CHAPTER - 1

INTRODUCTION
Introduction

The new drug development is most tedious, time-consuming and cost-intensive process. Drug discovery is a linear, consecutive process that involves several steps includes target identification, target validation, assay development, lead identification, lead optimization, pre clinical *in-vitro* and *in-vivo* studies and clinical phase.¹ In pharmaceutical industry, the number of years to bring a drug from discovery to market is approximately 12 to 14 years and costing up to $1.2 to $1.4 billion dollars. Most of the clinical candidates were discontinued because of poor pharmacokinetics, lack of efficacy, animal toxicity, and severe adverse effects in humans. Now a days, the drug discovery process has been revolutionized the advent of genomics, proteomics, bioinformatics and combinatorial chemistry, high throughput screening (HTS) and molecular modeling.¹ The role of molecular modeling in drug discovery process was represented in Fig. 1.1.
**MOLECULAR MODELING**

Molecular modeling is branch science that elucidates and validates experimental evidence through imagination, visualization, and rationalization. It uses sophisticated computers and computational methods permeate all aspects of drug design.

**Molecular modeling strategies**

The molecular modeling strategies vary depending upon the extent of structural information available for target (enzyme/receptor) and ligands. The Structure-based design, (SBD) and Ligand-based design (LBD) are two major modeling strategies currently used in the drug design process. In SBD, the three-dimensional features of the target are directly considered. In LBD approach the design is based on comparative analysis of the structural features of known active and inactive compounds. The molecular modeling strategies, techniques, capabilities and its success in drug discovery process were shown in Fig.1.2 to Fig 1.7.

**Molecular modeling techniques**

- **Structure-based design (SBD)**
  1. Sequence analysis and Homology modeling
  2. Active site analysis and Docking studies
  3. Core Hopping

- **Ligand-based design (LBD)**
  1. Phamacophore Modeling and 3D-QSAR analysis
  2. Combinatorial Library generation (CombiLib)
  3. Virtual screening
Fig. 0.2 Molecular modeling strategies in the drug discovery process

Fig. 0.3 Molecular modeling techniques used in drug design
Molecular modeling capabilities and its success in drug discovery

**Fig. 0.4** Molecular modeling capabilities

**Fig. 0.5** Success of Molecular modeling in drug discovery
QUANTITATIVE STRUCTURE ACTIVITY RELATIONSHIP (QSAR)

The pharmacophore based 3D-QSAR approach is used for deriving a QSAR model, which involves the generation of a robust pharmacophore hypothesis, followed by QSAR modeling using a training/test set structures that match the pharmacophore. The pharmacophore models provide a rational hypothetical picture of the primary chemical features that are responsible for activity. The atom-based 3D-QSAR is a rational drug design approach provides new insight into the dataset of molecules by correlating the biological activity with steric, electrostatic, hydrophobic and hydrogen bond fields sampled at lattice points of 3D-grid. The regression maps for the atom based 3D-QSAR model represented by color codes according to the sign of their coefficient values. The positive coefficients indicate an increase in activity; negative coefficients indicate decrease in activity. The 3D-QSAR regression maps provide valuable structural insights for better ligand design. The steps involved in 3D-QSAR modeling were shown in Fig. 1.6.

Fig. 0.6 Steps involved in 3D-QSAR modeling
HOMOLOGY MODELING

The human genome project estimates about 40000 proteins out of which typically 1400 are considered to be therapeutically relevant. Out of these 1400 targets, only 250 have crystal structure reported. The aim of protein homology modeling/comparative modeling is to predict a structure from its sequence with an accuracy that is comparable to the best results achieved experimentally. Homology modeling helps to build a three dimensional structure of a target protein based on template and used to identify the putative active sites and binding pockets, which further delves the probable ligand-protein interactions to understand the exact mechanism/function of the particular protein.

A typical homology modeling involves the following steps

- Identifying the target and template proteins
- Determine sequence similarity and align the sequences
- Identify structurally conserved and structurally variable regions
- Generate coordinates for core (structurally conserved) residues
- Generate conformations for the loops (structurally variable)
- Build the side chain conformations
- Refine and evaluate the target protein

Several programs such as PROCHECK, WHAT IF and verify 3D are used to evaluate and refine the structure of a model protein. The major parameters are considering in validation of protein includes stereochemical accuracy (bond length, bond torsion angles, chirality of Cα atoms, planarity of amide bonds), packing quality of secondary elements (α-α,β-β, α-β etc., inspecting buried residues) and folding reliability (comparing overall fold of template and model). Ramachandran plot is one of most
important validation tool used to check the quality of protein structure. The lines in the Ramachandran plot indicate preferred areas for phi and psi torsional angles of the protein backbones. The orthogonal box regions correspond to conformations where there are no steric clashes. Outside the orthogonal box areas shows the allowed region if slightly shorter van der Waals radii are used in the calculation, i.e. the atoms are allowed to come a little closer together. This brings out an additional region, which corresponds to the left-handed alpha helix. The dihedral angles $\Psi$ and $\Phi$ defined by the peptide main chain. $\Phi$ is the angle between the two planes defined by $C_{i-1}-N_i-C^\alpha_i$ and $N_i-C^\alpha_i-C_i$, whereas $\Psi$ is the angle between the two planes of $N_i-C^\alpha_i-C_i$ and $C^\alpha_i-C_{i+1}$.

**CORE HOPPING**

The aim of the structure based core hopping is to discover structurally novel diverse compounds starting from known active compounds by modifying the central core structure of the molecule and explore ligand-receptor interactions in the protein active site. Core hopping allows for the rapid screening of novel cores to help overcome unforeseen toxicity, selectivity and other unsuitable physicochemical properties by generating new lead compounds with improved core properties while preserving key R-group interactions.

**DOCKING**

Docking is a SBD technique aim to identify correct poses of ligands in the binding pocket of a protein and to predict the affinity between the ligand and the protein. Glide (Grid-based Ligand Docking with Energetics) module of Schrödinger suite searches favorable interactions between ligands and a receptor and generate possible ligand posed
depends on position and orientation of a ligand relative to the receptor. These ligand poses pass through a series of hierarchical filters include

- Stage 1. Site-point search
- Stage 2: Diameter test, Subset test, Greedy score, Refinement
- Stage 3. Grid minimization and final scoring (Glide Score)

The final GlideScore includes a steric-clash term and adds buried polar terms to penalize electrostatic mismatches used to prioritize the docked poses.

GScore = 0.065*vdW + 0.130*Coul + Lipo + Hbond + Metal + BuryP + RotB + Site

Where, vDW is van der Waals energy; Coul is coulomb energy; Lipo is lipophilic contact term, Metal is metal-binding term, BuryP is penalty for buried polar groups, RotB is penalty for freezing rotatable bonds and site is for polar interactions in the active site. The basic principles of molecular docking was shown in Fig.1.7.

![Principles of molecular docking](image_url)

**Fig. 0.7 Principles of molecular docking**