Chapter 7
Docking Studies

7.1 DRUG DESIGN

Drug discovery is a structured investigation partially associated with luck. It requires teamwork, the members of the team being specialists in various fields, such as medicine, biochemistry, chemistry, computerized molecular modeling, pharmaceutics, pharmacology, microbiology, toxicology, physiology and pathology.

7.1.1 GENERAL STEPS IN THE SYNTHESIS OF A NEW COMPOUND

- Assessment of the biochemical and biological processes of the disease and/or its cause.
- Basic research into the disease process and its causes
- Initial biological and toxicological testing, synthesis of analogues
- Team decides the structure of a suitable lead compound.
- Design of the synthetic pathway to produce the lead compound.
- Synthesis of analogues

7.1.2. STEREOCHEMISTRY AND DRUG DESIGN

It is well established that the structure of a molecule is normally one of the most important factors affecting drug activity. Consequently, the structure and overall shape of a molecule is an important consideration in designing an analogue. One must take into account all these stereo chemical features when proposing structures for potential leads and analogues.

7.1.3. THE TYPE OF GROUP

The incorporation of polar groups, such as the hydroxyl, amine, amide, carboxylic acid, sulphonic acid and phosphorus oxyacid groups, which either ionize or are capable of producing relatively strong intermolecular forces of attraction with water (hydrogen bonding), will usually result in analogues with an increased water solubility.
Acidic and basic groups are particularly useful, since these groups can be used to produce salts, which would give a wider range of dosage forms for the final product. However, the formation of zwitter ions can reduce water solubility. Introduction of weakly polar groups, such as carboxylic acid esters, aryl halides and alkyl halides, will not significantly improve water solubility and can result in enhanced lipid solubility. In all cases, the degree of solubility obtained by the incorporation cannot be accurately predicted since it also depends on other factors.

7.2  THE SARAND QSAR APPROACHES TODRUGDESIGN

7.2.1  STRUCTUREACTIVITY RELATIONSHIP (SAR)

Compounds with similar structures to a pharmacologically active drug are often themselves biologically active (Torres Piedra M et al., 2010). This activity may be either similar to that of the original compound, but different in potency and unwanted side effects or completely different to that exhibited by the original compound. These structurally related activities are commonly referred to as structure–activity relationships (SAR). A study of the structure–activity relationships of a lead compound and its analogues can be used to determine the parts of the structure of the lead compound that are responsible for both its beneficial biological activity, (its pharmacophore), and also the side effects.

At this stage, \textit{in silico} method is the right tool for further studies. The information obtained can be used to develop a new drug that has increased activity, a different activity from an existing drug and fewer/less side effects. Structure–activity relationships are usually determined by making minor changes to the structure of a lead to produce analogues and assessing the effect of these structural changes on biological activity. The investigation of numerous lead compounds and their analogues have made it possible to make some broad generalizations about the biological effects of specific types of structural change. These changes can be conveniently classified as:

1. The size and shape of the carbon skeleton.
2. The nature and degree of substitution.
3. The stereochemistry of the lead.
7.2.2 QUANTITATIVE STRUCTURE ACTIVITY RELATIONSHIP (QSAR)

QSAR is an attempt to identify the probably more effective structure by establishing a mathematical relationship in the form of an equation between biological activity and measurable physicochemical parameters. These parameters are used to represent properties such as lipophilicity, shape and electron distribution which are believed to have a major influence on the activity of the drug. They are normally defined so that they are in the form of numbers, which are derived from practical data that is thought to be related to the property the parameter represents. This makes it possible either to measure or to calculate these parameters for a group of compounds and relate their values to the biological activity of these compounds by means of mathematical equations using statistical methods such as regression analysis. These equations can be used by the medicinal chemist to decide on which analogues are to be prepared.

7.2.3 MOLECULAR DESCRIPTORS ANALYSIS

A descriptor is a number that describes a particular molecular property of a drug molecule. The main properties of a drug that appear to influence its activity are its, lipophilicity, the electronic effects within the molecule and the size and shape of the molecule (steric effects). Lipophilicity is a measure of the solubility of a drug molecule in lipid membranes. This is usually an important factor in determining how easily a drug passes through lipid membranes.

The electronic effects of the groups within the molecule will affect its electron distribution, which in turn has a direct bearing on how easily and permanently the molecule binds to its target molecule.

Drug size and shape will determine whether the drug molecule is able to get close enough to its target site in order to bind or not [http://www.rscb.org/pdb/home](http://www.rscb.org/pdb/home) (Ramanathan et al., 2008), (Bender A et al., 2009). The parameters commonly used to represent these properties are:

- Hydrophobicity/ lipophilicity (Log P): It is the molecular hydrophobicity, which effects drug interactions, metabolism of molecules as well as toxicity. Partition coefficient is determined using octanol/water system to measure the log P value.
High log P value indicates hydrophobicity of the compounds and low value indicates hydrophilicity. Normally it is considered that, log P value close to 2 indicates the ability of drug to reach CNS.

- **Electronic effects**: The important electronic parameters are pKa, Hammet constant (σ) and polarisability. The electronic effects of various substituents have an effect on the ionization and polarity of the drug. Hammet constant can be evaluated by measuring pKa values.

- **Steric effects**: The steric effects determine whether the drug molecule is able to get close enough to its target site in order to bind or not. The significant steric parameters are molar refractivity, molar volume, Taft’s steric constant etc.

- **Molecular volume**: determines the transport characteristics of molecules, such as intestinal absorption or blood brain barrier penetration. Various methods can be used to calculate molecular volume, including methods requiring generation of fragment contribution methods such as McGowan volume approximation.

- **Total molecular polar surface area (TPSA)**: QSAR derived equations take the general form: It is a parameter used to predict the drug transport properties. Polar surface area is defined as a sum of surfaces of polar atoms (usually oxygen, nitrogen and attached hydrogen) in that molecule. It is calculated based on the method published by Ertl et al as a sum of fragment contributions. TPSA has been shown to be a very good descriptor characterizing drug absorption including intestinal absorption and blood brain barrier penetration. Molecules with a polar surface area greater than 140 Å squared are poor at permeating cell membranes. Only molecules which have PSA less than 60 Å squared can cross the blood brain barrier.

- **Number of rotatable bonds (NROTB)**: This simple topological parameter is a measure of molecular flexibility. Rotatable bond is defined as any single non ring bond bound to non-terminal heavy (ie. non-hydrogen atom). Amine C-N bonds are not considered because of their high rotational energy barrier.
Biological activity = function \{\text{parameter(s)}\} in which the activity is normally expressed as log[1/(concentration term)], usually C, the minimum concentration required to cause a defined biological response.

7.3 LIPINSKI’S RULE OF 5

It provides a method of assessing the likelihood that a given molecule could be orally bioavailable, based on a series of physicochemical parameters, not more than one of which should be violated. Molecules violating any of these rules will have a problem with bioavailability (Lipinski C A et al., 2001).

The rule of five is so called because the numbers involved are multiples of 5. In general, an orally active drug has:

- Not more than 5 hydrogen bond donors (nitrogen or oxygen atoms with one or more hydrogen atoms)
- Not more than 10 hydrogen bond acceptors (nitrogen or oxygen atoms)
- A molecular weight not greater than 500 Daltons
- An octanol-water partition coefficient log P not greater than 5

7.4 IN SILICO DRUG DESIGN PROGRAMS AND RESOURCES

Several computational chemistry approaches are available based on the target protein structures and known active ligands.

7.4.1 COMPUTER AIDED DRUG DESIGN

Computer-aided drug design uses computational chemistry to discover, enhance, or study drugs and related biologically active molecules (Milano Chemometrics and QSAR Research Group). The most fundamental goal is to predict, whether a given molecule will bind to a target and if so how strongly. Molecular mechanics or molecular dynamics are most often used to predict the conformation of the small molecule and to model conformational changes in the biological target that may occur when the small molecule binds to it (Veer D F et al., 2002).

Drug design with the help of computers can be used at any one or more of the following stages of drug discovery:
• Hit identification using virtual screening (structure or ligand-based design)

• Hit-to-lead optimization of affinity and selectivity (structure-based design, QSAR, etc.).

7.5 TYPES OF DRUG DESIGN

There are three major types of drug design: Ligand-based drug design, receptor-based drug design and combinatorial based approach (Schrodinger Suite, 2012).

7.5.1. RECEPTOR BASED APPROACH

If both the receptor and ligand structures are known, the receptor-based docking approach is the most ideal situation. The ligand can be docked into the known receptor site and molecular mechanics can be used to simulate receptor-ligand interactions and dynamics. One can address highly specific receptor–ligand interactions using these techniques, but even then, there can be surprises in the form of alternative modes of binding and conformational changes in the receptor structure. It is initially necessary to study the binding of thousands of compounds at low resolution rather than fewer compounds at high resolution. Often this is the case when searching for lead compounds. The GRID approach maps the binding site by superimposing a grid and calculating the electrostatics, hydrogen-bonding sites, or lipophilicity at each grid point. Among the docking-related programs discussed above, classical DOCK program uses a similar approach for scoring and also provides other tools for scoring compounds, attempting alternate orientations, and performing geometry optimization to include the flexibility of the ligand. A grid-based approach significantly reduces the time required in computer studies. Several common programs, including AUTODOCK, Glide, LUDI, LigandFit, and FlexX/CSCores, also have been widely used in ligand docking and lead screening studies.

7.5.2 LIGANDBASED APPROACH

In the case, when ligand structures are known but the receptor structure is unknown, a ligand-based approach is used. This situation represents the most common case. An extension of the QSAR approach is taken to study the active ligands, also known as pharmacophore-based drug design. The pharmacophore refers to an ensemble of steric and electronic features (or 3D arrangement of the functional groups)
that enables a molecule to exhibit a specific biological activity. The ligand-based approach is basically an indirect method of pharmacophore identifications.

In general, pharmacophore-based virtual screening techniques depend on the application of descriptors of molecular structure and properties. They include

1. Structure or descriptor-based queries (or similarity search in two-dimensional (2D) and three dimensional (3D) substructures or pharmacophore models)

2. Fingerprint queries (or binary bit string-search based on Tanimoto coefficient, 2D/3D pharmacophore fingerprints)

3. Clustering and partitioning (or intermolecular distances in chemical reference spaces and reference frame).

7.5.3 COMBINATORIAL BASED APPROACH

Virtual combinatorial chemistry approaches are used when both the receptor and ligand structures are unknown. In this case, computational chemistry is used to generate structures and in parallel to perform chemical similarity and diversity search analysis before and after combinatorial chemistry-based experimental HTS. E.g. the combinatorial chemistry software package from Accelrys and Tripos companies provides the unique features to preanalyze and select diverse building blocks for the target libraries to help discovery of small molecule drugs. This promises a maximum range of drug-like compounds and increases the potential of finding active leads for the targets.

7.6 DOCKING

In the field of molecular modeling, docking is a method which predicts the preferred orientation of one molecule to a second molecule/receptor protein when bound to each other to form a stable complex. Knowledge of the preferred orientation in turn may be used to predict the strength of association or binding affinity between two molecules using scoring functions. Therefore docking is useful for predicting both the strength and type of bond produced and in turn the stability/efficiency of the new complex structure formed.
The aim of molecular docking is to achieve an optimized conformation for both the protein and ligand and relative orientation between protein and ligand such that the free energy of the overall system is minimized. Two approaches are particularly popular within the molecular docking community. One approach uses a matching technique that describes the protein and the ligand as complementary surfaces. The second approach simulates the actual docking process in which the ligand-protein pair wise interaction energies are calculated. In docking studies, the first requirement is the identification of the protein of interest and its structure. Usually the structure has been determined using a biophysical technique such as x-ray crystallography, or less often, NMR spectroscopy. This protein structure and a database of potential ligands serve as inputs to a docking program. Protein data bank of RSC Britain is a reliable and rich source of protein collection (http://www.rscb.org/pdb/home).

Materials and methods

7.7 SOFTWARE USED

Autodock 4.2

Method followed: Ligand based

Selection of protein from PDB of Royal society. Protein relevant to insulin receptor was selected.

Preparation of protein: performed by minimizing energy, elimination of unbound water, deletion of hetero atoms and finally checking of amino acid orientation.

Selected protein: 3SZ1

Ligand preparation: The molecule (Ligand) was drawn in 2D sketches, transformed to 3D structure and the charts of the ligands were prepared. Then local energy minimization with the help Algorithm software was followed.

The grid was generated as per the standard procedure. (Schrodinger Suite, 2012)

After the above steps the Global docking concept was followed.
## RESULTS

**Table 7.1** Glide score of molecules present in CMRH

<table>
<thead>
<tr>
<th>Sl.No</th>
<th>Name and structure of compound</th>
<th>Glide score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><img src="image" alt="Quercetin" /> Quercetin</td>
<td>-11.0796</td>
</tr>
<tr>
<td>2</td>
<td><img src="image" alt="Rutin" /> Rutin</td>
<td>-6.8463</td>
</tr>
<tr>
<td>3</td>
<td><img src="image" alt="Gallic acid" /> Gallic acid</td>
<td>-7.4402</td>
</tr>
<tr>
<td>Sl.No</td>
<td>Name and structure of compound</td>
<td>Glide score</td>
</tr>
<tr>
<td>-------</td>
<td>-------------------------------</td>
<td>-------------</td>
</tr>
<tr>
<td>1</td>
<td>[Structural diagram of Quercetin]</td>
<td>-11.0796</td>
</tr>
<tr>
<td>2</td>
<td>[Structural diagram of Rutin]</td>
<td>-6.8463</td>
</tr>
<tr>
<td>3</td>
<td>[Structural diagram of Apigenin]</td>
<td>-7.63314</td>
</tr>
<tr>
<td>4</td>
<td>[Structural diagram of Kaempferol]</td>
<td>-7.57302</td>
</tr>
</tbody>
</table>

Table 7.2 Glide score of molecules present in HALH
Table 7.3 Prediction of ADME profile of selected analogues

<table>
<thead>
<tr>
<th>Compound</th>
<th>H oral Abs</th>
<th>% Horal Abs</th>
<th>QlogKhsa</th>
<th>QPPCaco</th>
<th>QPlogBB</th>
<th>QPlogKp</th>
<th>QPlogHERG</th>
<th>QPlogS</th>
<th>Dipole</th>
<th>SASA</th>
<th>FOSA</th>
<th>FISA</th>
<th>#metab</th>
<th>QPPMDCK</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quercetin</td>
<td>2</td>
<td>53.65</td>
<td>-0.318</td>
<td>21.04</td>
<td>-2.42</td>
<td>-5.35</td>
<td>-5.36</td>
<td>-3.096</td>
<td>0</td>
<td>531.59</td>
<td>0</td>
<td>286.86</td>
<td>5</td>
<td>7.62</td>
</tr>
<tr>
<td>Rutin</td>
<td>1</td>
<td>4.99</td>
<td>-1.18</td>
<td>1.416</td>
<td>-4.06</td>
<td>-6.97</td>
<td>-4.47</td>
<td>-1.681</td>
<td>0</td>
<td>751.66</td>
<td>208.4</td>
<td>404.5</td>
<td>12</td>
<td>0.41</td>
</tr>
<tr>
<td>Gallic acid</td>
<td>2</td>
<td>41.49</td>
<td>-0.983</td>
<td>9.96</td>
<td>-1.67</td>
<td>-5.88</td>
<td>-1.43</td>
<td>-0.707</td>
<td>0</td>
<td>342.67</td>
<td>0</td>
<td>253.2</td>
<td>3</td>
<td>4.32</td>
</tr>
<tr>
<td>Apigenin</td>
<td>3</td>
<td>73.52</td>
<td>-0.029</td>
<td>116.38</td>
<td>-1.45</td>
<td>-3.96</td>
<td>-5.15</td>
<td>-3.372</td>
<td>0</td>
<td>492.56</td>
<td>0</td>
<td>203.5</td>
<td>3</td>
<td>48.38</td>
</tr>
<tr>
<td>Kaempferol</td>
<td>3</td>
<td>64.71</td>
<td>-0.198</td>
<td>58.25</td>
<td>-1.80</td>
<td>-4.54</td>
<td>-5.08</td>
<td>-3.051</td>
<td>0</td>
<td>500.79</td>
<td>0</td>
<td>235.2</td>
<td>4</td>
<td>22.89</td>
</tr>
</tbody>
</table>
7.8 DRUG MOLECULE-RECEPTOR PROTEIN INTERACTION - RIBBON STRUCTURE

**Fig. 7.1** Quercetin

**Fig. 7.2** Rutin

**Fig. 7.3** Gallic acid

**Fig. 7.4** Apigenin
**Fig. 7.5** Kaempferol

7.9 LIGAND MOLECULE INTERACTION DIAGRAM

**Fig. 7.6** Quercetin
Docking Studies

**Fig. 7.7** Rutin

**Fig. 7.8** Gallic acid
Docking Studies

Fig. 7.9 Apigenin

Fig. 7.10 Kaempferol
7.10 DISCUSSION

In the docking studies, the glide score of all the molecules were found to be in the range of -6 to -11. According to the principle of docking, values more than -5 is indicative of good activity. Out of the five molecules studied only rutin violated one among five rules of Lipinski’s rule of five. Thus, it was concluded that the results are promising.

Upregulation of PPARγ is happening through the 3SZ1 receptor selected for docking studies. Upregulation of PPARγ is one among the different mechanisms through which antidiabetic action is produced without insulin (Panda S, Kar A, 2007), (Mansor F et al., 2013). ME-CMRH fraction demonstrated PPARγ upregulation in the antidiabetic mechanism identification studies.

******