QSAR - Hansch Analysis and Related Approaches in Drug Design

Bharat Jhanwar[b], Vandana Sharma[a], Rajeev K Singla[a], Birendra Shrivastava[a`]*

[a]Department of Pharmaceutical Chemistry, School of pharmaceutical sciences , Jaipur national university, Jaipur-302025, Rajasthan, India

[b]Lachoo Memorial college of science and technology, Jodhpur, Rajasthan, India

Summary

QSAR (Quantitative Structure Activity Relationships) have been applied for decades in the development of relationships between physicochemical properties of chemical substances and their biological activities to obtain a reliable mathematical and statistical model for prediction of the activities of new chemical entities. (QSAR) have helped the scientists in the development of mathematical relationships linking chemical structures and pharmacological activity in quantitative manner of series of compound. The fundamental principle underlying the QSAR is that the difference in structural properties is responsible for the variations in biological activities of the compounds. In the classical QSAR studies, affinities of ligands to their binding sites, inhibition constants, rate constants, and other biological end points, with atomic, group or molecular properties such as lipophilicity, polarizability, electronic and steric properties (Hansch analysis) or with certain structural features (Free-Wilson analysis) have been correlated. QSAR certainly decreases the number of compounds to be synthesized by facilitating the selection of the most promising candidates. This review seeks to provide a view of the different QSAR approaches employed within the current drug discovery process to construct predictive structure–activity relationships and also discusses the limitations that are fundamental to these approaches, as well as those that might be overcome with the improved strategies.

Key words: Quantitative Structure Activity relationship, Hansch analysis, QSAR, Drug Design

Introduction

QSARs (Quantitative Structure–Activity relationships) are based on the assumption that the structure of a molecule (i.e. its geometric, steric and electronic properties) must contain the features responsible for its physical, chemical, and biological properties, and on the ability to represent the chemical by one, or more, numerical descriptor(s). By QSAR models, the biological activity (or property, reactivity, etc.) of a new or untested chemical can be inferred from the molecular structure of similar compounds whose activities (properties, reactivities, etc.) have already been assessed. The QSPR (Quantitative Structure–Property relationship) acronymous is used when a property is modeled. Simply it means that "The structure of chemical compound influences its properties and bioactivity. QSAR in simplest terms, is a method for building computational or mathematical models which attempts to find a statistically significant correlation between structure and function using a chemometric technique. In terms of drug design, structure here refers to the properties or descriptors of the molecules, their substituents or interaction energy fields, function corresponds to an experimental biological/biochemical

endpoint like binding affinity, activity, toxicity or rate constants. Various QSAR approaches have been developed gradually over a time span of more than a hundred years and served as a valuable predictive tool, particularly in the design of pharmaceuticals and agrochemicals. All one and two dimensional and related methods are commonly referred to as 'classical' QSAR methodologies. It is sometime used in more sense as a hansch analysis. The introduction of the Hansch model in 1964 enabled medicinal chemists to formulate their hypothesis of structure activity relationships in quantitative terms and to check these hypotheses by means of statistical methods. From such QSAR, it is possible to elucidate the influence of various physiological properties on drug potency and to predict activity values for new compounds within certain limits. QSAR is essentially a computerised statistical method which tries to explain the observed variance in the biological effect of certain classes of compounds as a function of molecular changes caused by the substituents. It assumes that the potency of a certain biological activity exerted by a series of congeneric compounds is a function of various physicochemical parameters of the compounds. Once statistical analysis shows that certain physico-chemical properties are favourable to the concerned activity, the concerned activity can be optimized by choosing such substituents which would enhance such physicochemical properties. Description of the molecular structure, electronic orbital reactivity and the role of structural and steric components have been the subject of mathematical and statistical analysis. The ultimate object of such studies is to understand the forces governing the activity of a particular compound or a class of compounds. QSAR vary to an appreciable extent in depth and sophistication based on the nature of evaluation of structure or activity. A purposeful relation of structural variables must include steric factors, electronic features of component functional groups and the molecule as a whole.

History of Q.S.A.R

It has been nearly 40 years since the quantitative structure-activity relationship (QSAR) paradigm first found its way into the practice of agrochemistry, pharmaceutical chemistry, toxicology, and eventually most facets of chemistry[1]. Crum-Brown and Fraser[2] (1868) expressed the idea that the physiological action of a substance in a certain biological system(Φ) was a function (f) of its chemical composition and constitution (C).

$$\Phi = f\,C \qquad \text{Equation [1]}$$

Thus, an alteration in chemical constitution, ΔC, would be reflected by an alteration in biological activity ΔΦ. Richardson[3] (1868) expressed the chemical structure as a function of solubility. Mills[4] (1884) developed a QSPR model for the prediction of melting and boiling points in homologous series, results were accurate to better than one degree. Richet[5] (1893) Correlated toxicities of a set of alcohols, ethers and ketones with aqueous solubility and showed that their cytotoxicities are inversely related to their corresponding water solubilities. Overton and Meyer[6,7] (1897, 1899) correlated partition coefficients of a group of organic compounds with their anesthetic potencies and concluded that narcotic (depressant) activity is dependent on the lipophilicity of the molecules. The seminal work of Hammett[8,9] (1935, 1937) gave rise to the σ-ρ culture correlated the effect of the addition of a substituent on benzoic acid with the dissociation constant, postulated electronic sigma-rho constants and established the linear free-energy relationship (LFER) principle.

Hammett found that a linear relationship resulted when substitutions of different groups were made to aromatic compounds.

$$\log \frac{K}{K_0} = \rho \log \frac{K'}{K_0} = \rho \sigma$$

K_0 and $K_{0'}$ are equilibrium constants for unsubstituted compounds and K and K' are the equilibrium constants for substituted compounds. Hammett used benzoic acid as reference compound yielding the σ. To interpret this equation, if the linear relation defines $\rho > 1$, then the effect of the substitutions is greater than making the same substitutions on benzoic acid. The σ describes the properties of the substitution groups. If σ is positive, the group is electron withdrawing. If σ is negative, the group is electron donating. The magnitude of σ indicates the degree of these effects. In 1939, Ferguson[10] correlated depressant action with the relative saturation of volatile compounds in their vehicle and introduced a thermodynamic generalization to the toxicity. Bell and Roblin[11] (1942) Studied antibacterial activities of a series of sulfanilamides in terms of their ionizations. Albert[12] (1948) examined the effects of ionization/electron distribution and steric access on the potencies of a multitude of aminoacridines. Taft[13] (1952) Postulated a method for separating polar, steric, and resonance effects and introduced the first steric parameter, E_S. Hansch and Muir[14] (1962) Correlated the biological activities of plant growth regulators with Hammett constants and hydrophobicity.

Using the octanol/water system, a whole series of partition coefficients were measured, and thus a new hydrophobic scale was introduced. The parameter π, which is the relative hydrophobicity of a substituent, was defined in a manner analogous to the definition of sigma[15].

$$\Pi_X = \log P_X - \log P_H \qquad \text{Equation [2]}$$

P_X and P_H represent the partition coefficients of a derivative and the parent molecule, respectively.

The contributions of Hammett and Taft together laid the basis for the development of the QSAR paradigm by Hansch and Fujita[16] (1964), which combined the hydrophobic constants with Hammett's electronic constants to yield the linear Hansch equation and its many extended forms.

$$\text{Log } 1/C = a\,\sigma + b\,\pi + ck$$

There is a consensus among current predictive toxicologists that Corwin Hansch is the founder of modern QSAR. In the classic article[17] it was illustrated that, in general, biological activity for a group of 'congeneric' chemicals can be described by a comprehensive model:

$$\text{Log } 1/C_{50} = a\,\pi + b\,\varepsilon + c\,S + d \qquad \text{Equation [3]}$$

In which C, the toxicant concentration at which an endpoint is manifested (e.g. 50% mortality or effect), is related to a hydrophobicity term, p, (this is a substituent constant denoting the difference in hydrophobicity between a parent compound and a substituted analog, it has been replaced with the more general molecular term the log of the 1-octanol/water partition coefficient, log K_{ow}), an electronic term, 1, (originally the Hammett substituent constant, s) and a

steric term, S, (typically Taft's substituent constant, ES). Due to the curvilinear, or bilinear, relationship between log1/C50 and hydrophobicity normally found in single dose tests the quadratic π^2 term was later introduced to the model. Hansch[18] (1969) Developed the parabolic Hansch equation for dealing with extended hydrophobicity ranges.

$$\text{Log } 1/C = - a (\log P)^2 + b. \log P + c \sigma + k$$

Free and Wilson[19] (1964) formulated an additive model, where the activity is discretized as a simple sum of contributions from different substituents.

$$BA = \sum a_i x_i + u$$

BA is the biological activity, u is the average contribution of the parent molecule, and a_i is the contribution of each structural feature; x_i denotes the presence $X_i = 1$ or absence $Xi = 0$ of a particular structural fragment. Fujita and Ban[20] (1971) simplified the Free-Wilson equation estimating the activity for the non-substituted compound of the series and postulated Fujita-Ban equation that used the logarithm of activity, which brought the activity parameter in line with other free energy-related terms.

$$\text{Log } BA = G_i X_i + u$$

In this equation, u is defined as the calculated biological activity value of the unsubstituted parent compound of a particular series.G_i represents the biological activity contribution of the substituents, whereas X_i is ascribed with a value of one when the substituent is present or zero when it is absent. Kubinyi[21] (1976) Investigated the transport of drugs *via* aqueous and lipoidal compartment systems and further refined the parabolic equation of Hansch to develop a superior bilinear (non-linear) QSAR model.

$$\text{Log } 1/C = a. \log P - b.\log (\beta. P + 1) + k$$

Hansch and Gao[22] (1997) Developed comparative QSAR (C-QSAR), incorporated in the C-QSAR program. Heritage and Lowis[23,24] (1997) Developed Hologram QSAR (HQSAR), where the structures are converted into all possible fragments, which are assigned specific integers, and then hashed into a fingerprint to form the molecular hologram. The bin occupancies of these holograms are used as the QSAR descriptors, encoding the chemical and topological information of molecules. Cho and workers[25] (1998) Developed Inverse QSAR, which seeks to find values for the molecular descriptors that possess a desired activity/property value. In other words, it consists of finding the optimum sets of descriptor values best matching a target activity and then generating a focused library of candidate structures from the solution set of descriptor values. Labute[26] (1999) Developed Binary QSAR to handle binary activity measurements from high-throughput screening (*e.g.*, pass/fail or active/inactive), and molecular descriptor vectors as input. A probability distribution for actives and inactives is then determined based on Bayes' Theorem.

QSAR Theory

The overall goals of QSAR retain their original essence and remain focused on the predictive ability of the approach and its receptiveness to mechanistic interpretation. Rigorous analysis and fine-tuning of independent variables has led to an expansion in development of molecular and atom-based descriptors, as well as descriptors derived from quantum chemical calculations and spectroscopy[27]. It is now possible not only to develop a model for a system but also to compare models from a biological database and to draw analogies with models from a physical organic database[28]. This process is dubbed model mining and it provides a sophisticated approach to the study of chemical-biological interactions. All QSAR analyses are based on the assumption of linear additive contributions of the different structural properties or features of a compound to its biological activity, provided that there are no nonlinear dependences of transport or binding on certain physicochemical properties. This simple assumption is proven by some dedicated investigations, for example the scoring function of the de novo drug design program LUDI (eqn.1)[29,30,31] ; in addition, the results of many Free Wilson and Hansch analyses support this concept.

$$\Delta G_{binding} = \Delta G_0 + \Delta G_{hb} + \Delta G_{ionic} + \Delta G_{lipo} + \Delta G_{rot} \qquad (1)$$

Overall loss of translational and rotational entropy,

$$\Delta G_{binding} = + 5.4 KJ\ mol^{-1}$$

Ideal neutral hydrogen bond, $\Delta G_{hb} = -4.7\ KJ\ mol^{-1}$

Ideal ionic interaction, $\Delta G_{ionic} = -8.3\ KJ\ mol^{-1}$

Lipophilic contact, $\Delta G_{lipo} = -0.17\ J\ mol^{-1}\ Å^{-2}$

Entropy loss per rotatable bond of the ligand, $\Delta G_{rot} = +1.4\ KJ\ mol^{-1}$

Eqn.1 correlates the free energy of binding, $\Delta G_{binding,}$ with a constant term, ΔG_0, that describes the loss of overall translational and rotational degrees of freedom and ΔG_{hb}, ΔG_{ionic} and ΔG_{lipo}, which are structure-derived energy terms for neutral and charged hydrogen bond interactions and hydrophobic interactions between the ligand and the protein; ΔG_{rot} describes the loss of internal rotational degrees of freedom of the ligand. Eqn1 1 holds for a wide range of energy values: the $\Delta G_{binding}$ of 45 different ligand-protein complexes ranges from -9 to -76 KJ mol^{-1}, which corresponds to binding constants between 2.5×10^{-2} M and 4×10^{-14} M; its standard deviation of 7.9 KJ mol-1 corresponds to a mean error of about 1.4 log units in the prediction of ligand binding constants from the mathematical model[29,30,31]. Because of the extrathermodynamic relationship between free energies ΔG and equilibrium constants K (eqn.2) or rate constants k (k_{on} = association constant, k_{off} = dissociation constant of ligand-receptor complex formation), the logarithms of such values can be correlated with binding affinities.

$$\Delta G = -2.303\ RT \log K = -2.303\ RT \log k_{on} / k_{off} \qquad (2)$$

Logarithms of molar concentrations C that produce a certain biological effect can be correlated with molecular features or with physiological properties that are also free-energy-related

equilibrium constants; normally the logarithms of inverse concentrations, log1/C, are used to obtain larger values for the more active analogs.

Development of Receptor Theory

The idea that drugs interacted with specific receptors began with Langley, who studied the mutually antagonistic action of the alkaloids, pilocarpine and atropine. He realized that both these chemicals interacted with some receptive substance in the nerve endings of the gland cells[32]. Paul Ehrlich defined the receptor as the binding group of the protoplasmic molecule to which a foreign newly introduced group binds[33]. In 1905 Langley's studies on the effects of curare on muscular contraction led to the first delineation of critical characteristics of a receptor: recognition capacity for certain ligands and an amplification component that results in a pharmacological response[34]. Receptors are mostly integral proteins embedded in the phospholipid bilayer of cell membranes. Rigorous treatment with detergents is needed to dissociate the proteins from the membrane, which often results in loss of integrity and activity. Pure proteins such as enzymes also act as drug receptors. Their relative ease of isolation and amplification have made enzymes desirable targets in structure based ligand design and QSAR studies. Nucleic acids comprise an important category of drug receptors. Nucleic acid receptors (aptamers), which interact with a diverse number of small organic molecules, have been isolated by in vitro selection techniques and studied[35].

Recent binary complexes provide insight into the molecular recognition process in these biopolymers and also establish the importance of the architecture of tertiary motifs in nucleic acid folding[36]. Groove-binding ligands such as lexitropsins hold promise as potential drugs and are thus suitable subjects for focused QSAR studies[37].

It is now possible to isolate membrane bound receptors, although it is still a challenge to delineate their chemistry, given that separation from the membrane usually ensures loss of reactivity. Nevertheless, great advances have been made in this arena, and the three-dimensional structures of some membrane- bound proteins have recently been elucidated. To gain an appreciation for mechanisms of ligand-receptor interactions, it is necessary to consider the intermolecular forces at play. Considering the low concentration of drugs and receptors in the human body, the law of mass action cannot account for the ability of a minute amount of a drug to elicit a pronounced pharmacological effect. The driving force for such an interaction may be attributed to the low energy state of the drugreceptor complex: KD = [Drug] [*Receptor*] / [Drug-Receptor Complex]. Thus, the biological activity of a drug is determined by its affinity for the receptor, which is measured by its K_D, the dissociation constant at equilibrium. A smaller K_D implies a large concentration of the drug-receptor complex and thus a greater affinity of the drug for the receptor. The latter property is promoted and stabilized by mostly non-covalent interactions sometimes augmented by a few covalent bonds. The spontaneous formation of a bond between atoms results in a decrease in free energy; that is, ΔG is negative. The change in free energy ΔG is related to the equilibrium constant K_{eq}.

$$\Delta G^\circ = -RT \ln K_{eq}$$

Thus, small changes in ΔG^0 can have a profound effect on equilibrium constants.

Types of Intermolecular Forces

Bond that formed between drug-receptor interactions include covalent, ionic, hydrogen, dipole-dipole, vander Waals, and hydrophobic interactions. Most drug-receptor interactions constitute a combination of the most of these bonds which are reversible under physiological conditions. Covalent bonds are not as important in drug-receptor binding as non-covalent interactions. Alkylating agents in chemotherapy tend to react and form an immonium ion, which then alkylates proteins, preventing their normal participation in cell divisions. Baker's concept of active site directed irreversible inhibitors was well established by covalent formation of Baker's antifolate and dihydrofolate reductase[38].

Ionic (electrostatic) interactions are formed between ions of opposite charge with energies that are nominal and that tend to fall off with distance. They are ubiquitous and because they act across long distances, they play a prominent role in the actions of ionizable drugs. The strength of an electrostatic force is directly dependent on the charge of each ion and inversely dependent on the dielectric constant of the solvent and the distance between the charges. Electrostatic interactions are generally restricted to polar molecules.

Hydrogen bonds are ubiquitous in nature. Their multiple presence contributes to the stability of the (α-helix and base-pairing in DNA. Hydrogen bonding is based on an electrostatic interaction between the non-bonding electrons of a heteroatom (e.g., N, O, S) and the electron-deficient hydrogen atom of an -OH, SH, or NH group. Hydrogen bonds are strongly directional, highly dependent on the net degree of salvation, and rather weak, having energies ranging from 1 to 10 kcal/mol[39,40]**.** Bonds with this type of strength are of critical importance because they are stable enough to provide significant binding energy but weak enough to allow for quick dissociation. The energy of dipole-dipole interactions can be described by Equation,

$$E = 2\mu_1\mu_2\cos\theta_1\cos\theta_2/Dr^3$$

where μ is the dipole moment, θ is the angle between the two poles of the dipole, *D* is the dielectric constant of the medium and r is the distance between the charges involved in the dipole.

There are also strong interactions between non-polar molecules over small intermolecular distances. Dispersion or London/van der Waals forces are the universal attractive forces between atoms that hold non-polar molecules together in the liquid phase. They are based on polarizability and these fluctuating dipoles or shifts in electron clouds of the atoms tend to induce opposite dipoles in adjacent molecules, resulting in a net overall attraction. The energy of this interaction decreases very rapidly in proportion to $1/r^6$, where r is the distance separating the two molecules. These vander Waals forces operate at a distance of about 0.4-0.6 nm and exert an attraction force of less than 0.5 kcal/mol.

Hydrophobicity refers to the tendency of non-polar compounds to transfer from an aqueous phase to an organic phase[41,42]. When a non-polar molecule is placed in water, it gets solvated by a sweater of water molecules ordered in a ice-like manner. This increased order in the water molecules surrounding the solute results in a loss of entropy. Association of hydrocarbon molecules leads to a squeezing out of the structured water molecules. The displaced water becomes bulk water, less ordered, resulting in a gain in entropy, which provides the driving force referred to as a hydrophobic bond.

OBJECTIVES OF QSAR

Mostly all the QSAR methods focus on the following goals:

1. Quantitative relationship between the structure and physiochemical properties of substances and their biological activity are being used as the foundation stone in search of new medicines. The mathematical and statistical analysis helps us to predict the drug activity.

2. QSAR makes it easy now to reach the conclusion for any of the congener that still not in process, in way that whether it will optimal and profitable or not.

3. To quantitatively correlate and recapitulate the relationships between trends in chemical structure alterations and respective changes in biological endpoint for comprehending which chemical properties are most likely determinants for their biological activities.

4. To optimize the existing leads so as to improve their biological activities.

5. To predict the biological activities of untested and sometimes yet unavailable compounds.

TECHNIQUES AND TOOLS OF QSAR

1. Compound Selection: In setting up to run a QSAR analysis, compound selection is an important angle that needs to be addressed. One of the earliest manual methods was an approach devised by Craig, which involves two-dimensional plots of important physicochemical properties. Care is taken to select substituents from all four quadrants of the plot[43]. The Topliss operational scheme allows one to start with two compounds and construct a potency tree that grows branches as the substituent set is expanded in a stepwise fashion[44]. Topliss later proposed a batchwise scheme including certain substituents such as the 3,4-Cl_2, 4-Cl, 4-CH_3, 4-OCH_3, and 4-H analogs[45].

2. Biological Parameters: In QSAR analysis, it is vital important that the biological data be both accurate and precise to develop a meaningful model. The equilibrium constants and rate constants that are used extensively in physical organic chemistry and medicinal chemistry are related to free energy values ΔG. Thus for use in QSAR, standard biological equilibrium constants such as K_i or K_m should be used in QSAR studies. Percentage activities (e.g., % inhibition of growth at certain concentrations) are not appropriate biological endpoints because of the nonlinear characteristic of dose-response relationships. These types of endpoints may be transformed to equieffective molar doses. Only equilibrium and rate constants pass muster in terms of the free-energy relationships or influence on QSAR studies. Biological data are usually expressed on a logarithmic scale because of the linear relationship between response and log dose in the midregion of the log dose-response curve. Inverse logarithms for activity (log $1/C$) are used so that higher values are obtained for more effective analogs. Various types of biological data have been used in QSAR analysis.

Types of Biological Data Utilized in QSAR Analysis

Source of Activity	Biological Parameters

1. Isolated receptors

Rate constants	Log k
Michaelis-Menten Constants	Log $1/K_m$
Inhibition constants	Log $1/K_i$
Affinity data	pA_2; pA_1

2. Cellular systems

Inhibition constants	Log $1/IC_{50}$
Cross resistance	Log CR
In vitro biological data	Log 1/C
Mutagenicity states	Log TA_{98}

3. In vivo systems

Biocencentration factor	Log BCF
In vivo reaction rates	Log I (Induction)
Pharmacodynamic Rates	Log T (total clearance)

3. Statistical Methods: Linear Regression Analysis

The most widely used mathematical technique in QSAR analysis is multiple regression (MRA). Regression analysis is a powerful means for establishing a correlation between independent variables and a dependent variable such as biological activity[46].

$Y_i = b + a X_i + E_i$

Certain assumptions are made with regard to this procedure[47]:

1. The independent variables, which in this case usually include the physicochemical parameters, are measured without error. Unfortunately, this is not always the case, although the error in these variables is small compared to that in the dependent variable.

2. For any given value of *X*, the *Y* values are independent and follow a normal distribution. The error term E_i possesses a normal distribution with a mean of zero.

3. The expected mean value for the variable *Y*, for all values of X, lies on a straight line.

4. The variance around the regression line is constant. The best straight line for model $Y_i = b + a Z_i + E$ is drawn through the data points, such that the sum of the squares of the vertical distances from the points to the line is minimized. Y represents the value of the observed data point and Y_{calc} is the predicted value on the line.

The sum of squares $SS = \sum (Y_{obs} - Y_{calc})^2$.

$$Y_{obs} = a X_i + b + E_i \quad (1)$$

$$Y_{calc} = a X_i + b \quad (2)$$

$$E = Y_{obs} - a X_i - b \quad (3)$$

$$\sum_{i=1}^{n} E_i^2 = \sum \Delta^2 = SS$$

$$= \sum (Y_{obs} - Y_{calc})^2 \quad (4)$$

$$\text{Thus, } SS = \sum_{i=1}^{n} (Y_{obs} - a X_i - b)^2 \quad (5)$$

Expanding Equation (5), will give

$$SS = \sum_{i=1}^{n} (Y_{obs}^2 - Y_{obs} a X_i - Y_{obs} b - Y_{obs} a X_i + a^2 X_i^2 + a X_i b - b Y_{obs} + a b X_i + b^2) \quad (6)$$

i=1

Taking the partial derivative of Equation (4) with respect to b and then with respect to a, results in Equations (7) and (8).

$$\frac{dSS}{db} = \sum(\text{taking summation from i = 1 to n}) - 2(Y_{obs} - b - a X_i) \quad (7)$$

$$\frac{dSS}{db} = \sum (\text{Taking summation from i=1 to n}) - 2 Xi (Y_{obs} - b - aX_i) \quad (8)$$

SS can be minimized with respect to b and a and divided by -2 to yield the normal Equations (9) and (10).

$$\sum_{i=1}^{n} = (Y_{obs} - b - a X_i) = 0 \quad (9)$$

$$\sum_{i=1}^{n} = X_i (Y_{obs} - b - a X_i) = 0 \quad (10)$$

These normal equations can be rewritten as follows:

$$b \sum_{i=1}^{n} X_i + a \sum_{i=1}^{n} X_i^2 = \sum_{i=1}^{n} X_i Y_{obs} \quad (11)$$

$$b + a \sum_{i=1}^{n} X_i = \sum_{i=1}^{n} Y_{obs} \quad (12)$$

The solution of these simultaneous equations yields *a* and *b*.

PARAMETERS USED IN QSAR

1. Hydrophobicity Parameters: More than a hundred years ago, Meyer and Overton made their discovery on the correlation between oil/water partition coefficients and the narcotic potencies of small organic molecules[48,49].

1.1 Estimation of hydrophobicity:-

Hansch established a model to measure the lipophilicity in term of partition coefficient. Drug travels to the site of action that means solubility in 1- octanol that simulate the lipid membrane then it goes to via cytoplasm that is simulated by Aqueous buffer "water"

Hansch proposed the lipophilicity measurement in term of partition coefficient "P"

$P = [C]_{octanol} / [C]_{water}$

It is called "Distribution coefficient**".**

With help of the partition coefficient we can determined the hydrophobic (π) like the difference caused in the partition coefficient of substituted and unsubstantiated compounds is relevant to the attached new susbtituent in it

Formula is

$$\pi = \log P_x - \log P_H$$

Px denotes for substituted compound by "x"

P_H denotes unsubstituted "x = H"

Eg.- Consider the log P values for benzene(logP = 2.13), chlorobenzene (logP = 2.84), and benzamide (logP = 0.64) , Since benzene is the parent compound, the substituents constants for Cl and $CONH_2$ are 0.71 and -1.49 respectively. Having obtained these values, it is now possible to calculate the theortical log P value for media **Chlorobenzamide.**

Π values of aromatic substituent

Substituents	Π values		
	Π_{ORTHO}	Π_{META}	Π_{PARA}
—H	0.00	0.00	0.00
—Cl	0.76	0.77	0.73
—Br	0.84	0.96	1.19
—I	0.93	1.18	1.43
—NH_2	-1.40	-1.29	-1.30
—OH	-0.41	-0.50	-0.61
—CHO	-0.43	-0.47	—0.47
—NO_2	—	0.11	0.22
—CH_3	0.84	0.52	0.60

—OCH_3	-0.02	—	-0.04

1.2 Partition coefficient :- It is the ratio of concentrations of a compound in the two phases of a mixture of two immiscible solvents at equilibrium.

$$P = \frac{\text{Concentration of drug in organic phase}}{\text{Concentration of drug in aqueous phase}}$$

Hydrophobic compounds will have a high P value, whereas hydrophilic compounds will have a low P value. The hydrophobic character of a drug can be measured experimentally by testing the drug's relative distribution coefficient. Octanol is a suitable solvent for the measurement of partition coefficients for many reasons[50,51]. It is cheap, relatively nontoxic, and chemically unreactive. The hydroxyl group has both hydrogen bond acceptor and hydrogen bond donor features capable of interacting with a large variety of polar groups. Despite its hydrophobic attributes, it is able to dissolve many more organic compounds than can alkanes, cycloalkanes, or aromatic hydrocarbons. It is UV transparent over a large range and has a vapour pressure low enough to allow for reproducible measurements. Varying substituents on the lead compound will produce a series of analogues having different hydrophobicities and therefore different P values.Various substituents make to hydrophobicity. This contribution is known as the substituent hydrophobicity constant. These π value are characteristic for the substituents and can be used to calculate how the partition coefficient of a drug would be affected by adding these substituents. QSAR would allow us to predict the most promising and satisfying structures (closest to the optimum value log P^o).The substituent hydrophobicity constant is a measure of how hydrophobic a substituent is, relative to hydrogen. A positive value of π indicates that the substituent is more hydrophobic then hydrogen. A negative value indicates that the substituent is less hydrophobic.

By plotting these P value against the biological activity of these drugs. It is possible to see if there is any relationship between the two properties. The biological activity is normally expressed as 1/C so a graph is drawn by plotting log 1/C versus log P values to correlate the activity and partition coefficient or hydrophobicity. In studies where the range of the log P values are ranges between 1 to 4 and a straight-line graph is obtained i.e. there is an existence of relation between hydrophobicity and biological activity.
As per the equation is
$\log (1/C) = K_1 \log P + K_2$

Eg. Binding of drug to serum albumin. It can be determined by their hydrophobicity, in study of 40 compounds they resulted in the following equation:

$\log (1/C) = 0.75\log P + 2.30$

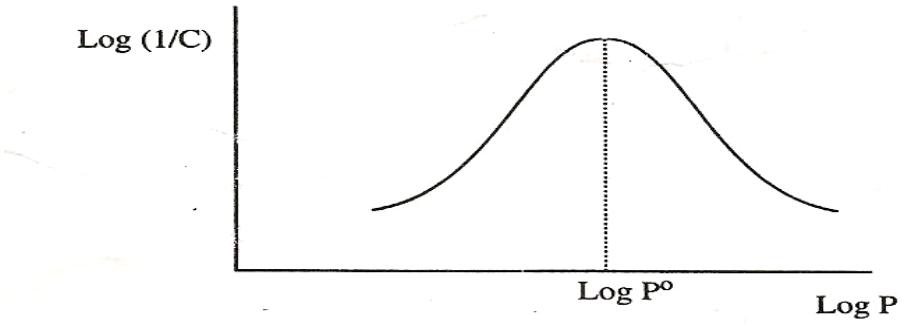
Conclusions

(i) Serum albumin binding increases as log P increases that mean hydrophobic drugs bind more strongly to serum albumin than hydrophilic drug.
(ii) It helps us to know how strongly a drug binds to serum albumin that can be important in estimating effective dose levels for that drug and drugs of similar structure and predict
Some time log P Values increases over the given ranges results in decreased activity

But drugs which are independent of cell target action like Eg.- General Anaesthetics. These are related to the log P factor alone to operate in cell membrane only no receptor interaction. These function by entering the central nervous system (CNS) and 'dissolving' into cell membranes where they affect membrane structure and nerve function.

Conclusion:-

(i) Anaesthetic activity increases with increasing hydrophobicity
(ii) they depend upon lipophilicity only
(iii) There is an optimum value for log P (log P^0), beyond which increasing hydrophobicity causes a decrease in anaesthetic activity.
Finally, hydrophobic drugs are often more susceptible to metabolism and subsequent elimination. Lipophilicity has a relationship with concentration and indirectly with biological activity .It can be concluded by this graph



[Transition of linear equational effect of "log P" into parabolic equational effect "log P^{2}"]

The value of log P at the maximum (log P^{0}) represents the optimum partition coefficient for biological activity. Beyond that point, an increase in log P results in a decrease in biological activity.

(A)SHAKE-FLASK METHOD

The shake-flask method, so-called, is most commonly used to measure partition coeffcients with great accuracy and precision and with a log *P* range that extends from -3 to +6[52, 53]. The procedure calls for the use of pure, distilled, deionized water, high-purity octanol, and pure solutes. At least three concentration levels of solute should be analyzed and the volumes of octanol and water should be varied according to a rough estimate of the log *P* value. The classical and most reliable method of log *P* determination is the, which consists of dissolving some of the solute in a volume of octanol and water, then measuring the concentration of the solute in each solvent.

Advantages:

1. Most accurate method

2. Accurate for broadest range of solutes (neutral and charged compounds applicable)

3. Chemical structure does not have to be known beforehand.

Disadvantages:

1. Time consuming (>30 minutes per sample)

2. Octanol and water must be premixed and equilibrated (takes at least 24 hours to equilibrate)

3. Complete solubility must be attained, and it can be difficult to detect small amounts of undissolved material.

4. The concentration vs. UV-VIS response must be linear over the solute's concentration range.

5. If the compound is extremely lipophilic or hydrophilic, the concentration in one of the phases will be exceedingly small, and thus difficult to quantify.

6. Relative to chromatographic methods, large amounts of material are required.

(B) By Chromatography

Chromatography provides an alternate tool for the estimation of hydrophobicity parameters. *Rm* values derived from thin-layer chromatography provide a simple, rapid, and easy way to ascertain approximate values of hydrophobicity[54, 55].

$R_m = \log (1/ R_f – 1)$

Recently, a rapid method for the determination of partition coefficients using gradient reversed phase/high pressure liquid chromatography (RP-HPLC) was developed. A faster method of log P determination makes use of high-performance liquid chromatography. The log P of a solute can be determined by correlating its retention time with similar compounds with known log P values. This method is touted as a high-throughput hydrophobicity screen for combinatorial libraries[56, 57]. A chromatography hydrophobicity index (CHI) was established for a diverse set of compounds. Acetonitrile was used as the modifier and 50mm ammonium acetate as the mobile phase[56]. A linear relationship was established between Clog *P* and CHIN for neutral molecules.

Clog *P* = 0.057 CHIN - 1.107
$n = 52$, $r^2 = 0.724$, $s = 0.82$, $F = 131$

Advantage:

Fast method of determination (5-20 minutes per sample)

Disadvantages:

1. The solute's chemical structure must be known.

2. Since the value of log *P* is determined by linear regression, several compounds with similar structures must have known log *P* values.

3. Different chemical classes will have different correlation coefficients, between-class comparisons are not significant.

(C) Fragmentation Methods and interaction factors (Computational)

Initially, the p-system was applied only to substitution on aromatic rings and when the hydrogen being replaced was of innocuous character. It was apparent from the beginning that not all hydrogens on aromatic systems could be substituted without correction factors because of strong electronic interactions. It became necessary to determine π values in various electron-rich and –deficient systems (e.g., X-phenols and X-nitrobenzenes). Correction factors were introduced for special features such as unsaturation, branching, and ring fusion. The proliferation of p-scales made it difficult to ascertain which system was more appropriate for usage, particularly with complex structures. The shortcomings of this approach provided the impetus for Nys and Rekker to design the fragmental method, a reductionist approach, which was based on the statistical analysis of a large number of measured partition coefficients and the subsequent assignment of appropriate values for particular molecular fragments[58, 59]. Hansch and Leo took a constructionist approach and developed a fragmental system that included correction factors for bonds and proximity effects[60]. Labor-intensive efforts and inconsistency in manual calculations were eliminated with the debut of the automated system CLOGP and its powerful SMILES notation [61-62].

Recent attempts to compute log *P* calculations have resulted in the development of solvatochromic parameters[63,64]. This approach was proposed by Kamlet et al. and focused on molecular properties. In its simplest form it can be expressed as follows:

$\text{Log } P_{oct} = a\,V + b\,\pi^* + c\,\beta_H + d\,\alpha_H + e$

V is a solute volume term; π* represents the solute polarizability; β_H and α_H are measures of hydrogen bond acceptor strength and hydrogen bond donor strength, respectively; and *e* is the intercept.

2. Electronic factors :- The extent to which a given reaction responds to electronic perturbation constitutes a measure of the electronic demands of that reaction, which is determined by its mechanism. The introduction of substituent groups into the framework and the subsequent alteration of reaction rates helps delineate the overall mechanism of reaction. Early work examining the electronic role of substituents on rate constants was first tackled by Burckhardt and firmly established by Hammett[8,65,66,67]. Hammett employed, as a model reaction, them ionization in water of substituted benzoic acids and determined their equilibrium constants K_a.

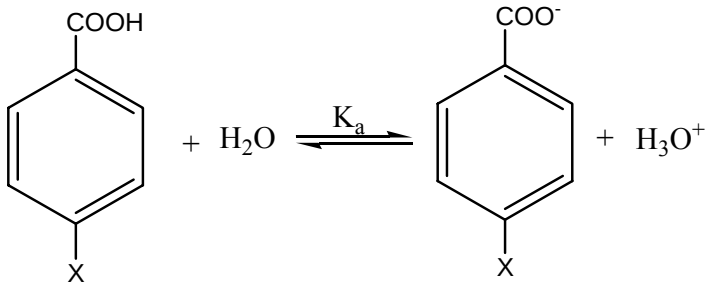


Figure-Ionization of Benzoic acid

This led to an operational definition of σ, the Hammett substitution constant. It is a measure of the size of the electronic effect i.e. whether a electron withdrawing or electron donating for a given substituent and represents a measure of electronic charge distribution in the benzene nucleus.

$\sigma_X = \log K_X - \log K_H$ or $\log(K_X / K_H) = - pK_X + pK_H$

If the substituent X is an electron donating group (I+)

Then the aromatic ring is less able to stabilize the carboxylate ion.

The equilibrium shifts to the left and a weaker acid is obtained with a smaller K_X value.

Substituents such as Methyl, Ethyl and t-Butyl have negative σ values.

If the substituent X is an electron withdrawing group (I-)

Smaller K_X values than benzoic acid itself and hence the value of σ_X for an electron withdrawing substituent will be positive.

Then the aromatic ring is able to stabilize the carboxylate ion.

The electronic effects of various substituent will clearly have an effect on a drug's ionization or polarity. This in turn may have an effect on how easily a drug can pass through cell membranes or how strongly it can bind to a receptor. It is therefore useful to have some measure of the electronic effect a substituent can have on a molecule.

Hammett also drew attention to the fact that a plot of log K_A for benzoic acids versus log k for ester hydrolysis of a series of molecules is linear, which suggests that substituents exert a similar effect in dissimilar reactions.

$$\log \frac{k_X}{k_H} \alpha \log \frac{K_X}{K_H} = \rho . \sigma$$

K_H = Dissociation constant of unsubstituted Benzoic acid.

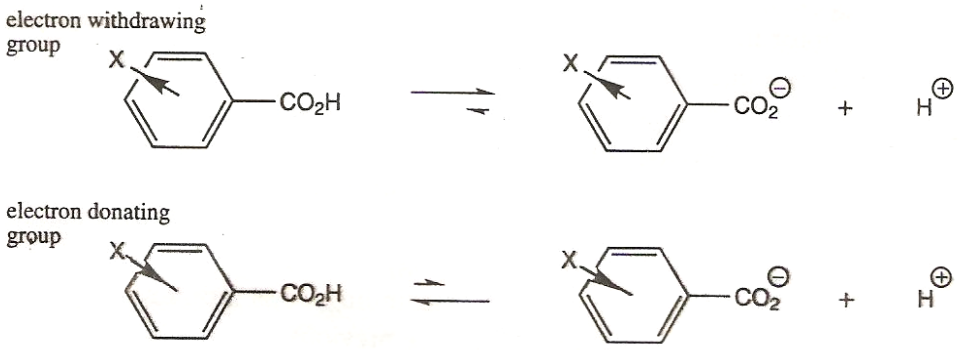
$\log (k_X / k_H)$ refers to comparing rates of reaction to the rate.

K_H = Ionization constant unsubstituted benzoic acid.

As far as substituents on an aromatic ring are concerned, the measure used is known as the Hammett substitution constant (σ). ρ (rho) is defined as a proportionality or reaction constant, which is a measure of the susceptibility of a reaction to substituent effects. A positive rho value suggests that a reaction is aided by electron withdrawal from the reaction site, whereas a negative rho value implies that the reaction is assisted by electron donation at the reaction site.

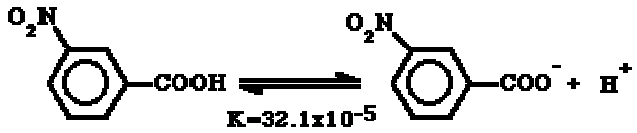
When a substituent is present on the aromatic ring. This equilibrium is affected. Electron withdrawing groups, such as a nitro group, result in the aromatic ring having a stronger electron withdrawing and stabilizing influence on the carboxylate anion. The equilibrium will therefore

shift more to the ionized form such that the substituted benzoic acid is a stronger acid and has larger K_X value (X represents the substituent on the aromatic ring)



Case--The Hammett substituent constant for H is zero.

Eg, the nitro substituent has σ_p = 0.78 and σ_m = 0.71. In the meta position, the electron withdrawing power is due to the inducting influence of the substituent, whereas at the para position inducting and resonance effects both play a part, and so the σ_p value is greater.



A nitro group in the meta position(σ_m)--K increases—because nitro group is electron-withdrawing, thereby stabilizing the negative charge that develops. Force working is the Inducting effect (I)

a nitro group in the para position (σ_p)

K--larger than for the nitro group in the meta position

indicating even greater electron-withdrawal because in case of para position inducting and resonance effects both play a part , whereas at the para position Inducting and Resonance effects(R) both take part .

The Hammett equation relates reaction rates and equilibrium constants that are described as

1. **Positive σ:-** If σ is positive, the substituent is electron attracting, that means it takes out the electrons from the ring . It favours the anion and shifts the equilibrium to right side. Ultimately results in increased K values.they couse the increment in values like $\log(K/K_0)$ is positive and K/K_0 will be greater than 1

Eg. Nitro, cyano ,chlorides carboxy etc.

2. **Negative σ:-** if σ is negative the substituent is electron donating, that means it gives out its own electron to the species. It favours the cationic one and neutral species. It increase the electron density on ring and shifts the equilibrium to left. Ultimately results in decresed K values like K/K_0 will be less than 1 and log $\log(K/K_0)$ will be negative.

Eg. Hydroxyl, alkyl groups, methoxy, ethoxy etc.

Substituents	(σ_m)	(σ_p)
—H (hydrogen)	0.0	0.0
—NO_2 (nitro)	0.71	0.78
—Cl (chloride)	0.37	0.23
—OCH_3 (methoxy)	0.12	0.27
—OH (hydroxyl)	0.12	0.37

These equation, $\log \frac{k_X}{k_H} \alpha \log \frac{K_X}{K_H} = \rho . \sigma$

has become known as the Hammett equation and has been applied to thousands of reactions that take place at or near the benzene ring bearing substituents at the meta and para positions. Because of proximity and steric effects, ortho-substituted molecules do not always follow this maxim and are subject to different parameterizations. Thus, an expanded approach was established by Charton[68] and Fujita and Nishioka[69]. Charton partitioned the ortho electronic effect into its inductive, resonance, and steric contributions; the factors α ,β , and *X* are susceptibility or reaction constants and *h* is the intercept.

$\text{Log } k = \alpha\, \sigma_1 + \beta\, \sigma_R + X\, r_v + h$

Fujita and Nishioka used an integrated approach to deal with ortho substituents in data sets including meta and para substituents.

$\text{Log } k = \rho\sigma + \delta\, E_S^{ortho} + f\, F_{otho} + C$

For ortho substituents, para sigma values were used in addition to Taft's E_S values and Swain-Lupton field constants F_{ortho}. The reason for employing alternative treatments to ortho-substituted aromatic molecules is that changes in rate or ionization constants mediated by meta or para substituents are mostly changes in ΔH^o because substitution does not affect ΔS^o. Ortho substituents affect both enthalpy and entropy; the effect on entropy is noteworthy because entropy is highly sensitive to changes in the size of reagents and substituents as well as degree of solvation.

3. Steric parameters

Quantifying steric properties is more difficult than quantifying hydrophobic or electronic properties. sterics are of overwhelming importance in ligand- receptor interactions as well as in transport phenomena in cellular systems. The first steric parameter to be quantified and used in QSAR studies was Taft's E_S constant[70]. E_S is defined as

$E_S = \log (k_X / k_H)_A$

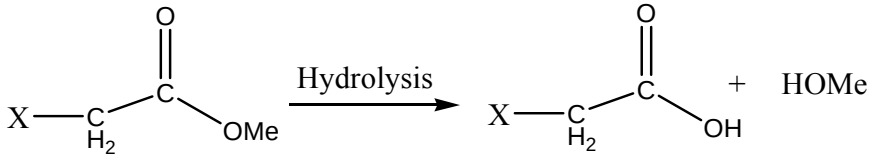
E_S is called Taft's steric factor.

Where k_X and k_H represent the rates of acid hydrolysis of esters, XCH_2COOR and CH_3COOR, respectively.

It is highly unlikely that a drug's biological activity will be affected by steric factors alone, but these factors are frequently to be found in Hansch equations.

Steric effect on the substituent can be describe by the following example

Eg. Ester hydrolysis



[Under acidic conditions where only steric factors are important.]

Bulky substituent may shield the ester from attack and lower the rate of hydrolysis. It is therefore necessary to separate out these two effects. This can be done by measuring hydrolysis rates under basic conditions and also under acidic conditions.

To correct for hyperconjugation in the a-hydrogens of the acetate moiety, Hancock devised a correction on E_S such that
$E_S^C = E_S + 0.306\ (n- 3)$
In this equation, n represents the number of α-hydrogens and 0.306 is a constant derived from molecular orbital calculations[71]. Unfortunately, the limited availability of E_S and E_S^C values for a great number of substituents precludes their usage in QSAR studies. Charton demonstrated a strong correlation between E_S and vander Waals radii, which led to his development of the upsilon parameter v_X[72].
$v_X = r_X - r_H = r_X - 1.20$
where r_X and r_H are the minimum vander Waals radii of the substituent and hydrogen, respectively.

One of the most widely used steric parameters is molar refraction (MR), which has been aptly described as a chameleon parameter by Tute[73]. However, the number of substituents which can be studied by this method is restricted. Although it is generally considered to be a crude measure of overall bulk, it does incorporate a polarizability component that may describe cohesion and is related to London dispersion forces as follows: $MR = 4\pi N\alpha/3$, where N is Avogadro's number and α is the polarizability of the molecule. It contains no information
on shape.
MR is also defined by the Lorentz-Lorenz equation:

The MR is obtained from the following equation:

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

MR is generally scaled by 0.1 and used in biological QSAR, where intermolecular effects are of primary importance. Where n is the refractive index, MW is the molecular weight, and d is the density. The term MW/d defines a volume, while the $(n^2 – 1) / (n^2 + 2)$ term provides a correction factor by determining how easily the substituent can be polarized. This is particularly significant if the substituent has π electrons or lone pairs of electrons.
The failure of the MR descriptor to adequately address three-dimensional shape issues led to Verloop's development of STERIMOL parameters[74], which define the steric constraints of a given substituent along several fixed axes. Five parameters were deemed necessary to define shape: *L*, B1, B2, B3, and B4. *L* represents the length of a substituent along the axis of a bond between the parent molecule and the substituent; B1 to B4 represent four different width parameters. However, the high degree of collinearity between B1, B2, and B3 and the large

number of training set members needed to establish the statistical validity of this group of parameters led to their demise in QSAR studies. Verloop subsequently established the adequacy of just three parameters for QSAR analysis: a slightly modified length *L*, a minimum width B1, and a maximum width B5 that is orthogonal to *L*[75]. The use of these insightful parameters have done much to enhance correlations with biological activities. Recent analysis in our laboratory has established that in many cases, B1 alone is superior to Taft's E_S and a combination of B1 and B5 can adequately replace E_S[76].

Eg.- Verloop steric parameters for a carboxylic acid group

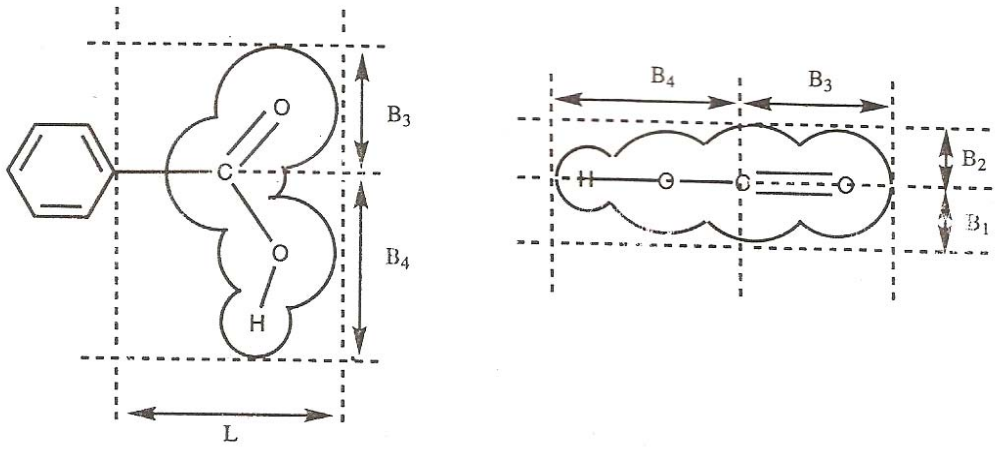


Fig- [L is the length of the substituent while $B_1 - B_4$ are the radii of the group.

QSAR METHOD: The QSAR method involves recognition that a molecule (organic, peptide, protein, etc.) is really a three-dimensional distribution of properties. The most important of these properties are steric (eg shape and volume), electronic (eg electric charge and electrostatic potential) and lipophilic properties(how polar or non-polar the sections of molecular are, usually exemplified by the log of the octanol-water partition coefficient, log P). Scientists are used to visualizing mainly steric properties of molecules. However, molecules look different when viewed in electrostatic or lipophilic space.

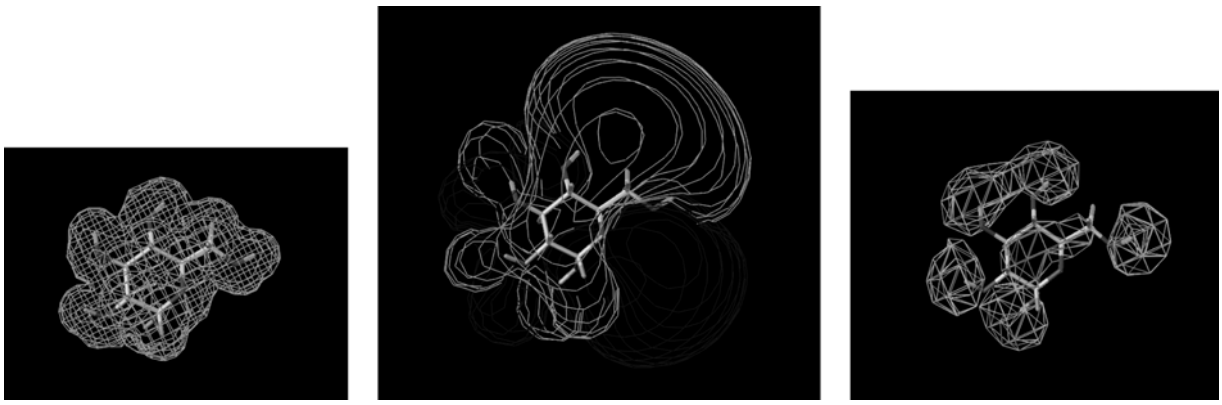


Figure: A small organic molecule (glucopyranose) viewed in steric (left), electrostatic (centre) and lipophilic (right) space

The QSAR method (and analogously QSTR And QSPR) involves a number of key steps:

1. Converting molecular structures into mathematical descriptors that encapsulate the key properties of the molecules relevant to the activity or property being modelled.

2. Selecting the best descriptors from a larger set of accessible, relevant descriptors.

3. Mapping the molecular descriptors into the properties, preferably using a model-free mapping system in which no assumptions are needed as to the functional form of the structure–activity relationship. These relationships are often complex, unknown and non-linear.

4. Validating the model to determine how predictive it is, and how well it will generalise to new molecules not in the date set used to generate the model (the training set).

The relationship between molecular structure and some biological response, BR (eg IC50, LD50, ED90) can be expressed as:[16]

$\log (BR) = f (x_1, x_2, \ldots , x_N)$

where f is usually an unknown complex, non-linear function, and $x_1, \ldots , x_N$ are molecular descriptors. Building of a QSAR model via the four steps outlined above involves finding the best form of function f.

VARIOUS DESCRIPTORS USED IN QSAR

MOLECULAR DESCRIPTORS: Molecular descriptors are numerical values that characterize properties of molecules.Molecular descriptors encoded structural features of molecules as numerical descriptors. Vary in complexity of encoded information and in compute time. These are truly structural descriptors because they are based only on the two-dimensional representation of a chemical structure.
Molecular descriptors can be of diverse types. It is categorized into fragment descriptors, involving properties of sections of molecules, and whole molecule descriptors, based on the properties of the intact molecule.

FRAGMENT DESCRIPTORS: The very earliest descriptors used in QSAR were of this type. QSAR was performed using substituent constants such as hydrophobic constants π, molar refractivity MR, Hammett constants σ and several other, less well-known constants. The recent explosion in the number of molecular descriptors is partly due to the ease by which they may be generated by computational methods, such as molecular orbital calculations[77-79]. There has also been a focus on developing fragment descriptors that are very computationally efficient. The reason is that rapid searching for leads in large chemical libraries (databases of real chemical compounds) or virtual libraries (databases of chemically reasonable molecules that have not yet been synthesized) require efficient information-rich descriptors. Surprisingly simple descriptors can yield useful models. For example, molecules may be represented simply by counting the numbers of atoms of specific elemental type, with specific numbers of connections (a measure of atomic hybridization). A current trend is to employ fragment descriptors based on important molecular properties such as hydrophobes (eg. Aromatic rings), hydrogen bond donors (eg amines), hydrogen bond acceptors (eg carbonyls), positive charges (eg NH_4^+) and negative charges (eg PO_3^-). The rationale for this was first described by Andrews and coworkers.[80] Other fingerprint and general fragment based methods such as molecular holograms[81,82] generalise this approach of breaking molecules into fragments. Another important class of fragment-based descriptors, the van der walla surface area descriptors (VSA), have been reported by Labute to have attributes that make them a widely applicable QSAR descriptors.[83] VSA descriptors are derived by adding together the van der walls surface area contributions of atoms exhibiting a given property (chosen from steric, electrostatic and lipophilic properties) within a given binned property range. Linear combinations of VSA descriptors correlate well with most commonly used descriptors. Fragment- based descriptors have advantages of being computationally efficient and independent of molecular confirmation or 3D structure.

Whole molecule descriptors: They typically capture information on molecular size and lipophilicity through properties such as the molecular weight or molecular volume and log of the octanol-water partition coefficient (logP). The relationship between log P and some biological responses was often inverse parabolic, in which a maximum in the biological response occurred at some optimum log P value. The explanation for this relationship was that it described the

partitioning of drug molecules into biological membranes. An important class of whole molecule descriptors are the topological descriptors.[84-88] These involve treating molecules as topological objects where atoms become the vertices, and bonds the edges, of a molecular graph.

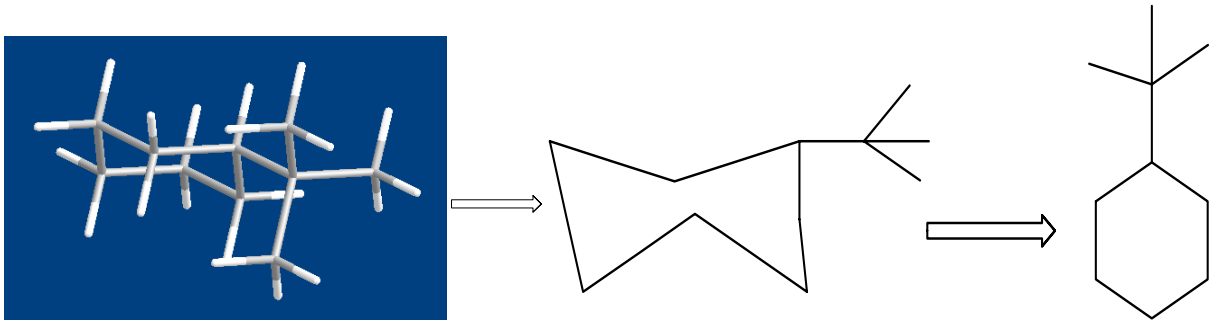


Figure shows the conversion of a molecular structure into a molecular graph. Three- dimensional structure(left), two-dimensional, hydrogen suppressed structure(centre) and hydrogen-suppressed molecular graph(right).

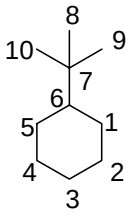
Topological indices are 2D descriptors based on graph theory concepts. These indices have been widely used in QSAR studies. They help to differentiate the molecules according mostly to their size, degree of branching, flexibility, and overall shape. The most widely known descriptors are those that were originally proposed by Randic[89] and extensively developed by Kier and Hall[90]. The strength of this approach is that the required information is embedded in the hydrogen-suppressed framework and thus no experimental measurements are needed to define molecular connectivity indices. For each bond the C_k term is calculated. The summation of these terms then leads to the derivation of X, the molecular connectivity index for the molecule.

$C_k = (\delta_i \delta_j)^{-0.5}$ where $\delta = \sigma - h$

δ is the count of formally bonded carbons and h is the number of bonds to hydrogen atoms. To correct for differences in valence, Kier and Hall proposed a valence delta (δ^v) term to calculate valence connectivity indices[91]. Molecular connectivity indices have been shown to be closely related to many physicochemical parameters such as boiling points, melting points, dipole moment, solubility, molar refraction, polarizability, and partition coefficients[92,93].

Recently, descriptors derived from eigenvalues of molecular matrices derived from graphs have shown promise in generating descriptors useful for QSAR[94-96] and for molecular diversity purposes (eg characterization of chemical libraries and databases, and for design of optimally diverse combinatorial libraries. Modified adjacency matrices describe how atoms in a molecule are connected. They provide a means of combining the molecular properties with topological information encoding the way a molecule is connected.

Figure below is showing an example of a modified adjacency matrix. Diagonalisation of these matrices provide eigenvalue descriptors. A modification of this eigenvalue approach has been particularly useful in the description of molecular diversity (dissimilarity between molecules).



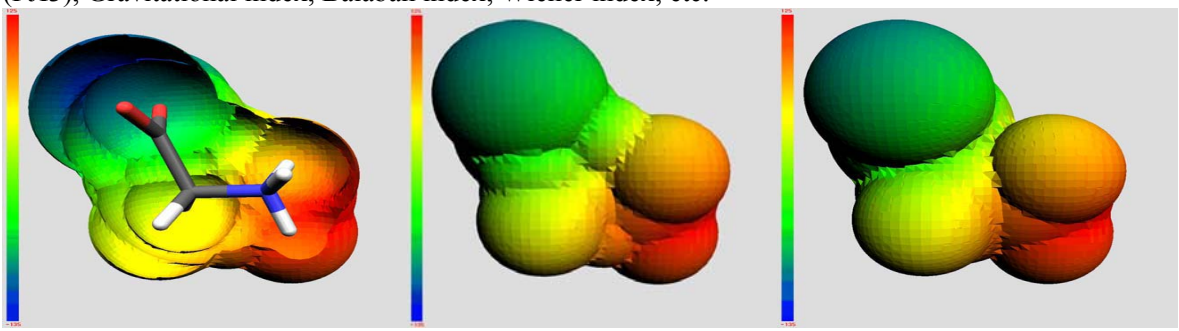
$$\begin{vmatrix} 0&1&0&0&0&1&0&0&0&0 \\ 1&0&1&0&0&0&0&0&0&0 \\ 0&1&0&1&0&0&0&0&0&0 \\ 0&0&1&0&1&0&0&0&0&0 \\ 0&0&0&1&0&1&0&0&0&0 \\ 1&0&0&0&1&0&1&0&0&0 \\ 0&0&0&0&0&1&0&1&1&1 \\ 0&0&0&0&0&0&1&0&0&0 \\ 0&0&0&0&0&0&1&0&0&0 \\ 0&0&0&0&0&0&1&0&0&0 \end{vmatrix}$$

Conversion of molecule into an adjacency matrix. Off-diagonal elements are 1 if the two atoms are bonded, 0 if not.

Brief description of some descriptors are[97]:

1. Constitutional Descriptor: Molecular weight, no. of atoms, no. of non-H atoms, no. of bonds, no. of heteroatoms, no. of multiple bonds (nBM), no. of aromatic bonds, no. of functional groups (hydroxyl, amine, aldehyde, carbonyl, nitro, nitroso, etc.), no. of rings, no. of circuits, no of H-bond donors, no of H-bond acceptors, no. of Nitrogen atoms (nN), chemical composition, sum of Kier-Hall electrotopological states (Ss), mean atomic polarizability (Mp), number of rotable bonds (RBN), mean atomic Sanderson electronegativity (Me), etc. Total number of atoms in the molecule. Numbers of atoms of certain chemical identity (C, H, O, N, F, etc.) in the molecule. Numbers of certain chemical groups and functionalities in the molecule. Total number of bonds in the molecule. Numbers of single, double, triple, aromatic or other bonds in the molecule. Number of rings, number of rings divided by the total number of atoms. Number of 6 membered aromatic rings. Molecular weight and average atomic weight.

2. Geometrical Descriptor: Descriptors using the atomic coordinates (x,y,z) of a molecules are therefore called 3D descriptors. Examples: vander Waals volume, molecular surface, polar surface, etc As a consequence they usually depend on the conformation. 3D petijean shape index (PJI3), Gravitational index, Balaban index, Wiener index, etc.



3. **Quantum Mechanical Descriptor:** Highest occupied Molecular Orbital Energy (HOMO) , Lowest Unoccupied Molecular Orbital Energy (LUMO), Most positive charge (MPC), Least negative charge (LNC), Sum of squares of charges (SSC), Sum of square of positive charges (SSPC), Sum of square of negative charges (SSNC), Sum of positive charges (SUMPC), Sum of negative charges (SUMNC), Sum of absolute of charges (SAC), Total dipole moment (DMt), Molecular dipole moment at X-direction (DMX), Molecular dipole moment at Y-direction (DMY), Molecular dipole moment at Z direction (DMZ), Electronegativity (χ= -0.5 (HOMO-LUMO)), Electrophilicity ($\omega = \chi 2/2\ \eta$), Hardness (η = 0.5 (HOMO+ LUMO)), Softness (S=1/ η).

4. Functional Group Descriptor: Number of total tertiary carbons (nCt), Number of H-bond acceptor atoms (nHAcc), number of total hydroxyl groups (nOH), number of unsubstituted aromatic C(nCaH), number of ethers (aromatic) (nRORPh), etc.

5. Chemical Descriptor: LogP (Octanol-water partition coefficient), Hydration Energy (HE), Polarizability (Pol), Molar refractivity (MR), Molecular volume (V), Molecular surface area (SA).

6. **Substituent Electronic Descriptors:** RMSQ (Root mean square error of charges), SPQ (Sum of positive charges), SNQ (Sum of negative charges), RMSDM (Root mean square of dipole moments at any Cartesian coordinate direction), TDM (Total dipole moment), FRMS (Root mean square force that any atom in constituent molecule see right before the optimization), FMAX (Maximum force on molecule), HOMO (Highest occupied molecular orbital), LUMO (Lowest unoccupied molecular orbital), HD (Hardness), SOF (Softness), EPH (Electrophilicity), EN (Electronegativity).

DESCRIPTOR SELECTION: To build a good QSAR model, a minimal set of information-rich descriptors is required. The large number of possible indices creates several problems for the modeller.[98,99]

1. Many descriptors do not contain molecular information relevant to the problem.

2. Many descriptors are linearly dependent (contain essentially the same information).

3. Use of poor descriptors in QSAR yields poor and misleading models.

4. Including too many descriptors in the model, even if they contain relevant information, can result in overfitting of the model, and loss of ability of the model to generalize to unseen molecules.

5. Many methods of screening this large pool of potential descriptors for relevant ones can lead to chance correlations (correlations that arise by chance because so many descriptors have been tried in models). In other words, if a large number of random numbers are generated as potential descriptors (which clearly do not contain any useful molecular information), and various subsets of these are used to build models, apparently significant models can arise by chance.

The earliest method of variable selection used stepwise regression. This was integrated with the model-building process and involved stepwise addition (or backwards elimination) of descriptors according to a statistical test, to find the best model. Another widely used variable reduction method is principal components analysis (PCA). This involves creating a smaller set of new orthogonal descriptors from linear combinations of the original descriptors and using these to generate QSAR models.

METHODS USED IN QSAR ANALYSIS:

Linear free energy related method[100]: Description of the molecular structure, electronic orbital distribution, reactivity, reaction rates and the role of structural and steric components and substituents of chemical compounds have been the subject of mathematical formulation by physical-organic chemistry for half a century. The most promising approach to the quantification of the interaction of drug molecules with biological systems involved the application of established thermodynamic principles. It is known as the linear free energy (LFE) or extra-thermodynamic method which assumes an additive effect of various substituents in electronic, steric, hydrophobic and dispersion data in the non-covalent interactions of a drug and biomacromolecules. This method is expressed as follows:

$\Delta \log 1/C = f(\Delta L/\Delta H, \Delta E, \Delta Es)$

Depending upon the circumstances, this equation can be modified as

$\log 1/C = b\pi + a$

$= c\ p\ Ka + a$

$= d\ Es + a$

$= b\pi + c\ p\ Ka + a$

$= b\pi + d\ Es + a$

$= b\pi + c\ p\ Ka + d\ Es + a$ (equation 1)

Hence the variance in the biological activity (ΔBA) is explained by the variance of linear free energy related substituent constants which describe the variance in lipophilic/hydrophilic (ΔL/ΔH), electronic (ΔE), steric (ΔEs) or other properties of the parent molecule induced by the subsituents.

Two important models included in LFER method are:

(1) HANSCH ANALYSIS

Introduction:-

Corwin Hansch[101] said "Similar compounds behave similarly" Corwin Hansch (born October 6, 1918, Kenmare, North Dakota) is Professor of Chemistry at Pomona College in California. Hansch taught Organic Chemistry for many years at Pomona College . His course in Physical Bio-Organic Medicinal Chemistry was ground-breaking at an undergraduate level.

Hansch may be best known as the father of the concept of Quantitative Structure-Activity Relationship (Q.S.A.R.), the quantitative correlation of the physicochemical properties of molecules with their biological activities.

He is also noted for the Hansch equation, which is used in

(1) Multivariate Statistics - Multivariate statistics is a set of statistical tools to analyse data (e.g., chemical and biological) matrices using regression and/or pattern recognition techniques.

(2) Hansch Analysis - Hansch analysis is the investigation of the quantitative relationship between the biological activity of a series of compounds and their physicochemical substituent or global parameters representing hydrophobic, electronic, steric and other effects using multiple regression correlation methodology.

(3) Hansch-Fujita π constant - The Hansch-Fujita π constant describes the contribution of a substituent to the lipophilicity of a compound.

Hansch proposed the action of a drug as depending on two processes. Firstly the movement of drug from the point of entry in the body to the site of action and secondly the interaction with the receptor site. Hansch suggested the linear and non-linear dependence of biological activity on difference parameters.

Importance of Lipophilicty:- Hansch visualized that diffusion into cell is slow process so as it is also important one to determine .It is highly dependent on molecular structure of the drug.

Drug must pass out two barriers to put out their effect at site of action ,lipophilliic barrier(cell membrane) and aqueous barrier(cytoplasm) as we know that cytoplasm is made of fatty acids and membrane is made of glycolipids and phospholipids, they have two ends -

(i)Lipophilicity or Hydrophobic—steroids and hydrocarbons

(ii)Hydrophillic end— hydroxyl group in cholesterol, sugar in glycolipids ammonia moiety in phospholipids

The cell is in lipid bilayer structure it is generally found that increasing the lipophilicity of a lead compound results in an increase in biological activity. This means fact that drugs have to cross hydrophobic barriers such as cell membranes to reach their target. But in vitro studies there is no such barriers have to be crossed and interact with a target system such as an enzyme or receptor where the binding site is usually hydrophobic. Therefore, increasing hydrophobicity must give positive results but it is not like we hope because highly hydrophobic means poorly soluble in the aqueous phase,Alternatively, it may got 'trapped' in fat depots and unable to reach the target site.

Conclusion-The drug must have a balance between hydrophilic and lipophilic properties to cross these barriers, that cell structure or target site approach make lipophilicity a crucial factor.

4.4 Assumptions in Hansch analysis:-

i. Conformational changes takes place in target site can be ignored

ii. Metabolism doesn't interferes in it

iii. Linear free energy terms are relevant to receptor's affinity and additive in nature

iv. Relationship between potency and lipophilicity is linear or parabolic

v. Hansch proposed the action of a drug as depending on two processes. Firstly the movement of drug from the point of entry in the body to the site of action and secondly the interaction with the receptor site. Hansch suggested the linear and non-linear dependence of biological activity on difference parameters.

Linear Hansch model:

The correlation of biological activity with physicochemical properties is often termed an Extrathermodynamic relationship. Because it follows in the line of Hammett and Taft equations that correlate thermodynamic and related parameters, it is appropriately labeled. The Hammett equation represents relationships between the logarithms of rate or equilibrium constants and substituent constants. The linearity of many of these relationships led to their designation as linear free energy relationships. The Hansch approach represents an extension of the Hammett equation from physical organic systems to a biological milieu. It should be noted that the simplicity of the approach belies the tremendous complexity of the intermolecular interactions at play in the overall biological response.

Interestingly, the concurrent considerations of π and σ (Hammett's constant) has evolved gainful vital correlations existing between the biological activities of quite a few drug substances with their corresponding physical properties and chemical structures.

Therefore, Hansch's correlations piece together valuable informations of a newly designed 'drug molecule' in a more plausible, predictive and quantifiable manner than before and apply it to a biological system more logistically and judiciously. This particular concept and idea was further substantiated and expanded by assuming that all the three substituents viz., π, σ and Es, exert a significant effect on the efficacy and hence the potency of a 'drug substance'; and are found to be additive in nature independently. Therefore, it has given rise to the underlying linear Hansch equation also called extrathermodynamic approach.

$$\log 1/C = a \log P + b\sigma + c E_S + d \qquad \text{(Equation 2)}$$

Where, C = Concentrations of drug producing the biological response being measured,

Log P = Substituent constant for solubility (i.e., π),

E_S = Taft constant (for steric effects),

σ = Hammett substitution constant

a, b, d = Constants of the system (which are determined by computer to obtain the 'best fitting line').

It is pertinent to state at this juncture that not all the parameters shall necessarily be significant.

Nonlinear Hansch models

Later on experiences results that

(i) Increase in log P value from log P_o does not linearly cause increase in biological activity some time its decreases.

(ii) If P values spread over a large range.

Thus, Hansch et al suggested that the compounds could be involved in a random-walk process: low hydrophobic molecules had a tendency to remain in the first aqueous compartment, whereas highly hydrophobic analogs sequestered in the first lipoidal phase that they encountered. This led to the formulation of a parabolic equation, relating biological activity and hydrophobicity[102].

$\log 1/C = -a (\log P)^2 + b. \log P +$ constant (k) (equation 3)

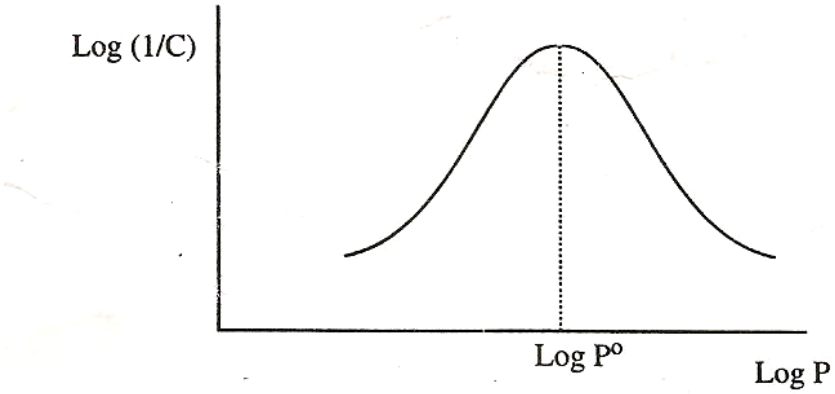
In the random-walk process, the compounds partition in and out of various compartments and interact with myriad biological components in the process. To deal with this conundrum, Hansch proposed a general, comprehensive equation for QSAR[103].

$\log 1/C = -a (\log P)^2 + b. \log P + \rho \sigma + \delta E_S + k$ (equation 4)

Where, P = n-octanol/ water partition coefficient, σ = Hammett electronic parameter, a,b,c = regression coefficients, E_S = taft's steric factor, k = constant term, ρ (rho) and δ are Proportionality constant designating the sensitivity of the reaction to electron density.

The optimum value of log P for a given system is log P_o and it is highly influenced by the number of hydrophobic barriers a drug encounters in its walk to its site of action.

The coefficients (a, b, c, d, e) are determined by multi-regression analysis.



Graph- When P is small, the $(\log P)^2$ term is very small and the equation is dominated by the log P term. This represents the first part of the graph where activity increases with increasing P. When P is large, the $(\log P)^2$ term is more significant and eventually 'overwhelms' the log P term. This represents the last part of the graph where activity drops with increasing P

Because of importance of electronic factors, steric effect and shape of molecules for receptor interaction, electric factor (σ) steric factor (Es) and topography terms (S) added to equation:-

If Linear, $\log 1/C = a \log P + b\,\sigma + c\,E_S + d$

If Parabolic, $\log 1/C = -a(\log P)^2 + b.\log P + c\,\sigma + d\,E_S + k$

Equation 4, was developed from the concept that the transport of a drug from the site of application to its site of action depends in a nonlinear manner on the lipophilicity of the drug, and that the binding affinity to its biological counterpart, such as an enzyme or a receptor, depends on the lipophilicity, the electronic properties and other linear free-energy-related properties. Equation 4, combines the description of both processes in one mathematical model. In addition to the introduction of a parabolic term for the nonlinear lipophilicity dependence and the combination of different physicochemical properties in one equation, Hansch and Fujita defined lipophilicity parameters π of substituents 'X'(Eqn. 5), in the same manner as Hammett had defined the electronic parameter σ (Eqn. 6). The partition coefficient P in equation 5 is an equilibrium constant, similar to the dissociation or reaction constants K in Eqn. 6. The absence of a 'reaction term' π in Eqn. 5 is explained by the fact that all π values refer to the n-octanol/water system.

$\Pi_x = \log P_x - \log P_{yH}$ (Equation 5)

$\rho\,\sigma = \log K_{RX} - \log K_{RH}$ (Equation 6)

Where, log P = logarithm of 1-octanol-water partition coefficient.

ρ (rho) = Proportionality constant designating the sensitivity of the reaction to electron density.

y = A parent compound (i.e., an unsubstituted reference compound/drug).

The hydrophobic characterictic, designated by π_x, may be correlated to a drug's distribution pattern, within which a given substituent 'x' affects molecular behavior and conduct with regard to its:

1. Distribution and transport, and

2. Drug-receptor activities.

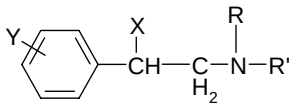
Salient Features: The various salient features of **Hansch equation** are as enumerated under:

(1)Value of π is indicative, to a certain extent, the behavioural pattern of a 'substituent' contributing to the solubility behavior of a molecule under investigation. It also reflects upon the manner it gets partitioned between lipoidal and aqueous interfaces in the reputed compartments it happens to cross as a 'drug' so as to reach the 'site of action' ultimately.

(2) It is, however, not very clear and definite whether the solid surface of a drug undergoes adsorption on colloidally suspended plasma proteins while establishing the hydrophobic characteristics π.

Example: β-Halo-arylamines: The adrenergic blocking profile of β-halo-arylamines was observed to be solely related to the constants, π and σ; and specifically excluded the steric factor altogether.

i.e., log 1/C = 1.22 π – 1.59 σ + 7.89

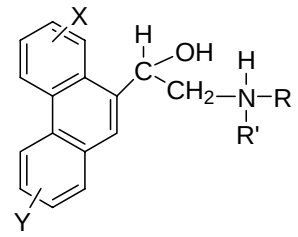


The aforesaid equation offers a dictum that the 'biological response'gets enhanced if the substituents possess a positive π value and a negative σ value; or more explicitely the substituents must preferentially be both hydrophobic in nature and electron donating in character.

It has been established beyond any reasonable doubt that there exists no correlation between the π factor and the P value; therefore, it is quite feasible to have Hansch equations essentially comprising of these two stated components:

Examples: Phenanthrene aminicarbinols: An analogues series of more than one hundred phenanthrene aminocarbinols were successfully synthesized and subsequently screened for their antimalarial profile. Interestingly, the analogues series fitted appropriately into the following version of Hansch equation:

$$\text{Log } 1/C = -0.015 (\log P)^2 + 0.14 \log P + 0.27 \square \pi_x + 0.40 \square \pi_y + 0.65 \square \sigma_x + 0.88 \square \sigma_y + 2.34$$



PHENANTHRENE AMINOCARBINOL

Salient features: The various salient features that may be derived from the above equation are, namely:

(1) As the hydrophobicity of the molecule (P) enhances there exists a very nominal increase in the antimalarial activity.

(2) The corresponding constant is low (0.14) which reflects that the increase in antimalarial activity is also low.

(3) The value of $(\log P)^2$ evidently reveals that there prevails a maximum P value for activity.

(4) Further the above equation suggests that the antimalarial activity gets enhanced appreciably when the hydrophobic moieties are strategically located either on ring 'X' or more specifically on ring 'Y'. It further ascertains that the hydrophobic interactions(s) are virtually taking place at these sites.

(5) The electron- withdrawing substituents on rings 'X' and 'Y' contribute enormously to the antimalarial activity; however, the effect is more on ring 'Y' than in ring 'X'.

Free-Wilson analysis (structure-property relationship)

In 1964, Free and Wilson derived a mathematical model that describes the presence and absence of certain structural features, i.e. those groups that are chemically modified, by values of 1 or 0 and correlate the resulting structural matrix with biological activity values, following Eqn. 7; the values a_i in Eqn.7 are the biological activity group contributions of the substituents X_1, X_2,…X_i in the different positions P of compound, the most often the unsubstituted parent structure of a series[105,106]**.** The method of Free and Wilson is based upon an additive mathematical model in which a particular substituent in a specific position is assumed to make an additive and constant contribution to the biological activity of a molecule in a series of chemically related molecules. This method is based on the assumption that the introduction of a particular substituent at a particular molecular position always leads to a quantitatively similar effect on biological potency of the whole molecule, as expressed by the equation

$$\log 1/C = \Sigma a_i + \mu \qquad \text{(Equation7)}$$

a_i = substituent group contributions, μ = activity contribution of reference compound

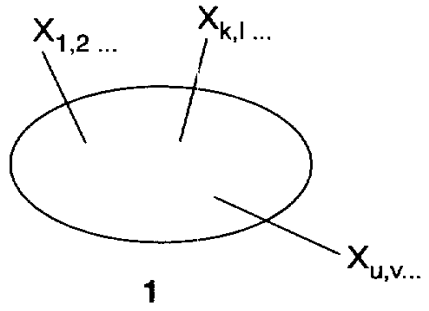


Figure 1: Schematic presentation of a molecule for Free Wilson analysis. A common skeleton bears subsituents Xi in different positions p; the presence or absence of these substituents is coded by the values 1 and 0, respectively.

Mixed Hansch/Free-Wilson model

The similarity in approaches of Hansch analysis and Free-Wilson analysis allows them to be used within the same framework. This is based on their theoretical consistency and the numerical equivalencies of activity contributions. This development has been called the mixed approach and can be represented by the following equation:

Log $1/C = \sum a_i + \sum c_j \Phi_j$ + constant (Equation 8)

The term a_i denotes the contribution for each ith substituent, whereas Φ_j is any physicochemical property of a substituent X_j.
Equation 8 combines the advantages of Hansch and FreesWilson analyses and widens the applicability of both methods. Physicochemical parameters describe parts of the molecules with broad structural variation, whereas indicator variables a_i (Free Wilson type variables) encode the effects of structural variations that cannot be described otherwise[104,105]. A recent study of the P-glycoprotein inhibitory activity of 48 propafenone-type modulators of multidrug resistance, using a combined Hansch/Free-Wilson approach was deemed to have higher predictive ability than that of a stand-alone Free-Wilson analysis[107].

Conclusion: It involves the mathematical and statistical analysis of SAR-data which helps to reduce the number of educated guesses in molecular modification. QSAR is thus a scientific achievement and an economic necessity to reduce an empiricism in drug design to ensure that every drug synthesized and pharmacologically tested should be as meaningful.

Refernces

1. C. Hansch and A. Leo. Substituent Constants for Correlation Analysis in Chemistry and Biology, John Wiley & Sons, New York, 1979.

2. A Crum-Brown, TR Fraser, Trans *R* Soc Edinburgh 1868–1869; *25:* 151.
3. Richardson BJ. Physiological research on alcohols. Med Times Gaz 1868; 2: 703-706.
4. Mills EJ. On melting point and boiling point as related to composition. *Philos Mag* 1884; *17:* 173-187.
5. Richet C. On the relationship between the toxicity and the physical properties of substances. Compt. Rendus Seances Soc Biol 1893; 9: 775-776.
6. Overton E. Osmotic properties of cells in the bearing on toxicology and pharmacology. *Z* Physik Chem 1897; *22:* 189-209.
7. Meyer H. On the theory of alcohol narcosis I. Which property of anesthetics gives them their narcotic activity? Arch Exp Pathol Pharmakol 1899; *42:* 109-118.
8. Hammett LP. Some relations between reaction rates and equilibrium constants. Chem Rev 1935; 17: 125-136.
9. Hammett LP. The effect of structure upon the reactions of organic compounds. benzene derivatives. J Am Chem Soc 1937; *59:* 96-103.
10. Ferguson J. The Use of Chemical Potentials as Indices of Toxicity. Proc R Soc Lond B 1939; 127: 387-404.
11. Bell PH, Roblin RO. Studies in chemotherapy. vii. a theory of the relation of structure to activity of sulfanilamide type compounds1. J Am Chem Soc 1942; 64: 2905-2917.
12. Albert A, Goldacre R, Phillips J. The strength of heterocyclic bases. J Chem Soc 1948: 2240-2249.
13. Taft RW. Polar and steric substituent constants for aliphatic and o- Benzoate groups from rates of esterification and hydrolysis of esters1. J Am Chem Soc 1952; *74:* 3120-3128.
14. Hansch C, Maloney PP, Fujita T, Muir RM. Correlation of biological activity of phenoxyacetic acids with hammett substituent constants and partition coefficients. Nature 1962; 194: 178-180.
15. G Klopman, JK Li, S Wang, M Dimayuga. J Chem Inf Comput Sci 1994; *34:* 752.
16. Hansch C, Fujita TP. Analysis: A Method for the correlation of biological activity and chemical structure. J Am Chem Soc 1964; *86:* 1616-1626.
17. C Hansch, T Fujita. J Am Chem Soc 1964; 86: 1616.
18. Hansch C. Quantitative approach to biochemical structure-activity relationships. Acc Chem Res 1969; *2:* 232-239.
19. Free SM, Jr Wilson JW. A Mathematical contribution to structure-activity studies. *J* Med Chem 1964; *7*: 395-399.
20. Fujita T, Ban T. Structure-activity study of phenethylamines as substrates of biosynthetic enzymes of sympathetic transmitters. J Med Chem 1971; *14:* 148-152.
21. Kubinyi H. Quantitative structure-activity relationships. IV. Nonlinear dependence of biological activity on hydrophobic character: a new model. Arzneimittelforschung 1976; *26:* 1991-1997.
22. Hansch C, Gao H. Comparative QSAR: Radical reactions of benzene derivatives in chemistry and biology. Chem Rev 1997; 97: 2995-3060.
23. Hurst T, Heritage T. HQSAR - A Highly Predictive QSAR Technique Based on Molecular Holograms. In: 213th ACS Natl. Meeting, San Francisco, CA, 1997.
24. Lowis DR. HQSAR: A New, Highly Predictive QSAR Technique. In: Tripos Technical Notes; Tripos Inc.: USA, 1997:1.
25. Cho SJ, Zheng W, Tropsha A. Rational combinatorial library design. 2. Rational design of targeted combinatorial peptide libraries using chemical similarity probe and the inverse QSAR approaches. J Chem Inf Comput Sci 1998; *38:* 259-268.

26. Labute P. Binary QSAR: a new method for the determination of quantitative structure activity relationships. Pac Symp Biocomput 1999; 444-455.
27. DJ Livingstone. J Chem Inf Comput Sci 2000; 40: 195.
28. C Hansch, A Kurup, R Garg and H Gao, Chem Rev 2001; 101: 619.
29. Kubinyi H, 3D QSAR in drug design. Theory, methods and Applications, ESCOM science publishers, 1993.
30. Bohm HJ. J Comput Aided Mol Design 1994; 8: 243-256.
31. Bohm HJ and Klebe G. Angew Chem 1996; 108: 2750-2778.
32. J Langley. J Physiol 1878; 1: 367.
33. P Ehrlich. Klin Jahr 1897; 6: 299.
34. J N Langley. J Physiol 1905; 33: 374.
35. M Famulok. Curr Opin Struct Biol 1999; 9: 324.
36. KY Wang, S Swaminathan and P H Bolton. Biochemistry 1994; 33: 7617.
37. JW Lown in S. Neidle and M.-J. Waring, Eds., Molecular Aspects of Anticancer Drug-DNA Interactions, Macmillan, Basinstoke, UK, 1993; 322.
38. AA Kumar, JH Mangum, DT Blankenship et al. J Biol Chem 1981; 256: 8970.
39. GD Rose and R Wolfenden. Annu Rev Biophys Biomol Struct 1993; 22: 381.
40. A T Hagler, P Dauber and S Lifson. J Am Chem Soc 1979; 101: 5131.
41. W Kauzmann. Adv Protein Chem 1959; 14: 1.
42. A Ben-Naim. Pure Appl Chem 1997; 69: 2239.
43. PN Craig. J Med Chem 1971; 14: 680.
44. JG Topliss. J Med Chem 1972; 15: 1006.
45. JG Topliss. J Med Chem 1977; 20: 463.
46. NR Draper and H Smith. Applied Regression Analysis, 2nd ed., John Wiley & Sons, New York, 1981.
47. Y Martin in G. Grunewald, Ed., Quantitative Drug Design, Marcel Dekker, New York, 1978; 167.
48. H Meyer. Arch Exp Pathol Pharmakol 1899; 42: 109.
49. E Overton. Studien Uber die Narkose, Fischer, Jena, Germany, 1901.
50. A Leo and C Hansch. J Org Chem 1971; 36: 1539.
51. B C Lippold and M S Adel. Arch Pharm 1972; 305: 417.
52. A Leo. J Pharm Sci 1987; 76: 166.
53. A Leo. Methods Enzymol 1991; 202: 544.
54. JC Dearden, AM Patel and JM Thubby. J Pharm Pharmacol 1974; 26 (Suppl.): 75P.
55. W Draber, K H Buchel and K Dickore. Proc Int Congr Pest Chem, 2nd ed., 1971; 5: 153.
56. K Valko, C Beran and D Reynolds. Anal Chem 1997; 69: 2022.
57. K Valko, CM Du, C Bevan et al. Curr Med Chem 2001; 8: 1137.

58. RF Rekker. The Hydrophobic Fragmented Constant. Its Derivation and Application: AMeans of Characterizing Membrane Systems,Elsevier, Amsterdam, 1977: 131-134.
59. GG Nys and RF Rekker. Chim Ther 1973; 8: 521.
60. A Leo, PYC Jow, C Silipo et al. J Med Chem 1979; 14: 865.
61. D Weininger. J Chem Int Comput Sci 1988; 28: 31.
62. CA Ramsden, Ed., Comprehensive Medicinal Chemistry. The Rational Design Mechanistic Study and Therapeutic Application of Chemical Compounds, Quantitative Drug Design, Pergamon, Elmsford, NY, 1990; 4: 315.
63. M J Kamlet, P W Carr, R W Taft and M H Abraham. J Am Chem Soc 1981; 103: 6062.
64. M J Kamlet, J L Abboud, M Abraham and R Taft. J Org Chem 1983; 48: 2877.
65. L P Hammett. Physical Organic Chemistry. 2nd ed., McGraw-Hill, New York, 1970.
66. G N Burckhardt, W G K Ford and E Singelton. J Chem Soc *1936:* 17.
67. L P Hammett. J Chem Ed 1966; 43: 464.
68. M Charton. Prog Phys Org Chem 1971; 8: 235.
69. T Fujita and T Nishioka. Prog Phys Org Chem 1976; 12: 49.
70. R W Taft in M S Newman, Ed., Steric Effects in Organic Chemistry, John Wiley & Sons, New York, 1956: 556.
71. K Hancock, E A Meyers and B J Yager. J Am Chem Soc 1961; 83: 4211.
72. M Charton in M Charton and I Motoc, Eds., Steric Effects in Drug Design, Springer, Berlin, 1983: 57.
73. MS Tute in C A Ramsden, Ed., Comprehensive Medicinal Chemistry. The Rational Design, Mechanistic Study and Therapeutic Application of Chemical Compounds, Quantitative Drug Design, Pergamon, Elmsford, NY, 1990; 4: 18.
74. A Verloop, W Hoogenstraaten and J Tipker in E. J. Ariens, Ed., Drug Design, Vol. VII, Academic Press, New York/London, 1976, 165.
75. A Verloop. The STERIMOL Approach to Drug Design. Marcel Dekker, New York, 1987.
76. Hansch CD, Hoekman A Leo, D Weininger, and C D Selassie. unpublished results.
77. Warne MA and Nicholson JK. Quantitative structure-activity relationships (QSAR's) in environmental research. Part II. Molecular orbital approaches to property calculation. Prog Environ Sci 2000; 2 (1): 31-52.
78. Karelson M, Lobanov VS and Katritzky AR. Quantum-chemical descriptors in QSAR/QSPR studies. Chem Rev 1996; 96(3): 1027-1043.
79. Carbo Dorca R, Amat L, Besalu E et al. Quantum mechanical origin of QSAR: theory and applications. Theochem 2000; 504: 181-228.
80. Andrews PR, Craik DJ and Martin JL. Functional group contributions to drug receptor interactions. J Med Chem 1984; 27: 1648- 1657.
81. Tong W, Lowis DR, Perkins R et al. Evaluation of quantitative structure-activity relationship methods for large-scale prediction of chemicals binding to the estrogen receptor. J Chem Inf Comput Sci 1998; 38: 669-677.
82. Winkler DA. Holographic QSAR of benzodiazepines. Quantitative structure-activity relationship 1998; 17: 224.
83. Labute P. A widely applicable set of molecular descriptors. J Mol Graph Mod 2000; 18: 464-477.
84. Randic M. Computation of optimum parameters for multivariate analysis of structure property relationship. J Comp Chem 1991; 12(8): 970-980.

85. Balaban AT. A personal view topological indices for QSAR/QSPR, QSPR/QSAR Stud Mol Descriptors 2001: 1-30.
86. Devillers J. New trends in QSAR modeling with topological indices. Curr Opin Drug Discovery Development 2000; 3(3): 275-279.
87. Estrada E. Novel strategies in the search of topological indices, Topol Indices Relat Descriptors QSAR QSPR 1999: 403-453.
88. Bonchev D. Overall connectivity and topological complexity. A new tool for QSPR/QSAR, Topol Indices Relat Descriptors QSAR QSPR 1999: 361-401.
89. Randic M. J Am Chem Soc 1975; 97: 6609-6615.
90. LH Hall and L B Kier. J Pharm Sci 1977; 66: 642.
91. LB Kier and M H Hall. J Pharm Sci 1983; 72: 1170.
92. L B Kier and L H Hall. Molecular Connectivity in Chemistry and Drug Research, Academic Press, New York/London, 1976.
93. L H Hall and L B Kier. J Pharm Sci 1975; 64: 1978.
94. Burden FR. A chemically intuitive molecular index based on eigenvalues of a modified adjacency matrix. J Chem Inf Comput Sci 1997; 16: 309-314.
95. Stanton DT. Evaluation and use of BCUT descriptors in QSAR And QSPR studies. J Chem Inf Comput Sci 1999; 39: 11-20.
96. Randic M, Vracko M and Novic M. Eigenvalues as molecular descriptors QSPR/QSAR Stud. Mol Descriptors 2001: 147-211.
97. Fassihi A and Sabet R. QSAR Study of $p56^{lck}$ Protein Tyrosine Kinase Inhibitory Activity of Flavonoid Derivatives Using MLR and GA-PLS. Int J Mol Sci 2008; 9: 1876-1892.
98. Topliss JG and Edwards RP. Chance factors in studies of quantitative structure-activity relationships. J Med Chem 1979; 22(10): 1238-1244.
99. Manallack DT and Livingstone DJ. Artificial neural networks: Application and chance effects for QSAR data analysis. Med Chem Res 1992; 2: 181-190.
100. Kadam, SS, Mahadik, KR and Bothra KG. Principles of Medicinal Chemistry 2004; 1: 375-376.
101. http://en.wikipedia.org/wiki/Corwin_Hansch
102. J T Penniston, L Beckett, D L Bentley, and C Hansch. Mol Pharmacol 1969; 5: 333.
103. C Hansch. Adv Chem Ser 1972; 114: 20.
104. Hanch C and Leo A. Exploring QSAR. Fundamentals and Applications in chemistry and biology, American Chemical Society 1995.
105. Kubinyi H. Burger's Medicinal Chemistry(Wolff, M.F., ed.), John Wiley & Sons 1995; Vol 1: 497-571.
106. Free SM, Jr and Wilson JW. J Med Chem 1964; 7: 395-399.
107. C Tmej, P Chiba, M Huber, E Richter, M Hitzler, K J Schaper, and G Ecker. Arch Pharm 1998; 331: 233.