Chapter-1

General Introduction
1.1 General

1.1.1 Nitric Oxide

Nitric oxide synthases catalyze the synthesis of nitric oxide. Nitric oxide (NO) is a free radical gas commonly referred to as nitrogen monoxide. Initially, it was identified by Robert F. Furchgott’s et al as a chemical responsible for vasodilation activity in vascular epithelium, called as endothelium-derived relaxing factor (EDRF) [1-3]. It is a highly chemically reactive environmental pollutant. In 1988, it was demonstrated that arginine is the biochemical precursor for NO [4-6]. Nitric oxide acts as an important messenger molecule in various pathological and physiological processes of the mammals. Lower levels of NO production are important to protect an organ (such as liver) from ischemic damage [7]. The synthesis of NO is regulated by the enzymes, nitric oxide synthases (NOSs), which are flavo-hemoproteins in nature. NO have dual biological activities: (1) it has a key signaling molecule for neurotransmission and vasodilation at low concentrations, and (2) it acts as a defensive cytotoxin at higher concentrations [8]. NO has a critical place in the fields of neuroscience, physiology, and immunology [9]. The half-life of NO is estimated to be about 30 seconds or less. Nitric oxide may inhibit platelet adhesion, aggregation, agglutination and severe left ventricular dysfunction [10-11]. This compound has also reported some role in the regulation of seminiferous tubules contraction, androgen biosynthesis, blood flow, and germ cells metabolism/apoptosis in testis [12]. Recently, the NO supplement is found to be used in body building, because it is helpful in internal wound and muscles injury healing [13]. NO plays the following roles in the nervous, immune and cardiovascular systems.
Along with positive beneficial effects, NO has various negative effects, too. NO and its derivative peroxynitrite can cause DNA damage, activate poly [ADP-ribose] polymerase-1 (PARP-1), and induce apoptosis-inducing factor (AIF) translocation. NO can also modify protein functions by nitration and nitrosylation, resulting in cell damage and death. Its involvement in several neurodegenerative diseases, such as Parkinson's disease, Alzheimer's disease, stroke, multiple sclerosis, etc., is a matter of great concern. The over production of nitric oxide causes some dysfunction of organs and affects lactate level, resulting in acute pulmonary toxicity, shortness of breath, hypoxemia, pulmonary edema, and even death [14-15]. The beneficial and negative effects are illustrated in Table 1.1.

### Table 1.1 The role of nitric oxide in various physiological and pathological processes.

<table>
<thead>
<tr>
<th>Physiological processes (Positive effects)</th>
<th>Pathological processes (Negative effects)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vasodilation</td>
<td>Septic shock</td>
</tr>
<tr>
<td>Neurotransmission</td>
<td>Rheumatoid arthritis</td>
</tr>
<tr>
<td>Macrophage-mediated cytotoxicity</td>
<td>Inflammatory bowel disease</td>
</tr>
<tr>
<td></td>
<td>Pulmonary fibrosis</td>
</tr>
<tr>
<td>Bronchodilation</td>
<td>Multiple sclerosis</td>
</tr>
<tr>
<td></td>
<td>Congestive heart failure</td>
</tr>
<tr>
<td></td>
<td>Diabetes</td>
</tr>
<tr>
<td></td>
<td>Atherosclerosis</td>
</tr>
<tr>
<td></td>
<td>Ischemia-reperfusion injury</td>
</tr>
<tr>
<td></td>
<td>Asthma</td>
</tr>
<tr>
<td></td>
<td>Alzheimer’s disease</td>
</tr>
<tr>
<td></td>
<td>Huntington’s disease</td>
</tr>
<tr>
<td></td>
<td>Parkinson’s disease</td>
</tr>
</tbody>
</table>
1.1.2 Nitric Oxide Synthase Enzyme

Nitric oxide synthases (NOSs) catalyze the production of nitric oxide (NO) from L-arginine [16-18]. They exist mainly in three isoforms: neuronal, inducible, and endothelial. Neuronal NOS (nNOS), also known as nitric oxide synthase 1 (NOS-1), is the first NOS to be isolated. It generates the NO in nervous tissues of nervous system (central as well as peripheral) and in the skeletal muscles. Neuronal NOS also play a role in cell communication (communications between cells) and is associated with plasma membranes. The 7-nitroindazole is a specific inhibitor for nNOS. The inducible NOS (iNOS), also called nitric oxide synthase 2 (NOS-2), is found in cytosol of cells. The iNOS has ability to protect the body from invading pathogens as NO interferes the DNA of host cells. These properties describe the role of iNOS in various fields such as host immunity, anti-tumor activities and enabling its participation in anti-microbial in the form of oxidative burst of macrophages. The endothelial NOS (eNOS) enzyme or called as nitric oxide synthase 3 (NOS-3) produce the nitric oxide (NO) in blood vessels. It is involved in regulating vascular function such as vasodilation in smooth muscles. It is also responsible for inhibition of platelets aggregation and in wounds healing. A constitutive Ca$^{2+}$-dependent NOS provides a basal release of NO. The structures of all these three isoforms are given in Fig. 1.1 [19]. All the three isoforms of NOS are homodimer and hemoprotein in nature. The molecular weight of their each monomer is in between 110 kDa to 160 kDa, depending on the isofrom. Each monomer contains one oxidase domain, called amino-terminal end, and one reductase domain, called carboxy-terminal end [20].
Fig. 1.1 The schematic diagram of structures of three isoforms of NOS

All nitric oxide synthases exhibit a bi-domain structure as shown in Fig. 1.1, in which the N-terminal or oxygenase domain is present which is responsible for binding of haem, BH$_4$ and L-arginine [19]. It is linked by a CaM-recognition site to another C-terminal or reductase domain that is responsible for binding of FAD (flavin adenine dinucleotide), FMN (flavin mononucleotide) and NADPH (nicotinamide adenine dinucleotide phosphate). All three isoforms of NOS share a carboxyl-terminal reductase domain homologous to the cytochrome P$_{450}$ reductase. They also share an amino-terminal oxygenase domain containing a haem prosthetic group, which is linked in the middle of the protein to a calmodulin-binding domain. Binding of calmodulin appears to act as a “molecular switch” to enable electron flow from flavin prosthetic groups in the reductase domain to haem [21]. The three isoform of NOS are distributed in human cells as shown in Table 1.2
**Table 1.2:** Cellular distribution of NOS isoforms

<table>
<thead>
<tr>
<th>Types of NOS</th>
<th>Occurrence</th>
</tr>
</thead>
<tbody>
<tr>
<td>nNOS</td>
<td>Neurons, skeletal muscle, hepatocytes, keratinocytes, macula densa, neutrophils, gastrointestinal smooth muscle, tubular epithelium, vascular smooth muscle cells, cardiomyocytes.</td>
</tr>
<tr>
<td>iNOS</td>
<td>Constitutive expression: hepatocytes, neurons, airway epithelium, cortical tubules, keratinocytes colon mucosae. Inducible expression: vascular smooth muscle cells, macrophages and airway smooth muscle cells, lung fibroblasts, alveolar macrophages, chondrocytes, Kupffer cells, endothelial cells, mast cells, skeletal muscle, type II epithelial cells, neutrophils.</td>
</tr>
<tr>
<td>eNOS</td>
<td>Endothelial cells, eosinophils, epithelial cells of human, bronchial epithelial cells, nasal mucosa, fibroblasts, gastrointestinal mucosae, hepatocytes, lymphocytes, neutrophils, syncytiotrophoblasts of human placenta, skeletal muscle, type II alveolar cells</td>
</tr>
</tbody>
</table>

### 1.1.3 Synthesis of NO

These NOSs can synthesize the nitric oxide in any part of body from the amino-acid L-arginine (ARG) illustrated in **Fig 1.2**.
The electron flow in the NO synthase reaction: Electrons are donated by NADPH to the reductase domain of the enzyme and proceed to the oxygenase domain via FAD and FMN redox carriers. In oxygenase domain, they interact with the haem iron and BH$_4$ at the active site to catalyse the reaction of oxygen with L-arginine, generating citrulline and NO as products (NADPH → FAD → FMN → Haem → O$_2$). Electron flow through the reductase domain requires the presence of bound Ca$^{2+}$/CaM (Fig. 1.3) [19]. This pathway has been found useful for characterization and localization of arginine and tetrahydrobiopterin (BH$_4$) sites, and their interaction.
1.1.4 Pathogenesis

All three Nitric oxide synthase enzyme isoforms have some pathophysiological or physiological implications in the cardiovascular system of human body. The eNOS plays a major physiological role in keeping the vasculature in the dilated form, preventing from platelet aggregation, vascular homeostasis, leukocyte adhesion, and preventing smooth muscle proliferation. The nNOS may also contribute to vascular disease associated with hypercholesterolemia, blood pressure regulation, hypertension and diabetes, due to the reduced endothelium-mediated vasodilatation [22] and endothelial dysfunction. The inactivation of NO by superoxide anions and the oxidative stress plays the vital role for the above mentioned disease conditions [23, 24]. The endothelial dysfunction in human beings and animals can be improved by giving supplements of L-arginine (NOS substrate). In addition, the NOS cofactor (6R)-5, 6, 7, 8-tetrahydrobiopterin helps in improving the endothelium mediated vasodilatation during cerebrovascular stroke [25]. Inducible NOS and neuronal NOS play an important role in neurodegeneration, but, endothelial NOS has the important role in maintaining cerebral blood flow as well as preventing from the neuronal injury [26]. The synthesis of nitric oxide by three NOS isoenzymes is shown in Fig. 1.4.
Fig. 1.4 Nitric oxide synthesis by NOS isoenzymes: (a) NO synthesis by eNOS in the vascular endothelial cell stimulated by acetylcholine; (b) NO synthesis by nNOS in the neuronal dendrite stimulated by glutamate; and (c) NO synthesis by iNOS in the macrophage following induction of iNOS mRNA and enzyme by cytokines [27].

1.1.5 Nitric Oxide Synthase Inhibitors

Inhibitors of NOSs inhibit the synthesis of NO by the competitive inhibition of NOSs in mammalian [28]. Some specific NOS inhibitors are as shown in Table 1.3, which are $N^\omega$-monomethyl-L-arginine (L-NMMA) and $N^\omega,N^\omega$-dimethyl-L-arginine (L-ADMA) [29], both naturally occurring compounds derived from methylated protein residues during protein turnover, $N^\omega$-nitro-L-arginine and L-NAME which are methyl ester of L-NNA, $N^\omega$-iminoethyl-L-ornithine (L-NIO), some selective inhibitors like L-canavanine, and $N^\omega$-amino-L-arginine that inhibits the inducible macrophage NO synthases more potently than the constitutive brain and endothelial NO synthases.
Some nitric oxide inhibitors also act as antidepressants. They are, for example, L-N\textsubscript{G}Propyl-L-arginine (N-PLA), N\textsubscript{G}.N\textsubscript{G}-Dimethyl-L-arginine (ADMA), L-N6(1-iminoethyl)lysine (L-NIL), L-Thiocitrulline, S-Methyl-L-Thiocitrulline, Agmatine (1-Amino-4-guanidinobutane), Ethyl-L-NIO, Vinyl-L-NIO, 7-Nitroindazole (7-NI), 7-Nitro-3-bromo-indazole, 1-[2-Trifluoromethyl]phenyl]-imidazole, 2-Imino-4-methylpipridine, Aminoguanidine, S-(2-Aminoethyl)isothiourea, S,S-(1,3-Phenylene-bis(1,2-ethanediyl)bis-isothiourea (1,3-PBIT), S,S-(1,4-Pheny-lene-bis(1,2-ethanediyl)bis-isothiourea) [31].

Table 1.3: Nitric oxide synthase inhibitors.

<table>
<thead>
<tr>
<th>Name/ Abbreviation</th>
<th>Structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>L-NMMA</td>
<td><img src="image" alt="L-NMMA_structure" /></td>
</tr>
<tr>
<td>L-ADMA</td>
<td><img src="image" alt="L-ADMA_structure" /></td>
</tr>
<tr>
<td>L-NAME</td>
<td><img src="image" alt="L-NAME_structure" /></td>
</tr>
<tr>
<td><strong>L-NIO</strong></td>
<td>![L-NIO structure]</td>
</tr>
<tr>
<td>-----------</td>
<td>-------------------</td>
</tr>
<tr>
<td>L-Canavanine</td>
<td>![L-Canavanine structure]</td>
</tr>
<tr>
<td><strong>N-PLA</strong></td>
<td>![N-PLA structure]</td>
</tr>
<tr>
<td><strong>ADMA</strong></td>
<td>![ADMA structure]</td>
</tr>
<tr>
<td><strong>L-NIL</strong></td>
<td>![L-NIL structure]</td>
</tr>
<tr>
<td><strong>L-Thiocitrulline</strong></td>
<td>![L-Thiocitrulline structure]</td>
</tr>
<tr>
<td><strong>Agmatine</strong></td>
<td>![Agmatine structure]</td>
</tr>
<tr>
<td>Name</td>
<td>Structure</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>------------------------------------------------</td>
</tr>
<tr>
<td>Ethyl-L-NIO</td>
<td><img src="image" alt="Ethyl-L-NIO" /></td>
</tr>
<tr>
<td>Vinyl-L-NIO</td>
<td><img src="image" alt="Vinyl-L-NIO" /></td>
</tr>
<tr>
<td>7-Nitroindazole</td>
<td><img src="image" alt="7-Nitroindazole" /></td>
</tr>
<tr>
<td>7-Nitro-3-bromoindazole</td>
<td><img src="image" alt="7-Nitro-3-bromoindazole" /></td>
</tr>
<tr>
<td>1-[2-Trifluoromethylphenyl]-imidazole</td>
<td><img src="image" alt="1-[2-Trifluoromethylphenyl]-imidazole" /></td>
</tr>
<tr>
<td>Aminoguanidine</td>
<td><img src="image" alt="Aminoguanidine" /></td>
</tr>
<tr>
<td>S-(2-Aminoethyl)isothiourea</td>
<td><img src="image" alt="S-(2-Aminoethyl)isothiourea" /></td>
</tr>
</tbody>
</table>
1.2 Quantitative Structure-Activity Relationship

Quantitative structure–activity relationship (QSAR) study is a method for developing mathematical or computational models. It is used to find out a statistically significant correlation between the structure and function by using a chemometric technique. The QSAR studies help to accelerate the drug discovery process at this stage by reducing wastage of money and particularly the time. The basis of this approach is that there is a relationship between structural or physicochemical properties of a compound and its biological activity. On that basis, an equation is generated which correlates biological activity of compounds with their physicochemical parameters [32]. The physiochemical descriptors such as hydrophobicity, electronic, and steric effects are calculated empirically or by using various latest computational methods. A QSAR generally express its result in the form of a linear equation as:

$$\text{Biological activity} = (P_1 \times C_1) + (P_2 \times C_2) + \ldots + (P_n \times C_n) + \text{constant} \quad \text{(Eq. 1.1)}$$

Where, the parameters $P_1$ to $P_n$ are calculated for every molecule present in the series. The coefficients $C_1$ to $C_n$ are calculated by putting the changes in the parameters and the biological activity.

QSAR initially started in 19th century. The scientists, Crum-Brown and Fraser in 1868, gave an equation which describes that physiological activity ($\Phi$) must be a function of the chemical structure c (Eq. 1.2).

$$\Phi = f(c) \quad \text{(Eq. 1.2)}$$

In 1893, Richet expressed that the water solubility of a series of simple organic compounds (alcohol, ether) is inversely related their corresponding cytotoxicities. In the 20th century, Hans Horst Meyer (University of Marburg) and Charles Ernest
Overton (University of Zurich) expressed the linear relationships between lipophilicity (expressed as oil-water partition coefficients) and narcotic activities. After that, Lazarev (St. Petersburg University) continued the research of Overton and Meyer. He applied the partition coefficients to make the hygiene standards for the development of industries. He designed a system for calculating partition coefficients from chemical structure by expressing the correlations on a log scale. In 1939, Ferguson developed a mathematical interpretation for evaluating the toxicity. By using physical properties like solubility in water, distribution between phases, capillarity and steam pressure he observed the increase in anaesthetic activity rising in a homologous series of alkanols or n-alkanes at the point where no further increase or loss of potency occurred [33]. In 1937, Louis Hammett found that a quantitative effect on the dissociation constant was observed by adding the substituents on the aromatic ring of benzoic acid. A correlation of equilibrium constants and reactivity of organic acid/bases with their electronic properties was given by him. An equation known as Hammett equation, linear relationship, was derived by him. The equation is:

\[ \log \left( \frac{k}{k_0} \right) = \rho \sigma \]  

(Eq. 1.3)

Where, \( \sigma \) is the parameter that describes the electronic properties of aromatic substituents, i.e., electron-donating or electron-withdrawing power. \( \rho \) is the slope, a proportionality reaction constant for the given equilibrium. It is related to the effect of substituents on that equilibrium to the effect on the benzoic acid equilibrium. On the basis of Hammett’s linear relationship equation, the electronic properties were considered as the descriptors of structure. Taft designed a method for differentiating the effects of steric, polarity, steric and resonance. He introduced the first steric parameter (Es). After that, Swain studied the effects of resonance and field. He
observed the variations in the reactivity of a particular electrophilic substrate towards the variety of nucleophilic reagents.

Afterwards, Free Wilson developed a relationship between the substituents, which are added to the ring (alongwith their type, numbers and position in the ring), and the biological activity of the series of molecules. In 1962, on the structure-activity relationship, Hansch and Muir published a remarkable work on plant growth regulators. This work was dependant on hydrophobicity and Hammett constant [34]. The parameter $\pi$ was described in a manner analogous to the definition of sigma, it is the relative hydrophobicity of the substituents:

$$\pi_x = \log P_X - \log P_H$$  \hspace{1cm} (Eq. 1.4)

Where, $P_H$ and $P_X$ represent the partition coefficients of parent molecule and the derivative respectively. The combination of these hydrophobic constants with Hammett’s electronic constants was given by Hansch and Fujita (1964) to yield the modified linear relationship, the Hansch equation:

$$\log (1/C) = \pi + \sigma + cE_s + d$$  \hspace{1cm} (Eq. 1.5)

Where, $C$ is the molar concentration of compounds. $\sigma$, $\pi$ and $E_s$ are the electronic, hydrophobic and steric components respectively. $a$, $b$, $c$ and $d$ are the regression coefficients. This relationship is known as linear free energy relationship (LFER) or Hansch analysis. The application of QSAR became wide by the combination of Hansch approach and Free-Wilson analysis.

The Hansch analysis or linear free energy relationship (LFER) was the first approach used for the prediction of the biological activity of a given compound from the analysis of its structure. To develop the quantitative models for energy-based
properties like binding constants, partition coefficients or reaction rate constants, linear free energy relationship methods are widely used. Hammett developed the method to predict the chemical reactivity, the LFER method is based on it. The assumption for a congeneric series of compounds is that the effect of a structural feature on the free energy change of a chemical process is constant. The LFER approach was further extended for the basic concept of fragmentation. The molecules were broken into substructures, each substructure contributed in a constant increase in the percentage of the free energy [35, 36]. The results given by this approach were not found accurate many times, so, correction methods are needed to apply in majority of fragmentation approach based methods. Corrections are generally concerned with the long range interactions, such as steric or resonance effect. A more advanced drug design tool related to the computer applications was developed as Computer-assisted drug design (CADD) which also known as computer-assisted molecular design (CAMD. However, computers are not able to get the whole findings for a system which is being studied. Therefore, a computer is just an additional tool to achieve the desired goals in the investigations for the combined chemistry and biological system. Computer tool has the ability to produce large number of molecules rapidly. It can produce a huge amount of data which describes everything about the molecules in a very short duration of time.

Two classical QSAR approaches were developed in the middle of the 20th century, [37]:

1. The regression analysis approach, it was developed for identification of molecular features (fragments, groups or sites) and evaluation of their effect on the biological activity of a compound.
2. The physicochemical parameters based approach as structural descriptors. This approach was developed to check the role of electronic, lipophilicity, and steric properties of the drugs on the biological responses of the living organism.

1.2.1 Objectives of QSAR [38]

The QSAR methods generally have the following objectives:

- To correlate the relationships between chemical structure modifications and the quantitative changes occur in the biological activity. To modify the existing drug molecules to increase their biological activities.

- To predict the biological activities of untested compounds.

1.2.2 Requirements of QSAR

All the molecules being considered for QSAR study should belong to a congeneric series and act through the same mechanism of action. Normally molecules in a series of compounds have the similar basic moiety but they have different substituents, they can bind to the same target in the same manner and hence all analogues are supposed to interact in a comparable manner.

For a successful QSAR on any series, the requirements for biological activities are as follows:

1) Wide range in observed activities.

2) Identical mode of action.

3) Concentration in molar units.
4) Activity data as a function of concentration (IC\textsubscript{50}).

5) Activity data in percentage.

6) Possible time dependency.

1.2.3 Training set and test set design

The data set consists of the compounds with their biological activities, which is divided into two subset test set and training set. Approximately 20\% are selected by seeing wide span in activity and assigned as test set, with the remaining 80\% assigned to training set. The training set is used for QSAR model development and test set is used for model validation.

The QSAR in a generalised way consists of following steps:

(a) Calculation of various physiochemical and structural parameters.

(b) Correlation and regression analysis.

(c) Activity prediction of new designed compounds.

1.2.4 Molecular Descriptors

Molecular descriptors are the numerical values which represent the chemical information of a molecular structure which is calculated by the mathematical procedure. The dimension of molecular descriptors used describes the type of QSAR. The descriptors are given below:
0D These descriptors are calculated by using molecular formula e.g. atomic number, type of atom, molecular weight etc.

1D It is the one-dimensional molecular descriptor which consists of a list of the molecular fragments (e.g. bonds, rings, substituents, functional groups etc.).

2D It is the two dimensional molecular descriptor (2D orientation of a molecule) that provides the topological information about the molecule. It gives the chemical information of the structure of a molecule such as the interaction of particular atoms, the bonding pattern of the atoms in a molecule and the type of bonding (e.g. molecular connectivity indices, total path count etc.).

3D It is the three dimensional molecular descriptor, a molecule which has 3D orientation. It includes molecular volume, molecular surface and other geometrical properties. The various 3D descriptors are available like steric, electronic, shape etc.

4D In this descriptor the 4th dimension is referred to as any other experimental condition or different conformations.

1.2.5 Two-Dimensional (2D) QSAR

In 2D QSAR, the biological activity of a compound is co-related to molecular properties. The molecular properties are expressed or modelled in the form of molecular descriptors. Standard QSAR analysis is the method which expresses the biological activity in the linear combination with the descriptors. It shows the presence of a linear relationship between the relevant molecular properties and activity. For predicting the activity, coefficients in the regression analysis equations give weights to the descriptors according to the calculated importance of their contributions. Some of the two- dimensional QSAR methodologies are:
(a) Hansch Approach           (b) Free-Wilson Model
(c) Discriminant Analysis    (d) Pattern Recognition
(e) The Distance Geometry Approach (f) Cluster Analysis
(g) Fibonacci Approach      (h) Molecular Shape Analysis
(i) Minimal Steric Difference (j) Hypothetical Active site lattice

1.2.6 Physicochemical properties

There are many structural, chemical and physical properties which have been used in QSAR study. The most important are described in Table 1.3.

**Table -1.4** Major parameters used in QSAR approach.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P$</td>
<td>partition coefficient</td>
<td>hydrophobicity of the molecule is measured in terms of LogP.</td>
</tr>
<tr>
<td>$\Pi$</td>
<td>hydrophobic constant</td>
<td>$\pi = \log P_X - \log P_H$, where, $P_X$ is the partition coefficient of the derivative and $P_H$ is the partition coefficient of the unsubstituted compound.</td>
</tr>
<tr>
<td>$R_M$</td>
<td>hydrophobic constant from chromatography</td>
<td>LogP is linearly related to $R_M$ as $\log P = R_M +$ constant.</td>
</tr>
<tr>
<td>$K'$</td>
<td>hydrophobic constant from high-pressure</td>
<td>Log $P$ linearly related to $\log k'$ as $\log P = \log k' +$ constant.</td>
</tr>
<tr>
<td>Symbol</td>
<td>Description</td>
<td>Explanation</td>
</tr>
<tr>
<td>----------</td>
<td>--------------------------------------------------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>$\sigma_{m}, \sigma_p$</td>
<td>Hammett constant</td>
<td>$\sigma$ represents electronic character of meta and para substituents. Positive value of $\sigma$ shows electron-withdrawing nature, negative value of $\sigma$ shows the electron-donating nature.</td>
</tr>
<tr>
<td>$\sigma_o$</td>
<td>apparent Hammett constant for ortho molecule substituents</td>
<td>Unreliable and used rarely, hydrogen bonding to the solvent decreases $\pi$, so covariance may occur.</td>
</tr>
<tr>
<td>$\sigma^*$</td>
<td>Taft constant</td>
<td>Measures the electronic effect generated by aliphatic substituents.</td>
</tr>
<tr>
<td>$\sigma_I$</td>
<td>Inductive constant</td>
<td>Related to $\sigma^<em>$ as $\sigma_I = 6.43 \sigma^</em>$.</td>
</tr>
<tr>
<td>$R, F$</td>
<td>resonance &amp; field constants</td>
<td>$R$ and $F$ represent the internal resonance and field effects of a substituent, sign of $R$ and $F$ indicate the sign of the charge placed by substituents on the ring.</td>
</tr>
<tr>
<td>p$K_a$</td>
<td>ionization constant</td>
<td>$pK_a = -\log K_a$. $K_a$ is the ionization constant of an acid.</td>
</tr>
<tr>
<td>$Q (Q^\sigma, Q^\pi)$</td>
<td>charge of an atom</td>
<td>$Q^\sigma$ represents the $\sigma$ charge and $Q^\pi$ represents the $\pi$ charge. It is calculated by quantum mechanical methods.</td>
</tr>
<tr>
<td>Symbol</td>
<td>Term</td>
<td>Definition and Notes</td>
</tr>
<tr>
<td>----------</td>
<td>-----------------------------------------</td>
<td>--------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>$f^E_E$, $f^N_N$</td>
<td>electrophilic &amp; nucleophilic frontier orbital densities</td>
<td>Quantum mechanical descriptors used to describe the electrophilic and nucleophilic reactions at a particular centre in the molecule.</td>
</tr>
<tr>
<td>$S^E_E$, $S^N_N$</td>
<td>electrophilic &amp; nucleophilic super delocalizabilities</td>
<td>Involves the charge-transfer capabilities along with an orbital charge at a particular atom.</td>
</tr>
<tr>
<td>$E_{\text{HOMO}}$</td>
<td>energy of the highest occupied molecular orbital</td>
<td>Quantum mechanical parameter related to the energy of the ionization potential of the molecule.</td>
</tr>
<tr>
<td>$E_{\text{LUMO}}$</td>
<td>energy of the lowest empty molecular orbital</td>
<td>Quantum mechanical parameter related to the electron affinity of the molecule.</td>
</tr>
<tr>
<td>$E_s$</td>
<td>Taft steric constant</td>
<td>Concerned with the acid-catalyzed hydrolysis of $\alpha$-substituted acetates ($X\text{CH}_2\text{COOR}$). It represents the steric effect affecting intramolecular and intermolecular hindrance during the action or binding.</td>
</tr>
<tr>
<td>MR</td>
<td>molar refractivity</td>
<td>$\text{MR} = \left[\frac{(n^2 - 1)}{(n^2 + 2)}\right] \frac{\text{MW}}{\text{d}}$, where, $n$ is the refractive index for the sodium D line, MW is the molecular weight, and $d$ is the density of the compound. It may be used as a steric parameter in the absence of $E_s$. It calculates the electronic effect as well. It may show the dipole-dipole interaction at the active site.</td>
</tr>
</tbody>
</table>
MV, R, L, MW, \( V_w \), r and B etc., have also been used as steric parameters.

<table>
<thead>
<tr>
<th>MV</th>
<th>molar volume</th>
<th>MV, R, L, MW, ( V_w ), r and B etc., have also been used as steric parameters.</th>
</tr>
</thead>
<tbody>
<tr>
<td>( V_w )</td>
<td>van der Waals volume</td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>van der Waals radius</td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>interatomic distance</td>
<td></td>
</tr>
<tr>
<td>L</td>
<td>length parameter</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>width parameter</td>
<td></td>
</tr>
<tr>
<td>X</td>
<td>molecular connectivity index</td>
<td></td>
</tr>
</tbody>
</table>

It is a topological parameter measures the effects of kinds of atoms, branching patterns, bonding type, unsaturation and heteroatom content in a molecule on its reactivity or activity.

1.2.7 Applications of QSAR

Today QSAR study is used widely. It is the most significant tool in the research of drug discovery and environmental toxicology. It is applicable for predicting the biological activities of untested compounds. It can be used for:

- differentiating drug molecules from nondrug molecules
- physicochemical properties prediction (e.g. lipophilicity, water solubility)
- ADME (adsorption, distribution, metabolism, and excretion) prediction
- drug resistance
- toxicity prediction

### 1.2.8 Three-Dimensional Quantitative Structure-Activity Relationships (3D QSAR) [39]

3D-QSAR is used for the calculation of force field of three-dimensional structures. It depends on the protein crystallography and molecule superimposition. It provides the information of whole molecule instead of a single substituent and requires computed potentials (Lennard-Jones potential) instead of experimental constants. It calculates the electrostatic fields and the steric fields (shape of the molecule), based on the applied energy function. By using statistical correlation methods, the 3D-QSAR provides the analysis of quantitative relationship between the three-dimensional properties of a set of compounds and their biological activity. The 3D QSAR approach is classified into the ligand-based and structure-based approaches, respectively. The ligand-based 3D QSAR is applicable, when experimentally the fine 3D structure of the particular target protein is not obtained or unknown. The structure-based 3D QSAR is applicable for those cases, where the 3D structures of a target protein (or its homologue) bound to the active compound is experimentally known by using the X-ray crystal structure analysis. In this, a homology-modeled target protein is used for the analysis and in the ligand-binding site, the lot of combinations of side chain rotamers have to consider in order to achieve a linear relationship between the
in vitro activity of the compounds and the interaction energy (or score) of the complexes [40].

1.2.9 3D QSAR Programs [41]

There are many programs available for QSAR and molecular modeling. One of them is discussed below:

1.2.9.1 SYBYL 6.9

Tripos group of company developed this program. It contains lot of tools for molecular modeling such as visualization of structures and associated data, structure building, annotation, optimization and comparison, screen capture capabilities, hardcopy and a wide range of force fields. It has following modules:

- **QSAR** alongwith CoMFA & CoMSIA provide tools which are capable to build graphical and statistical models of biological activity from the structure of molecules. It can be used to build the models for prediction the biological activity of untested compounds.

- **Advance CoMFA** has the specialized CoMFA fields and includes the set of tools that increase the capabilities of QSAR with CoMFA.

- **Genefold** is used to analyse the protein function from sequence by threading method.

- **Composer** used to create 3D modeling when the homology modeling is not possible. It is based on Genefold’s sequence alignments.

- **UNITY’s 3D Flexible Search** is used to find out the potential lead compounds
from the corporate or vendor database.

- **Diverse solutions** applicable to the chemical diversity of libraries and used for selecting diverse or representative subsets.

- **FlexX and C Score**, in the combined form, prioritize compounds for synthesis or screening by performing virtual high throughput screening. *FlexX* rapidly docks the ligands into receptor binding site and C Score integrates multiple types of scoring function for ranking the binding affinity of ligand-receptor complex.

- **CONCORD** used to convert the 2D chemical structures into 3D structures.

- **Biopolymer** provides a set of tools for visualization, building, prediction and manipulation in the 3D structure and biological molecules, including peptides, nucleic acids, proteins and polysaccharides.

- **MOPAC** is semi-empirical molecular orbital program used for the study of chemical structure. It is used to optimize the geometry of a molecule (quantum-mechanical approach based). It calculates the charge on various atoms and it can be used for all the molecules (molecules contain formal charge and free radicals).

- **DISCO** is used for pharmacophore analysis from the conformations (precomputed) of active compounds which bind to the same target site.
1.3 Docking

In the drug discovery, high-throughput screening requires the screening of millions of compounds for a particular protein target. The biological research has become increasingly data sensitive. The biomedical projects require latest informatics tools. Molecular docking and database mining are the important tools for improving such screenings. The structure of receptor ligand complexes is predicted by molecular docking. The receptor is generally a protein or a protein oligomer and the ligand is a small molecule or a protein. Initially, molecular docking was used to predict protein-ligand complexe structures. To increase the application of molecular docking in different fields many simplifications were done. [42-46].

Formally, the restrictions in protein-ligand docking came when the description (Geometrical and Chemical) of the protein and the small organic molecule was not known. Therefore, computational technique is required to check the binding of small molecule to the protein. If it occurs, then we need to know the geometry and the binding affinity of the complex. Mostly algorithms have two components, first, a search technique to find out the best possible position of the ligand in the binding pocket of the protein and second, a scoring function to rate each position, as well as to rank candidate ligands adjacent to each other. Alternatively, we can say that the docking is the process to predict of conformation and orientation (or posing) of a ligand within a targeted binding site. Structural modeling and prediction of activity are two aims of docking studies. But, the prediction of modifications in compound that improve effectiveness and identification of molecular features which are responsible for specific biological identification, are much more complex issues that are usually difficult to understand and simulate on the computer [47].
1.3.1 Basic requirement for docking: 3D Atomistic representation of the receptor

A receptor structure of interest at hand is the precondition so as to perform a structure-based screening task. Most commonly, experimental techniques such as X-ray crystallography or NMR is used for determination of structure of the receptor. For obtaining protein structures the protein structure prediction techniques are used [48] such as homology modeling and ‘threading’ [49, 50]. When the three dimensional structure of the proteins achieved experimentally or by prediction, it may be analyzed by using various computational techniques. If the function of the protein is unknown, then its structure finding is required for recognized binding sites [51]. Usually the binding site or the function of the protein is identified through reference (e.g. the protein can be chosen from a protein family with known function), or by the crystal of ligand-protein complex. The analysis of the binding-site characteristics and the interactions of the protein with a given ligand provide chances to design the novel ligands [52, 53].

1.3.2 Docking Scenarios

Docking scenarios typically fall into one of the following categories:

1) When the active site of the protein is unknown, the searching of the binding site as well as the binding mode of the ligand is defined as Blind Docking. It is used for identifying the protein interactions also.

2) When the site of the binding is known from X-ray diffraction or from NMR studies, the docking into the active site is called Direct Docking. In such cases it is possible that:

- The active site is overlapping with the binding site of a cofactor.
• The active site contains discrete crystal molecules of water, mediating the binding between the ligand and the target.

• The active site contains catalytic metal ions.

3) Recent virtual screening techniques deal with some factors such as tautomeric or ionized states of the compounds as well as the effect of the temperature, $pK_a$, pH, or micro-environment of the active site on ligand binding which are missing in most of the docking approaches.

4) Reproducing the conformational space accessible to a macromolecule is a very difficult task and always necessitates an approximation level:

• Rigid body docking: both ligand and protein are considered as rigid bodies.

• Semi flexible docking: the ligand only is considered flexible.

• Fully flexible docking: both ligand and protein are treated as flexible molecules.

Docking is usually a process which has multiple steps and in each step begins one or more supplementary degrees of complexity [46]. Docking algorithms are the initial step of the process which place (POSE) small molecules in the active site. ‘POISING’ is defined as the process to determine the specified conformation and orientation of a ligand that placed the active site of the protein. It is generally an indistinct procedure which returns numerous alternative results. Scoring functions are involved with various docking algorithms which are designed to predict the biological activity by evaluating the interactions molecules with the possible targets. Initially scoring functions evaluate the compound interactions with target on the basis of calculations.
of predicted shape and electrostatic complementarities. Scoring plays its role in both the posing and ranking. The pose score is basically a rough idea to measure of the binding or fixing of a ligand into the active target site. The binding energies may be calculated by rank score, but it is more complex. Ranking is the process that typically takes numerous results from the starting scoring phase and re-calculates them. It is more advanced than pose scoring process. This process is generally used to calculate the free energy of binding as precisely as possible. The posing phase uses simple energy calculations (van der Waals and electrostatic) but the ranking methods involves more detailed calculations (properties such as explicit solvation or entropy). Basically, there are three types of scoring functions which are currently in use: (a) Force field-based, (b) empirical, (c) knowledge-based. Here we describe briefly all the three types of scoring functions with an emphasis on force fields.

1.3.3 Force Fields

Force fields define the parameters which coupled with the simulation equations allow for simulation of molecules. In molecular mechanics, simple equations of motion are used to describe the motions of atoms within a molecule. A ‘good’ force field reveals maximum possible different classes of forces with reasonable precision. The subsequent reliability of the calculations done depends on the potential energy functions and the quality of the parameters included in these functions. Force fields can be categorised into different classes on the basis of their nature and complexity of the force field equation:

- Those with rigid or partially rigid geometries - e.g. ECEPP, JUMNA
- Those without electrostatics - e.g. SYBYL, YETI
• Simple diagonal force fields - e.g. GROMOS, CHARMM, OPLS/AMBER

• More complex force fields - e.g. MM2/MM3, MMFF

The GROMOS, CHARMM and OPLS/AMBER force fields fall under the category of the simple diagonal force fields. All of these force fields have a simple harmonic diagonal representation in the terms of bond and angle. Some other common force fields in application are:

a) CVFF (Consistent Valence Force Field) - The CVFF force field was parameterized using protein and peptide structures. It is commonly used to minimize these structures.

b) AMBER (Assisted Model Building with Energy Refinement) Force Field – The AMBER force field is used when dealing with compounds concerned with nucleic acids, proteins and polysaccharides.

c) AMBER95 Force Field - This is the recent version of AMBER

d) CFF91 (Consistent Force Field) - This force field is more accurate than CVFF for most proteins as it was parameterized using a wider range of experiments.

e) ESFF (Extensible Systematic Force Field) - This force field supports a wide variety of atom types but to use it one needs to work in the Discover3, Accelrys Software Inc. [54] environment.

Force field geometry optimization method has many advantages such as maximum accuracy, speed and wide applicability on small as well as large molecule which
makes it the most important standard method. When different force fields are used possibilities of big variations in geometry should not be expected. The energy terms such as bond stretching, angle, bending, etc. are altered in the process but torsional angles are affected very less. It is also to be realized that crystal structures are not "bad" geometries.

1.3.4 Pose Prediction versus Affinity Prediction

The ‘docking’ is referred as the placing of the molecules in the receptor-binding site, but the ‘scoring’ is referred as the prediction of affinity of the binding. Now a day the docking programs are designed to perform both the actions [55]. Technically the two major challenges are faced by docking programs due to the variations in docking and scoring functions: First, the prediction of the molecule’s binding mode accurately called as ‘pose prediction’ (Pose refers to the molecule’s conformation and orientation at the receptor target binding site) [56]. Second, the binding affinity prediction of the compounds or to generate the comparative rank-ordering for a number of compounds used to target binding site in a consistent manner [57].

1.3.5 Searching for the Right Pose

Searching for the accurate binding pose of a molecule can be achieved by attempting the number of trials and the poses which are logical in terms of energy, keep those poses for the further process. The search process ends when a number of trials have been attempted and/or an adequate number of poses achieved for a molecule. Docking algorithms have been developed, as discussed in following sections, to maintain the track for previously discovered minima and to find out the new regions.
1.3.6 Docking Algorithms

The accurate binding mode can be discovered by finding the exact conformation of the docked molecule and the exact orientation because most of the ligand molecules are flexible. Therefore ligand flexibility needs to be handled very carefully. The docking algorithms can be categorized into three types of searches: (a) Systematic, (b) Stochastic and, (c) Deterministic approaches [46, 47, 55]. Some search algorithms use a combination of these approaches.

1.3.6.1 Systematic Algorithms

In systematic search method, conformations of a molecule are generated by systematically assigned values to the torsion angles of the rotatable bonds. A predetermined set of possible values is produced by each torsion angle. The grid search is the simplest form of systematic search algorithm. It is based on a grid of values for each formal degree of freedom. During the search each of these grid values are explored in a combinatorial manner and conformations related to all promising combinations of torsion angle values are generated. For example, suppose a molecule has two different torsion angles, the first torsion angle has the values 60°, 180° and -60° and the second torsion angle has the values 0° and 180°. The grid search will generate six conformations (i.e. 60,0; 60,180; 180,0; 180,180; -60,0; -60,180). These conformations are then one by one subjected to energy minimization. A mathematical equation expresses the grid search, if the number of values permissible to torsion angle $i$ is $n_i$, and the molecule has $N$ variable torsion angles then the total number of conformations $C$ generated by the grid search would be:

$$C = \prod_{i=1}^{N} n_i$$
So, in a molecule a number of structures will grow exponentially with the number of rotatable bonds. This whole process is sometimes known as a *Combinatorial Explosion* [58]. The algorithms face the problem of combinatorial explosion while they try to search all the degrees of freedom in a molecule. When the number of degrees of freedom increases, simultaneously, the number of evaluations required is also increases. To overcome this situation, a termination criteria is incorporated to avoid the algorithm to get involved in sampling space that leads to the wrong solution.

Thus with the aim to prevent combinatorial explosion in some of these algorithms, the ligand is developed incrementally in the active site, starting from a docked ‘base fragment’. A systematic stepwise search can be performed by two different ways: firstly, by docking the active-site region with different kind of molecular fragments and connecting them covalently (It is referred as most popular *de novo* ligand-design strategy), secondly, by differentiating docked ligands into core fragment (rigid) and side chains (flexible). In the second case, the identified core fragments which are docked into the active site and the flexible side chains are added in an incremental manner [59-61]. Representatives of this approach are Hammerhead [62], DOCK [63] and Flex X [64]. In other approaches, such as Auto Dock [65], Gold [66], ICM-Dock [67] and QXP [68], the ligand is treated in its entirety.

### 1.3.6.2 Random Search

The random search algorithms are often called stochastic methods which include some kind of repetitious procedure. In this procedure, the structure is selected on the
basis of previously generated and randomly modified search and then it is minimized. If this generates a new logical conformation then it is considered in the list of structures found and the process is repeated again. In contrast to the systematic search methods, there is no natural endpoint for a random search or we can say that there is an uncertainty of convergence, this process will continues until either a pre-defined number of trials have been carried out and/or until there is no new conformations can be produced. In other words we can say that in order to improve convergence, multiple, independent runs should be performed. Various random search procedures can be differentiated in two ways: by the structural modification method and by the method of selection of the structure for the next repetition. Modifications can be achieved most commonly either by changing the torsion angles (keeping the bond angles and lengths fixed) or by changing the (xyz) coordinates of the atoms. The simplest method to select the structure for the next repetition is to use the structure generated in the earlier step, however, other methods are also available. An alternative way is to choose the structures randomly which are generated already. This was all about the random search algorithm and it provides two most popular random approaches which are Monte Carlo and genetic algorithms.

1.3.6.2.1 The Metropolis Monte-Carlo Scheme

The Monte-Carlo scheme can also be used for the selection. In this, when the new structure has the lower energy ($V_{\text{new}}$) than its predecessor energy ($V_{\text{old}}$) then it can be used as the next starting structure. The Boltzmann factor, $\exp[-(V_{\text{new}} - V_{\text{old}})/kT]$, is calculated (k is the Boltzmann Constant and $T$ is the temperature) when $V_{\text{new}}$ is higher than the $V_{\text{old}}$. The new structure is selected, if the Boltzmann factor is larger than a random number between zero and one and if it is not, then the previous structure is
taken. Thus the Monte-Carlo Metropolis method makes it possible that the structures having higher energies can be selected and these can be correlated to the previously unknown areas of conformational space.

Simulated annealing is the most commonly used version of *Monte Carlo* method [69]. In this method, the temperature decreases steadily from a high value to the lower value. The system can overcome the high energy barriers at the high temperatures due to the presence of temperatures in the denominator of the Boltzman factor and this result to discover the search space widely. When the temperature decreases, the lower energy states become viable. At absolute zero temperature, the computational system should exist in the global minimum-energy state by using the physical process of annealing in the manufacture of very large single crystals of materials, for instance, silicon. The global minimum energy conformation is related to the conformational analysis, though, it requires an infinite number of temperature decreases and for each of them the system has to come in equilibrium. Therefore practically it is not mandatory that a system will find the global energy minimum. Thus several simulated annealing runs have to be performed. Alternative implementations of Monte Carlo search have been reported [70] earlier including a popular type in Auto Dock [71].

### 1.3.6.2.2 Genetic Algorithms

Genetic algorithms are the set of computational problem-solving approaches which involves the principles of population dynamics and biological struggle. ‘Chromosome’ encoded the model parameters which are differentiated stochastically. Chromosomes provide the feasible solutions to a given problem which can be analysed through the fitness function. The chromosomes, responsible for the best
intermediate solutions, subjected to crossover dry mutation operations (similar to gene recombination and mutation) to produce the next generation. The genetic algorithm solution is a collection of possible ligand conformations for docking applications. Several programs for example DOCK and GOLD have implemented genetic algorithms [65, 66, 72].

### 1.3.6.3 Tabu Search Algorithm

The tabu search algorithm technique is applicable for the already explored areas of conformational space [73]. The root mean square deviation is calculated between current molecular coordinates and every molecule's previously recorded conformation. It is required to know that a molecular conformation is accepted or not, PRO_LEADS is an example for making the use of a tabu search algorithm [74].

Tabu search algorithm usually:

- Make little random changes to the existing conformation.
- Rank each change according to the value of the selected fitness function.
- Determine which changes are ‘tabu’ (that is, earlier rejected conformations).
- If any modification gives the lower value than any other accepted modification then accept it, even if it is in ‘tabu’. Or else accept the best ‘non-tabu’ modification.
- Add the accepted modification to the ‘tabu’ list and record its score.
- Go to the first step.
1.3.6.4 Deterministic Search/Simulation Methods

In deterministic searches, the initial state determines the progress that can be made to produce the next state. This state should have equal to or lower in energy than the initial state. Deterministic searches performed precisely on the same starting system, including each degree of freedom, with the same parameters which produce exactly the same final state [46]. A problem which usually comes in deterministic algorithms is that they frequently get trapped in local minima because they cannot cross the barriers. To increase capability, for crossing the barriers or to decrease the height of the energy barriers various approaches were tried. Instances of deterministic methods are molecular dynamics (MD) simulations and energy minimization methods.

1.3.6.4.1 Molecular Dynamics

Molecular Dynamics is concerned with the atoms of the system and solves the Newton’s equations of motion. It is related to the positions of the atoms, how they changes with times. Pictures taken from the sequence may be considered in the minimization, in order to make a sequence of minimum energy conformations. The MD docking algorithm is the combination of changes in temperatures and the degrees of freedom of the system. The two temperatures which are responsible for calculating the flexibility of the ligand can be vary at the time of simulation process. It helps the system to avoid the trapping in a local minimum [75, 76].
1.3.7 Scoring

Ligand scoring method is applicable for calculating the binding affinity of a ligand. The binding affinity of candidate ligand is evaluated when it is docked in the binding site of the target receptor structure. The computed ligand–receptor interaction energy (score) of a pose takes the decision for trial pose. ‘Dock score’ is calculated by many programs to recognize and giving rank order to various poses of a molecule during the search. It is a score based on a simple energy function such as a force field with an electrostatic term, repulsive and attractive van der Waals terms. This can be calculated very easily during the docking process, whereas, more sophisticated function is used to calculate the final ‘affinity score’ for that molecule.

1.3.7.1 Scoring Methods and Scoring Functions

Wide ranges of scoring function systems are available now a day [77]. They mostly divided into two main categories (Fig. 1.5). The first category includes scoring schemes based on force field physical interaction terms (i.e. Vander Waals interactions or hydrogen bonding) [78]. This category is further divided in two groups: (a) the ‘empirical’ scoring functions in which each of the terms are multiplied with a coefficient, the resultant products are summed up to provide the final score. The coefficients are optimized to provide a good fit to the training set of molecules. (b) the ‘first-principle-based’ scoring functions, in this, the terms are directly derived from physico-chemical theory which are not included in experimental data [79]
The second category includes ‘knowledge-based’ scoring functions. These functions can be obtained by applying statistics in the frequencies of the observed inter–atomic contact and distances in crystal structures of protein–ligand complexes. The scoring functions which fall into this category are such as PMF [80] and Drug Score [81]. Many scoring functions can be combined to make a single scoring scheme, this method simplifies the process and this is known as 'consensus score' [82, 83]. In this particular case, a molecule has to score well across the many different scoring schemes to find out an active target binder. Scoring methods typically use empirical functions, as discussed before, which are developed by fitting various functional forms and characterize various aspects of the ligand-receptor interactions against binding affinity data. However, for statistical analysis of known ligand–receptor structures, knowledge based approach is used. It is also used to estimate the frequency of occurrence of specific receptor–ligand interactions without the help of binding affinity information.
The types of scoring functions (from both categories) are as follow:

- Jain
- Lig Score 1
- Lig Score 2
- Ludi
- Piecewise Linear Potential (PLP)
- Potential of Mean Force (PMF)

The first four scoring functions come under the empirical based approach and the PMF function come under the knowledge-based statistical approach. The scoring function which is correlated well with binding affinities is PLP. It was originally developed as a docking function.

1.3.7.1.1 Jain Scoring Function

A.N. Jain developed an empirical scoring function for the evaluation of the structures as well as binding affinities of protein–ligand complexes. The Jain score is a sum of five interaction terms [84]. These terms describe:

1. Lipophilic interactions
2. Polar repulsive interactions
3. Polar attractive interactions
4. An entropy term for the ligand
5. Solvation of the ligand and protein

For the pairwise interactions, only proximate ligand-protein atoms are considered. The sum of a Gaussian and a sigmoidal function represents the lipophilic and polar interactions. This functional form is short-ranged with a pronounced maximum that occurs at close surface contacts and this also incurs a significant penalty for short contacts between ligand and protein atoms. Options are provided to ignore non-polar hydrogens of the receptor or ignore water molecules when calculating the Jain score.

1.3.7.1.2 LigScore Scoring Function

Lig Score scoring functions are further of two types Lig Score 1 and Lig Score 2.

a) **LigScore 1** is the simple, fast, scoring function used for predicting the binding affinity of ligand-receptor complex. Three descriptors are involved to calculate Lig Score1 and expressed in units of pKₐ (-log Kₐ). These descriptors are:

- **vdW:** It is a softened Lennard–Jones 6-9 potential. It is expressed in units of kcal/mol.

- **C+pol:** It calculates the buried polar surface area between a ligand and receptor which holds attractive ligand-receptor interactions. It expressed in units of Å².

- **TotPol²:** It is the squared sum of the total polar surface area of the ligand and receptor and expressed in units of Å².

The individual inputs of these descriptors are also available along with the overall LigScore 1 value during the scoring. LigScore1 calculated by using two different kinds of equation depending upon the forcefield (CFF or DREIDING) [85] employed
for the calculation of the vDW descriptor which corresponds to the charge model (Gastieger or CFF) used to assign atoms as polar or nonpolar. The corresponding LigScorel equations are:

\[
\text{LigScorel\_CFF} = 0.4896 - 0.04551*\text{vdW} + 0.1439*\text{C+pol} - 0.001010*\text{TotPol}^2
\]

(Eq. 1.6)

\[
\text{LigScorel\_Dreiding} = -0.3498 - 0.04673*\text{vdW} + 0.1653*\text{C+pol} -0.001132*\text{TotPol}^2
\]

(Eq. 1.7)

Where, the coefficients were achieved by regression analysis of the binding affinities of a series of ligand-protein- complexes [86].

b) **LigScore 2** is also the simple and fast scoring function for predicting the binding affinity of ligand-receptor complex. In this also, three descriptors are involved to calculate LigScore 2 and expressed in the units of $pK_i$ ($-\log K_i$). These descriptors are:

- **vdW**: It is a softened Lennard–Jones 6-9 potential. It is expressed in units of kcal/mol.

- **C+pol**: It calculates the buried polar surface area between a ligand and receptor which holds attractive ligand-receptor interactions. It expressed in units of Å$^2$.

- **Bury Pol$^2$**: It is the squared sum of the buried polar surface area of the ligand and receptor and expressed in units of Å$^2$. It is used for the desolvation penalty invited by desolvating water molecules from the receptor and ligand binding space so that the ligand can bind to the receptor.
The individual inputs of these descriptors are also available along with the overall LigScore 2 value during the scoring. Like LigScore1, LigScore2 is calculated by using two different kinds of equation depending upon the forcefield (CFF or DREIDING) employed for the calculation of the vdw descriptor. The current LigScore2 equations are:

\[
\text{LigScore2}_{\text{CFF}} = 1.900 - 0.0730 \times \text{vdW} + 0.06246 \times \text{C+pol} - 0.00007324 \times \text{BuryPol}^2
\]  
(Eq. 1.8)

\[
\text{LigScore2}_{\text{Dreiding}} = 1.539 - 0.07622 \times \text{vdW} + 0.6501 \times \text{C+pol} - 0.00007821 \times \text{BuryPol}^2
\]  
(Eq. 1.9)

Where, the coefficients were achieved by regression analysis of the binding affinities of a series of ligand-protein- complexes [86].

1.3.7.1.3 Ludi Scoring Function

The empirical scoring functions provided for the Ludi algorithm are used to evaluate the receptor-ligand interactions [87]. The Ludi score is the sum of five contributions:

- Ideal hydrogen bonds contribution.
- The ionic interaction contribution (It is the interaction of donor/acceptor in the receptor [e.g. COO, or -NH3+]).
- Lipophilic interactions contribution
- The freezing of internal degrees of freedom of the ligand contribution.
- The loss of translational and rotational entropy of the ligand contribution
This scoring function known as *Energy_Estimate_1* derived from the empirically fitting a set of ligand-protein complexes with the binding affinities, which are experimentally achieved. A second scoring function known as *Energy_Estimate_2* calculated by using additional complexes and refitting the weights related previously described terms. A final scoring function known as *Energy_Estimate_3* derived from the same set of complexes used for *Energy_Estimate_2* and including the additional contribution of the aromatic-aromatic interactions in the function [88].

1.3.7.1.4 Piecewise Linear Potential (PLP)

Piecewise Linear Potential is also a simple and fast docking function which evaluates the binding affinities of ligand-protein complexes. PLP scores are measured in arbitrary units. The negative PLP scores are used in the calculations of consensus score. The higher PLP values represent the stronger ligand-receptor binding (larger pKᵢ values). Two types of the PLP function are available: PLP1 [89] and PLP2 [90].

(a) PLP1

In this scoring function, the every non-hydrogen receptor or non-hydrogen ligand atom is considered as a PLP atom. Hydrogens are excluded at the time of scoring. There are four types of PLP atoms:

1. H-bond acceptor only

2. Hydrogen bond (H-bond) donor only.


4. Non-polar.
Note: When PLP1 is used, for every new ligand conformation the internal energy will be calculated. This energy is used only to prevent the vander Waals clashes and is not used for the PLP scoring. The forcefield used in this for providing the atomic parameters is generally the robust DREIDING [85].

The PLP1 score is the sum of the function values of all pairwise interactions in a ligand-receptor complex. H-bond and steric interaction are two types of pairwise interactions included in PLP1 as shown in Table 1.5. These interactions can be explained by the same functional form, but with different parameters (Table 1.6).

<table>
<thead>
<tr>
<th>Interaction type</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>H-bond</td>
<td>2.3</td>
<td>2.6</td>
<td>3.1</td>
<td>3.4</td>
<td>-2.0</td>
<td>20.0</td>
</tr>
<tr>
<td>Steric</td>
<td>3.4</td>
<td>3.6</td>
<td>4.5</td>
<td>5.5</td>
<td>-0.4</td>
<td>20.0</td>
</tr>
</tbody>
</table>

Table 1.5 Interaction types of PLP1

<table>
<thead>
<tr>
<th>Ligand PLP type</th>
<th>Receptor PLP type</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Donor</td>
</tr>
<tr>
<td>Donor</td>
<td>Steric</td>
</tr>
<tr>
<td>Acceptor</td>
<td>Steric</td>
</tr>
<tr>
<td>Both</td>
<td>Steric</td>
</tr>
<tr>
<td>Non-polar</td>
<td>Steric</td>
</tr>
</tbody>
</table>
(b) PLP2

In this function, the PLP atom typing will remain the same as in PLP1, but the atomic radius for hydrogen atom is not considered. Three types different radii are:

1. Small : a value of 1.4 for F and metal ions (including Zn, Mn, Mg, and Fe).
2. Medium: a value of 1.8 for C and N.
3. Large : a value of 2.2 for S, P, Cl, and Br.

The PLP2 score is the sum of the function values of all pairwise interactions in a receptor-ligand complex. The H-bond, dispersion, and repulsion are three types of pairwise interactions present in PLP2 as shown in Table 1.7. The H-bond and dispersion interactions have the different parameters, but same functional form. A scaling factor is used for H-bond and repulsion interactions based on the angle formed by the corresponding ligand-receptor atoms.

<table>
<thead>
<tr>
<th>Ligand PLP type</th>
<th>Receptor PLP type</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Donor</td>
</tr>
<tr>
<td>Donor</td>
<td>Repulsion</td>
</tr>
<tr>
<td>Acceptor</td>
<td></td>
</tr>
<tr>
<td>Both</td>
<td></td>
</tr>
<tr>
<td>Non-polar</td>
<td></td>
</tr>
</tbody>
</table>

Table 1.7 Interaction types of PLP2
1. Pairwise potential for H-bond and dispersion terms.

2. Pairwise potential for repulsion terms.

3. Scaling factor for H-bond and repulsion interactions based on the angle formed by the receptor and ligand atoms.

1.3.7.1.5 Potential of Mean Force (PMF)

The PMF [80] and PMF04 [91] scoring functions are designed on the basis of statistical analysis of the 3D structures of ligand-protein complexes. They are simple and fast scoring functions concerned with the protein-ligand binding free energies calculations. The scores are the total sum of pairwise interactions with over all inter-atomic pairs of the ligand-receptor complex. PMF04 score is the upgraded version of the previous PMF score. For the evaluation a large set of data is required and additionally, halogen potentials and metal ions are also involved. The PMF scores are measured in arbitrary units. The negative score value used in the calculation of consensus score. The higher values represent the stronger ligand-receptor binding affinity. To ensure accurate atom typing for the PMF calculation, the correct bond order information is required and the hydrogen atoms are excluded from the ligand and receptor structures. While the hydrogen atom type was included in the parameterization scheme for PMF, it can be ignored during the score calculation. Options are also available for setting distance cutoffs for the calculation of carbon-carbon interactions and other interactions as discussed by Muegge and Martin [91].
1.3.7.2 Influence of Forcefield, Partial and Formal Charges on Scoring Functions

Several of the available scoring functions depend on the values of atom charges (both partial and formal) and also the choice of forcefield. Atom partial charges are automatically recalculated when performing a scoring calculation, and these partial charges depend on the forcefield choice. Any previously calculated partial charges are ignored in scoring calculations. Two forcefields are available when calculating ligand-protein scores, CFF and DREIDING. If CFF is chosen, then partial charges will be assigned to all atoms of the ligand and protein based on the CFF charging rules. If DREIDING is selected, then the Gasteiger charging method will be employed to calculate the partial charges of ligands and proteins. The forcefield selected will also affect the vander Waals energy term used in both LigScore1 and LigScore2.

- LigScore1 uses the partial charges on the atoms of both ligand and receptor to determine whether an atom is polar or nonpolar based on a cutoff threshold. This influences the computation of C+Pol and TotPol2.

- LigScore2 types atoms of ligands and proteins as polar or nonpolar based on rules that employ only formal charges, ignoring partial charges.

- PLP1 and PLP2 do not use any charge information, so partial charges, formal charges, and forcefield have no effect on the computed scores.

- The Jain scoring function depends explicitly on the formal charge values in the polar attractive and repulsive interaction terms. The function is independent for both partial charges and forcefield.
• The PMF and Ludi scoring functions depend implicitly on formal charges in their atom typing assignment rules, but are independent of partial charges or forcefield.

The major aim at the end of docking is to reach a druggable site and in order to assess it following are the major parameters which need to be considered as shown in Table 1.8 [92].

**Table 1.8** Parameters considered for identifying a druggable site

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Ideal Value</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Shape</strong></td>
<td>Deep or enclosed</td>
<td>The lined-receptor interaction energy roughly correlates with the surface contact area.</td>
</tr>
<tr>
<td><strong>Size</strong></td>
<td>Fits ligands of 300 – 600 Da</td>
<td>Small molecules require enveloping cavities to attain sufficient binding affinity. Small cavities may not be able to accommodate drug like molecules. Very large cavities may not provide sufficient surface contact area.</td>
</tr>
<tr>
<td><strong>Chemical Character</strong></td>
<td>Mix of hydrophobic and hydrophilic</td>
<td>Drug molecules present a balance between lipophilicity (low logP) and hydrophilicity (H-bond donor/acceptors, PSA).</td>
</tr>
<tr>
<td><strong>Flexibility</strong></td>
<td>Rigid</td>
<td>Binding to very flexible binding sites involves an entropic penalty. Flexibility of the receptor is a difficult property to handle.</td>
</tr>
</tbody>
</table>
Besides the difficulties associated with the scoring of compound conformations, there are other complications exist which demands to predict the accurate conformations for binding and biological activity of compound. These are such as limited resolution of crystallographic targets, induced fit, inherent flexibility, other conformational changes which occur on binding, and the participation of water molecules in ligand-protein interactions. The docking is scientifically complex process.

1.3.8 Analysis of Poses

There are number of protocols available in DS2.5 for analyzing the poses, all found under Receptor Ligand Interactions. Some of them are Analyse Ligand Poses where calculation of RMSD of the poses and the analysis of contacts between the ligand and the receptor is done; Calculate Binding Energies; Calculate Interaction Energies etc. All the protocols are run in similar manner by entering the required parameters in the specified protocol as discussed above for docking and scoring protocols.

1.4 Synthesis of Designed Compounds.

The various compounds showed iNOS and nNOS inhibitory activities as reported earlier. After applying the QSAR studies on the series of the compounds, some new compounds were designed which have better predictive IC$_{50}$ value than the observed IC$_{50}$ value of compounds existed in the series. QSAR and molecular modeling includes analysis of all theoretical aspects drug design. To evaluate their NOS inhibitory activity in-vitro or in-vivo these new compounds need to be synthesised. In view of this, we have synthesised a designed compound in our lab and evaluated its iNOS inhibition activity in-vitro.
1.5 Estimation of NOS Inhibitory Activity.

Nitric oxide synthase enzyme is available in three different isoforms, eNOS and nNOS being constitutive in nature whereas iNOS being inducible one is an inducible enzyme and expressed during inflammation. Inhibition of iNOS is being sought in various inflammatory conditions. However, their selective inhibition has more clinical significance. Various methods are provided by different researchers for the estimation NOS inhibition, they all measure L-citrulline levels after incubating the L-Arginine with enzyme and its possible inhibitors. The L-citrulline and L-arginine used here are radioligands and requires scintillographic method. Other most commonly used method available in labs is based upon Griess reaction, where, NO formed by NOS is measured as a function of Nitrate/Nitrite obtained due to oxidation of NO. The Nitrate/Nitrite formed are subjected to the Griess reagent which eventually turns it into a colored compound (Azo dye) which is estimated with the help of a plate reader at 540 nm.

The enzyme is sourced from serum/plasma or tissue homogenate in such reactions. The incubation is done with the potential enzyme inhibitors. A standard curve is also prepared with known amount of nitrate so that unknown readings can be intrapolated from them. This method being easier and does not require any sophisticated instrumentation. It is more popular among the researchers to find out NOS inhibition.
REFERENCES


36. Hammett, L.P. Some relations between reaction rates and equilibrium constants. 

37. Gozalbes, R.; Doucet, J. P.; Derouin, F. et.al. Application of topological
descriptors in QSAR and drug design: History and New trends. *Current Drug

38. Jhanwar, B.; Sharma, V.; Singla, R.K. et.al. QSAR - Hansch analysis and related
approaches in drug design *Pharmacology Online*, 2011, 1, 306-344.


40. Podlogar, B. L. and Ferguson, D. M. QSAR and CoMFA: a perspective on the


42. Kuntz, I. D.; Blaney, J. M.; Oatley, S. J. A geometric approach to

43. Levinthal, C.; Wodak, S. J.; Kahn, P. et al. Hemoglobin interaction in sickle cell
1975, 72, 1330-34.

44. Salemme, F. R. A hypothetical structure for an intermolecular electron transfer


