Computer aided drug design

A Jain

Department of Pharmaceutical Sciences, Jyoti Mahila Vidya Peeth, Jaipur, India

*E-mail: anurekhajain1@gmail.com

Abstract. Computer based method can help in discovery of leads and can potentially eliminate chemical synthesis and screening of many irrelevant compounds, and in this way, it save time as well as cost. Molecular modeling systems are powerful tools for building, visualizing, analyzing and storing models of complex molecular structure that can help to interpret structure activity relationship. The use of various techniques of molecular mechanics and dynamics and software in Computer aided drug design along with statistics analysis is powerful tool for the medicinal chemistry to synthesis therapeutic and effective drugs with minimum side effect.

1. Introduction

Drugs are the chemicals that prevent, treat, and make a diagnosis disease-restoring health to diseased individuals. As such, they play a central role in new medicine. For many years the strategy for discovering new drug consisted of taking a lead structure and developing a chemical programme for finding analog molecule exhibiting the desired biological properties. The process involved several trial and error cycles patiently developed and analyzed by medical chemist utilizing their experience and chemical intuition to ultimately select a candidate analog for further development. The entire process is laborious, expensive, and perhaps when looked today, conceptually inelegant [1]. Medicinal chemistry is the branch of science, which provides these drugs through discovery or design. As ever increasing understanding of the nature of the disease how cell works and how drug influence these processes has in the last two decades led more and more to the purposeful design, synthesis and evaluation of the candidate drug molecules [2]. The strategies of drug design have altered significantly within the past few decades. Whereas chemistry, biological activity hypotheses, and animal experiments dominated drug research, especially in its "golden age," from the sixties to the eighties of the last century, a lot of new technologies have developed over the past 20 years [3].

It was recognized in the 60s, that computer based method can help in discovery of leads and can potentially eliminate chemical synthesis and screening of many irrelevant compounds, and in this way, it save time as well as cost [4-6]. Molecular modeling has opened the way to the discovery of lead structure by a rational approach, and its central role in rational drug design has become fully apparent. The putative ligand can either be extracted from the large library of compounds [7-8] or can be obtained by joining the molecular fragment [9-11] or atom computationally. After that second pre requirement is an accurate prediction of the binding affinities which are a much more difficult task to achieve. A computational method that attempts to design leads, different in nature and in the degrees of the simplifying assumption they use [12].
Step involved in CADD are summarized in the following Figure 1 [13].

![Diagram of CADD process]

**Figure 1.** Interdisciplinary approach to rational computer aided drug design

2. **Computer Aided Drug Design**

Computer aided drug design CADD can contribute not only the design of potent compounds but too many of the step of going ‘from concept to clinic.’ In the development of the new drug, CADD methodologies and technology are used to calculate molecular properties to aid in drug design process [14]. CADD methods are aimed at using the information which is in the three-dimensional structure of
the unliganded target to design completely new lead compounds de novo, as well as to construct large virtual combinatorial libraries of compounds that then can be screened computationally before going to attempt and expense of actually synthesizing and testing them [15].

Some example of successful use of CADD for drug discovery

- Design of thymidylate synthases inhibitors as anticancer agents [16],
- HIV protease inhibitors as antiviral agents [17],
- Neutrophil elastase inhibitors [18],
- Carbonic anhydrase inhibitor as antiglaucoma agent [19],
- Discovery of novel sweetners [20]

**Molecular modeling**

Molecular modeling [21] has become a well-established research area during the last decade due to advances in computer hardware and software. Molecular modeling systems are powerful tools for building, visualizing, analyzing and storing models of complex molecular structure that can help to interpretate structure activity relationship.

**Molecular modeling functions**

There are many possible method and applications of molecular modeling

1) **Structure generation or retrieval**
   Molecular structure may be generated by a variety of procedures. If a crystal structure exists, it can be found in the Cambridge crystallographic data file and turned into molecular coordinates by a standard method. CONCORD [22] programme is used to convert the 2D structure into 3D molecular coordinates. LUDI [23] is programme to generate novel structure quickly according to some algorithm.

2) **Structure visualization**
   One of the most popular uses of molecular modeling system is to visualize molecular structure and interaction by a different method [24]. Representation of the structure is done with the stick model, Ball, and Stick model, Space–filling and Surface model.

3) **Conformation generation**
   Molecular modeling to calculate possible conformation and deduce which conformation(s) is important for the property of interest. Monte Carlo technique produced conformation that can be analyzed statistically or energetically [25].

4) **Deriving bioactive conformations**
   Given a molecule with biologically active, it is necessary to determine the conformation(s) of the molecule associated with its activity. If one assumes that two or more molecule binds in common modes, then the search for a common conformation may be made. For searching bioactive conformer involves the calculation of energetic since only low energy conformations and the second approach involves the generation of possible conformations in a systematic way. So that all possible binding patterns can be explored on the basis of criteria that one would be interested in include energy, allowed bonding and nonbonding distance, torsional angular increment, and other factors. Gundet al. have performed systematic conformations searching coupled with distance and energy parameters to determine bioactive conformations of a series of semi rigid nicotinic agonists [26].
5) Molecule superposition and alignment
Computing properties of molecules often involve comparisons across a homologous series, and that is done by superimposition or alignment of the molecules so that their differences become obvious and interpretable. When molecules possess a large, rigid common substructure, their alignment is relatively easy. In general case, the molecules are sufficiently different in structure or conformation that their alignment is unobvious and perhaps not unique [27]. Molecular modeling used for superimposition by different programs like DISCO and Aladdin [28,29].

Crippen[29] superimposed molecules of the different structure using 3-D quantitative structure-activity relationship (QSAR) method based on distance geometry.

6) Deriving the pharmacophoric pattern
A pharmacophore may be defined as the essential geometric arrangement of atoms or functional groups necessary to produce a given biological response. A given set of biologically active molecules which produces activity by the same mechanism are assumed to have the same essential pharmacophore. Pharmacophores have been described as topologic (graph-theoretic or connectivity-based structural fragment) and topographic (geometric, usually 3-D patterns) [29,30].

Automatic generation of pharmacophoric patterns is the goal of several modeling packages [31]. Automated pharmacophore mapping is now available in programs from Biocad, Biosyn, Chemical Design Ltd., and Tripos.

7) Receptor mapping
When the structure of the receptor is not known, a “receptor map” may be constructed using the principle of receptor ligand complementarity.

A number of algorithms are available to compute unions of molecular volumes. Pseudoelectron density functions calibrated to reproduce Van der Waals radii have been mapped to a 3-D grid to compute the union, intersection, and subtraction of volumes. Analytical volume representation by Connolly may be an alternative which would allow optimization of volume overlap [32,33].

8) Estimating biological activities
In reality, drugs vary in activity from being very active or potent to being inactive. QSAR is a technique that quantifies the relationship between structure and biological data and is useful for optimizing the groups that modulate the potency of a molecule [34].

Hansch et al. suggested that if one employed cluster analysis on all possible substituent’s and included one member from each of the clusters of similar substitutions, the designed ligands would show more independent variation in physical properties [35].

9) Molecular interactions
Modeling the interaction of a drug with its receptor is a complex problem because there are too many degrees of freedom and insufficient knowledge of the effect of solvent on the binding association and is stereospecific determined by a fit of the molecule to the receptor. Many forces are involved in the intermolecular association: hydrophobic, dispersion or Van der Waals, hydrogen bonding and electrostatic [36]. This type of interaction is determined by the structure or fit of the drug molecule to the receptor site and induces a common biological response.

10) Calculation of molecular properties
Molecular properties are important indicators of the utility of various molecules and categorized as physical (electronic, thermodynamic, physical state), chemical (reactivity, solubility, dynamics, explosively) or biological (enzyme inhibition, receptor toxicity, metabolism).
11) Energy calculations
Geometry optimization and energy minimization of the substrates can be carried out using quantum chemical ab initio and or semi empirical methods. For larger systems, molecular mechanics methods (force field techniques) can be employed. The three major theoretical, computational methods of calculating properties of molecules include empirical, semiempirical and ab initio (QM) methods. Molecular mechanics \[37\] and semiempirical methods rely on embedded empirical parameters, whereas, ab initio quantum mechanical methods are potentially capable of reproducing an experiment without such parameters.

a) Molecular mechanics
Molecular mechanics method is less complicated, fast and is able to handle very large systems including enzymes \[38,39\]. Molecular mechanics methods are widely used to give accurate structure and energies for molecules.

Molecular Mechanics energy minimization involves successive iterative computations where an initial conformation is submitted to full geometry optimization. All parameters defining the geometry of the system are modified by small increments until the overall structural energy reaches a local minimum. The local minimum, however, may not be the global minimum.

b) Semi empirical calculation
The semi empirical \[40\] method uses the mathematical formulation of the wave function which describes hydrogen like orbits. Semi empirical calculations differ in the approximations that are made concerning repulsion between electrons in a different orbital. The approximations are adjusted by parameterizing values to corresponding to either ab initio data or available experimental data. The earliest methods used were the extended Huckel theory and CNDO. Improved method includes MINDO-3, MNDO, and AMI. MOPAC contains the MINDO/3 and MNDO programs and has been parameterizing.

c) Quantum mechanics : \textit{ab} initio
The solution of the Schrödinger equation with approximations is the basis of semiempirical calculation, whereas the solution of the Schrödinger equation without approximations is the basis of the highest level quantum mechanical calculation so-called “ab initio” method. Abintio method are most useful for case where there is no experimental data to draw from but suffer from the disadvantages that much computer power is need and therefore the method is not routinely for systems with more than 50 heavy atoms \[41\].

Two major modeling strategies currently used in the conception of new drugs:
- Direct drug design
- Indirect drug design

Direct drug design
In this approach, the three dimensional features of known receptor sites are directly considered which are obtained from the X-ray crystallography or NMR analysis for the design of the lead structure. The X-ray crystal structure of macromolecules and their complexes with ligand may be obtained from protein data bank. If a protein structure is known, it can be used for explicit docking of ligands.

Docking
One goal of computational chemistry is to predict the binding interactions of molecules. Docking is a direct drug design approach which deals study of interactions between drug, and receptor \[42\]. Docking studies may help to increase ligand specificity; and also better therapeutic index can be achieved. Docking plays a role in the QSAR methods and homology modeling and is useful in the
structure-based drug design. Various docking programs are used such as DOCK, FLOG, ADAM, and UGIN.

Computational docking always requires two components, which may be briefly characterized as "searching" & "scoring." The searching is to find the orientation and conformation of the interacting molecules corresponding to the global minimum of the free energy of binding. "Scoring" [43] refers to the fact that any docking procedure must evaluate and rank the configurations generated by the search process. Scoring is actually composed of three different aspects relevant to the docking and design:

- Ranking of the configurations generated by the docking search for one ligand interacting with a given protein; this aspect is essential to detect the binding mode best approximating the experimentally observed situation.

- Ranking different ligands with respect to the binding to one protein, that is, prioritizing ligands according to their affinity; this aspect is essential in virtual screening.

- Ranking one or different ligands with respect to their binding affinity to different proteins; this aspect is essential for the consideration of selectivity and specificity [44].

- Various approaches to scoring aspects [45] that try to capture the essential elements of protein-ligand interactions have been described by several workers as outlined in Table 1.

**Table 1.** Name and type of scoring function for docking studies

| S.No. | Type of scoring function                      | Name of function          |
|-------|-----------------------------------------------|----------------------------|
| 1     | Force field                                   | CHARMM                     |
| 2     | Force field + desolvation                     | AMBER + desolvation        |
| 3     | Linear response                               | LIE                        |
| 4     | Simplified free energy perturbation           | OWFEG GRID                 |
|       |                                               | SCORE1                     |
|       |                                               | GOLD SCORE                 |
|       |                                               | PLP                        |
|       |                                               | Flex X SCORE               |
|       |                                               | VALIDATE                   |
| 5     | Empirical                                     | CHEM SCORE                 |
|       |                                               | SCORE2                     |
|       |                                               | SCORE                      |
|       |                                               | AUTODOCK 3                 |
|       |                                               | FRESNO                     |
|       |                                               | SCREEN SCORE               |
| 6     | Desolvation terms                             | HINT                       |
|       |                                               | SMOG                       |
| 7     | Knowledge based                               | BLEEP / PMF                |
|       |                                               | DRUGSCORE                  |
Types of docking
Following different types of docking [46] have been reported:

- Flexible protein-ligand docking
- Flexible protein-protein docking
- Hydrophobic docking

When X-ray crystallography structure of receptor is not available for docking then homology modeling is used to develop homologous model discussed as below:

Homology modeling
Homology modeling [47, 48, 49] involves taking a known sequence of an unknown structure and mapping it against a known structure of one or several similar (homologous) proteins. It would be expected that two proteins of similar origin and function would have reasonable structural similarity. Therefore, it is possible to use the known structure as a template for modeling the structure of the unknown structure [48]. Homology modeling approaches consists of following steps [49, 50] as shown in Figure 2.

![Homology Modeling Diagram](image)

Figure 2. Ray diagram defining the steps of homology modeling

Indirect drug design
The design is based on the comparative analysis of the structural features of known active and inactive molecules that are complementary to a hypothetical receptor site. A successful application of this technique has been in the designing of drug Sildenafil by Pfizer laboratory, England [51].

Quantitative structure activity relationship (QSAR)
Quantitative structure activity relationship (QSAR), a method introduced in 1960, is a useful tool to establish quantitatively the relationship between various physicochemical properties and biological activity of compounds. One of the most important goals of new age of QSAR as an integral part of
drug design and discovery is a unique growth of bimolecular database which contained data on chemical structure and in some cases, biological activity of chemical [52].

**Role of QSAR in design of better drugs**
- QSAR allows making a quantitative prediction of the potency of new analogs [53].
- It is used for the design of series based on a lead molecule.
- One can also find a new region on the receptor and thus start a whole new series.
- QSAR helps to decide to stop synthesis in a series.
- Helps to find the mechanism of action of the molecule and thus complement receptor mapping technique.
- The QSAR result can be used to understand interactions between functional groups in the molecules of greatest activity, with those of their target.
- Quantifying the relationship between structure and activity provides an understanding of the effect of structure on activity.

**Drawback of QSAR**
- Due to too heavy a reliance being placed on biological data, which, by its nature, is subject to considerable experimental error, thefalse correlation may take place.
- Various physicochemical parameters are known to cross correlated. Only variables or their combination that have little covariance should be used in QSAR analysis; similar considerations apply when correlation are sought for a different set of biological data.
- Many QSAR cannot be used to confidently predict the most likely compounds of best biological activity because frequently experiments upon which QSAR analysis depend, lack design in the strict sense of experimental design.

**QSAR involves four main steps**
- Converting molecular structures into mathematical descriptors that review the key properties of the molecules applicable to the activity or property being modeled.
- Selecting the finest descriptors to form a large set of available, significant descriptors.
- Mapping the molecular descriptors into the properties, preferably using a model free mapping system in which no hypothesis are needed as to the functional form structure activity relationship are often complex, unknown and non-linear.
- Validating the model to decide how predictive it is, and fine it will generalize to new molecules not in the dataset in used to generate the model (The training set).

**Quantitative model**

1) The Extra thermodynamic approach (Hansch Analysis)

Hansch analysis [55] correlates biological activity value with physiochemical properties, thus Hansch analysis really a property- property relationship model. All parameter used in the Hansch analysis are linear free energy value, the term linear free energy linked move towards or extra thermodynamic are used as synonyms for Hansch analysis. The general from of Hansch QSAR equation is:

\[
\log \left( \frac{1}{c} \right) = a \pi + b \pi^2 + c \delta + d E_s + e
\]

Where,
- \(c\) = Concentration of drug necessary to release a specific biological response.
- \(\pi\) = Hydrophobicity constant that can be replaced or complemented with log P
- \(\delta\) = Hammett substitution constant
- \(E_s\) = Taft steric constant
2) Free Wilson analysis (additivity model or denovo approach)
It is true Structure activity relationship model [56]. It is based on the following assumption:

- The entire drug listed should have the same parent structure.
- The substitution pattern in various derivatives has to be same.
- The substitutions have to contribute to the biological activity additively and in the same position with a constant amount being independent of the presence or absence of other substitution in the molecules.

So the total activity ($A$) of a certain derivative is the sum of constant independent partial contribution.

$$ A = \sum (a_i I_i) + \mu $$

Where,

- $I_i$ = the substituent, $a_i =$ the contribution of substituent $I$, $I_i =$ the contribution of parent structure.

The equation is solved by multilinear regression using the presence (1) or absence (0) of the different substituents as independent dummy parameters, while the measured activity served as the dependant variable.

3) Mixed approach
It is Intercorrelation between Free Wilson parameters and physicochemical properties of the substituent’s used [57]. Form the general formulation of linear Hansch equation ($\Phi_i$ is any physicochemical property) group contribution $a_i$ can be derived for each substituent’s under consideration (Eq.1 $\Phi_{ij}$ is a physicochemical property $j$ of the substituent $X_i$).

$$ \log \frac{1}{c} = K_1\Phi_1 + K_2\Phi_1 + \ldots + K_n\Phi_1 + C \quad \text{---------Eq.1} $$

$$ a_i = \sum K_i\Phi_{ij} $$

It was a powerful tool for the quantitative description of large and structurally diverse data sets.

Descriptors used in QSAR
The various physicochemical properties of the compound can be represented by means of descriptors in QSAR [58]. Some descriptors used in QSAR are shown in Table 2.
Table 2. Descriptors in QSAR

| S. No. | Descriptor Type | Descriptors                                      | Symbol |
|--------|----------------|-------------------------------------------------|--------|
| 1      | Hydrophobic    | Partition coefficient                           | log p  |
|        |                | Pi substituted constantly                       | $\pi$, $\pi^2$ |
|        |                | Chromatographic parameter                       | log Rm |
|        |                | Solubility                                       | $\delta$ |
|        |                | Parachor                                         | [P]    |
|        |                | Elution time in HPLC                             | log $K$ |
|        |                | Ionization constant                              | $p_ka$, $\Delta p_ka$ |
|        | Electronic     | Sigma substituted constant                       | $\sigma$, $\sigma^2$, $\sigma^*$ |
|        | Experimental   | Spectroscopic chemical shift                     | Fr, ppm |
|        |                | Resonance effect                                 | $R$    |
|        |                | Field effect                                     | $F$    |
| 2      | Theoretical    | Ionization potential                             | $I$    |
|        |                | Quantum mechanical                               | $C$    |
|        |                | Atomic charge densities Indices                  | $q_i$  |
|        |                | Atomic net charges                               | $Q^r_i$, $Q_i$, $Q^a_i$, $Q^p_i$ |
|        |                | Super delocalizability                            | $S_{r_i}^N$, $S_{r_i}^E$, $S_{r_i}^R$ |
|        |                | Taft’s substituent constant                      | $E_s$  |
|        |                | Vander Waals radii                               | $\Gamma$ |
|        | Steric         | Inter atomic distances                           | $B_i$, $L$ |
|        |                | Molar refractivity                               | $MR$   |
|        |                | Molar volume                                     | $MV$   |
|        |                | Sterimol                                         | $L_i$, $B_1$, $B_3$, ($B_2$, $B_3$, $B_5$) |

Different approaches that can be used in QSAR

1) **1D QSAR**

1D QSAR [59] makes use of structural descriptors only, like functional groups or atom centered fragments and is empirical in nature. It uses indicator feature and zeroes for its absence as sole parameters. It can be used in conjunction with other 2D or 3D variables. QSAR model is generated by some statistical method, and the parameter is contributing to the biological activity then the functional group is essential if negatively contributing then not essential to biological activity.

2) **2D QSAR**

2D QSAR [60-66] was started in the early sixties by the pioneering work of Hansch-Fujita and Free-Wilson. Hansch and Fujita proposed to express the biological activity as a function of molecular or fragment descriptor.

\[
\text{Biological activity} = f (\text{molecular or fragment descriptor})
\]

3) **3D QSAR**

3D QSAR [67-68] is a well recognized method at a molecular level the interaction that produced an observed biological response are usually non covalent and that such steric and electrostatic interaction can account for many of the observed molecular properties. Extension of the traditional QSAR approaches has been developed which explicitly uses the 3D geometry of structure during the development of the QSAR model.

The fundamental problems when trying to develop a good and predictive 3D QSAR model is the identification of the bioactive conformation(s) of the investigated compounds and how to align them. 3D QSAR is quantitative models that relate the biological activity of small molecules with their properties calculated in 3D space.

A variety of techniques can broadly be used in 3D QSAR:
a. Comparative molecular field analysis (COMFA)

COMFA slowly developed from the first attempt to place molecule in the grid and to correlate properties of the molecules with biological activities. COMFA [69-71] introduced in 1988, depends on the fact that biological activity is especially sensitive to spatially localized differences in molecular field intensities. It is the most common approach for the determination of steric and electronic features of series. It is also identifying a region in space that is favorable or unfavorable for ligand receptor interactions. Molecules are described with molecular interaction fields similar to those computed by PLS, and cross-validation [72, 73] and the output is displayed as contours superimposed on the molecules. Contour maps are helpful in suggesting new compound likely to have higher property values and potential for synthesis.

b. Molecular shape analysis

It was introduced in 1980. Since the shape of the molecule is important in the receptor interaction, hence MSA is used to incorporate shape data in QSAR equations. The basic goal of MSA [74] is to identify the biologically relevant conformation without knowledge of the receptor geometry and in a quantitative fashion, explain the activity of a series of analogues using only the structure activity table between a reference structure and the compound of data set.

c. 3D-QSAR based on molecular similarity metrics

The basic underlying principle of both active subset selection and structure activity relationship studies assumes that compounds similar to biologically active ones should also be active and vice versa. The use N x N similarity matrices as the input vector [76]. During the process of molecule alignment, one of the compounds is very potent ligand, and the other are superimposed on it by maximizing the degree of steric and electronic similarity [77-78]. A the matrix can be generated by pair wise compound shape similarity comparison, in which each matrix entry is a measure of shape similarity between the corresponding pair of molecules. Analysis can be done by using PLS.

d. Distance geometry method

Distance geometry represents molecule in term of interatomic distances rather than certain coordinates [79-83]. Because the distance between atom does not change with rotation and translation of the whole molecule, the analysis of distance matrix can detect if and how a certain atom of two or more molecules overlap without performing a superposition. Distance geometry offers not only the projection of important intratomic distance that relates to biological activity but also to postulate critical intermolecular-binding distance that may be involved in the ligand-receptor interaction.

e. Quantitative binding site model, COMPASS

COMPASS algorithm that automatically deduces interchemotype relationships and generates predictive quantitative models of receptor binding based solely on structure-activity data [84]. It predicts the bioactive conformation, alignment and binding affinities of a series of ligands in an automated procedure based on surface type properties. COMPASS uses steric, hydrogen bond donor and hydrogen bond acceptor distances [85].

f. Receptor surface model

Receptor surface model (RSM) [86] generates a surface loosely enclosing the common volume of the most potent ligands. Points on this surface are described by the proficient of the average partial charges, electrostatic potential, and hydrogen bonding ability or by the average hydrophobic of the most active compounds. By this method user decide which active compound to use and how close to the Vander walls envelope of the overlapped molecule the surface should be. Genetic function algorithms [87] can be used to detect which set of molecule produces the most predictive models.
g. 3D QSAR from GRID probe comparison
GRID probe force field to compute the interaction energy of a series of compound and a probe over a regular 3D grid around the target molecule to develop 3D QSAR used by Davis et al. Along with other descriptor for the ligands the total energy of interaction equal to sum of electrostatic, steric and hydrogen bonding terms, at each point are calculated. A regression analysis, such as PLS, indicate the GRID points with high PLS components are useful for defining that component [88-89].

h. Pharmacophore Mapping
It is a collection of chemical feature distribution in 3D space that is intended to represent the group in the molecule that participates in important binding interaction between drugs and their receptor [89]. Pharmacophore model is spatial arrangement of atoms are functional group believed to be responsible for biological activity [90].

i. 3D QSAR based on the intermolecular contribution to binding energy
In this method use of intermolecular binding energies of ligand structure and receptor interaction independent and biological activity values as dependent variables to derive a significant QSAR model to predict the binding potency of another ligand.

j. Comparative binding energy analysis
Comparative binding energy analysis (COMBINE) [92-93] can be regarded as the receptor dependent analogue method to COMFA. The different region of the receptor serves as a probe for the elucidation of major interaction sides. For each ligand, the intermolecular and interamolecular energies are calculated for the ligand receptor complexes, the unbound ligand, and the receptor.

k. De NOVO ligand Design approaches
De novo ligand [94] design algorithms are applicable when the structure of the receptor is available, and they generally involve predicting a ligand’s complementary to the receptors active side by using a geometry or partial force field based scoring function.

l. Hypothetical active site lattice
Hypothetical active site lattice (HASL) was described in 1988, the same year as COMFA. HASL is related to the COMFA methodology and also to MSA [95,96]. The HASL approach represents each of the shapes of the molecules as a collection of 3D grid points. It determines the number of lattice points that represent a molecule and also the resolution of the generated receptor map.

m. Genetically Evolved receptor models
Genetically evolved receptor models (GERM) is based on the principle to produce atomic level models of receptor sites based on atrial set of ligands [97]. It optimizes the correlation of the biological potency of molecules with the potential energy of the interaction of the ligands with probes placed on the union surface of the superimposed molecules. The novel feature of GERM is the character of the probes is not fixed, but rather evolves during the calculation using a genetic algorithm [98, 99].

4) 4D QSAR
4D QSAR tool with the fourth dimension being the possibilities to represent each molecule by an ensemble of conformations, orientation, prolongation states and thereby reducing the bias associated with ligand alignment [100].

5) 5D QSAR
The Concept of 4D QSAR (Software Quasar) has been extended to 5D QSAR [101] by an additional degree of freedom – the fifth dimension which allows for multiple representations of the topology of the quasi atomistic receptor surrogate. While this entity may be generated using up to six different induced fit protocols, the simulated evolution converges to a single model and that 5D QSAR due to
the fact that model selection may vary throughout the entire simulation — yields less biased results than 4D QSAR where only a single induced fit model can be evaluated at a time.

**Statistical basis of QSAR studies**

1. **Simple linear regression**
   
   Simple linear regression (SLR) describes the relationship between a single independent variable (X) and a single dependent variable (Y), assuming a linear (straight line) relationship between the dependent variable and the independent variable.

   \[ Y = a + bx ± e \]

   Where \( a \) is the intercept, \( b \) is the slope, \( e \) is an error term.

2. **Multiple linear regressions**
   
   Multiple linear regressions (MLR) calculates QSAR equation by using multiple variables in a single equation. The variable should be independent and to minimize the possibility of chance correlations. The number of independent variables can not be more than the one fifth of the compound in the training set. It describes the relationship between two or more (m) independent variables (X1, X2, Xm) and a single dependent variable (Y), assuming a linear (straight line) relationship between the dependent variable and the m independent variables.

   \[ Y = a + b1X1 + b2X2 + …. bmXm± e \]

   Where \( a \) is the intercept, \( b \)'s are the slopes, \( e \) is an error term.

3. **Stepwise multiple linear regressions**
   
   It calculates QSAR equation by adding one value at a time testing each addition for significance. Only variable found to be significant are used in the final QSAR equation. It is a multiple regression procedures where independent variables (Xi) are entered into the multiple regression equations. Each independent variable is tested to see whether it meets specific selection criteria before it is allowed to enter (forward stepwise) or remain in (backward stepwise) the regression.

4. **Sequential linear multiple regression**
   
   It is a process in which all the permutations and combinations of the parameters are tested sequentially to get the best set of n parameters which contribute significantly to the dependent variable optimally.

5. **Partial least square analysis**
   
   It is the most promising new approach in the multivariate statistic. Hundred or even thousand of the independent variable in (the X block) can be correlated with one of several dependant variables (the Y block). PLS is used when the X data contain colinearities or when N is less than 5 M, where N is the number of compounds and M is the number of dependant variable. Often perfect correlations are obtained in PLS analysis due to a large number of X variable. It can be used to solve the problems like the ratio of compounds to descriptor should greater than five, and the descriptor should not be intercorelated. PLS can handle numerous and even collinear variables.

6. **Discriminate analysis**
   
   Discriminate analysis separates things with different properties e.g. active and inactive compound, by deriving a function of other features which gives the best separation of individual class. So it gives bioactivity classification such as active Vs inactive.
7. Cluster analysis
The non statistical method that is used to study the relationship between compounds based on their physical and biological properties. The cluster was found on the basis of the distance between compounds in a space formed by the physical or biological properties. Each compound is a point in this space [104].

8) Principle component analysis
PCA [104] analysis has recently been used to describe the biological activities and molecular diversity of heterocyclic aromatic ring fragment. The aim of this study was to identify principal component which correlates the chemical structures with biological activities and to enable medicinal chemists to rationally select which heterocyclic rings to synthesize in order to optimize biological activities.

Statistical parameter used in QSAR
1) QSAR model building
   a) Model equation:
   The model [105] expresses the value of a predictand variable as a linear function of one or more predictor variables and an error term: Equation 1:
   \[ Y_i = b_0 + b_1x_{i1} + b_2x_{i2} + \ldots + b_kx_{ik} + e_i \]
   \( x_{ik} \) = value of kth predictor in year i
   \( b_0 \) = regression constant
   \( b_k \) = coefficient on the kth predictor
   \( k \) = Total number of predictors
   \( Y_i \) = predictand in year i
   \( e_i \) = error term
   
   b) Residuals
   The error term in equation 1 is unknown because the true model is unknown.
   \[ e = \hat{Y}_i - \hat{Y}_i \]
   \( \hat{Y}_i \) = Observed value of predictand in year i
   \( \hat{Y}_i \) = Predicted value of predictand in year i
   The residuals measure the closeness of fit of the predicted values and actual predictand. The variance of the residuals measures the “size” of the error and is small if the model fits the data well.

2) Correlation coefficients
   a) Pearsonian correlation coefficient (r)
   It is the multiple correlation coefficients and is the maximum correlation between the dependent variable and any linear function of the independent variables; it ranges from 0 (no correlation of Y with the multiple X variables) to +1 (perfect positive correlation of Y on the multiple X variables).
   \[ r = \sqrt{(SSR/SST)} \]
   SSR: Sum of squares, regression
   SST: Sum of squares, total
b) Coefficient of determination (R\(^2\))
   It is the fraction of the variability in the dependent variable that is explained by the variability in the dependent variable(s).
   The explanatory power of the regression is summarized by its “R-squared” value, computed from the sums-of-squares terms as -
   \[
   R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}
   \]
   If the regression is “perfect,” all residuals are zero, SSE is zero, and R\(^2\) is 1. If the regression is a total failure, the SSR equals the SST, no variance is accounted by regression, and R\(^2\) is zero.

c) Adjusted coefficient of determination (R\(^2\)adj)
   It provides an unbiased estimate of R\(^2\). The R\(^2\) value for a regression can be made arbitrarily high simply by including more and more predictors in the model. The adjusted R\(^2\) is one of the several statistics that attempts to compensate for this artificial increase in accuracy. The adjusted R\(^2\) is given by
   \[
   R^2_{adj} = 1 - \frac{(n-1)}{(n-p)}(1 - R^2)
   \]
   Where, \(n\) = sample size, \(p\) = number of predictors in the model, not counting the constant term, R\(^2\)adj value is always lower than R\(^2\)

d) Standard error of the estimate (SEE, RMSE, std)
   The residual mean square (MSE) is the sample estimate of the variance of the regression residuals.
   The square root of the residual mean square is called the root-mean-square error (RMSE), or the standard error of the estimate.
   \[
   s_e = \sqrt{s_e^2} = \sqrt{MSE} = RMSE
   \]

e) F and t statistics
   i) F ratio
      It estimates the statistical significance of the regression equation. The advantage of the F- ratio over R\(^2\) is that the F- ratio takes into account the degrees of freedom, which depend on the sample size and the number of predictors in the model. A model can have a high R\(^2\) and still not be statistically significant if the sample size is not large compared with the number of predictors in the model. The significance of the F- ratio is obtained by referring to a table of the F distribution, using degrees of freedom \{df1, df2\}, where df1 and df2 are K and n-K-1 respectively.
   ii) t statistic
      The t value for each coefficient tests the null hypothesis that the true value of the coefficient does not differ from 0. It tests the hypothesis that there is no relationship between the dependent variable and the independent variable associated with a given data.
      Looking for a t-ratio greater than 2 in absolute value is a common rule of thumb for judging significance because it approximates the 0.05 significance level.
      t statistic for the regression slope \(\beta\) tests whether it is significantly different from 0 (or any particular value, \(b\)); it is calculated as
      \[
      t_{n-k-1} = (\beta - b)/(Se/\sqrt{\sum x^2})
      \]
Where,

Se is the standard error of estimate
K is the number of predictors in the model
n is the sample size

3) Parameters that define fitness of the model

a) FIT (Kubinyi function)

It is closely related to the Fisher ratio (F).

\[
FIT = R^2. \frac{(n-K-1)}{(n+K^2)(1-R^2)}
\]

Where, R is the correlation coefficient.

The best model will present the highest value of the FIT function.

b) Z-score

\[
Z_i = \frac{x_i - \bar{x}}{s}
\]

Where,

\[
s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}}
\]

It is used to label the outliers in the dataset.

Where,

n = Number of data points
x_i = estimated value of predictand
\bar{x} = Mean of observed values of predictand

4) Parameters that define fitness of the predicted results

a) Reduction of error (RE, R^2(pred))

It is similar to R^2 of the model building and q^2 of LOO (Leave One Out) cross-validation.

In the R^2(pred) calculation, predicted value is replaced by estimated value.

5) Cross-Validation

a) Leave One Out

In the case of leave one out (LOO) [106,107] where by sequentially removing one data point at a time. The training dataset is divided into subsets (number of subsets = number of data points) of equal size. The model is build using these subsets, and the predict and (dependent variable) value of the data point that was not included in the subset is determined, this is the predicted value. Mean of predict and will be same for R^2 and LOO.q^2 (since all the data point will be sequentially considered as predict and in the LOO subset)
b) Cross-validated correlation coefficient value (q^2)
   It is obtained by LOO method, where a model is built with N-1 compounds, and the N^{th}
   compound is predicted. Each compound is left out from the model derivation and predicted in
   turn.
   In LOO calculation, predicted value is replaced by LOO predicted value.
   The condition q^2>0.5 is the basic requirement for declaring a model to be a valid one.

c) Predicted residual sum of squares
   The predicted residual sum of squares (PRESS) procedure is equivalent to “leave-one-out”
   cross-validation, as described previously. The PRESS statistic is defined as
   \[ \text{PRESS} = \sum_{i=1}^{n} \hat{e}_{(i)}^2 \]
   \( \hat{e}_{(i)} \) is the residual for observation i computed as the difference between the observed
   value of the predictand and the prediction from a regression model calibrated on the set of
   n−1(i) observations from which observation i was excluded.

d) Standard error of prediction
   The standard error of prediction (S\(_{\text{PRESS}}\)) is the sample prediction of the variance of the
   regression predicted residuals.
   \[ \text{S}_{\text{PRESS}} = \frac{(\text{PRESS}/n-K-1)}{n-K-1} \]

e) Standard deviation of prediction
   Standard deviation of prediction (S\(_{\text{DEP}}\)) is described as
   \[ \text{S}_{\text{DEP}} = \sqrt{\frac{\text{PRESS}}{n}} \]
   Internal consistency of the model is supported by S\(_{\text{PRESS}}\) and S\(_{\text{DEP}}\). These values are calculated
   from LOO cross validation. Lower values of these parameters describe the better predictability
   of this model. This value depends upon the variability (log difference) in the observed value
   of predictand (e.g. for two log difference it should be below 0.3).

f) Bootstrapping
   Bootstrapping analysis was performed for further access to the robustness and statistical
   confidence. In bootstrapping repetitively analyzed sub-samples of the data. Each sub-sample is
   a random sample with replacement from the full sample. One data point can be represented
   more than once or not at all, but the total number of data points should remain constant.
   The bootstrapping analysis gives an overview of the contribution of individual molecules
   to the QSAR model.

i. The Bootstrapping squared correlation coefficient
   Bootstrapping squared correlation coefficient (R^2_{bs}) is the average squared correlation
   coefficient calculated during the validation procedure that is computed from a subset
   of data points used one at a time for the validation procedure.

g) Chance
   Chances of fortuitous correlation were tested with the help of Chance statistics. It is evaluated
   as the ratio of the equivalent regression equations (equal or greater than conventional R^2) to
   the total number of randomized sets. In the randomized set, Independent parameters values are
   frozen, and dependent parameter values are shuffled. Later R^2 values for these randomized
   sets are calculated. A chance value of 0.001 corresponds to 0.1% chance of fortuitous
   correlation.
   \[ \text{Chance} = \frac{(\text{Nos of } R^2 >= \text{ conventional } R^2)}{\text{Number of randomized sets}} \]
References

[1] Cohen N C 1996 Guidebook on Molecular Modeling in Drug Design. p 1-17.
[2] Mitscher L A 2014 Drug Design and Discovery an overview. In Handbook of Drug Design and Discovery in Larsen P K, Liljefors T, Madsen U eds. (New York: CRC Press) p 1-32.
[3] Donald B B 1990 Successes of Computer-Assisted Molecular Design. In reviews in computational chemistry In Lipkowitz K B, Boyd D, eds., (New York : VCH Weinheim) p 55-371.
[4] Topliss J G 1972 Utilization of operational schemes for analogue synthesis in drug design. J. Med. Chem. 15 1006-11.
[5] Redl G, Cramer R D, Berkoff C E 1974 Quantitative Drug Design. Chem. Soc. Rev. 3 273-92.
[6] Ewing T JA, Kuntz I D 1997 Critical evaluation of search algorithms for automated molecular docking and database screening. J. Comput. Chem. 18 1175-89.
[7] Fountain E 1992 Application of genetic algorithms in the field of constitutional similarity. J. Chem. Inf. Comput. Sci. 32 748-52.
[8] Moon J B and Howe W J 1991 Computer design of bioactive molecules a method for receptor based de novo ligand design. Proteins. 111 314-28.
[9] Bohme H J 1995 Site directed structure generation by fragment joining. Perspect. Drug. Discuss. Des. 3 21-33.
[10] Bohacek R S and McMartin C 1997 Modern computational chemistry and drug discovery: structure generating programs. Curr. Opin. Chem. Biol. 1 157-61.
[11] Bohacek R S and McMartin C 1995 De novo design of highly diverse structures complementary to enzyme binding sites – application thermolysin. Comput. Aid. Mol. Design. 3 21-33.
[12] Ajay A and Marko M A 1995 Computational methods to predict binding free energy in ligand receptor complexes. J. Med.Chem. 38 4953-67.
[13] Leow D H, Vilar H O and Alkorta I 1993 Strategies for indirect computer Aided Drug Design. Pharma. Res. 10 475-86.
[14] Liljefors T and Petterson I 2004 Computer-Aided Development and Use of Three Dimensional Pharmacophore Models In Handbook of Drug Design and Discovery; Larsen P K, Liljefors T and Madsen U., Eds, (New York: CRCPress)p 86-116.
[15] McCarthy J D 1999 Computational approaches to structure-based ligand design Pharmacology & Therapeutics. 84 179–91.
[16] Appelt K, Becquet R J and Barlett C A 1991 Design of enzyme inhibitors using iterative protein crystallographic analysis. J. Med. Chem. 34 1834-925.
[17] Thaisrivongs S 1994 HIV protease inhibitors. Annu. Rep. Med. Chem. 29 133-44.
[18] Hlasta D J and Pagani E D 1998 Human leukocyte elastase inhibitors. Annu. Rep. Med. Chem. 29 195-204.
[19] Baldwin J J, Panticello GS and Anderson P S 1989 Novel topically active carbonic anhydrase inhibitors for the treatment of glaucoma. J. Med. Chem. 32 2510-13.
[20] Worthy W Synthetic Sweeteners Developed Design. Chem. Eng. News. 8
[21] Cohen N C, Blaney J M, Humblet C, Gund P and Barry DC 1990 Molecular modeling software and method for medicinal chemistry. J. Med. Chem. 33 883-94.
[22] Pearlman R S 1987 Rapid generation of high quality approximate 3D molecular structures. Chem. Des. Auto. News. 2 17.
[23] Boehm H J 2004 A novel computational tool for automated structure-based drug design. *J. Molec. Recogn.* **6** 131-37.

[24] Marshall G R, Naylor C B and Rams den C A 1990 The Rational Design, Mechanistic Study and Therapeutics Application of Chemical compounds. In Comprehensive Medicinal Chemistry, Quantitative Drug Design in Hansch C, Sammes P G, Taylor J B and Ramsden CA, Eds. (New York: Pergamon Press) **4** p 431-58.

[25] Hagler A T, Osguthorpes D J, Dauber O P and Hampel J C 1985 Dynamics and conformational energetics of a peptide hormone: vasopressin. *Sci.* **227** 1309-12.

[26] Spivac C E, Waters J A, Hermosmeier M, Yadav J S, Liang R F and Gaud T M 1988 Synthesis, pharmacology, and molecular modeling studies of semirigid, nicotinic agonists. *J. Med. Chem.* **31** 545.

[27] Bures G M, Danaher E, DeLaZzer J and Martin C Y 1994 New molecular modeling tools using three-dimensional chemical substructures. *J. Chem. Inf. Comput. Sci.* **34** 218–23.

[28] Brint A T and Willet P 1987 Algorithms for the identification of three-dimensional maximal common substructures. *J. Chem. Info. Comput. Sci.* **27** 152.

[29] Crippen G M 1987 Voronoi binding Sites Models. *J. Comput. Chem.* **8** 943-55

[30] Gund P 1977 3D Pharmacophore pattern searching. *Progress in Molecular and Subcellular Biology.* **11** 117-143.

[31] James P S 2006 Computer-assisted drug design. PartI. Conditions in the 1980s. *Med. Research. Rev.* **6** 641-62.

[32] Conolly M L 1983 Solvent-accessible surfaces of proteins and nucleic acids. *Sci.* **221** 709.

[33] Conolly M L 1983 Analytical molecular surface calculation. *J. Appl. Cristallogr.* **16** 548.

[34] Wold S and Dunn J W III 1983 Multivariate quantitative structure-activity relationships (QSAR): conditions for their applicability. *J. Chem. Inf. Comput. Sci.* **23** 6-13.

[35] Hopfinger A J 1985 Computer-assisted drug design. *J. Med. Chem.* **28** 1133.

[36] Fersht A R 1984 Basis of biological specificity. *TIBS.* **9** 144-7.

[37] Buckert U and Allinger N L 1982 *Molecular Mechanics.* (Washington D.C.: American Chemical Society) p 157.

[38] Osawa E and Musso H 1982 Application of molecular mechanics calculations to organic chemistry. In *Topics in Stereochemistry* **1** (New York: Wiley New York)p 117.

[39] Stewart J J P 1990 Semi empirical Molecular Orbital Methods. In Reviews in computational chemistry; Lipkowitz K. B, Boyd, D. Eds. (New York: VCH Weinheim) p 45-81.

[40] Zerner M C 1991 Semi empirical Molecular orbital Methods. In Reviews in *Computational Chemistry* **2**; Lipkowitz, K. B, Boyd, D. B. Eds. (New York: VCH Weinheim) p 313.

[41] Feller D and Davidson E R 1980 Basis sets for ab initio molecular orbital calculations and intermolecular interactions. In reviews in *computational chemistry*; Lipkowitz K. B, Boyd, D. Eds. (New York: VCH Weinheim) p1-43.

[42] Thomas G 2002 *Medicinal Chemistry* (New York: John Wiley Sons) p 67.

[43] Wang R, Liu L, Lai L and Tang Y 1998 SLORO: A New empirical method for structure based estimation of binding affinities. *J. Mol. Modeling.* **4** 379.

[44] Lewis R A, Pickett S D and Clark D E 2000 Computer-aided Molecular diversity analysis and combinatorial library design. *Rev. Compute. Chem.* **16** 1-5

[45] Jain A, Jain A, Bharadwaj S, Agrawal R, Jain A. and Pancholi S S 2005 Docking: A direct drug design approach. *Indian Drugs* **42** 777-86.
[46] Crippen G M 1979 Distance geometry Approach to Rationalizing Binding Data. J. Med. Chem. 22 988-97
[47] Bourne E P and Weissig H 2003 Homology Modeling. Structural Bioinformatics. 25 507.
[48] Mosoberg I H and Pogozheva D I 2005 Homology modelling of opioid receptor ligand complexes using experimental constraints. AAPS J 7 E434.
[49] Szymoszek A and Zacharias M 2003 Reduced Protein Model as A Tool in the homology Modeling, Internet Electronic Conference of Molecular Design Biochem Press.
[50] Available from: http://swissmodel.expasy.org (Accessed in June 2010).
[51] Chun M S, Shen J L and Joo C T 2009 Recent Advances in Computer-Aided Drug Design. Briefings in Bioinformatics. 10 579-91.
[52] Tropsha A 2003 Recent trends in quantitative structure activity relationship. In Burger’s Medicinal Chemistry and drug discovery 1 in Abraham, D.J., Ed, (New York: Willey Intersciences Publication)p 49-76.
[53] Jain A, Chaturvedi S C 2000 3D QSAR Pharmacophoric mapping of some Alpha-Thiol Dipeptides as angiotensin converting enzyme inhibitor. M. Pharm Thesis, DAVV Indore, MP. 28.
[54] Winkler D 2002 The role of Quantitative structure relationship in bimolecular discovery. Briefing in bioinformatics 3 73-86.
[55] Kubinyi H QSARhansch analysis and related approaches; Ed, (New York: VCH Weinheim/ New York)p 21-56.
[56] Free S M and Wilson N J 1964 A Mathematical Contribution to Structure Activity Studies. J. Med. Chem. 7 395-99.
[57] Kubinyi H 1990 The Free-wilson method and its relationship to the extrathermodynamic approach. In Comprehensive Medical Chemistry, Quantitative drug design 4 In Hansch, C. Sammes, P.G., Tylor, J.B, Ramdsen, C.A. Eds (New York : PragramonPress) p 589-643.
[58] Norinder U and Hogberg T 2004 Quantitative structure-activity relationship and experimental design. In Handbook of Drug Design and discovery: Larsen, P.K. Liljefors, T., Madsen, U. Eds, (London and New York: Tylor and Francis) p 117-55.
[59] Victor E K and Anatomy G A 2005 Hirarchic system of QSAR models (1D-4D) on the base of simplex representation of molecular structure. J. Mol. Model. 32 475-67.
[60] Hansch C and Klein T E 1986 Molecular graphics and QSAR in the study of enzyme ligand interactions. On the Definition of Bioreceptor. Acc. Chem. Res. 19 392-400.
[61] Fujita T 1990 The Extrathermodynamic Approach to Drug Design. In Comprehensive Medical Chemistry; Quantitative Drug Design. 4;Hansch, C., Samms, P.G, Tatlor, J.B, Ramdsen, C.A. Eds, (New York : Pergamon Press)p 497-560.
[62] Gould R G 1972 Biological Correlations-The Hansch Approach. In Advances in Chemistry. (New York: ACS New York) p 114.
[63] Goldberg L 1983 Structure Activity Correlation as a Predictive Tool in Toxicology Fundamental, Methods and Application, Hemisphere. 204
[64] Hall L H and Kier L B 1991 The molecular connectivity chi indexes and kappa shape indexes in structure property modeling. In Reviews in Computational Chemistry; Likowitz, K.B, Boyd, D.B. Eds, (New York: VCH) p 367-422.
[65] Bersuker I B and Dimoglio A S 1991 The Electron Topological Approach to the QSAR Problem. In Reviews in Computational Chemistry; Lipkowitz, K.B, Boyd, D.B. Eds, (New York: VCH) p 423-479.

[66] Weinstein H, Osman R and Green J P 1979 The molecular basis of structure activity relationship: quantum chemical recognition mechanism in drug receptor interaction. In Computer Assisted Drug Design; Olson, R.E, Christofferson, R.E. Eds, American Chemical Society. 161-187.

[67] Martin Y C 1978 Quantitative Drug Design. (New York: Dekker) p 425.

[68] Burger A 1994 Medical Chemistry - the First Century. Med. Chen. Res. 4 3-15.

[69] Kim K H, Greco G, Novellino E A 1998 Critical Review of Recent CoMFA Applications. In 3D QSAR in Drug Design: Kubinyi, H, Folkers, G, Martin, Y. C. Eds, (Great Britain: Kluwer Academic Publishers) p 257.

[70] Cramer R D, Patterson D E and Bunce J D 1988 Comparative Molecular Field Analysis 1. Effect of Shape on Binding of Steroids to carriers Proteins. J. Am. Chem. Soc. 110 5959-67.

[71] Kim K H 1992 3D Quantitative structure activity relationships: non Linear Dependence Directly from 3D structures using Comparative Molecular Field Analysis (CoMFA). Quant. Struct. Act. Relat. 11 309-17.

[72] Wold S, Albano C, Dunnill W J, Edlund U, Esbensen K, Geladi P, Hellberg S, Johansson E, Lindberg W and Sjostrom 1987 Mathematics and Statistic in Chemistry; Kowalski, B.R. Ed, D. ReidelDordrect 17.

[73] Cramer R D, Bunce J D, Patterson D E and Frank I E 1988 Crossvalidation, bootstrapping and partial least square regression with multiple regression in conventional QSAR studies. Quant. Struct. Act. Relat. 7 18-25.

[74] Hopfnger J A 1980 QSAR Investigation of Dihydrofolatereductase inhibition by bakers trazine based upon molecular shape analysis. J. Am. Chem. Soc. 102 7196-206.

[75] Good A C, Se S and Richards WG 1993 Structure activity relationships from molecular similarity matrices. J. Med. Chem. 36 433-8.

[76] Rita S and Gerhard F E 2008 The similarity principle – new trends and applications in ligand-based drug discovery and ADMET profiling. Sci Pharm. 76 5–18.

[77] Good A C, Peterson S J and Richards W B 1993 QSAR’s from similarity matrices. Technique Validation and Application in the Comparison of Different Similarity Evaluation Methods. J. Med. Chem. 36 2929-37.

[78] Kearsley S K and Smith G M 1990 An alternation method for the alignment of molecular structures: maximizing electrostatic and steric overlap. Tetrahedron Comput. Methodol. 3 615-33.

[79] Ghose A K and Crippen G M 1985 Use of physicochemical parameters in distance geometry and related three dimensional quantitative structure activity relationship: a demonstration using eschericha coli dihydrofolatereductase inhibitors. J. Med. Chem. 28 333-46.

[80] Ghose A K, Crippen G M, Revankar G, Mckernan P, Smee D and Robins R 1989 Analysis of the in vitro activity of certain ribonucleosides against parainfluenza virus using a novel computer aided molecular modeling procedure. J. Med. Chem. 32 746-56.

[81] Crippen G M 1979 Distance geometry analysis of the benzodiaazepine binding site. J. Med. Chem. 22 11-9.
[82] Ghose A K and Crippen G M 1982 Quantitative structure- activity relationship by distance geometry: quinazolines as dihydrololate reductase inhibitors. *J. Med Chem.** 25* 892-9.

[83] Crippen G M 1977 A novel approach to calculation of conformation: distance geometry. *J. Comput. Phys.** 24* 96-107.

[84] Hoplinger A J and Tokarski J S 1997 Three Dimensional Quantitative Structure Activity Relationship Analysis. In *Practical Application of Computer Aided Drug Design*; Charifson, P.S. Ed, Marcel Dekker: Inc. p 141-144.

[85] Jain A N, Harris N L and Park J Y 1995 Quantitative binding site model generation: compass applied to multiple chemotypes targeting the 5-HT1A receptor. *J. Med. Chem.** 38*, 1295-1308.

[86] Hahn M 1995 Receptor surface Models.1. definition and construction. *J. Med. Chem.** 38* 2080-90.

[87] Fukunaga J Y, Hansch C and Stellar E E 1976 Inhibition of dihydrofolate reductase. structure – activity correlation of quinazolines. *J. Med. Chem.** 19* 605-11.

[88] Davis A M, Gensmantel N P, Jahnsson E and Marriott D P 1994 The use of GRID Program in the 3D QSAR analysis of a series of calcium channel agonists. *J. Med. Chem.** 37* 963-72.

[89] Hoplinger A J and Tsokars J S 1997 Three Dimensional QSAR analysis. In *Practical Application of Computer Aided Drug Design*; Charifson, P.S. Ed, Marcel Dekker: Inc. p 144-145.

[90] Bures M G 1997 Recent Techniques and application in Pharmacophore mapping. In *Practical Application of computer aided Drug design*; Charifson, P.S. Ed, Marcel Dekker: Inc., 1997, p 39-72.

[91] Holloway M K, Wiz J M, Halgreen T A, Fitzgerald P M D, Desolms SJ, Gaffin N, Ghosh A K, Gillani E A, Graham S L, Grare J P, Hungate R W, Lyle T A, Sanders WM, Tucker TJ, Wiggins M, Wisonow CM, Woltersdolt OW, Young SC, Darke PL and Zagay, J.A. A Prior Prediction of Activity for HIV-1 Protease Inhibitors Employing Energy Minimization In the Active site.*J. Med. Chem.** 38* 305-317.

[92] Ortiz A R, Pisabarro M T, Gago F and Wade R C 1995 Predictions of drug binding affinities by comparative binding energy analysis. *J. Med. Chem.** 38* 2681-91.

[93] Lazono J J, Pastor M, Cruciani G, Gaedt K, Centeno N B, Gago F and Sanz F 2000 3D-QSAR Methods on the Basis of Ligand-Receptor Complexes. Application of COMBINE and GRID/GOPE Methodology to a Series of CYP1A2 Ligands. *J. Computer Aided Mol. Des.*14 341-353.

[94] Murcko M A 1997 An Introduction to De Novo Ligand Design. In *Practical application of computer aided drug design*; Charifson, P.S. Ed, Marcel Dekker: Inc. 305-354.

[95] Doweyka A M 1989 New Tool for the study of structure Activity Relationship in Three Dimensions. Probing Bioactive. *ACS Symp. Ser.* 413 82.

[96] Doweyka A M 1988 The hypothetical Active site Lattice. An approach to Modeling Active sites from data on inhibitor Molecules. *J. Med. Chem.** 31* 1396-1406.

[97] Walters D E and Hinds R M 1994 Genetically evolved receptor models: a computational approach to construction of receptor models. *J. Med. Chem.** 37* 2527-36.

[98] Jones P, Willet P and Glen R C 1995 A genetic algorithm for flexible molecular overlay and pharmacophore elucidation. *J. Comput. Aided Mol. Des.* 9 532-49.

[99] Payne A W R and Glen R C 1993 Molecular recognition using a binary genetic search algorithm. *J. Mol. Graphics.** 11* 74-91.
[100] Vedani A and Dobler M 2000 Multidimensional QSAR in drug Research: predicting Binding affinities, toxicity and Pharmacokinetic Parameters. *Prog. In Drug. Res.* **55** 105-35.

[101] Vedani A and Dobler M 2002 5D QSAR: The key for Simulating induced Fit. *J. Med. Chem.* **45** 2139-49.

[102] Jain A 2000 3D QSAR *Pharmacophoric Mapping of Some ‘Alpha-Thiol Dipeptide’ as Angiotensin Converting Inhibitors M.Pharm Thesis Department of Pharmacy, S.G.S.I.T.S Indore, India* p 24-25.

[103] Martin Y C 1978 *Quantitative Drug Design, a Critical Introduction, Medicinal Research Series*, (New York:Marcel Dekker)p 425.

[104] Wold S, Ruhe A, Wold H and Dunn W J 1984 The collinearity problem in linear regression. the partial least squares (pls) approach to generalized inverses. *Siam. J. Sci. Statist. Comput.* **5** 735-43.

[105] Daniel C and Wood F S 1980 Fitting Equations to Data, Computer analysis of multifactor data.(New York:VCH) **18** 458.

[106] Wold S 1991 Validation of QSAR. *Quant. Struct. Act. Relat.* 1991 10191-3.

[107] Kulkarni V M 1997 Design of new chemical entities as therapeutic agent. *Ind. J. Pharma. Sci.*, **59** 205-14.