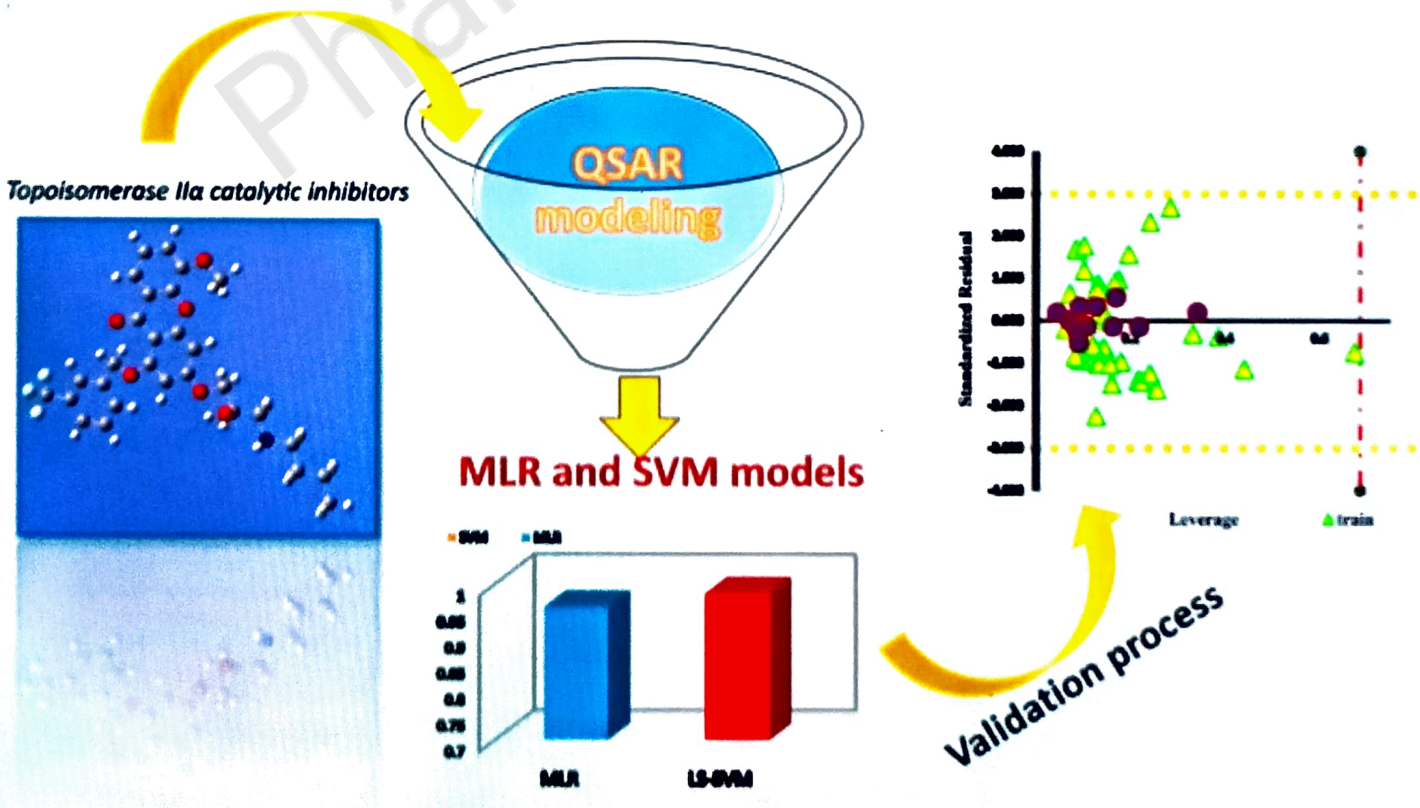
UNIT-II

Quantitative structure activity relationship (QSAR)

Points to be covered in this topic

- ❑ SAR Vs. QSAR
- ❑ HISTORY AND DEVELOPMENT
- ❑ DRUG DESIGN IN QSAR
- ❑ PHYSICOCHEMICAL PARAMETERS
- ❑ 3D-QSAR APPROACHES



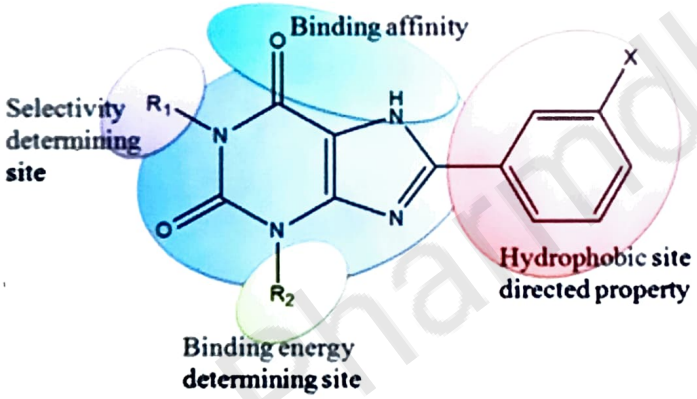
❑ SAR Vs. QSAR

❖ SAR (Structure activity relationship)

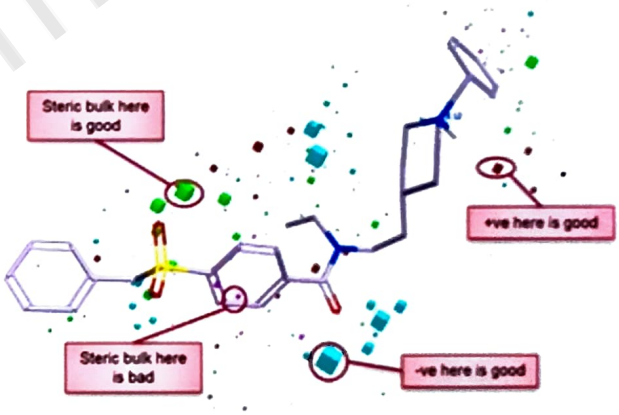
- The analysis of the **dependence of biological effects** of a chemical upon its molecular structure.
- **Molecular structure and biological** activity are correlated by observing the results of systematic structural modification on defined **biological endpoints.**

❖ QSAR (Quantitative structure activity relationship)

- It is defined to be a mathematical relationship in the form of an equation **between the biological activity and measurable physiochemical parameters.**
- QSAR attempts to **identify and quantify** the physicochemical properties of a drug and to see whether any of these property has an effect on the drugs biological activity.



SAR



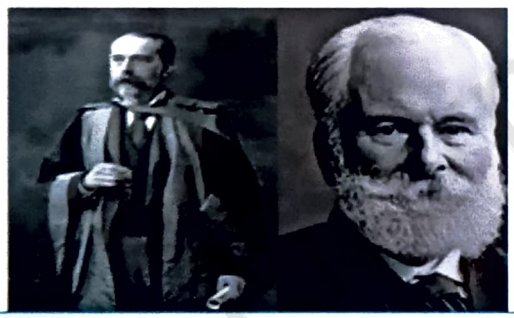
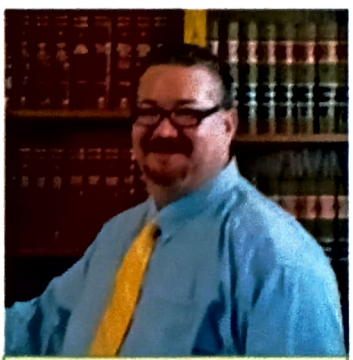
QSAR

❖ Differences between SAR and QSAR

SAR	QSAR
SAR **relationship between the chemical or 3D structure** of a molecular and its biological activities	QSAR is a mathematical **relationship between the biological activities and measurable physiochemical parameter**
It is used to develop a new drug that **has increased activity**	Its mainly help in **drug designing purpose**
SAR is **mainly done by lead molecule**	QSAR **optimize the properties of a lead compound**

SAR is not special case of qsar	QSAR is a special case of sar
In SAR we don't consider the amount of the change of the potency	QSAR dealing with all the potency and change potency or efficacy

❑ HISTORY AND DEVELOPMENT

SCIENTIST NAME AND YEAR	DISCOVERY
 D. Mendeleev 1868	• Russian chemist (D. Mendeleev) who arranged the 63 known elements into a periodic table based on atomic mass, which he published in Principles of Chemistry in 1869. • Mendeleev left space for new elements, and predicted three yet-to-be-discovered elements: Ga (1875), Sc (1879) and Ge (1886).
 A. Crum-Brown and T.R. Fraser 1868	• Formulated a suggestion that physiological activity of molecules depends on their constitution: Activity = F(structure) • They studied a series of quaternized strychnine derivatives, some of which possess activity similar to curare in paralyzing muscle.
 B.J. Richardson 1869	• Narcotic effect of primary alcohols varies in proportion to their molecular weights.

C. Richet 1893	• He shown that toxicities of some simple organic compounds (ethers, alcohols, ketones) were inversely related to their solubility in water.
H. Meyer 1899 and E. Overton 1901	• They have found variation of the potencies of narcotic compounds with LogP.
J. Traube 1904	• Found a linear relation between narcosis and surface tension.
L.P. Hammett 1937	• Studied chemical reactivity of substituted benzenes: Hammett equation, Linear Free Energy Relationship (LFER)
J. Fergusson 1939	• Formulated a concept linking narcotic activity, LogP and thermodynamics.
R.W. Taft 1952-1956	• Devised a procedure for separating polar, steric and resonance effects.
C. Hansch and T. Fujita 1964	• The biologist's Hammett equation.
Free and Wilson 1964	• QSAR on fragments
1970s-1980s	• Development of 2D QSAR (descriptors, mathematical formalism).
1980s - 1990s	• Development of 3D QSAR (pharmacophores, CoMFA, docking).
1990s	• Present, virtual screening.

❑ DRUG DESIGN IN QSAR

❖ Steps involved in drug designing through QSAR modeling:-

➢ **Step 1: Selection of biologically active series (Structure + Biological activity):-**

- **A series of already synthesized compounds is selected with their particular biological activity.**

➢ **Step 2: Calculation of Various Physico Chemical Descriptors:-**

- **The Molecular Descriptor is the final result of a logic and mathematical procedure which transforms chemical information encoded within a symbolic representation of a molecule into a useful number or the result of some standardized experiment.**

➢ **Step 3: Correlation of Physico-chemical properties with Biological activity by QSAR methods:-**

- **Widely used are: Partial Least Square(PLS), Multiple Linear Regression (MLR), Artificial Neural Networks (ANNs), k-Nearest Neighbor (kNN) and Support Vector Machine (SVM) methods.**

➢ **Step 4: Getting equation.**

➢ **Step 5: Designing of the compounds based upon QSAR equation.**

➢ **Step 6: Predicting the biological activity of the designed compounds.**

➢ **Step 7: Toxicity Prediction.**

➢ **Step 8: Synthesis of the compounds.**

❑ PHYSICOCHEMICAL PARAMETERS

- **Many physical, structural and chemical properties have been studied by the QSAR approach, but the most common are Hydrophobic, Electronic, and Steric properties, this is because it is possible to quantify these effects.**

❖ Hydrophobicity:-

- **Hydrophobic character of a drug is crucial to how easily it crosses the cell membrane and may also important in receptor interactions.**

- Hydrophobicity of a drug is measured experimentally by testing the drugs relative distribution is known as partition coefficient

Partition coefficient:

- Partition coefficient P usually expressed as logP.
- It is defined as

$$P = \frac{(x)\text{octanol}}{(x)\text{aqueous}}$$

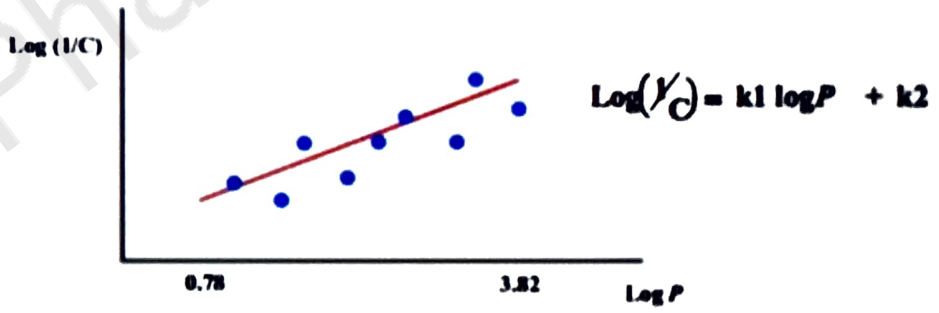
- P is a measure of the relative affinity of a molecule for the lipid and aqueous phase in the absence of ionization.
- 1-Octanol is a most frequently used lipid phase in pharmaceutical research.
- LogP for a molecule can be calculated from a sum of fragmental or atom based terms plus various corrections.

$$\text{LogP} = \text{fragments} + \Sigma \text{ corrections}$$

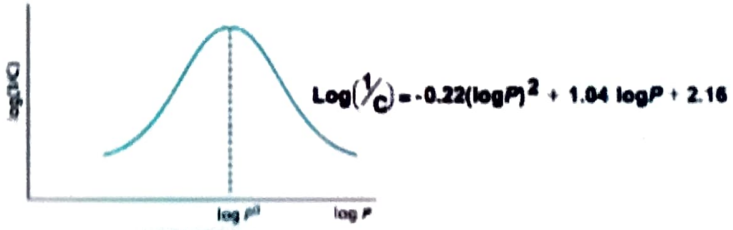
Relationship between LogP and Log1/C

- Activity of drugs is often related to *P*

e.g. binding of drugs to serum albumin
(straight line - limited range of log *P*)



- Binding increases as log *P* increases
- Binding is greater for hydrophobic drugs



➢ Molar Refractivity

- **Molar refractivity is a term which determines the size and polarizability of molecules as it is closely related with molar properties and refractive index of particular substance which is being tested.**
- **Molar refractivity is a measure of molar volume corrected by refractive index independent of physical factors and is useful in differentiating the structurally different compounds.**
- **(MR) is obtained from the following equation:**

$$MR = \frac{(n^2 - 1)}{(n^2 + 2)} \times \frac{MW}{d}$$

MW = Molecular weight, d = Density, n = Refractive index

❖ Electronic effect

- **The electronic effect of various substituent will clearly have an effect on drug ionization and polarity.**
- **Have an effect on how easily drug can pass through the cell membrane or how strongly it can interact with a binding site.**

➢ The Hammett constant (σ)

- **This value is a measure of the electron- withdrawing or electron-donating ability of a substituent and has been determined by measuring the dissociation of a series of substituted benzoic acid comparing to the dissociation of benzoic acid itself.**
- **Equilibrium shifts Right & $K_X > K_{benzoic}$**
- **Benzoic acid is a weak acid and only partially ionizes in water.**
- **An equilibrium is set up between the ionized and non-ionized forms where the relative proportions of these species are known as the equilibrium or dissociation constant K_H**
- **Since $S_1 = \log K_1 - \log K_{benzoic}$, then S will be positive .**
- **Hammett constant takes into account both resonance and inductive effects; thus, the value depends on whether the substituent is para or**

meta substituted.

- -ortho not measured due to steric effects.

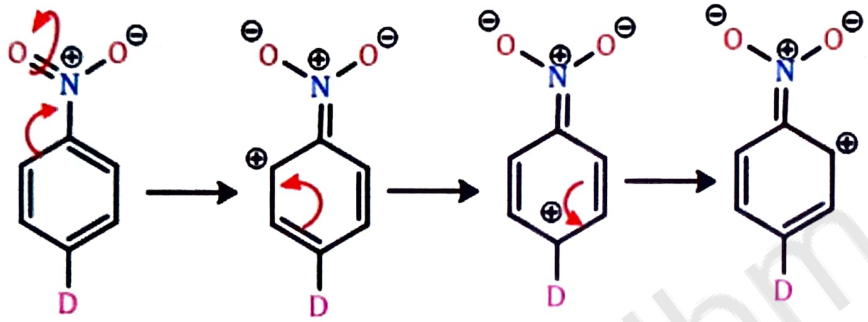
✓ **Example:-**

$\sigma_P(NO_2) = 0.78$ $\sigma_m(NO_2) = 0.71$

- ***meta*-Substitution**

e-withdrawing (inductive effect only)

- ***para*-Substitution**



e-withdrawing (inductive+ resonance effects)

❖ Steric effect

- **The bulk, size and shape of a drug will influence how easily it can approach and interact with a binding site.**
- **A bulky substituent may act like a shield and hinder the ideal interaction between a drug and its binding site.**
- **Alternatively, a bulky substituent may help to orientate a drug properly for maximum binding and increase activity.**

➢ Taft's steric factor (ES)

- **The value for (E_S) can be obtained by comparing the rates of hydrolysis of substituted aliphatic ester against a standard ester under acidic conditions.**

$$E_S = \log K_X - \log K_O$$

- K_X - **Rate of hydrolysis of an aliphatic ester having substituent X.**
- K_O - **Rate of hydrolysis of the reference ester.**

➢ Disadvantages

- **They are a measure of an intermolecular steric effect, whereas drugs interact with target binding sites in an intermolecular manner.**

➢ Molar refractivity [MR]

- **This is a measure of the volume occupied by an atom or a group of atoms. The MR is obtained by following equation:**

$$MR = \frac{(n^2 - 1)}{(n^2 + 2)} \times \frac{MW}{d}$$

n = index of refraction; MW = molecular weight; d = density

$\frac{MW}{d}$ = volume

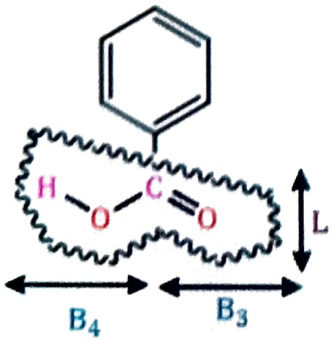
$\frac{(n^2 - 1)}{(n^2 + 2)}$ = term provides a correction factor by defining how easily the substituent can be polarized.

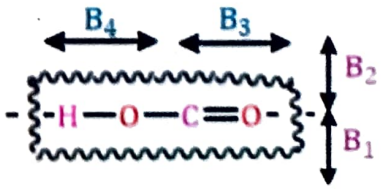
➢ Verloop Steric Parameter:

- **Calculated by software STERIMOL.**
- **Which calculate steric substituent values from standard bond angles, Van der Waals radii, bond lengths, and possible conformations for the substituent.**
- **The Verloop steric parameters can be measured for any substituent.**

✓ **Example**

- **The Verloop steric parameters for a carboxylic acid group are demonstrated in fig.**





L = Length of substituent, B_1 - B_4 are radii of the group in different dimensions

➢ Hansch equation

- The biological activity of most drugs is related to a combination of physicochemical properties.
- These equations are known as Hansch equations and they relate the biological activity to the most commonly used physicochemical properties (log P or p, s, and a steric factor).
- If the range of hydrophobicity values is limited to a small range, then the equation will be linear.

$$Log(1/C) = K_1 logP + K_2\sigma + K_3E_S + K_4$$

- If the log P values are spread over a large range, then the equation will be parabolic.

$$Log(1/C) = -K_1 (logP)^2 + K_2 logP + K_3 \sigma + K_4E_S + K_5$$

- The constants K_1-K_2 are determined by computer software in order to get the best fitting.
- Accuracy depends on using enough analogs, accuracy of data, & choice of parameters.

➢ Applications:

- Used to predict the activity of an as yet unsynthesized analogue.

➢ Free-wilson approach

- The biological activity of the parent structure is measured and compared with the activity of analogues bearing different substituents.
- An equation is derived relating biological activity to the presence or absence of particular substituents.
- Activity = $K_1X_1 + K_2X_2 + K_3X_3 + \text{----------} + K_nX_n + Z$
- X_n is defined as an indicator variable.
- It is given the value 1 or 0, depending on whether the substituent (n) is present or not.
- The contribution that each substituent makes to the activity is determined by the value of K_n.
- Z is constant representing the overall average activity of the structure studies.

➢ Advantages

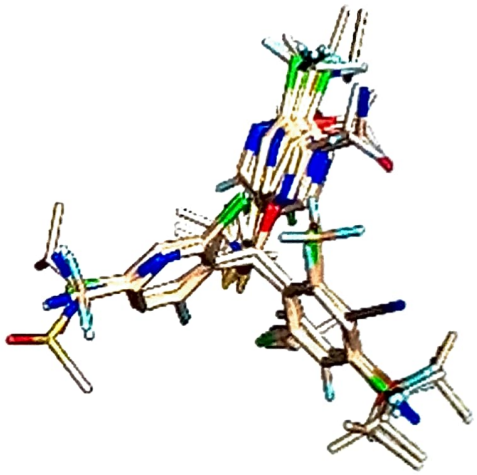
- Since the approach considers the overall effect of a substituent to biological activity rather than its various physicochemical properties, there is no need for physicochemical constants and tables.
- This method only requires experimental measurements of biological activity.

➢ Disadvantages

- The large number of analogues have to be synthesized and tested to make equation meaningful. Example:- Each of the term K_nX_n refers to a specific substituent at a specific position in the parent structure.
- Difficulty in rationalizing the results and explaining why a substituent at a particular position is good or bad for activity.
- Finally, the effects of different substituent may not be additive.
- There may be intermolecular interactions which affect activity.

➢ Types of QSAR

a) 1D-QSAR
b) 2D-QSAR
c) 3D-QSAR
d) 4D-QSAR
e) 5D-QSAR
f) 6D-QSAR



❖ 3D-QSAR:-

- A method known as 3D-QSAR has been developed in which 3D properties of molecule are considered as a whole rather than by considering individual substituents or moieties.

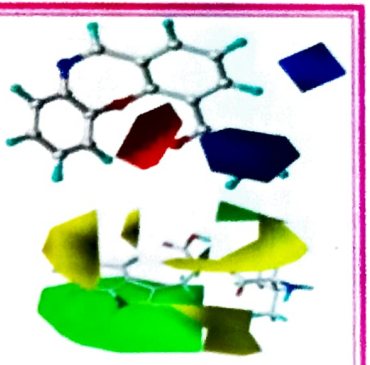
❖ CoMFA

- CoMFA (Comparative Molecular Field Analysis) is a 3D QSAR technique based on data from known active molecules.
- The aim of CoMFA is to derive a correlation between the biological activity of a set of molecules and their 3D shape, electrostatic and hydrogen bonding characteristics.

- A set of molecules is first selected which will be included in the analysis.

  As a most important precondition, all molecules have to interact with the same kind of receptor (or enzyme, ion channel, transporter) in the same manner, i.e., with identical binding sites in the same relative geometry.
- A certain subgroup of molecules is selected which constitutes a training set to derive the CoMFA model.
- Atomic partial charges are calculated and (several) low energy conformations are generated.
- A pharmacophore hypothesis is derived to orient the superposition of all individual molecules and to afford a rational and consistent alignment.
- A sufficiently large box is positioned around the molecules and a grid distance is defined.
- Different atomic probes, e.g., a carbon atom, a positively or negatively charged atom, a hydrogen bond donor or acceptor, or a lipophilic probe, are used to calculate field values in each grid point, i.e., the energy values which the probe would experience in the corresponding position of the regular 3D lattice.
- PLS analysis is the most appropriate method for this purpose.
- The result of the analysis corresponds to a regression equation with thousands of coefficients.

  Most often it is presented as a set of contour maps. These contour maps show favourable and unfavourable steric regions around the molecules as well as favourable and unfavourable regions for electropositive or electronegative substituents in certain position.
- Predictions for the test set (the compounds not included in the analysis) and for other compounds can be made, either by a qualitative inspection of these contour maps or, in a quantitative manner, by calculating the fields of these molecules and by inserting the grid values into the PLS model.



- The color coding's indicate regions where electronegative substituents enhance (blue) or reduce (red) the binding affinity.
- Regions where substitution enhances (green) or reduces (yellow) the binding affinity

➢ Application:-

- Predict the properties and activities of untested molecules
- Compare different QSAR models statistically and visually
- Optimize the properties of a lead compound
- Validate models of receptor binding sites
- Generate hypotheses about the characteristics of a receptor binding site
- Prioritize compounds for synthesis or screening
- There are now a few hundred practical applications of CoMFA in drug design.
- Most applications are in the field of
  a) Ligand protein interactions
  b) Describing affinity or inhibition constants
  c) Correlate steric and electronic parameters

➢ Advantages of COMFA

- Very generally applicable
- Robust, widely used and accepted
- Models easy to understand, interpret
- Excellent record for predicting potency

➢ Disadvantages of COMFA

- Input: "alignment" of 3D models is ill-defined
- Output: does not select, only predicts

❖ CoMSIA

- CoMSIA was developed to overcome certain limitations of CoMFA.
- CoMSIA is less affected by changes in molecular alignment and provides smoother and interpretable contour maps as a result of

employing Gaussian type distance dependence with the molecular similarity indices.

- Also, in addition to the steric and electrostatic fields, CoMSIA defines hydrophobic, HBD and HBA descriptor fields.
- CoMSIA study helps to obtain a more statistically robust model.
- The maps obtained by the new CoMSIA approach are superior and easier to interpret.
- The CoMSIA maps highlight those regions within the area occupied by the ligand skeletons that require a particular physicochemical property important for activity.