J Comput Aided Mol Des (2013) 27:347–363
DOI 10.1007/s10822-013-9650-x

In silico model for P-glycoprotein substrate prediction: insights from molecular dynamics and in vitro studies

Rameshwar Prajapati · Udghosh Singh · Abhijeet Patil · Kailas S. Khomane · Pravin Bagul · Arvind K. Bansal · Abhay T. Sangamwar

Received: 8 February 2013 / Accepted: 16 April 2013 / Published online: 24 April 2013

Abstract P-glycoprotein (P-gp) is a plasma membrane efflux transporter belonging to ATP-binding cassette 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 *C. elegans* 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 (NSC745689). As per our observations, the molecule could be a potential lead for anticancer agents devoid of P-gp mediated multiple drug resistance. The in silico results were further validated experimentally using Caco-2 cell lines studies, where NSC745689 exhibited poor permeability (P_{app} $1.03 \pm 0.16 \times 10^{-6}$ cm/s) and low efflux ratio of 0.26.

Keywords P-glycoprotein · Substrate binding · Molecular dynamics · Molecular modeling · MDR · Caco-2

Introduction

P-glycoprotein (P-gp), a member of ATP-binding cassette (ABC) superfamily membrane transporters, was first isolated by Ling and Juliano in 1976. P-gp was shown to efficiently "pump" substrates out of tumor cells through an ATP-dependent mechanism in an unidirectional fashion [1]. It is also known as ABCB1 or MDR1. P-gp is an endogenous xenobiotic monitor expressed at cellular level. P-gp is most widely expressed ABC efflux transporter and is known to play a crucial role in multi drug resistance (MDR) [2]. Continuous administration of chemotherapeutic agents results in development of natural and acquired resistance in tumor cells, which is imputable to over expression of P-gp in the tumour cells. Due to the substrate promiscuity of P-gp, the drug resistance is not only limited to a single chemotherapeutic agent, instead many of the chemotherapeutic agents are actively effluxed out of the cancer cells resulting in the MDR. For instance, over-expression 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)

Electronic supplementary material The online version of this article (doi:10.1007/s10822-013-9650-x) contains supplementary material, which is available to authorized users.

R. Prajapati · U. Singh · A. Patil · A. T. Sangamwar (✉)
Department of Pharmacoinformatics, National Institute of Pharmaceutical Education and Research (NIPER), Sector-67, S.A.S. Nagar 160 062, Punjab, India
e-mail: abhays@niper.ac.in

K. S. Khomane · P. Bagul · A. K. Bansal
Department of Pharmaceutics, National Institute of Pharmaceutical Education and Research (NIPER), Sector-67, S.A.S. Nagar, Punjab, India

[3]. Apart from this, P-gp is also expressed in epithelial cells of normal tissues involved in drug disposition including liver, intestine, and kidney, while providing a barrier to sites such as the brain and testes [4]. P-gp is characterized by a broad structural and functional variety of its substrates. P-gp substrate encompasses a prominent component of therapeutics arsenal like anticancer agents, calcium channel blockers, neuroleptics, antiarrythmics, antimalarials, antifungals, and many other drugs. Although most of these drugs share some common physicochemical and structural features, the structural parameters required for a molecule to be recognized and transported by P-gp are still not well known [5]. Identification of potential P-gp substrate liability early in the drug discovery process would be advantageous in increasing efficacy of anticancer drugs.

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 [6–8]. The TM segments of P-gp are believed to be the putative site for substrate recognition and the NBDs are the site for ATP binding/hydrolysis [9–11].

Prior to 2007, two, high-resolution bacterial ABC efflux transporter crystal structures were published (Sav1866, a *Staphylococcus aureus* transporter and MsbA) [12, 13]. In the absence of human P-gp crystallographic structure, the bacterial crystal structure templates were widely used for homology modeling of human P-gp. Several homology models have been published that employed bacterial crystal structure as template. MsbA crystallographic structure was used to model human P-gp in outward-open conformation. It was also used to model apo-open conformation, when co-crystallized structure with nucleotide (AMPPNP) was taken as template, and apo-closed conformations, when co-crystal structure without nucleotide was taken as template. After the retraction of MsbA crystal structure, Sav1866 was the most promising template for model building. Sav1866 template was utilized to model an outward-facing conformation of P-gp as its crystallographic structure was obtained in the presence of ATP in outward facing conformation, with NBDs in close proximity [12]. In 2009, Aller et al. [14] published three, murine P-gp crystallographic structures, which were made available in the Protein Data Bank (PDB), PDBIDs: 3G5U, 3G60 and 3G61. Out of these crystallographic structures, 3G5U was reported with a resolution of 3.80 Å and 87 % sequence identity with human P-gp, providing a better template for homology modeling than Sav1866, with sequence identity of only 17 %. However, certain register errors have been reported in the murine P-gp crystallographic structure [15]. In the recently published *C. elegans* P-gp crystal structure, Jin et al. have pointed out significant register errors in TM3, TM4 and TM5 regions. A region of TM3 (residues 184–200 in mouse P-gp) was shifted by one amino acid, the entire TM4 helix (residues 217–251) was shifted by four amino acids, and TM5 was shifted by three amino acids. Furthermore, residues flanking IH2 (residues 251–255 and 265–268) were built incorrectly as loops rather than α-helices in the mouse model [15]. The corrections in these errors are relevant to the identification of drug-interacting amino acids and an accurate definition of the NBD–TMD interface. Thus a model build from murine P-gp may not be much reliable for substrate binding studies. However, the crystallographic structure of *C. elegans* P-gp, with a resolution of 3.4 Å, shares less sequence identity (49 %) to human P-gp as compared to that of murine P-gp (87 %), but the similarities between their amino-acid sequences and functional properties suggest that the structure of *C. elegans* P-gp would be a reasonable starting point for substrate binding studies and mechanistic understanding of human P-gp [14, 16, 17].

A plethora of molecular dynamics studies have been published till date employing above mentioned crystal structures information to elucidate translocation mechanism and substrate binding of P-gp. Studies reported for isolated NBDs [7, 18–28] gave detailed depictions of ATP hydrolysis mechanisms and conformational changes taking place in response to hydrolysis. On the other hand, full-length ABC transporter molecular dynamics simulations [29–38] gave insights into conformational changes in TMDs and substrate binding site during substrate translocation. Though these studies gave valuable information on P-gp translocation mechanism and substrate binding site none has reported an analysis of substrate tendency of novel anticancer compounds.

In the present study, we have generated a homology model of human P-gp using *C. elegans* P-gp crystal structure as template. The model was further embedded into Palmitoyl oleoyl phosphatidylcholine (POPC) lipid bilayer to simulate the lipophilic environment surrounding protein. The model was utilized to evaluate the potential of an in-house collection of anticancer compounds [39, 40] as putative P-gp substrates which could therefore give rise to P-gp-mediated drug resistance. The compounds were subjected to induced fit docking (IFD) and molecular dynamics simulations. In-depth analyses of binding interactions of these compounds were carried out in the substrate binding site of P-gp during molecular dynamics simulations. The binding affinity of these compounds in substrate binding site was assessed by binding free energy calculation using MMPBSA method. The positive results obtained from molecular dynamic simulations prompted us to validate our results with in vitro studies.

Materials and methods

Homology modeling

The amino acid sequences of human P-gp and *C. elegans* P-gp were retrieved from UniProtKB/TrEMBL database [41], primary accession number **P08183** and **P34712**, respectively. Sequence alignment of *C. elegans* versus human P-gp sequence was performed on ClustalW [42] (Fig S1), using a gap penalty of 10 and a gap extension penalty of 0.05. Subsequently, homology modeling of human P-gp was performed on MODELLER9v8 [43]. The linker region (60 amino acids) was modeled, since its functional role in maintaining structural stability of P-gp had been reported [17, 44, 45]. Secondary structure prediction of linker region was performed on HMMTOP and SAM-T02 servers. MODELLER builds model by satisfying different types of spatial restraints that includes homology-derived restraints, stereochemical restraints obtained from CHARMM22 force field and statistical preferences for dihedral angles and non bonded distances obtained from a representative set of protein structures. MODELLER uses conjugate gradient and simulated annealing molecular dynamics to optimize the positions of heavy atoms. Simulated annealing incorporates constraint derived from the template structure preventing the homology model from unfolding in vacuum. The initially generated models were evaluated by Ramachandran plot and Errat plot using PROCHECK program in SAVES server. The iterative loop refinement and energy minimization were performed to refine the obtained homology model.

Induced fit docking

Preparation of model and ligands for the IFD protocol

The 3D structures of the bioactive compounds (NSC745677, NSC745678 and NSC745689), along with reference compound verapamil, and 5F-203 (benzothiazole compound-Phortress®, presently in clinical trials) were built on Sybyl7.1 software package. Figure 1 shows 2D structure of the compounds. All 3D structures were minimized by Powell method. The minimization was terminated when the energy gradient convergence criteria of 0.05 kcal/mol*Å was reached or 2,000 steps of minimization cycle were exceeded. All compounds were taken in the unionised state (more hydrophobic), considering hydrophobic vacuum cleaner model [46]. In order to gain access to the drug binding pocket, a molecule must be hydrophobic so that it permeates into the inner leaflet of the membrane. Verapamil though exist as quaternary ammonium salt at physiological pH, tends to exist in neutral form at membrane interface [47].

Protein preparation wizard of Schrödinger suite 9.0.211 was used for structural preparations of P-gp homology model. The protein preparation includes addition of hydrogen atoms, optimization of hydrogen bonding networks, and minimization of protein to RMSD 0.30 Å, using Optimized Potential for Liquid Simulations (OPLS) 2005 force field. The selected ligands were prepared in LigPrep and minimized using OPLS 2005 force field.

Induced-fit docking protocol

The IFD protocol as mentioned in Schrödinger 9.0.211, IFD manual was followed [48, 49]. In brief, the IFD utilize two modules of Schrödinger suite, Glide and Prime. The docking program Glide accounts for ligand flexibility and the Prime program accounts for receptor flexibility. The steps followed for IFD were:

1. Constrained minimization of the receptor (Glide protein preparation, refinement only) with an RMSD cutoff of 0.18 Å.
2. Initial Glide docking of each ligand using a softened potential (van der Waals radii scaling). By default, a maximum 20 poses per ligand were retained, and poses that were retained had a Coulomb–vdW score less than 100 and an H-bond score less than 0.05.
3. One round of Prime side chain prediction for each protein–ligand complex. In our study, residues of active site were defined instead considering residues within the cutoff distance from the ligand.
4. Prime minimization of the same set of residues and the ligand for each protein–ligand complex pose. The receptor structure in each pose reflected an induced fit to the ligand structure and conformation.
5. Glide re-docking of each protein–ligand complex structure within a specified energy of the lowest-energy structure (default 30 kcal/mol). The ligand was then rigorously docked, using default Glide settings, into the induced-fit receptor structure.
6. Estimation of the binding energy (IFD Score) for each output pose.

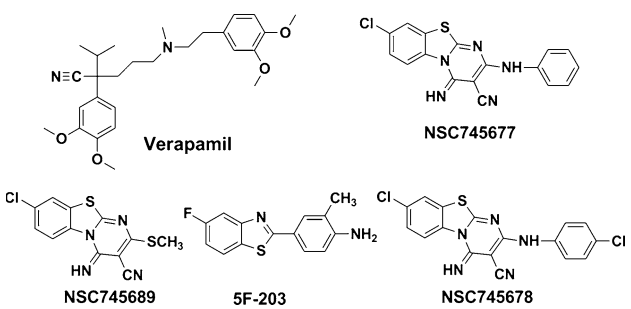

Fig. 1 2D structures of benzthiazole molecules and verapamil

The active site grid was generated by specifying the experimentally reported residues involved in substrate recognition, in TM4, TM5, TM6, TM9, TM10, TM11 and TM12 regions (listed in supplementary information, Table S1) [50–54]. The ligand docking calculations were done on the standard precision mode of Glide software.

Molecular dynamics simulations

Lipid bilayer membrane and protein system was prepared using VMD, GROMACS4.5.5 package and AMBER11. POPC membrane was used as lipid bilayer. Membrane Builder package of VMD [55], was used to prepare POPC membrane and the alignment of membrane and protein was done using *editconf* program of GROMACS4.5.5 [56]. The size of membrane was kept ~20 Å from edge of the P-gp model in all directions. Lipid residues lying within 8 Å from protein backbone were removed using INFLATE-GRO perl script [57]. The lipid–protein system was then solvated by TIP3P water model, using *genbox* program of GROMACS4.5.5 package. Water molecules, lying in the hydrophobic chain region of lipid, were removed. Then, the system was neutralized by adding Cl^- counter ions by replacing water molecules.

The ligand from best docked pose of each protein was extracted. The atomic charges for ligands were derived with AM1-BBC method implemented in Antechamber program of AmberTools. General amber force field (GAFF) was used to generate missing parameters of ligands using Parmchk module of Antechamber. The geometry of lipid molecule was optimized with 6-31G* basis set by Hartree–Fock method using GAUSSIAN03 program [58]. GLYCAM_06_lipid, parameter set was used for parameterization of lipid. AMBER99SB force field was used for parameterization of protein. The complete P-gp-membrane system is shown in figure S2.

Minimization of the protein–lipid system was carried out in two successive steps. Firstly, the protein and ligand were restrained with a force constant of 50 kcal/mol/Å^2, and only solvent and lipid phase were minimized with 500 cycles of each steepest descent and conjugate gradient methods. In the second step, whole system was minimized with 500 cycles of both methods mentioned above. The system was gradually heated to 310 K on NVT ensemble for 50 ps, where the protein–ligand complex was restrained with a force constant of 20 kcal/mol/Å^2. Equilibration was performed on NPT ensemble for 2 ns with a restraint force constant of 20 kcal/mol/Å^2 on the protein–ligand complex.

Production phase, the final phase was performed on NTP ensemble at 310 K temperature and 1 atm pressure for 10 ns. For entire simulation the step size was kept 2 fs. The trajectories were saved after every 1,000 steps (2 ps). During the MD run all bonds containing hydrogen atoms were constrained using SHAKE algorithm. Langevin thermostat and barostat was used for temperature and pressure coupling. The non-bonded cutoff was kept at 8 Å, and long-range electrostatic interactions were treated by the Particle-Mesh Ewald (PME) method. *Ptraj* module of Amber tools, VMD and xmgrace were used for trajectory analysis.

Binding free energy calculations

The trajectories obtained from molecular dynamics were used for binding free energy calculations. The calculations were performed using Molecular Mechanics Poisson Boltzmann Surface Area (MM-PBSA) method, implemented in Amber 11 [59]. The binding free energy for a protein–ligand complex is given as:

$$\Delta G_{binding} = \bar{G}_{complex} - (\bar{G}_{receptor} + \bar{G}_{ligand}) \quad (1)$$

where, $\Delta G_{binding}$ is an estimate of absolute free energy of binding and $\bar{G}$ is the average free energy of complex, receptor and ligand respectively. The average free energy on the other hand is defined as:

$$\bar{G} = -\bar{E}_{MM} + \bar{G}_{solv} - T\bar{S} \quad (2)$$

where, $T\bar{S}$ is the solute entropic contribution to the system at temperature T.

$\bar{E}_{MM}$ is the molecular mechanical energy obtained from bonded and non bonded interactions within the system and can be represented by following equation:

$$\bar{E}_{MM} = \bar{E}_{bond} + \bar{E}_{angle} + \bar{E}_{dihedral} + \bar{E}_{VWD} + \bar{E}_{elec} \quad (3)$$

$\bar{G}_{solv}$ is the average solvation free energy and is equal to the sum of electrostatic and non polar terms and represented as:

$$\bar{G}_{solv} = \bar{G}_{PB} + \bar{G}_{non-polar} \quad (4)$$

$\bar{G}_{PB/GB}$ is the electrostatic contributions to solvation free energy which is evaluated by Finite Differential Poisson Boltzmann (FDPB) in case of PBSA method. $\bar{G}_{non-polar}$ is the hydrophobic non polar contributions to solvation free energy, calculated as:

$$\bar{G}_{non-polar} = \gamma(SASA) + b \quad (5)$$

where SASA is the solvent accessible surface area and, γ and b are constants.

For the MM-PBSA calculations the molecular dynamic trajectories were split into "snapshots". A total 1,000 snapshots were extracted from 5,000 frames obtained in 10 ns of production run. This allows not only to sample the flexibility of the binding site, but also to obtain a more reliable free energy estimate of binding than compared to a single snapshot calculation. The molecular mechanical

energies were calculated by *anal* program of Amber11. The Poisson Boltzmann solvation energies were calculated with the *pbsa* program of Amber11 using a grid spacing of 0.5 Å, dielectric constant for solute and solvent, 1 and 80 respectively [60]. For the calculation of non polar contributions the SASA was calculated with *MSMS* program of Amber11, by taking the solvent probe of 1.4 Å and the values for γ and b were set to 0.0072 kcal/mol/Å^2 and 0.0, respectively.

In vitro studies

Chemicals

NSC745689 was synthesized by Labhsetwar et al. and evaluated at NCI. Dulbecco's modified Eagle's medium (DMEM), Hank's balanced salts solutions (HBSS), fetal bovine serum-heat inactivated (FBS), non-essential amino acids (NEAA), trypsin-ethylenediamine tetra acetic acid (Trypsin–EDTA) solution, penicillin–streptomycin–amphotericin solution, lucifer yellow (LY), dimethyl sulfoxide (DMSO) were obtained from Sigma-Aldrich (St. Louis, MO, USA). 2-[4-(2-hydroxyethyl)-1-piperazinyl] ethanesulphonic acid (HEPES), 2-(*N*-morpholino) ethanesulphonic acid (MES), phosphate-buffered saline (PBS) and 3-[4, 5-dimethylthiazol-2-yl]-2, 5-diphenyl tetrazolium bromide (MTT) were acquired from Himedia Laboratories Pvt. Ltd. (Mumbai, India). Propranolol and furosemide were kindly gifted by IPCA Laboratories (Mumbai, India). Verapamil was a gift from Nicholas Piramal India Ltd. (Mumbai, India). Absolute ethanol was procured from Hong Yang Chemical Co. Ltd (Jiangsu, China). Milli-Q grade water purified by a Milli-Q UV Purification System (Millipore, Bedford, MA, USA) was used.

Cell culture

The Caco-2 cell line was obtained from American Type Culture Collection (ATCC, Manassas, VA,USA) and was grown in DMEM with 4500 mg/L D-Glucose, 110 mg/L Sodium Pyruvate and L-Glutamine and supplemented with 15 % of FBS, 1 % penicillin–streptomycin–amphotericin solution and 1 % NEAA solution. Cells were cultured in T-75 cm^2 tissue culture flasks obtained from Cellstar®, Greiner Bio-One (Germany). The cell cultures were maintained at 37 °C in a carbon dioxide (CO_2) incubator, water jacketed with HEPA Class 100 (Forma Series II, Thermo electron Corporation, USA) and atmospheric of 95 % air/5 % CO_2 and 95 % humidity. The cells became 80–85 % confluent in 4–7 days, after which they were harvested with trypsin–EDTA prior to seeding. The cells were grown on polycarbonate filters of 0.4 µ pore size (Millicell® 24-well Cell culture plate, Millipore, Billerica, MA, USA) at a seeding density of 75,000 cells per well for 20–21 days to achieve a consistent monolayer. The growth media was changed and the transepithelial electrical resistance (TEER) value was measured every alternate day. Cells from passage number 40–50 were used for the experiments.

MTT cytotoxicity assay

The cells were harvested and seeded in 96-well plates at a seeding density of 2×10^4 cells per well. Drug solutions were prepared at a concentration range of 10–200 µM and incubated for 24 h. MTT solution (5 mg/ml in PBS) was added to each well and incubated for 4–5 h (37 °C, 5 % CO_2) to allow MTT to be metabolized. The media was dumped off and formazan (metabolic product of MTT) was re-suspended in 100 µl of DMSO and incubated for 1 h to enable thorough mixing of formazan into the solvent. Optical density was read at 560 nm using ELISA Plate Reader (Bio-Tek Instruments, Inc, USA) and background was subtracted at 670 nm [61]. The percent cell viability was measured from the Eq. (6).

$$Cell\ viability\ (\%) = \frac{Signal - Background}{Blank - Background} \times 100 \qquad (6)$$

Stability study in HBSS

The study was carried out by incubating the NSC745689 solution (100 µM) in HBSS at pH 6.5 ± 0.02 for 3 h at 37 °C and 60 rpm. Samples were withdrawn at specific time intervals and analyzed using HPLC method.

Permeability experiments

The transepithelial electrical resistance (TEER) value was measured with a Millipore ERS voltameter (Millipore, Billerica, MA, USA) in order to evaluate and determine the monolayer integrity. The TEER value was measured from the Eq. (7).

$$TEER = (R_{monolayer} - R_{blank}) \times A \qquad (7)$$

where $R_{monolayer}$ is the resistance of the cell monolayer along with the filter membrane, R_{blank} is the resistance of the filter membrane and A is the surface area of the membrane (0.7 cm^2 in 24-well plates).

A stock solution (2 mM) of NSC745689 in DMSO was prepared and diluted with HBSS to achieve a working concentration of 100 µM. The final concentration of DMSO did not exceed 5 %.

Permeability experiments were performed in a shaker incubator (SW 23, Julabo, Seelbach, Germany) maintained at 37 °C and 60 rpm. Transport studies were carried out

using pH gradient (apical pH 6.5 and basolateral pH 7.4). For apical to basolateral transport studies (A → B), 400 µL of the drug solution in HBSS was added to apical side and 800 µL of the blank transport buffer was added to the basolateral side. A sample volume of 600 µL was withdrawn from the basolateral side at 15, 30, 45, 60, 90, and 120 min and the volume withdrawn was replaced with blank transport buffer each time. The basolateral to apical transport studies (B → A) were also carried out wherein initial drug solution was added to the basolateral side and the concentration in the apical side was measured. The apparent permeability coefficients, P_{app} (cm/s), for both A → B and B → A studies were calculated from the Eq. (8).

$$P_{app} = \frac{dQ}{dt} \Big/ C_0 A \tag{8}$$

where dQ/dt is the cumulative transport rate (µM/min) defined as the slope obtained by linear regression of cumulative transported amount as a function of time (min), A is the surface area of the filters or inserts, C_0 is the initial concentration of the compounds on the donor side (µM). The efflux ratio (ER) was calculated from the Eq. (9).

$$ER = P_{app(BA)} / P_{app(AB)} \tag{9}$$

Permeability studies were also carried out in the presence and absence of verapamil (100 µM), a P-gp efflux inhibitor to investigate the role of P-gp efflux system in the transport mechanism of NSC745689.

Monolayer integrity test

At the end of the experiment, the monolayer integrity test was done by analysing the concentration of LY in the apical and basolateral compartments. An initial stock solution of LY (50 mM) was prepared and diluted to 100 µM working solution. Working solution (400 µL) of LY was added to the apical side of Caco-2 cell monolayer (in the wells in which drug transport study was performed) and 800 µL of HBSS buffer (pH 7.4) was added to the basolateral side. The plate was then kept in a shaker incubator at 37 °C and 60 rpm. After 120 min, 700 and 300 µL of the samples were withdrawn from the basolateral side and apical side respectively. The samples were analysed by fluorescence spectroscopy at an excitation wavelength (λ_{ex}) of 485 nm and emission wavelength (λ_{em}) of 535 nm using a Spectrofluorophotometer (RF-5301-PC, Shimadzu, Japan) [61].

HPLC analysis

The NSC745689 samples from the receiver compartment were analysed using a Shimadzu HPLC system (Shimadzu, Japan) equipped with a model series SPD-M20A photodiode array detector, a gradient elution pump with degassing device DGU-$20A_5$, a cooling autosampler SIL-20AC, a column heater/cooler CTO-10A VP, and a system controller CBM-20A. Separations were performed by isocratic elution using a reverse phase LiChrospher100 C_{18} analytical column (4.6 × 250 mm, 5 µm particles; Merck, KgaA, Darmstadt, Germany) maintained at 35 °C. The column was preceded by a 3 mm × 4 mm C_{18} guard column (LichroCART, Merck). Mobile phases consisted of acetonitrile as the organic phase and 0.03 % trifluroacetic acid (pH 2.2) as the aqueous phase in the ratio of 80:20, respectively. The flow rate and detection wavelength (λ_{max}) were 1 mL/min and 315 nm, respectively. Data was acquired and analyzed using Class VP data acquisition software, version 6.12 SP1.

Result and discussions

Homology modeling

The P-gp structure undergoes several conformational changes during its catalytic cycle, due to this reason it is very important to decide which conformation to model, and carefully sort out the experimental data describing it. In our study we used *C. elegans* P-gp crystal structure as template to model human P-gp. The crystal structure of *C. elegans* P-gp was obtained in the absence of nucleotide and transport substrate, and was determined at a resolution of 3.4 Å. The ATP devoid, open conformation structure has been extensively studied by cysteine scanning mutagenesis and cysteine disulphide cross-linking experiments [50–52, 54, 62–68], which permits to obtain important structural information about the relative positions of different parts of the TM domains, NBDs, and residues on the substrate binding site(s).

C. elegans P-gp shared 49 % sequence identity to human P-gp, the similarities between their amino-acid sequences and functional properties suggested that the structure of *C. elegans* P-gp would be a reasonable starting point for 3D structural studies of human P-gp. The sequence alignment of *C. elegans* and human P-gp obtained from ClustalW showed a good alignment of conserved residues in TM and NBD regions (Fig. S1). ClustalW alignment was used to generate homology model using MODELLER9v8 software, 100 structures were generated and ten best modeled structures were selected on the basis of dope score, and RMSD with respect to crystal structure. The quality of these models was further analyzed with Errat plot and their dihedral angle distributions were analyzed using Ramachandran plot.

Moreover, it has been reported that linker plays a crucial role in maintaining stability of P-gp structure[17, 44, 45]. Thus, the structure was modeled including linker region. The secondary structure prediction from HMMTOP and SAMT02 servers showed helical structure from Thr 627 to Glu 636 and Ser 658 to Arg 666 in linker region. The secondary structure of linker was thus incorporated into the best model. After loop refinement and energy minimization with Amber force field, the Ramachandran plot (Fig. S3) showed that 5 residues were in the disallowed region. The overall quality factor of model was 89 % in Errat plot (Fig. S4). Of the five outliers in Ramachandran plot two (Arg 741 and Ser 671) were in TM region. Thus, the model was further embedded into POPC lipid bilayer and the surrounding lipid was allowed to equilibrate around it. The system was then subjected to 10 ns MD simulation. This improved the side chain conformations and dihedral angle distribution, the Ramachandran plot (Fig. S5) showed two residues in the disallowed region, 86.3 % (992) residues in the most favoured region, 12.1 % (139) residues in allowed region and 1.5 % (17) in generously allowed region. The overall quality factor of homology model was improved to 97.91 % in Errat plot (Fig. S6). Total RMSD between template structure and final homology model was found to be 1.43 Å. The PROSA program uses potential of mean force to locate regions of the protein structure that may contain improper or unsuitable geometries. PROSA derives the PMFs by monitoring the C_α to C_α and the C_β to C_β interactions. The method allows to quickly evaluate the C_α trace (or backbone) of a proposed protein structure, which in turn provides a fast way to appraise the quality of a proposed model. The energies of the protein structure are converted in PROSA into Z-score. The PROSA energy plot (Fig. S7) of template and final homology model showed that the modelled structure was satisfactory. The standard RMS Z-score of a good model is always negative. The RMS Z-scores for model and template were −13.36 and −14.35, respectively. As the scores of both the model and template were negative, it could be inferred that the overall quality of model was acceptable. Figure 2 depicts the model and its trans-membrane arrangement.

The consistency of modeled structure was further validated in accordance with the following biochemical and biophysical data:

Cysteine scanning mutagenesis experiments [62, 63]: Cysteine mutagenesis data permits to obtain important structural information about the relative positions of different parts of the TM domains, NBDs, and drug binding residues on the substrate binding site(s). The TM residues which formed disulphide bond during mutagenesis studies were located at close proximity in the modeled structure. The residues of TM2 (Val 133, Cys 137, Leu 138) and TM11 (Gly 939, Ala 935) were mutated to Cysteine using Pymol, the inter-residue sulphur distance ranged from 3.3 Å to 5.2 Å (Fig. 3a), which was within the prerequisite for disulhide bond formation [69]. Similar interactions were noticed between TM8 (Phe 766, Phe 770, Phe 771, Gly 774, Phe 775) and TM5 (Gly 300, Ile 299, Ile 297, Asn 296, Ile 293) residues (Fig. 3d).

Conserved salt bridge interactions [37, 65, 70]: The structure was also found consistent with highly conserved salt bridge interactions in NBD1-TMD and NBD2-TMD interfaces. Figure 4a shows salt bridge interactions between three highly conserved residues, Arg 905 (IH4), Asp 164 (IH1) and Tyr 444 (NBD1). Similar atomic interactions were also observed in the NBD2–TMD interface, involving residues Arg 262 (IH2), Asp 805 (IH3) and Tyr 1087 (NBD2) (Fig. 4b).

Cysteine disulphide cross-linking experiments [65]: The residues Arg 905 and Ser 909 of IH4 are in close proximity to Leu 443 and Ser 474 of NBD1. These residues when mutated to Cys using Pymol (Fig. 3b), the distance between sulphur atom of Ser 909 and Leu 443 was 4.5 Å, and Ser 474 and Arg 905 was 5.3 Å, indicating the formation of disulphide cross-linking between these residues. Similar interactions were noticed between Ala 266 (IH2) and Phe 1086 (NBD2) (Fig. 3e).

Arginine-scanning mutagenesis experiments [52, 53, 67, 68, 71]: Loo et al. performed arginine-scanning mutagenesis studies to find out substrate binding residues along the translocation pathway. The study suggested that the residues protected by drug substrates from inhibition by thiol-reactive analogues form the drug-binding region. The modelled structure was consistent with location and conformation of drug binding residues (Fig. S8). The structure also showed the inward-facing stabilizing interaction between Phe 335, Tyr 310, Phe 314 and Ile 218. The Phe 335 was located at the apex of the drug transport pathway, making van der Waals interactions with Tyr 310 and Phe 314 in TM5, and Ile 218 in TM4 (Fig. 3c).

Molecular interactions of docked compounds

IFD protocol was validated by docking co-crystallized ligand into *C. elegans* crystal structure (PDB ID: 4F4C). The IFD-generated docked pose of co-crystallized ligand (OSA) showed similar conformation and interactions to that noticed in crystal structure of *C. elegans* P-gp (Fig. 5a). The docked pose of co-crystallized ligand formed H-bonding interactions with Thr 1025 and Thr 1028 similar to that in crystal structure. The atomic RMSD of co-crystallized ligand in the IFD structure with respect to that of crystal structure was 0.68 Å.

Further, IFD studies were performed on benzthiazole molecules and verapamil into the P-gp homology model. The docked pose of co-crystallized ligand was also replicated in case of IFD in homology model (Fig. S9). The in-depth

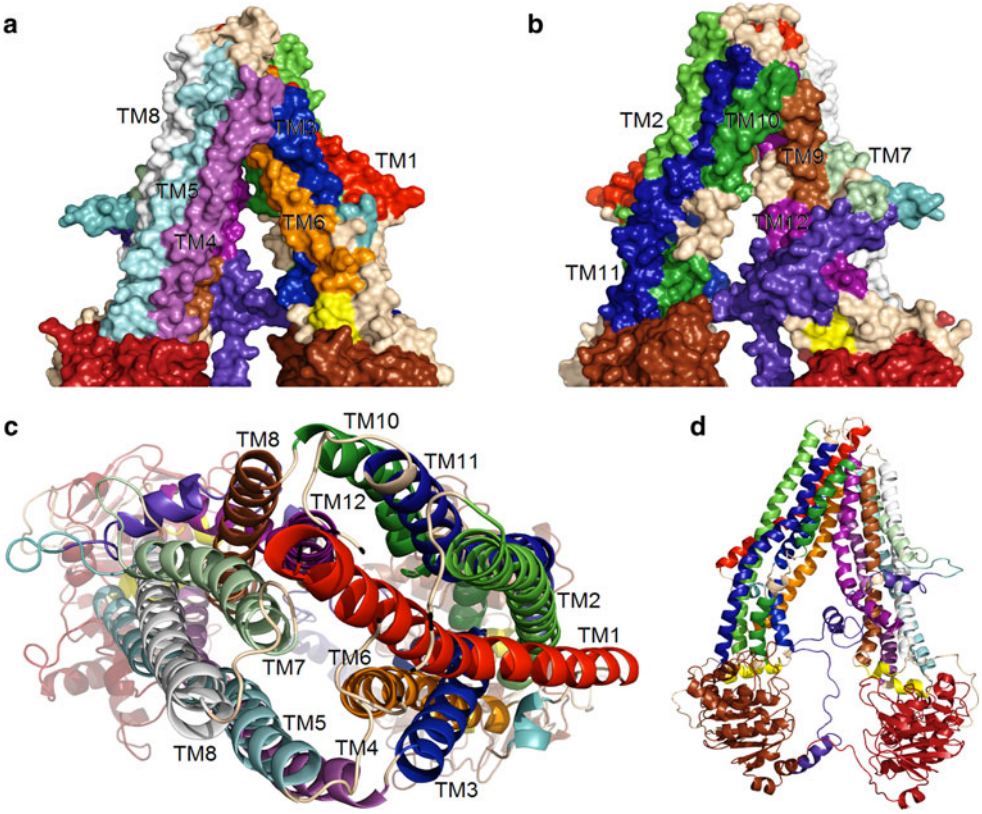

Fig. 2 Human P-gp homology model: **a** front view (surface representation) **b** back view (surface representation) **c** top view (cartoon representation) **d** complete view of homology model (cartoon representation)

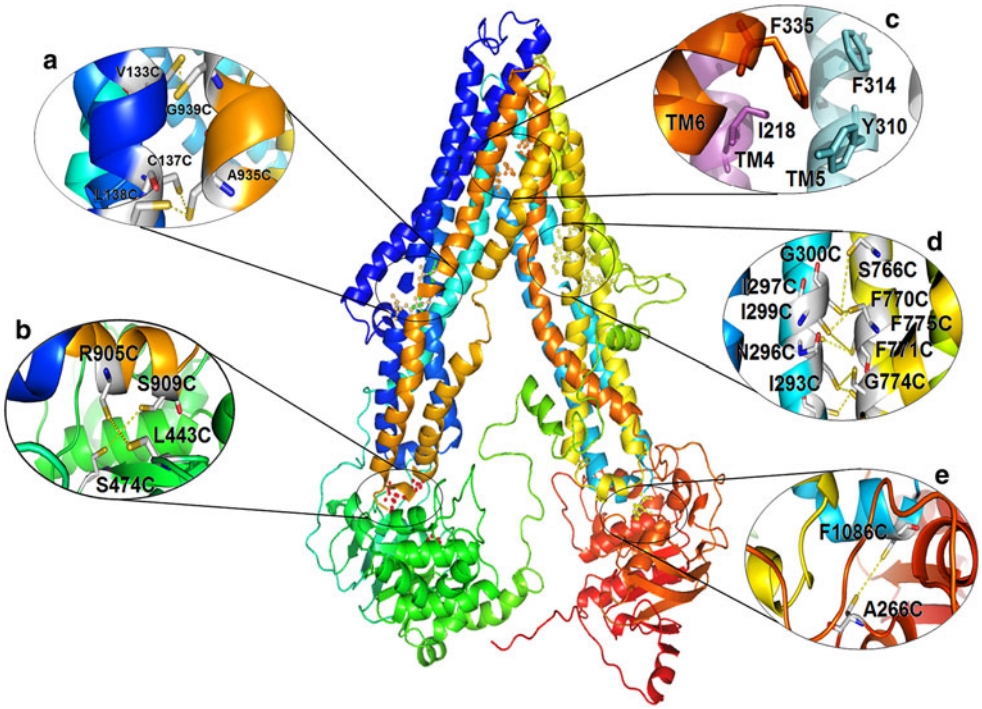

Fig. 3 Validation of human P-gp homology model against biochemical and biophysical data: **a, d** pairs of residue that form disulphide bond when mutated to Cysteine [60, 61], **b, e** NBD-TM residue pairs that cross-linked, when mutated to Cysteine [63] **c** interaction stabilizing inward facing conformation of P-gp, Phe 335 is located at the apex of the drug transport pathway, making van der Waals interactions with Tyr 310 and Phe 314 in TM5, and Ile 218 in TM4

investigation of IFD poses showed that most of the anticancer compounds occupied the similar binding site that has been reported in the experimental studies [52]. A summary of the IFD results for each anticancer compound is given in Table 1. A schematic 2D representation of the best IFD poses for each compound is depicted in figure S10. Table 1 also shows the IFD scores that were used to rank docked poses; IFD score for best pose has been mentioned. In addition, IFD solutions as 3D structures are illustrated in Fig. 5.

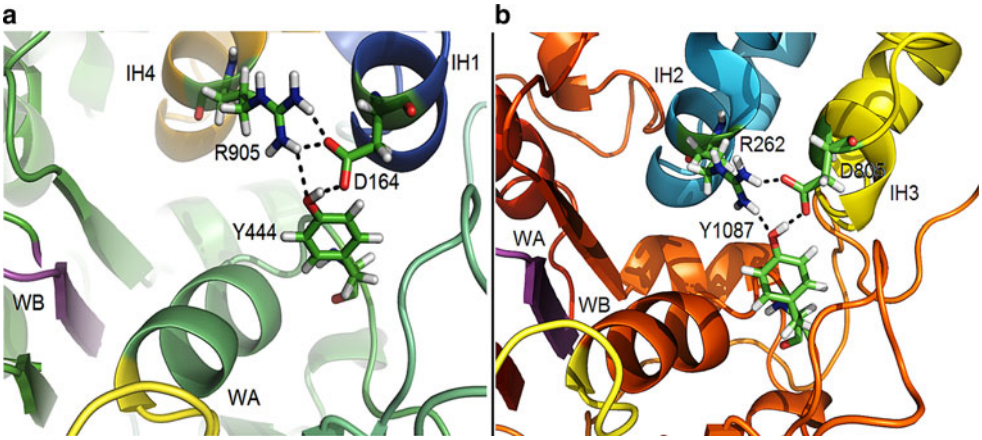

Fig. 4 Salt bridge interactions in NBD1-TMD and NBD2-TMD interface. **a** Interactions between three highly conserved residues, Arg 905 (IH4), Asp 164 (IH1) and Tyr 444 (NBD1) **b** interactions between the NBD2–TMD interface, involving residues Arg 262 (IH2), Asp 805 (IH3) and Tyr 1087 (NBD2)

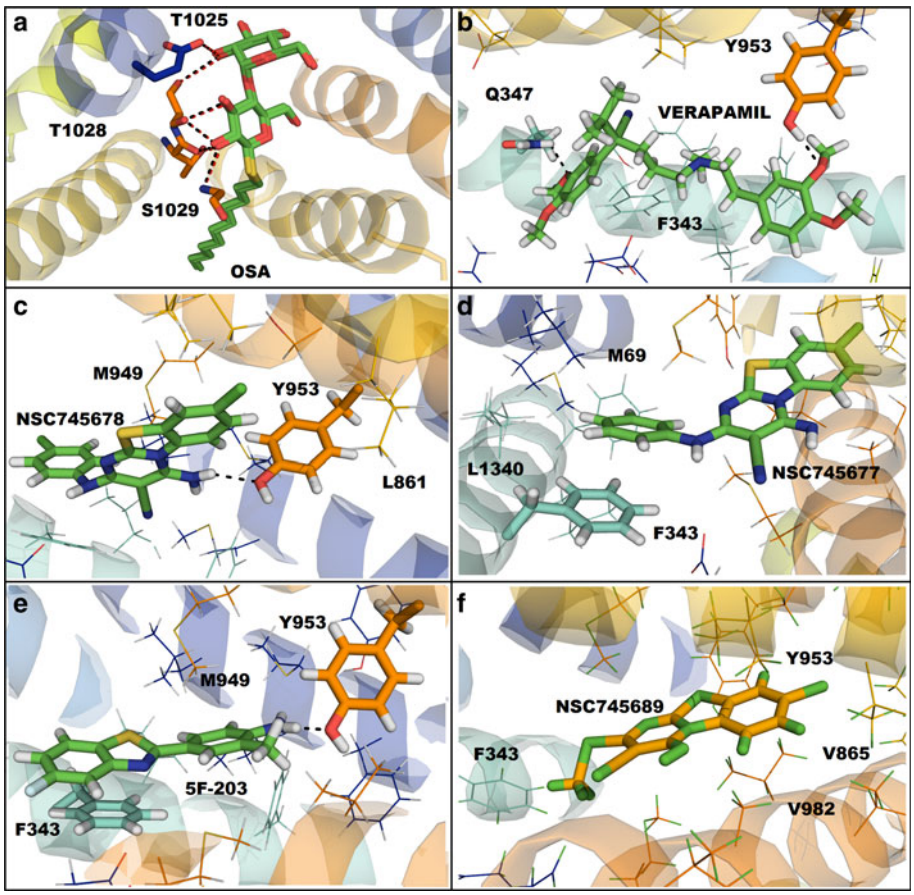

Fig. 5 Molecular interactions of best docked poses obtained from IFD. **a** IFD pose of co-crystallized ligand (*green*) showing similar conformation and interactions to that in crystal structure (*blue*) **b** docking pose of verapamil showing H-bonding interaction with Gln 347 and Tyr 953, and π–π interaction with Phe 343 **c** docking pose NSC745678 forming H-bonding interaction with Tyr 953 and hydrophobic interaction with Val 981 **d** π–π interaction of Phe 343 with phenyl group of NSC745677 **e** 5F-203 showing π–π interaction with Phe 343 and H-bonding interaction with Tyr 953 **f** NSC745689 showing hydrophobic interactions with Phe 343 and Val 982

IFD analysis showed that each compound occupied the binding site in the hydrophobic cavity of P-gp. Hydrophobic interactions were the prominent interactions responsible for ligand binding in most of the cases. The detailed interaction study of each compound is as follows:

Verapamil: Out of 20 IFD docked poses obtained for verapamil, the best pose showed H-bonding interactions with Gln 347 and Tyr 953. Additionally the benzene ring of verapamil formed π-π interaction with Phe 343 and hydrophobic interaction with methyl group of Met 949 (Fig. 5b). Thus, verapamil formed substantial stable binding interactions in the active site and had a highest docking score of −6.150, indicating its P-gp substrate properties.

NSC745678: The NH_2 group of NSC745678 formed H-bonding interaction with Tyr 953 and showed hydrophobic interaction with Val 981 (Fig. 5c). The docking score of NSC745678 was −5.892.

Table 1 Induced fit docking results of test compounds

Compound name	Glide docking score	E-model score	Hydrogen bonding	π–π interaction	IFD score
Verapamil	−6.150	−33.689	Yes	No	−2,652.674
NSC745678	−6.338	−53.253	Yes	No	−2,648.507
NSC745677	−5.892	−50.856	No	Yes	−2,649.763
5F-203	−4.872	−34.597	Yes	Yes	−2,645.709
NSC745689	−3.583	−32.543	No	No	−2,646.774

NSC745677: NSC745677 though had H-bond donor and acceptor groups did not form any H-bonding interaction in all of the analyzed 20 IFD poses. The prominent binding interactions in this case were hydrophobic interactions. The phenyl group of NSC745677 formed π–π interaction with Phe 343 (Fig. 5d). The docking score of NSC745677 was −5.892.

5F-203: The best docked pose of 5F-203 also showed H-bonding and π–π interaction. The benzene ring of 5F-203 formed π–π interaction with Phe 343 and the NH_2 group formed H-bonding interaction with Tyr 953 (Fig. 5e). The docking score of 5F-203 was −4.872. It could be inferred from docking analysis that 5F-203 formed significant interactions in P-gp binding site and it may be a P-gp substrate.

NSC745689: Out of 20 IFD docked poses analyzed for NSC745689, none of them showed substantial interaction at the binding site. Investigation of all poses showed that the conformations of –NH and –CN groups of NSC745689 was such that no H-bonding interactions were formed. Only few hydrophobic interactions were observed in the best docked pose, with Phe 343 and Val 982 (Fig. 5f). The docking score (−3.583) was least in this case. This clearly indicated the unfavorable binding of NSC745689 within the active site of P-gp.

Molecular dynamics analysis

The aim of this study was to evaluate the potential of an in-house collection of anticancer compounds as putative P-gp substrates which could therefore give rise to P-gp-mediated drug resistance. To confer this objective the compounds were subjected to molecular dynamics simulations. The actual biological environment was simulated by embedding P-gp model into lipid bilayer of POPC.

To mimic the biological conditions the lipid bilayer membrane must be in liquid crystalline phase. The orientation of POPC acyl chain, size of head group and the thickness of membrane are important parameters to be accessed to confer liquid crystalline state. In membrane protein simulations there are several methods to analyze the membrane characteristics, some particularly essential are deuterium order parameters of the acyl chains, density of the membrane environment, area per lipid headgroup, bilayer thickness (vertical dimension) and lateral diffusion of the lipids. The modeled system was validated on all these accounts (details discussed in supplementary information, Fig. S11–S14).

The best docked pose of each compound obtained in IFD was subjected to molecular dynamics simulation for in-depth binding interactions study in dynamic state. To analyze the average deviations in the atomic positions and stability of molecular dynamics trajectories, atomic RMSD of each system was plotted (Fig. 6a) with respect to the initial structure. The trajectories stabilized after 3 ns of production run and fluctuated within a range of 2 Å. On analyzing RMSD of loop and linker regions separately (Fig. S15), it was found that the higher RMSD of system (2 Å) was contributed to higher flexibility of loop and linker regions. The RMSD plot of TM (Fig. 6b) and drug binding residues (Fig. 6c) showed that the molecular dynamics trajectories are stable to perform binding energy and ligand binding interactions analyses.

The conformational changes associated with each compound in the active site and their atomic deviations in the active site were analyzed. Figure 6d shows RMSD plot of compounds. Most of the compounds maintained stable interactions within the active site showing deviation within 1 Å, except NSC745689. NSC745689 showed higher average atomic deviation (2 Å), indicating its unfavorable binding in the active site. A relative shift in atomic position was noticed after 6 ns, in the case of NSC745689. On visual inspection of trajectories it was noticed that NSC745689 was displaced from binding site with respect to its initial position (Fig. 7). The distance plot between S-atom of NSC745689 and H-atom of active site residue Val 982 (Fig. 6e), clearly shows the relative shift of NSC745689 from the active site. Verapamil also showed a noticeable higher deviation which was consistent throughout the trajectories, owing to higher flexibility in the structure. The H-bonding interaction of verapamil with Gln 347 and Tyr 953 were maintained throughout the trajectory (Fig. S16a), indicating its favorability at the binding site. A noticeable deviation was also observed in case of NSC745677 initially at 1 ns, owing to conformational change of phenyl ring which maintained hydrophobic

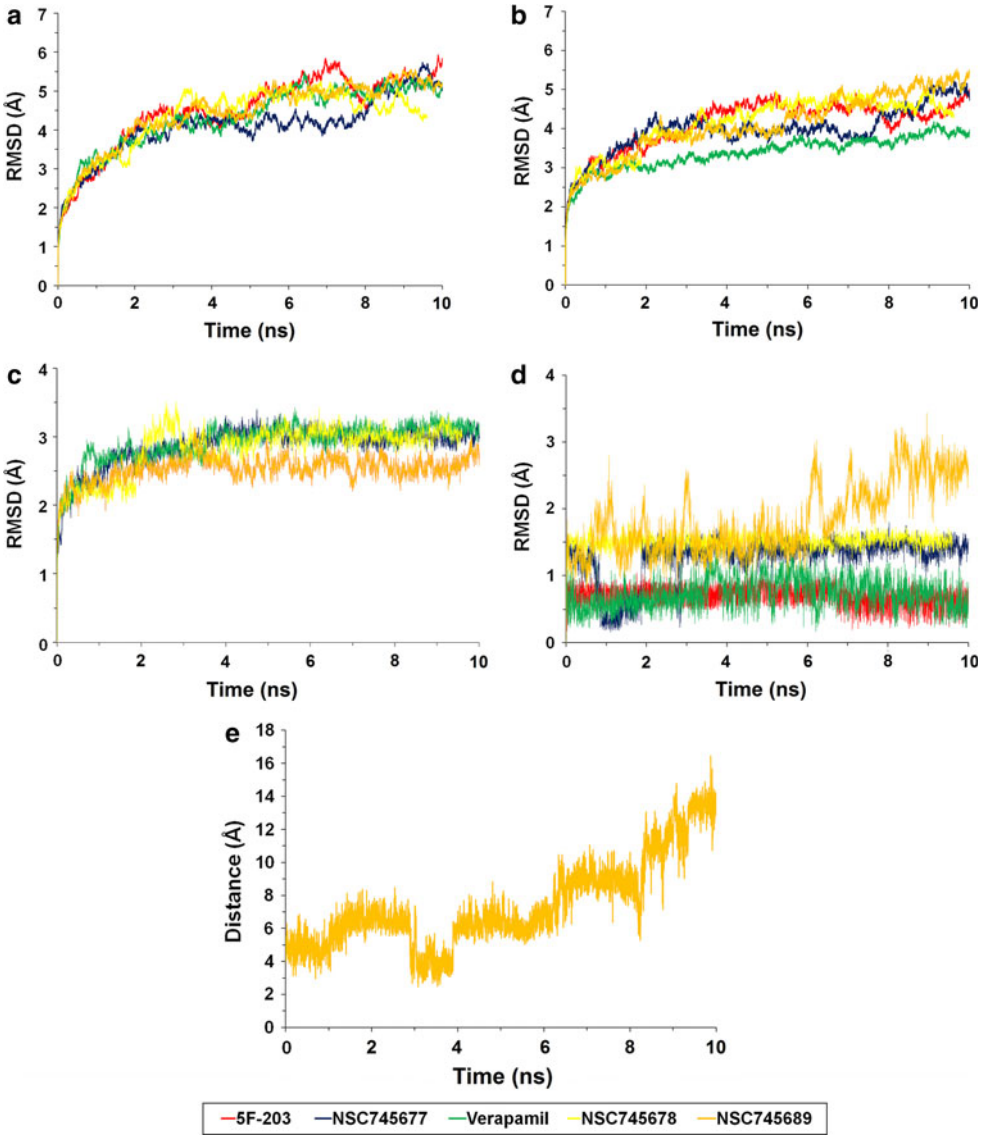

Fig. 6 Plots of RMSD versus time (ns) obtained after 10 ns of production run. **a** Protein–ligand systems **b** TMDs **c** drug binding residues **d** ligands **e** plot of distance (between S-atom of NSC745689 and H-atom of active site residue Val 982) versus time (ns) obtained after 10 ns of production run; the color coding of the graphs is: 5F-203 (*red*), NSC745677 (*blue*), verapamil (*green*), NSC745678 (*yellow*), NSC745689 (*orange*)

interaction with Ile 340 for about 1 ns but retained its initial conformation thereafter for rest of the simulation, maintaining π–π interaction with Phe 343. No, significant deviation in atomic positions was noticed for 5F-203 and NSC745678. The H-bond graph of 5F-203 and NSC745678 (Fig. S16b and S16c) shows that their H-bonding interactions were maintained throughout the simulation run.

Further to analyze the change in protein conformation in response to that of ligand, we plotted the RMSD plot of individual TM. The RMSD was plotted for TM4, TM5, TM6, TM9, TM10, TM11 and TM12 which form the drug translocation pathway. The RMSD plot gave important insights about the deviation in atomic positions of TM helices along the course of simulation and guided to seek for conformational changes in different TM. A similar kind of trend was observed in RMSD of TM4 in all cases (Fig. S16d), since none of the residues of TM4 were involved in binding interactions with the test compounds. In case of NSC745689 a significant increase in RMSD was observed in TM9 (Fig. 8b) and TM12 (Fig. 8e). This corresponds to shift in atomic position and change in side chain conformation of TM9 and TM12. Since, TM9 and TM12 are juxtapositioned and most of the drug binding residues of active site reside on them, suggested that this change in atomic position is associated with unfavorable binding and instability of NSC745689 at the active site. The other significant fluctuation on RMSD was noticed in case of 5F-203 in TM5, (Fig. S16e) this was due to conformational changes in the side chain residues. However, these changes did not affect the binding interaction of 5F-203, since residues of TM5 were not involved in ligand binding. No significant deviation was observed for TM6 (Fig. 8a), TM10 (Fig. 8c) and TM11 (Fig. 8d).in all cases. The stable RMSD plot of ligands and TM depict that verapamil,

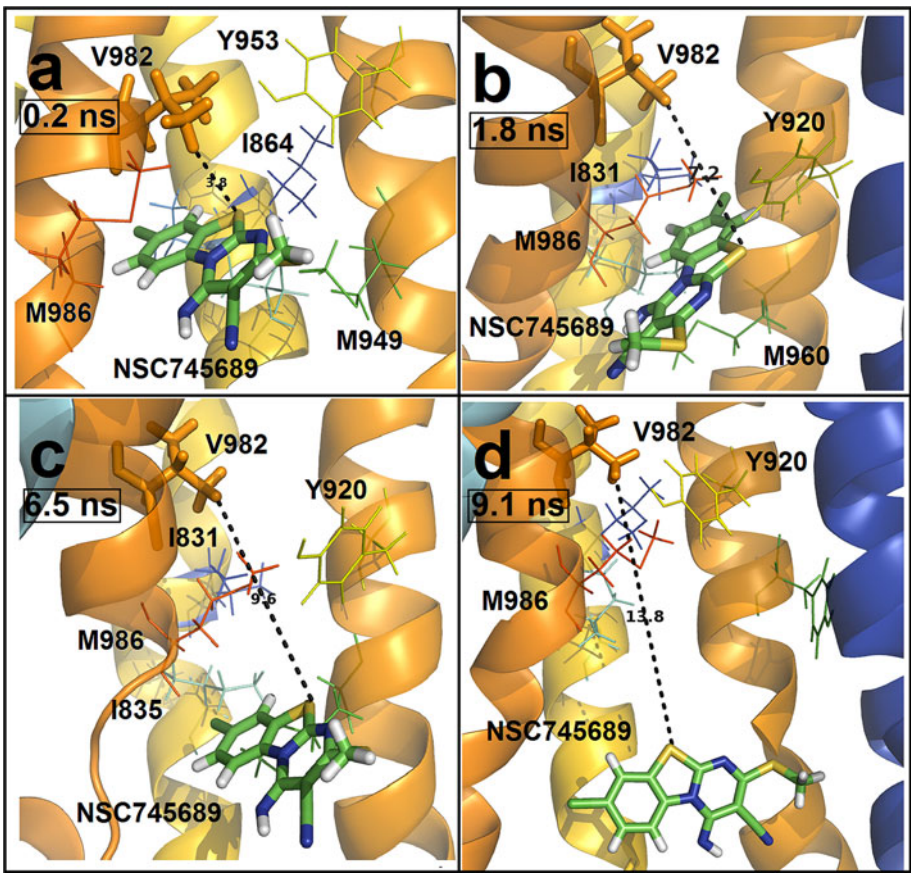

Fig. 7 Figure showing relative shift of NSC745689 from active site, the distance (d) between S-atom of NSC745689 and H-atom of active site residue Val 982 increases gradually from 3.8 Å to 13.8 Å. **a** Snapshot at 0.2 ns, d = 3.8 Å **b** snapshot at 1.8 ns, d = 7.2 Å **c** snapshot at 6.5 ns, d = 9.6 Å **d** snapshot at 9.1 ns, d = 13.8 Å

NSC745677 and NSC745678 held their conformation and maintained their binding interaction in the active site during molecular dynamics simulations.

The results of molecular dynamics simulations suggested that as compared to other compounds NSC745689 showed least affinity at the active site of P-gp. Lack of stable interactions and unfavorable binding indicated that NSC745689 may not be a P-gp substrate. The analysis was further potentiated by performing binding free energy calculation using MMPBSA method.

Binding energy calculations

As discussed previously the prediction of binding affinity of compounds was done on the basis of binding free energy, using MMPBSA protocol. In general, only relative binding free energy was calculated and entropic calculations were skipped due to their high demand of computational power. These calculations were canceled on the assumption that ligands are binding in same active site pocket of the protein.

Herein, we estimated the different components of interaction energy that contributed to ligand binding. These include van der Waals, electrostatic, polar solvation, and nonpolar solvation interaction energies. Table 2 shows the different component energies. The *in vacuo* non-bonded electrostatic energy, $\bar{E}_{elec}$ (Fig. S17a), though favorable in most of the cases, but did not showed much contribution towards ligand binding. Exceptionally, it was unfavorable for NSC745689. Figure S17b shows plot of *in vacuo* non-bonded van der Waals energy, $\bar{E}_{VWD}$. The graph clearly indicated that $\bar{E}_{VWD}$ was favorable for binding in all protein–ligand complexes. It was highest for verapamil (known substrate of P-gp) and lowest for NSC745689. Thus, predicting that van der Waals contributions were more favorable for ligand binding than electrostatic contribution. The *in vacuo* binding enthalpic contribution, $\bar{G}_{gas}$ ($\bar{E}_{MM}$) values showed strong favoring toward ligand binding, having high negative values on the account of $\bar{E}_{VWD}$. Figure S17c shows plot of electrostatic contributions to solvation free energy ($\bar{G}_{PB}$). $\bar{G}_{PB}$ was unfavorable for ligand binding in case of all protein–ligand complexes. The hydrophobic contribution of solvation free energy, was favorable for ligand binding. Figure S17d shows $\bar{G}_{non-polar}$ for each system. The hydrophobic contribution of solvation free energy was highest in case of verapamil, explaining its strong binding in the active site of P-gp. $\bar{G}_{non-polar}$ was least in case of NSC745689. $\bar{G}_{solv}$ was mainly contributed by hydrophobic term, since electrostatic term is highly

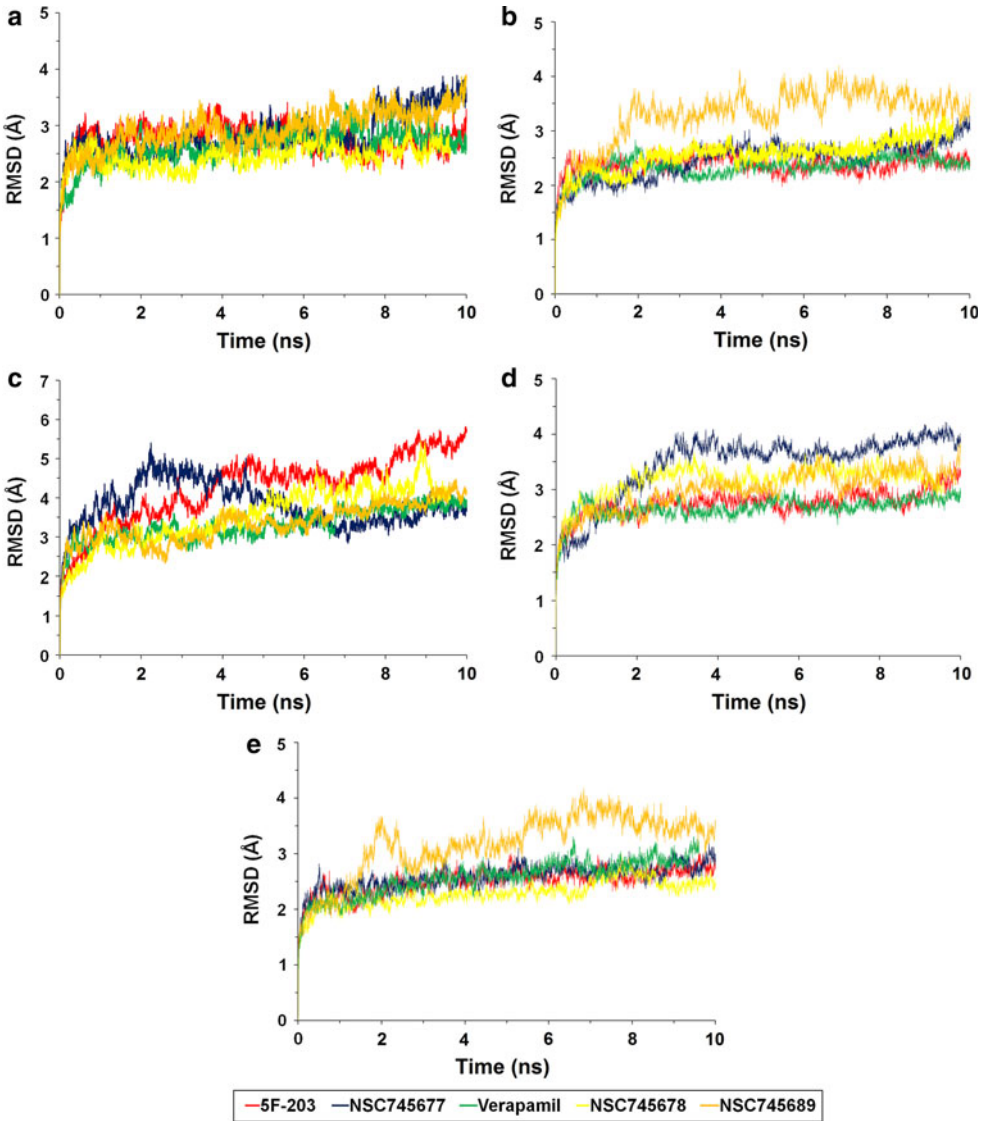

Fig. 8 Plots of RMSD versus time (ns) obtained after 10 ns of production run. **a** TM6 **b** TM9 **c** TM10 **d** TM11 **e** TM12; the color coding of the graphs is similar to Fig. 6

unfavorable. Finally, $\Delta\bar{G}_{binding}$ accounts of total relative binding free energy. It was highest for verapamil and lowest for NSC745689. The compound 5F-203 (presently in Phase IV clinical trial) also showed high value, predicting its propensity to be P-gp substrate.

Out of three bioactive compounds only NSC745689 showed exceptional results. The molecular dynamics and binding free energy calculations indicated that NSC745689 had least possibility to be a P-gp substrate. This incited us to verify our results with in vitro studies.

Table 2 Binding free energies and all its components energies for membrane-protein-complexes

Energy (kcal/mol)	Verapamil	5F-203	NSC745689	NSC745678	NSC745677
$\bar{E}_{VWD}$	−59.4908	−49.8695	−33.9434	−53.3981	−41.6617
$\bar{E}_{elec}$	−8.9111	−8.8152	2.6403	−9.3555	−9.0302
$\bar{G}_{PB}$	37.2979	27.9909	14.9247	34.3991	28.4254
$\bar{G}_{non-polar}$	−5.766	−2.7886	−2.8762	−3.3983	−3.1806
$\bar{G}_{gas}$	−68.4031	−58.6847	−31.3032	−62.7535	−50.6919
$\bar{G}_{solv}$	31.5319	25.2024	12.0485	31.0008	25.2449
$\Delta\bar{G}_{binding}$	−36.8712	−33.4823	−19.2547	−31.7528	−25.4471

Cytotoxicity assay of NSC745689

Cytotoxicity of NSC745689 was evaluated using MTT assay. MTT is a yellow tetrazolium salt that is oxidized by the mitochondrial dehydrogenase in living cells to give a dark blue formazan product. Damaged or dead cells show reduced or no dehydrogenase activity. Cell viability value of less than 50 % indicates reduced mitochondrial activity while >80 % indicates non toxic nature of the compound towards Caco-2 cells. The proposed working concentration of NSC745689 (100 μM) displayed approximately 88.5 % cell viability after 24 h incubation indicating its non-toxic nature. Although cell viability values were decreased after 24 h incubation at higher concentrations, concentrations up to 200 μM can be used as it did not have any effect on cell viability after 2 h incubation (the course of the experimentation).

Evaluation of Caco-2 cell monolayer

The TEER values were employed as an integrity evaluator to assure that the Caco-2 cells had formed a proper monolayer with efficient tight junctions that allowed the passage of the compounds by the transcellular route. Monolayer integrity was controlled by ensuring a TEER value of $>300\ \Omega.cm^2$ [61]. The apparent permeability coefficients were also determined for the permeability markers, furosemide and propranolol to validate proper monolayer formation. Furosemide (poorly permeable drug) showed a P_{app} value of 4.21 ($\pm$0.07) $\times$ 10^{-6} cm/s whereas propranolol (highly permeable drug) showed a P_{app} value of 33.4 ($\pm$0.08) $\times$ 10^{-6} cm/s. These P_{app} values were compared well with the reported values [72–74]. These two studies ensured the proper monolayer formation and further confirmed the validity of the protocol used for the permeability study. The monolayer integrity for permeability studies was confirmed by the transport of Lucifer yellow at the end of the experiment. The amount of Lucifer Yellow that permeated was less than 2 % after 2 h.

Stability of NSC745689 in HBSS

NSC745689 was stable in HBSS buffer during the time period required for the permeability studies (at least for 2 h). The RSD of the peak area were found to be less than 2 %.

Table 3 Permeability coefficients for NSC745689 in Caco-2 monolayer under different transport conditions

Conditions	Permeability coefficients ($\times 10^{-6}$ cm/s)
Control (apical to basolateral)	1.03 ± 0.16
Control (basolateral to apical)	0.27 ± 0.01
With verapamil (apical to basolateral)	1.77 ± 0.21
With verapamil (basolateral to apical)	0.15 ± 0.03

Permeability studies

It was found that NSC745689 is poorly permeable drug with P_{app} 1.03 ± 0.16 × 10^{-6} cm/s (Table 3). Apical to basolateral flux of NSC745689 was found higher than basolateral to apical flux. The efflux ratio was 0.26 indicating that NSC745689 was not actively transported from the basolateral to apical compartment. This was confirmed by the use of the P-gp inhibitor, verapamil, which failed to increase the flux rate significantly. Decrease in basolateral to apical flux of NSC745689 in presence of verapamil was also not significant. These findings ruled out the possibility of efflux mechanism by P-gp in transport of NSC745689. However, higher absorptive transport of NSC745689 could be because of carrier mediated transport and need to be investigated further.

Conclusions

The reported P-gp transmembrane model gave good correlations and could be reliably utilized in P-gp substrate prediction. The Ramachandran plot, Errat plot and PROSA II energy plot indicated good reliability of the model. The IFD of co-crystallized ligand showed good correlation with the interacting residues of the template as reported in the crystal structure, indicating the correct modeling of active site. The binding mode of verapamil also supported the experimental depiction of binding site, which were well maintained during 10 ns of molecular dynamics simulations. Thus, the membrane-protein model could be used to study P-gp-substrate interactions as well as P-gp substrate identification.

The anticancer compounds taken under investigation showed varied interactions in the active site. 5F-203 showed π–π interaction with Phe 343 and H-bonding interaction with Tyr 953, NSC745677 showed π–π interaction with Phe 343, and NSC745678 showed H-bonding interaction with Tyr 953 and hydrophobic interaction with Val 981. The RMSD plot pointed out that these interactions were maintained during MD simulations, thus indicating their possibility to be P-gp substrate. However, the compound NSC745689 did not show any key interaction in the active site, predicting it as a potential lead for anticancer drug devoid of MDR. The binding free energy studies also showed consensus result to IFD and molecular dynamics prediction, as binding free energy was found to be least in case of NSC745689. The in silico studies were strongly supported by in vitro studies and clearly indicated that NSC745689 was poorly permeable with P_{app} 1.03 ± 0.16 × 10^{-6} cm/s and with low efflux ratio

of 0.26. Studies in conjunction with verapamil also showed that NSC745689 would not be a characteristic substrate of P-gp.

Moreover, the study also indicated that compounds with higher number of hydrophobic features were more prone to P-gp binding since the active site is highly hydrophobic. The docking studies showed that the TM regions forming active site is highly populated with hydrophobic residues. Compounds with rigid scaffold are less venerable to be a P-gp substrate as compared to flexible molecule. This is because the compounds with rigid scaffold are unable to modify their conformation in accordance to binding site and does not make a good fit in the binding cavity. The binding free energy calculations also indicated that the ligand binding was mainly contributed by hydrophobic term. However, the hydrophilic residues within the active site were also found to be responsible for substrate binding (as in docking study of test compounds), which explained the binding of hydrophilic drugs to P-gp. Thus, on the basis of our studies we conclude that though lipophilicity is crucial to designing anticancer drugs but hindered hydrophilic features within the structure could help in bypassing P-gp efflux.

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, Ondrusek N, Yeger H, Ling V (1991) P-glycoprotein expression as a predictor of the outcome of therapy for neuroblastoma. N Engl J Med 325(23):1608–1614
3. Van der Valk P, Van Kalken CK, Ketelaars H, Broxterman HJ, Scheffer G, Kuiper CM, Tsuruo T, Lankelma J, Meijer C, Pinedo HM (1990) Original article: distribution of multi-drug resistance-associated P-glycoprotein in normal and neoplastic human tissues. Ann Oncol 1(1):56–64
4. Ambudkar SV, Kimchi-Sarfaty C, Sauna ZE, Gottesman MM (2003) P-glycoprotein: from genomics to mechanism. Oncogene 22(47):7468–7485
5. Seelig A (1998) A general pattern for substrate recognition by P-glycoprotein. Eur J Biochem 251(1–2):252–261
6. Higgins CF, Callaghan R, Linton KJ, Rosenberg MF, Ford RC (1997) Structure of the multidrug resistance P-glycoprotein. Semin Cancer Biol 8:135–142
7. Jones PM, George AM (1999) Subunit interactions in ABC transporters: towards a functional architecture. FEMS Microbiol Lett 179(2):187–202
8. Sharom FJ (1997) The P-glycoprotein efflux pump: how does it transport drugs? J Membr Biol 160(3):161–175
9. Crowley E, Callaghan R (2010) Multidrug efflux pumps: drug binding-gates or cavity? FEBS J 277(3):530–539
10. Didziapetris R, Japertas P, Avdeef A, Petrauskas A (2003) Classification analysis of P-glycoprotein substrate specificity. J Drug Target 11(7):391–406
11. Doppenschmitt S, Spahn-Langguth H, Regardh CG, Langguth P (1999) Role of P-glycoprotein-mediated secretion in absorptive drug permeability: an approach using passive membrane permeability and affinity to P-glycoprotein. J Pharm Sci 88(10): 1067–1072
12. Chang G (2003) RETRACTED: structure of MsbA from Vibrio cholera: a multidrug resistance ABC transporter homolog in a closed conformation. J Mol Biol 330(2):419–430
13. Dawson RJP, Locher KP (2006) Structure of a bacterial multidrug ABC transporter. Nature 443(7108):180–185
14. Aller SG, Yu J, Ward A, Weng Y, Chittaboina S, Zhuo R, Harrell PM, Trinh YT, Zhang Q, Urbatsch IL, Chang G (2009) Structure of P-glycoprotein reveals a molecular basis for poly-specific drug binding. Science 323(5922):1718–1722
15. Jin MS, Oldham ML, Zhang Q, Chen J (2012) Crystal structure of the multidrug transporter P-glycoprotein from *Caenorhabditis elegans*. Nature 490(7421):566–569
16. Pajeva IK, Globisch C, Wiese M (2009) Comparison of the inward and outward-open homology models and ligand binding of human P-glycoprotein. FEBS J 276(23):7016–7026
17. Sato T, Kodan A, Kimura Y, Ueda K, Nakatsu T, Kato H (2009) Functional role of the linker region in purified human P-glycoprotein. FEBS J 276(13):3504–3516
18. Campbell JD, Deol SS, Ashcroft FM, Kerr ID, Sansom MSP (2004) Nucleotide-dependent conformational changes in HisP: molecular dynamics simulations of an ABC transporter nucleotide-binding domain. Biophys J 87(6):3703–3715
19. Campbell JD, Sansom MSP (2005) Nucleotide binding to the homodimeric MJ0796 protein: a computational study of a prokaryotic ABC transporter NBD dimer. FEBS Lett 579(19): 4193–4199
20. Damas JM, Oliveira ASF, Baptista AM, Soares CM (2011) Structural consequences of ATP hydrolysis on the ABC transporter NBD dimer: molecular dynamics studies of HlyB. Protein Sci 20(7):1220–1230
21. Jones PM, George AM (2002) Mechanism of ABC transporters: a molecular dynamics simulation of a well characterized nucleotide-binding subunit. Proc Natl Acad Sci USA 99(20): 12639–12644
22. Jones PM, George AM (2004) The ABC transporter structure and mechanism: perspectives on recent research. Cell Mol Life Sci 61(6):682–699
23. Jones PM, George AM (2007) Nucleotide-dependent Allostery within the ABC transporter ATP-binding cassette. J Biol Chem 282(31):22793–22803
24. Jones PM, George AM (2009) Opening of the ADP-bound active site in the ABC transporter ATPase dimer: evidence for a constant contact, alternating sites model for the catalytic cycle. Proteins Struct Funct Bioinf 75(2):387–396
25. Jones PM, George AM (2011) Molecular-dynamics simulations of the ATP/apo state of a multidrug ATP-binding cassette transporter provide a structural and mechanistic basis for the asymmetric occluded state. Biophys J 100(12):3025–3034
26. Oliveira ASF, Baptista AM, Soares CM (2010) Insights into the molecular mechanism of an ABC transporter: conformational changes in the NBD dimer of MJ0796. J Phys Chem B 114(16): 5486–5496
27. Oloo EO, Fung EY, Tieleman DP (2006) The dynamics of the MgATP-driven closure of MalK, the energy-transducing subunit of the maltose ABC transporter. J Biol Chem 281(38): 28397–28407
28. Wen PC, Tajkhorshid E (2008) Dimer opening of the nucleotide binding domains of ABC transporters after ATP hydrolysis. Biophys J 95(11):5100–5110

29. Becker JP, Depret G, Van Bambeke F, Tulkens PM, Pravost M (2009) Molecular models of human P-glycoprotein in two different catalytic states. BMC Struct Biol 9(1):3
30. Ivetac A, Campbell JD, Sansom MSP (2007) Dynamics and function in a bacterial ABC transporter: simulation studies of the BtuCDF system and its components. Biochemistry 46(10): 2767–2778
31. Kandt C, Tieleman DP (2010) Holo-BtuF stabilizes the open conformation of the vitamin B12 ABC transporter BtuCD. Proteins Struct Funct Bioinf 78(3):738–753
32. Oliveira AS, Baptista AM, Soares CM (2011) Conformational changes induced by ATP-hydrolysis in an ABC transporter: a molecular dynamics study of the Sav 1866 exporter. Proteins: Struct, Funct, Bioinf 79:1977–1990
33. Oloo EO, Tieleman DP (2004) Conformational transitions induced by the binding of MgATP to the vitamin B12 ATP-binding cassette (ABC) transporter BtuCD. J Biol Chem 279(43):45013–45019
34. Sonne J, Kandt C, Peters GH, Hansen FY, Jensen MÃ, Tieleman DP (2007) Simulation of the coupling between nucleotide binding and transmembrane domains in the ATP binding cassette transporter BtuCD. Biophys J 92(8):2727–2734
35. St-Pierre JF, Bunker A, Rog T, Karttunen M, Mousseau N (2012) Molecular dynamics simulations of the bacterial ABC transporter SAV1866 in the closed form. J Phys Chem B 116(9):2934–2942
36. Sun TG, Liu M, Chen WZ, Wang CX (2010) Molecular dynamics simulation of the transmembrane subunit of BtuCD in the lipid bilayer. Sci China Life Sci 53(5):620–630
37. Wen PC, Tajkhorshid E (2011) Conformational coupling of the nucleotide-binding and the transmembrane domains in ABC transporters. Biophys J 101(3):680–690
38. Weng JW, Fan KN, Wang WN (2010) The conformational transition pathway of ATP binding cassette transporter MsbA revealed by atomistic simulations. J Biol Chem 285(5):3053
39. Labhsetwar LB, Shendarkar GR, Kuberkar SV (2010) Synthesis and in vitro anticancer activity of 8-chloro-3-cyano-4-imino-2-methylthio-4-H-pyrimido [2, 1-B][1, 3] benzothiazoel and its 2-substituted derivatives. JPRHC 3:273–278
40. Nandekar P, Tumbi K, Bansal N, Rathod V, Labhsetwar L, Soumya N, Singh S, Sangamwar A (2012) Chem-bioinformatics and in vitro approaches for candidate optimization: a case study of NSC745689 as a promising antitumor agent. Med Chem Res 1–15. doi:10.1007/s00044-012-0364-8
41. Apweiler R, Martin MJ, O'Donovan C, Magrane M, Alam-Faruque Y, Antunes R, Barrell D, Bely B, Bingley M, Binns D (2010) The universal protein resource (UniProt) in 2010. Nucleic Acids Res 38:D142–D148
42. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215(3):403–410
43. Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, Shen MY, Pieper U, Sali A (2007) Comparative protein structure modeling using Modeller. Curr Protoc Protein Sci 2(12):15–32
44. Ferreira RJ, Ferreira MJU, dos Santos DJVA (2012) Insights on P-glycoprotein's efflux mechanism obtained by molecular dynamics simulations. J Chem Theory Comput 8(6):1853–1864
45. Hrycyna CA, Airan LE, Germann UA, Ambudkar SV, Pastan I, Gottesman MM (1998) Structural flexibility of the linker region of human P-glycoprotein permits ATP hydrolysis and drug transport. Biochemistry 37(39):13660–13673
46. Raviv Y, Pollard HB, Bruggemann EP, Pastan I, Gottesman MM (1990) Photosensitized labeling of a functional multidrug transporter in living drug-resistant tumor cells. J Biol Chem 265(7):3975
47. Retzinger GS, Cohen L, Lau SH, Kezdy FJ (1986) Ionization and surface properties of verapamil and several verapamil analogues. J Pharm Sci 75(10):976–982
48. Schrödinger Suite 2009 Induced Fit Docking protocol; Glide version 5.5, Schrödinger, LLC, New York, NY, 2009; Prime version 2.1, Schrödinger, LLC, New York, NY, 2009
49. Sherman W, Beard HS, Farid R (2006) Use of an induced fit receptor structure in virtual screening. Chem Biol Drug Des 67(1):83–84
50. Loo TW, Bartlett MC, Clarke DM (2006) Transmembrane segment 1 of human P-glycoprotein contributes to the drug-binding pocket. Biochem J 396(Pt 3):537
51. Loo TW, Bartlett MC, Clarke DM (2006) Transmembrane segment 7 of human P-glycoprotein forms part of the drug-binding pocket. Biochem J 399(Pt 2):351
52. Loo TW, Bartlett MC, Clarke DM (2009) Identification of residues in the drug translocation pathway of the human multidrug resistance P-glycoprotein by arginine mutagenesis. J Biol Chem 284(36):24074–24087
53. Loo TW, Clarke DM (1997) Identification of residues in the drug-binding site of human P-glycoprotein using a thiol-reactive substrate. J Biol Chem 272(51):31945–31948
54. Loo TW, Clarke DM (2000) Identification of residues within the drug-binding domain of the human multidrug resistance P-glycoprotein by cysteine-scanning mutagenesis and reaction with dibromobimane. J Biol Chem 275(50):39272–39278
55. Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. J Mol Graph 14(1):33–38
56. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJC (2005) GROMACS: fast, flexible, and free. J Comput Chem 26(16):1701–1718
57. Kandt C, Ash WL, Peter Tieleman D (2007) Setting up and running molecular dynamics simulations of membrane proteins. Methods 41(4):475–488
58. Bayly CI, Cieplak P, Cornell W, Kollman PA (1993) A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. J Phys Chem 97(40):10269–10280
59. Wang J, Morin P, Wang W, Kollman PA (2001) Use of MM-PBSA in reproducing the binding free energies to HIV-1 RT of TIBO derivatives and predicting the binding mode to HIV-1 RT of efavirenz by docking and MM-PBSA. J Am Chem Soc 123(22):5221–5230
60. Hou T, Wang J, Li Y, Wang W (2011) Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. J Chem Inf Model 51(1):69–82
61. Wahlang B, Pawar YB, Bansal AK (2011) Identification of permeability-related hurdles in oral delivery of curcumin using the Caco-2 cell model. Eur J Pharm Biopharm 77(2):275–282
62. Loo TW, Bartlett MC, Clarke DM (2004) Disulfide cross-linking analysis shows that transmembrane segments 5 and 8 of human P-glycoprotein are close together on the cytoplasmic side of the membrane. J Biol Chem 279(9):7692–7697
63. Loo TW, Bartlett MC, Clarke DM (2004) Val133 and Cys137 in transmembrane segment 2 are close to Arg935 and Gly939 in transmembrane segment 11 of human P-glycoprotein. J Biol Chem 279(18):18232–18238
64. Loo TW, Bartlett MC, Clarke DM (2007) Suppressor mutations in the transmembrane segments of P-glycoprotein promote maturation of processing mutants and disrupt a subset of drug-binding sites. J Biol Chem 282(44):32043–32052
65. Loo TW, Bartlett MC, Clarke DM (2008) Processing mutations disrupt interactions between the nucleotide binding and transmembrane domains of P-glycoprotein and the cystic fibrosis transmembrane conductance regulator (CFTR). J Biol Chem 283(42):28190–28197
66. Loo TW, Clarke DM (2000) The packing of the transmembrane segments of human multidrug resistance P-glycoprotein is

revealed by disulfide cross-linking analysis. J Biol Chem 275(8): 5253–5256
67. Loo TW, Clarke DM (2001) Determining the dimensions of the drug-binding domain of human P-glycoprotein using thiol cross-linking compounds as molecular rulers. J Biol Chem 276(40): 36877–36880
68. Loo TW, Clarke DM (2001) Defining the drug-binding site in the human multidrug resistance P-glycoprotein using a methanethiosulfonate analog of verapamil, MTS-verapamil. J Biol Chem 276(18):14972–14979
69. Indu S, Kumar ST, Thakurela S, Gupta M, Bhaskara RM, Ramakrishnan C, Varadarajan R (2010) Disulfide conformation and design at helix N-termini. Proteins: Struct, Funct, Bioinf 78(5): 1228–1242
70. Newstead S, Fowler PW, Bilton P, Carpenter EP, Sadler PJ, Campopiano DJ, Sansom MSP, Iwata S (2009) Insights into how nucleotide-binding domains power ABC transport. Structure 17(9):1213–1222
71. Loo TW, Clarke DM (1999) Identification of residues in the drug-binding domain of human P-glycoprotein Analysis of transmembrane segment 11 by cysteine-scanning mutagenesis and inhibition by dibromobimane. J Biol Chem 274(50):35388–35392
72. Korjamo T, Honkakoski P, Toppinen MR, Niva S, Reinisalo M, Palmgrénb JJ, Mönkkönena J (2005) Absorption properties and P-glycoprotein activity of modified Caco-2 cell lines. Eur J Pharm Sci 26(3):266–279
73. Uchida M, Fukazawa T, Yamazaki Y, Hashimoto H, Miyamoto Y (2009) A modified fast (4 day) 96-well plate Caco-2 permeability assay. J Pharmacol Toxicol Methods 59(1):39–43
74. Yamashita S, Furubayashi T, Kataoka M, Sakane T, Sezaki H, Tokuda H (2000) Optimized conditions for prediction of intestinal drug permeability using Caco-2 cells. Eur J Pharm Sci 10(3):195–204