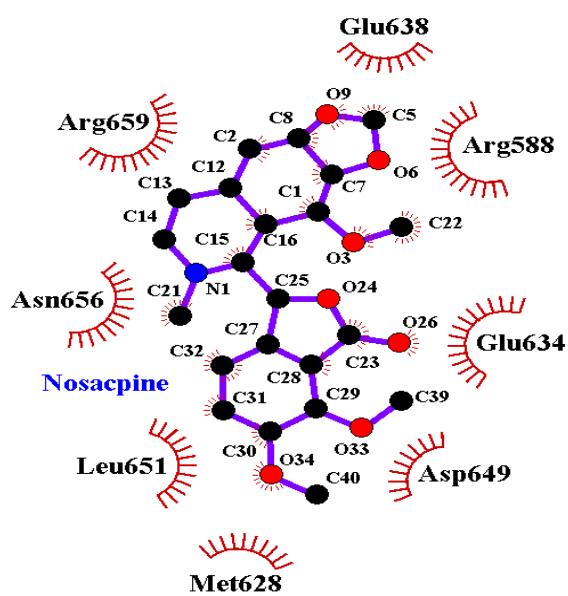
Fig 3.4.3(g) The complex of Human p-gp (site3) with Noscapine and the ligplot for the same to show the interaction between Human p-gp and Noscapine.

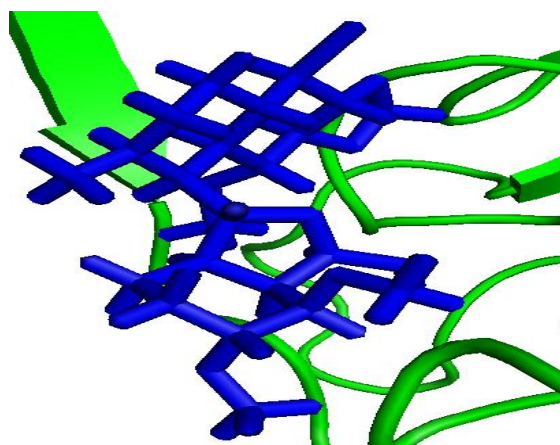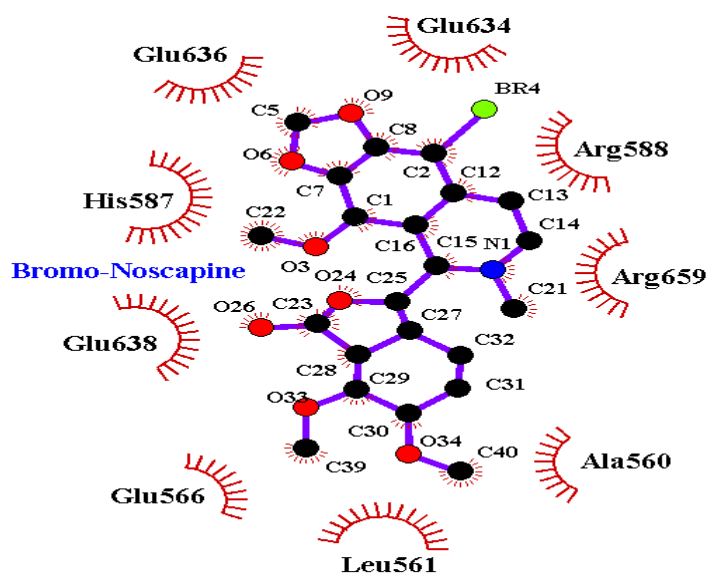Table 3.4.2 (a) The various scores calculated from glide docking and binding energy value for the site2.

S.no	Ligand	glide gscore	Docking score	glide energy	glide emodel
1	Amino	-3.15485	-1.75095	-28.798972	-31.480977
2	Bromo	-2.87285	-1.24685	-18.556206	-14.780011

Table 3.4.2 (b) The various scores calculated from glide docking and binding energy value for the site2.

S.no	Ligand	glide gscore	Docking score	glide energy	glide emodel
1	Amino	-4.92305	-3.59845	-30.2629	-17.350935
2	Noscapine	-5.20436	-3.16576	-33.4375	-44.766834
3	Bromo	-3.91615	-2.29015	-35.6901	-48.314415

Table 3.4.2 (c) The various scores calculated from glide docking and binding energy value for the site5.

S.no	Ligand	glide gscore	Docking score	glide energy	glide emodel
1	Bromo	-4.47681	-4.43741	-26.125277	0
2	Noscapine	-3.91097	-3.89167	-28.073464	0

Table 3.4.2 (d) The various scores calculated from glide docking and binding energy value for the site 3.

S.no	Ligand	glide gscore	Docking score	glide energy	glide emodel
1	Bromo	-0.289218	-3.57629	7.102693	-35.349033
2	Noscapine	-0.171603	-3.221391	3.214607	-31.98798
3	Amino	-0.276374	0.752808	0	-32.075149

Table 3.4.2 (e) The various scores calculated from glide docking and binding energy value for the site 4.

S.no	Ligand	glide gscore	Docking score	glide energy	glide emodel
1	Noscapine	-1.035092	-3.969132	2.184257	-29.499429
2	Bromo	-0.329943	-2.016378	0.87195	-27.384395

Chapter 4: Conclusion

We have been successfully able to model the structure of Human p-glycoprotein using Homology Modelling which is further refined using MD Simulation with good Errat score and very few residues in disallowed region were present in Ramachandran Plot. We have build the ligand molecules successfully and optimize them using Jaguar. We used these optimized ligands to further study their molecular interactions with P-gp using molecular docking. We obtained reasonable binding scores for all the three ligands with the Human p-gp in two binding sites. This study indicates that Noscapinoids are not good substrates for P-gp.

References

1. Juliano RL, Ling V (1976) A surface glycoprotein modulating drug permeability in Chinese hamster ovary cell mutants. Biochim Biophys Acta 455(1):152–162

2. Chan HSL, Haddad G, Thorner PS, DeBoer G, Lin YP, OndrusekN, Yeger H, Ling V (1991) P-glycoprotein expression as a predictorof the outcome of therapy for neuroblastoma. N Engl J Med 325(23):1608–1614

3. Hung LW, Wang I, Nikaido K. Crystal Structure of the ATP-binding subunits of an ABC transporter. Nature 1998;396:703-7

4. Ramachandran, G.N., Ramakrishnan, C. & Sasisekharan, V. *Stereochemistry of polypeptide chain configurations.* J Mol Biol, 7 pp 95–99. 1963.

5. Case, D.A., Walker, R.C. et al. *AMBER 11 University of California SanFrancisco* .2010.

6. Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz, K.M., Jr Ferguson, D.M., Spellmeyer, D.C., Fox, T., Caldwell, J.W. & Kollman, P.A. *A second generation force field for the simulation of proteins nucleic acids and organic molecules.* J Am Chem Soc, 117 pp 5179–5197. 1995.

7. Wang, J.M., Wolf, R.M., Caldwell, J.W., Kollamn, P.A. & Case, D.A. *Development and testing of a general amber force field.* J Comput Chem, 25 pp 1157–1174. 2004.

8. Ryckaert, J.P., Ciccotti, G. & Berendsen, H.J.C. *Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes.* J Comput Phys, 23 pp 327–341. 1977.

9. Daren, T., York, D. & Pedersen, L. *Particle mesh Ewald: An N-log(N) method for Ewald sums in large systems.* Journal of Chemical Physics 98 pp 10089-10092. 1993.

10. Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A. & Simmerling, C. *Comparison of multiple Amber force fields and development of improved protein backbone parameters.* Protein, 65 pp 712-25. 2006.

11. Berendsen, H.J.C., Postma, J. P., van, M., Gunsteren, W.F., DiNola, A. & Haak, J.R. *Molecular dynamics with coupling to an external bath*. J Chem Phys, 81 pp 3684 – 3691. 1984.

12. Becke, A.D. *A new mixing of Hartree-Fock and local density-functional theories.* J Chem Phys, 98 pp 1372-1377. 1993.

13. Lee, C., Yang, W. & Parr, R.G. *Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density.* Phys Rev B, 37 pp 785-789. 1988.

14. Binkley, J.S., Pople, J.A. & Hehre, W.J. *Self-consistent molecular orbital methods 21 Small split-valence basis sets for first-row elements.* J Am Chem Soc, 102 pp 939-947. 1980.

15. Gordon, M.S., Binkley, J.S., Pople, J.A., Pietro, W.J. & Hehre, W.J. *Self-consistent molecular-orbital methods 22 Small split-valence basis sets for second-row elements.* J Am Chem Soc, 104 pp 2797-2803. 1982.

16. Pietro, W.J., Francl, M.M., Hehre, W.J., Defrees, D.J., Pople, J.A. & Binkley, J.S. *Self-consistent molecular orbital methods 24 Supplemented small split-valence basis sets for second-row elements.* J Am Chem Soc, 104 pp 5039-5048. 1982.

17. Friesner, R.A., Banks, J.L., Murphy, R.B., Halgren, T.A., Klicic, J.J., Mainz, D.T., Repasky, M.P., Knoll, E.H., Shelley, M., Perry, J.K., Shaw, D.E., Francis, P. & Shenkin, P.S. *Glide: a new approach for rapid accurate docking and scoring 1 method and assessment of docking accuracy.* J Med Chem, 47 pp 1739–1749. 2004.

18. Halgren, T.A., Murphy, R.B., Friesner, R.A., Beard, H.S., Frye, L.L., Pollard, W.T. & Banks, J.L. *Glide: a new approach for rapid accurate docking and scoring 2 Enrichment factors in database screening.* J Med Chem, 47 pp 1750-1759. 2004.

19. Srinivasan, J., Cheatham, T.E., Cieplak, P., Kollman, P.A. & Case, D.A. *Continuum Solvent Studies of the Stability of DNA RNA and Phosphoramidate−DNA Helicesm.* J Am Chem Soc, 120 pp 9401–9409. 1998.

20. Kollman, P.A., Massova, I., Reyes, C., Kuhn, B., Huo, S., Chong, L., Lee, M., Lee, T., Duan, Y., Wang, W., Donini, O., Cieplak, P., Srinivasan, J., Case, D.A. & Cheatham, T.E. *Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models.* Acc Chem Res, 33 pp 889-97. 2000.