4. Designing of New Drug Molecule By Selecting Appropriate Series For Computer Aided Molecular Modeling

The computational process of searching for a ligand that is able to fit both geometrically and energetically into the binding site of a protein is called molecular docking. Molecular docking is an efficient tool for investigating receptor-ligand interactions and for virtual screening, which plays a key role in rational drug design, especially when the crystal structure of a receptor or enzyme is available. It is widely accepted that drug activity is obtained through the molecular binding of one molecule (the ligand) to the pocket of another, usually larger, molecule (the receptor), which is commonly a protein. In their binding conformations, the molecules exhibit geometric and chemical complementarily, both of which are essential for successful drug activity. Molecular docking can be thought of as a problem of “lock-and-key”, where one is interested in finding the correct relative orientation of the “key” which will open up the “lock” (where on the surface of the lock is the key hole, which direction to turn the key after it is inserted, etc.). Here, the protein can be thought of as the “lock” and the ligand can be thought of as a “key”. Molecular docking may be defined as an optimization problem, which would describe the “best-fit” orientation of a ligand that binds to a particular protein of interest. However since both the ligand and the protein are flexible, a “hand-in-glove” analogy is more appropriate than “lock-and-key”.

Molecular docking helps in studying drug/ ligand or receptor/ protein interactions by identifying the suitable active sites in protein, obtaining the best geometry of ligand–receptor complex and calculating the energy of interaction for different ligands to design more effective ligands. The target or receptor is either experimentally known or theoretically generated through knowledge based protein modeling or homology modeling. The molecular docking tool has been developed to obtain a preferred geometry of interaction of ligand - receptor complexes having minimum interaction energy based on different scoring functions viz. only electrostatics, sum of steric and electrostatic (parameters from MMFF force field) and Dock Score. VLife MDS uses genetic algorithm (GA), piecewise linear pair wise potential (PLP) and grid algorithms to minimize the interaction energy between ligand – receptor.

One key aspect of molecular modeling is calculating the energy of conformations and interactions using methods ranging from quantum mechanics to purely empirical energy functions. Molecular docking energy evaluations are usually carried out with
the help of a scoring function. Developing these scoring functions is a major challenge in structure based drug design. Efficiency and accuracy of geometric modeling of the binding process to obtain correct docking solutions depends on scoring function. Usually scoring functions are based on force fields that were initially designed to simulate the function of proteins (based on enthalpy). Many techniques have been proposed that address specific parts of this challenge. Among the first were methods that simply evaluate whether a particular ligand can fit into the receptor pocket under the assumption of both rigid ligands and a rigid protein. This problem allows an enumerative approach; because there are only six degrees of freedom that completely specify the relative position of the ligand with respect to the receptor. Such techniques are reasonably fast. Possible geometries can be scored by force field, empirical or knowledge-based methods. (MDS allows user to select different intermolecular interactions viz. steric, electrostatic).

In addition, a flexible ligand docking includes molecules internal degree of freedom along with values of translation and rotation in search of its suitable bound conformation that makes it computationally more expensive than rigid ligand docking. Distinction of good or bad docked conformation is based on scoring or fitness function. (MDS uses fitness functions on only electrostatic and both steric and electrostatic interactions between receptor-ligand as well as Dock Score scoring function). The Dock score or XScore as it is called compute binding affinity of a given protein ligand complex with known 3-D structure. Dock/X-Score scoring function include terms for van der Waals interaction, hydrogen bonding, hydrophobic effects. Genetic algorithm (implemented in VLifeMDS) offers a successful strategy for globally searching the docked conformers’ space. Such an algorithm mirrors Darwinian evolution, representing the solution as a 'chromosome'. Genetic algorithms allow a population of solutions to exist and in each 'generation' these can evolve by processes such 'breeding' and 'mutation'. Poor solutions are killed off, while good ones leave their offspring in future generations. Such algorithms may typically reach an excellent solution is a few tens of generations. The grid based docking is a rigid and exhaustive docking method. In this method, after unique conformers of the ligand are generated, the receptor cavity of interest is chosen by the user and a grid is generated around the cavity (default grid interval size 1 Å). Cavity points are found and the centre of mass of the ligand is moved to each cavity point. All rotations of ligand are scanned at each cavity point where ligand is placed (step size of rotation could be
typically 100-150 as an example). For each rotation a pose of the ligand is generated and the corresponding bumps are checked for each pose of ligand. The X-Cscore is calculated for each valid pose (determined by the cut off criteria fed by user in terms of max no of allowed bumps) and the pose of the ligand with the best score is given as output to user.

4.1 Target Site

The structure of target was obtained from the protein data bank (PDB code 1T3L) (figure 11). The target is voltage-dependent calcium channel β subunit functional core and its complex with α₁ interaction domain.

Figure 11. Target structure (PDB code 1T3L).

4.2 Target structure

Voltage-dependent calcium channels (VGCCs) comprise a pore-forming α₁ subunit and smaller auxiliary subunits known as β, α₂ and δ. β subunits are members of the membrane-associated guanylate kinase (MAGUK) family of proteins. β subunit contains five distinct modular domains regions, including the variable N-and C-terminus, a conserved Src homology 3 (SH3) domain, a conserved guanylate kinase (GK) domain, and a connecting variable and flexible hook region (figure 12).
Figure 12. Structure of VDCC β functional core in complex with the AID. Domain I (SH3), domain II (GuK), and the AID are represented in red, blue, and green, respectively.

The primary site on the α₁ subunit for interaction with the β subunit is regions of intracellular loop between domains I and II labeled the alpha interaction domain (AID). The AID of the α₁ subunit binds to a hydrophobic binding pocket within the modified GK domain of the β subunit (figure 13).

Figure 13. The AID (green) is represented as bonds and β is represented in blue ribbons.

4.3 Molecular Modeling: Method and Procedure

The access to experimental structural data of membrane proteins like G-protein coupled receptors (GPCRs) or ion channels by X-ray crystallography or NMR spectroscopy is still limited. For X-ray crystallography of these proteins it is rather difficult to obtain crystals, as the structure is dependent on the surrounding, the membrane. Separation of the proteins from the membrane can lead to irreversible structural changes, which inhibit ordered crystallization. Limitations for NMR
spectroscopy are the size of the systems, as NMR-spectroscopy can be applied only to smaller protein systems. Up to now, no transmembrane domains of calcium channels were crystallized. Many techniques such as patch clamp, fluorescence resonance energy transfer (FRET) used lead to significant new insights about the structure, function and drug interaction with the channels. Computational methods are also additional useful method to investigate these research areas. In general, the protein structure prediction using computational methods can be divided into three approaches
1) Homology modeling
2) Threading or fold recognition
3) Ab Initio.

For this work, homology modeling using multiple templates and modeling the loops using an Ab Initio method were applied. Homology modeling based on the crystal structure of potassium and sodium channels are used to create models for the diverse states of the voltage gated calcium channel Cav1.2. Besides the model of L-type calcium channel Cav1.2, the T-type calcium channel Cav3.1 was investigated. Due to the lack of conserved residues at the loop region, the Ab Initio method was applied to predict the loop structure of calcium channel Cav 3.1 combined with the results of experimental investigations. Subsequently, molecular docking studies were used to investigate ligands interacting with the internal pore of the Cav1.2 channel in the open state. Molecular dynamics simulations (MD) are important methods which give insight into the molecular basis of the biological activities. In this work, MD was also used to stabilize the model structures. Further, the MD of ligands with Cav1.2 channel was investigated for better understanding of the action mechanism and to increase knowledge about drug interaction mechanisms.9

4.4 Homology modeling
For unknown protein structures such as membrane proteins, homology modeling was introduced to construct the three-dimensional structure of a known atomic-resolution model of the protein (target) and related homologous protein (template). The procedure consists of three steps.10

The first step: To select similar proteins (known 3D structures) as templates and to align between the target (unknown structure) and one or more templates. The success of this method relies on the sequence alignment between target and suitable templates. In this work, sequence searches from Expasy11 database using BLAST12 were
performed to identify related sequences. The quality of the models depends on the sequence identity of the sequence alignment. Unfortunately, voltage-gated calcium channel shares in general low sequence identity with K+ channel crystal structure templates. Thus, it is necessary to find possible homolog’s for a protein by comparing several sequences. In this work, the multiply sequence alignment was built from ClustalX\textsuperscript{13} program combined with manual intervention based on sequence conservation information.

The second step: To construct the 3D model of Cav1.2 closed conformation, the software package MODELLER\textsuperscript{14} was used to generate the model from the alignment of a sequence. This program builds a structure with an extended strand for target and fold by satisfaction of spatial restraints\textsuperscript{15} from the alignment of the target and its templates. The hydrogen bonding features and main chain dihedral angles are preserved from the template structure. The model was manually adjusted according to the side-chains to optimize hydrophobic contacts, salt bridges, hydrogen bond formation and aromatic – aromatic interactions.

The third step: To evaluate the quality of the model. It is important to check the model structure using the quality assessment tools. There are various quality assessment programs available to evaluate for correctness of the overall fold, errors over localized regions and stereochemical properties of the model e.g. Prosa2003,\textsuperscript{16} Verify 3D, WHAT_CHECK,\textsuperscript{17} Procheck. A simple preliminary check can also obtain by a Ramachandran plot. The entire process of homology modeling may be necessary to repeat until a satisfactory model is obtained. In this work, molecular dynamic simulation of the satisfactory model in lipid bilayer environment is also used to evaluate the quality and to stabilize the model.

**Ab Initio method:**

Another approach for 3D structure prediction is the Ab Initio method. This method predicts the native state of a protein structure from the sequence information only. In general, Ab Initio methods consist of 3 steps:

1) Try to retrieve template proteins of similar folds from the database. In case that no appropriate template is identified in the database, the structure will be build by Ab Initio modeling.

2) Define the energy function compatible with the predicted structure.
3) Apply efficient and reliable algorithms to search the global predicted structure. The structure prediction ultimately is done through physical forces acting on all-atom of the model.

Ab Initio method to predict the loop structures of calcium channel Cav 3.1 using ROSETTA program and iTasser server. The ROSETTA method is a distributed-computing implementation based on Rosetta algorithm which tested, developed and succeed on globular proteins in Critical Assessment of Techniques for Protein Structure Prediction (CASP). For I-Tasser, the method is based on threading fragment structure reassembly.

4.5 Model Evaluation

To evaluate the homology models, quality assessment method are necessary to check the quality of the constructed models. The quality check should be able to verify the reliability of the model. It should be able to distinguish between properly versus improperly folded models, and evaluate steric and geometric properties of the models. There are numerous quality assessment programs which different criteria methods available. Most of the methods have been developed using empirical data from globular proteins of known structure. In our work, the models quality was assessed using Prosa2003, Verify3D, WHAT_CHECK Packing, Procheck, ProQresLG and ProQres MaxSub. Prosa2003 is a program to check the potential error in 3D models of protein structures. It uses statistical potential of distance and surface-dependent statistical for Cα atoms to obtain the model. Verify3D is the method that analyzes the 3D atomic model with its own amino acid sequence 1D. It will provide with a statistical potentials from real proteins. WHAT_CHECK Packing checks all possible atom types in all possible positions around the fixed fragments. Procheck addresses the stereochemical parameter of a protein. The structure is classified into highly populated to forbidden regions by the Ramachandran plot. It shows the torsion angles for all residues in the structure. ProQres analyzes contact between atom-atom, residue-residue, solvent-accessible surfaces and secondary structure.

4.6 Molecular Docking

Molecular docking is a method which gives insight into the molecular basis of the biological activities for better understanding of the action mechanism and increased knowledge about drug interaction mechanisms. A basic component of every docking program is a search algorithm and an energy scoring function which can be based on

2. Empirical data: functions fitted to experimental data which based on the knowledge of known protein-ligand interaction.

3. Knowledge-based methods: capture the knowledge of receptor-ligand binding by statistical data alone.

The general docking procedure generates docked models by finding the binding region on macromolecules (proteins) where the ligand most likely interacts. Then, the results of docking for groups of ligands are clustered based on the location and examining energetic or ranking score. Additionally, the experimental data are important in determining the probable binding sites. Then, the best models and experimental data should be compared to filter results. We can perform refinement docking where the ligand is restricted to a specified region based on these filtered results.

4.7 Docking programs used in thesis

Glide (Grid-based Ligand Docking with Energetics) and MOE, which have a different approach are the docking programs used in this work. Glide program places ligands in the receptor by grid docking alignment. This program computes the grid in terms of position, orientation and conformation space available. Each ligand atom is matched with the grid point with the lowest energy within its neighborhood. The scoring function is reported in term of GlideScore.

\[ \text{GScore} = a \times \text{vdW} + b \times \text{Coul} + \text{Lipo} + \text{Hbond} + \text{Metal} + \text{Rewards} + \text{RotB} + \text{Site} \]

Where,
\[ \text{vdW} = \text{van der Waals interaction energy.} \]
\[ \text{Coul} = \text{Coulomb interaction energy.} \]
\[ \text{Lipo} = \text{Lipophilic-contact plus phobic-attractive term.} \]
\[ \text{HBond} = \text{Hydrogen-bonding term.} \]
\[ \text{Metal} = \text{Metal-binding term.} \]
\[ \text{Rewards} = \text{various reward or penalty terms.} \]
\[ \text{RotB} = \text{Penalty for freezing rotatable bonds.} \]
\[ \text{Site} = \text{Polar interactions in the active site.} \]
\[ \text{a,b} = \text{the contribution from the Coulomb term is capped at -4 kcal/mol} \]
\[ a = 0.050, \ b = 0.150 \text{ for Glide 5.0.} \]
4.8 Docking methodology

From experiment it is known that ligands preferably to block the open conformation. Thus, the molecular docking of this work was focused on the open conformation of the ion-gated calcium channel and the improved open conformation with various ligands.

4.9 Ligand preparation

Coordinates of ligands were generated with GaussView and the geometry optimized with the Hartree-Fock method using 3-21G basis set implemented in Gaussian03.

4.10 Glide docking program

Protein Preparation Wizard workflow implemented in Maestro 8.5 was used to prepare the protein using default settings. Both protein and ligands were parameterized with the OPLS force field. The Receptor Grid Generation panel performed the grid map generation of the receptor. The grid was generated overall binding residues known the experimentally. Docking Conformation analysis calculations performed in Standard Precision (SP) mode using the Ligand Docking panel. The Receptor Grid Generation and Ligand Docking panels are functions in the Glide module.

4.11 MOE docking program

The standard protocol of the procedure in MOE 2008 was applied in this work. The Alpha Triangle placement which derives poses by random superposition of ligand atom triplets alpha sphere dummies in the receptor site is to determine the poses. The London dG scoring function estimates the free energy of binding of the ligand from a given pose.

\[ \Delta G = c + \sum_{h-bonds} C_{HB} f_{HB} + \sum_{m-lig} C_{M} f_{M} + \sum_{atom} \Delta P_{i} \]

Where,

c = the average gain/loss of rotational and translational motion.

Eflex = the energy due to loss of flexibility of the ligand.

CHB = an hydrogen bond energy

fHB = measures geometric imperfections of hydrogen bonds

CM = a metal ligation energy

fM = measures geometric imperfections of metal ligations

Di = the desolvation energy of each atom i
4.12 Molecular dynamic simulations

Molecular Dynamic Simulation (MD) is a theoretical and computational method based on solving the Newton’s equation of motion. This method is used to mimic the behavior of the system as a function of time. MD provides a basis for a more complete understanding of biological systems and aids in the interpretation of experiments concerned with their properties. In a molecular dynamics simulation, the trajectory of the molecules and atoms for choosing the potential function $U(r_1,...,r_N)$ of the position of the nuclei represent the potential energy of the system when the atoms are arranged in specific configuration. The potential energy is usually constructed from the relative positions of the atoms with respect to each other. Forces are derived as the gradients of the potential with respect to atomic displacement as shown in below formula. This form implies the presence of a conservation law of the total energy, where is the instantaneous kinetic energy.

![Cartesian coordinate system](image)

Figure 14. Cartesian coordinated laboratory-fixed reference frame used to define a position vector.

The translational motion of spherical molecules is caused by a force exerted by some external agent. The motion and the applied force are explicitly interpreted by Newtonian. Newton’s equation of motion of a particle system is written in a set of
coupled second order differential equation in time. The functional form is a sum of terms:

\[ m_i \frac{d^2 r_i}{dt^2} = -\nabla_i \left[ U(r_1, r_2, \ldots, r_N) \right] \quad i = 1, N \]

Where, \( m \) is the mass of the molecule, \( r_i \) is a vector that locates the atoms with respect to a set of coordinate axes as shown in figure 14.

The force fields describe atomic interactions with contributions from bonded (bond length, bond angle and bond torsion) and non-bonded (van der Waals and electrostatic) interactions. Several force fields have been developed such as AMBER, GROMOS, CHARMM and OPL.\(^{20,21}\)

4.13 Molecular Dynamics simulation studies of the open conformation of Cav1.2 calcium channel with Verapamil, D619, T13 and qDitiazem

The simulation setup of all ligands with open Cav1.2 is the same procedure as following. Molecular dynamics simulations have been carried out for the open conformation of Cav1.2 embedded in DOPC\(^{22}\) lipid bilayer. These simulations were performed with the Gromacs software version 4.0.4 using the Amber-03 force-field. The topology of ligands was generated with antechamber.\(^{23,24}\) The ligand charges were taken from the quantum chemical calculation (Gaussian 03) with the Hartree-Fock 3-21G basis set. The TIP3P water and 4 Ca2+ ions were placed along the z-axis. Cl- ions were added randomly within the solvent to neutralize the system. Snapshots of the trajectory were written out every 20 ps. The system was energy minimized with the steepest descent algorithm, followed by positional restrained MD for 2 ns. Subsequently, 40 ns of unrestrained MD simulation were carried out using the NVT ensemble. The V-rescale thermostat and Parrinello-Rahman barostat algorithms were used. Electrostatic interactions were calculated explicitly at a distance smaller than 1 nm, and long-range electrostatic interactions were calculated at every step by particle-mesh Ewald summation. Lennard-Jones interactions were calculated with a cut-off of 1 nm. All bonds were constrained by using the LINCS algorithm, allowing for an integration time step of 2 fs. The simulation temperature was kept constant by weakly \((\tau = 0.1 \text{ ps})\) coupling the lipids, protein, and solvent (water + counter ions) separately to a temperature bath of 300 K. The pressure was kept constant by weakly coupling
the system to a pressure bath of 1 bar with semi-isotropic pressure coupling. More details of parameter files are shown in Appendix I.

**4.14 Molecular Dynamics simulation studied of closed and open conformation Cav1.2 calcium channel in pure POPC and content of cholesterol in the membrane**

Molecular dynamics simulations of Cav1.2 in a closed and an open conformation have been carried out in pure POPC and POPC/CHOL environment. The system was set up by the following steps:

1. Generate the topology file of Cav1.2 of closed and improved conformation obtained from homology modeling.
2. For POPC environment (figure 15), replicate the starting configuration of a 128-lipid POPC bilayer in X and Y axis to create a bilayer of 512 lipids. For POPC/CHOL environment (figure 16), generate the mixture lipid bilayer POPC and 25% cholesterol from single POPC and cholesterol molecules. Randomly add cholesterol into the POPC bilayer.

Figure 15. The setup of the molecular dynamics simulation of Cav1.2 channel in pure POPC.
3. Superimpose the Cav1.2 channel with POPC and POPC/CHOL bilayer and use INFLATEGRO tool to expand the POPC for POPC system. In case of the POPC/CHOL system, a minimal number of overlapping POPC and CHOL molecules was removed.

4. Slowly compress only the POPC using the INGLATEGRO tool and POPC/CHOL using the tool included in Gromacs. Locate the channel in the center of bilayer and minimize the whole system. Repeat the compress and minimize whole system step until the area per lipid of POPC is around 65 Å².

5. Insert the 5 Ca²⁺ along the pore.

6. Solvate water TIP3P, add NaCl. Neutralize system by adding Na⁺ ion.

---

Figure 16. The setup of the molecular dynamics simulation of Cav1.2 channel in POPC/CHOL.
References


2. VLife MDS 3.5, Molecular design suite, VLife sciences technologies, Pvt. Ltd Pune, India..


