QUANTATIVE STRUCTURE ACTIVITY RELATIONSHIP (QSAR)

- Navneet F. Chauhan, Ph.D
- Associate Professor,
- Saraswati Institute of Pharmaceutical Sciences, Gandhinagar

QUANTATIVE STRUCTURE ACTIVITY RELATIONSHIP (QSAR)

Many physical, structural, and chemical properties have been studied by the QSAR approach, but the most common are hydrophobic, electronic, and steric properties.

QSAR relates the physicochemical properties of a series of drugs to their biological activity by means of a mathematical equation.

The commonly studied **physicochemical properties** are
1. Hydrophobicity
2. Electronic factors
3. Steric factors

QSAR study considers how the hydrophobic, electronic, and steric properties of the substituents aff ect biological activity.

The three most studied physicochemical properties are now considered in detail.

Navneet F. Chauhan, Ph.D.
Saraswati Institute of Pharmaceutical Sciences

Physicochemical Parameter (Hydrophobicity)

The hydrophobic character of a drug is crucial to how easily it crosses cell membranes and may also be important in receptor interactions.

Changing substituents on a drug may well have significant effects on its hydrophobic character and, hence, its biological activity.

Therefore, it is important to have a means of predicting this quantitatively.

1) The Partition Coefficient (P)
   The hydrophobic character of a drug can be measured experimentally by testing the drug's relative distribution in an n-octanol/water mixture.

   Hydrophobic molecules will prefer to dissolve in the n-octanol layer of this two phase system, whereas hydrophilic molecules will prefer the aqueous layer.

   The relative distribution is known as the partition coefficient (P) and is obtained from the following equation:

$$P = \frac{\text{Concentration of drug in octanol}}{\text{Concentration of drug in aqueous solution}}$$

Navneet F. Chauhan, Ph.D.
Saraswati Institute of Pharmaceutical Sciences

Physicochemical Parameter (Hydrophobicity)

Hydrophobic compounds have a high P value, whereas hydrophilic compounds have a low P value.

Varying substituent's on the lead compound will produce a series of analogues having different hydrophobicities and, therefore, different P values.

By plotting these P values against the biological activity of these drugs, it is possible to see if there is any relationship between the two properties.

In studies where the range of the log P values is restricted to a small range (e.g. log P = 1–4), a straight-line graph is obtained showing that there is a relationship between hydrophobicity and biological activity.

The biological activity is normally expressed as 1/C , where C is the concentration of drug required to achieve a defined level of biological activity. Such a line would have the following equation:

$$\log\left(\frac{1}{C}\right) = -k_1 \log P + k_2$$

The reciprocal of the concentration (1/C) is used, as more active drugs will achieve a defined biological activity at lower concentration.
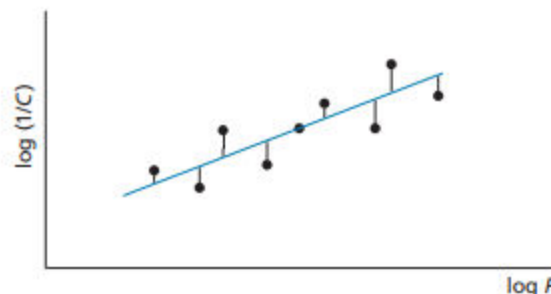

Proximity of data points to line of best fit.

The biological activity ad infinitum . In fact, this does not happen. Th ere are several reasons for this. For example, the drug may become so hydrophobic that it is poorly soluble in the aqueous phase. Alternatively, it may be 'trapped' in fat depots and never reach the intended site.

Finally, hydrophobic drugs are oft en more susceptible to metabolism and subsequent elimination.

A straight-line relationship between log P and biological activity is observed in many QSAR studies because the range of log P values studied is often relatively narrow.

If these studies were to be extended to include compounds with very high log P values, then we would see a different picture. The graph would be parabolic,
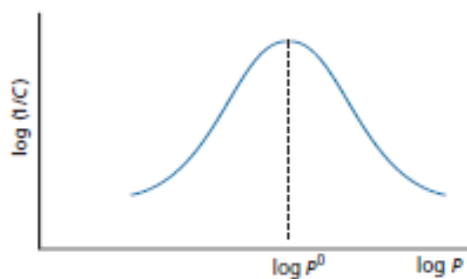
Here, the biological activity increases as log P increases until a maximum value is obtained. The value of log P at the maximum (log P^0) represents the optimum partition coefficient for biological

log (1/C)

log P^0 log P

Parabolic curve of log (1/C) vs. log P.

Beyond that point, an increase in log P results in a decrease in biological activity.
If the partition coefficient is the only factor influencing biological activity, the parabolic curve can be expressed by the equation:

$$\log\left(\frac{1}{C}\right) = -k_1(\log P)^2 + k_2 \log P + k_3$$

Navneet F. Chauhan, Ph.D.
Saraswati Institute of Pharmaceutical Sciences

Substitutent Hydrophobicity Constant (π)

Partition coefficients can be calculated by knowing the contribution that various substituents make to hydrophobicity.

This contribution is known as the substituent hydrophobicity constant (π) and is a measure of how hydrophobic a substituent is relative to hydrogen. The value can be obtained as follows.

Partition coeffi cients are measured experimentally for a standard compound, such as benzene, with and without a substituent (X). The hydrophobicity constant (π_X) for the substituent (X) is then obtained using the following equation:

$$\pi_X = \log P_X - \log P_H$$

where P_H is the partition coefficient for the standard compound and P_X is the partition coefficient for the standard compound with the substituent.

A positive value of π indicates that the substituent is more hydrophobic than hydrogen; a negative value indicates that the substituent is less hydrophobic.

Navneet F. Chauhan, Ph.D.
Saraswati Institute of Pharmaceutical Sciences

Case Study

Altering logP to remove central nervous system side effects
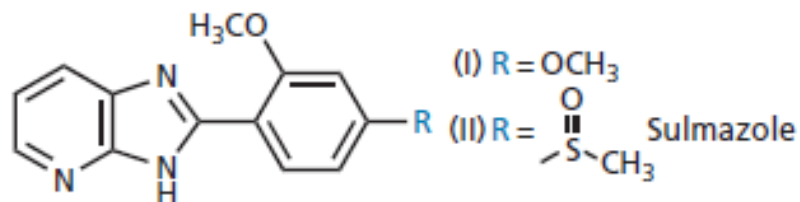
Cardiotonic agents.

The cardiotonic agent (I) was found to produce 'bright visions' in some patients, which implied that it was entering the central nervous system (CNS).

This was supported by the fact that the log *P value of the drug was 2.59. In* order to prevent the drug entering the CNS, the 4-Ome group was replaced by a 4-S(O)Me group. This particular group is approximately the same size as the methoxy group, but more hydrophilic.

The log *P value of the new* drug (sulmazole) was found to be 1.17. The drug was now too hydrophilic to enter the CNS and was free of CNS side effects.
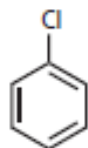
Navneet F. Chauhan, Ph.D.
Saraswati Institute of Pharmaceutical Sciences

Values of π for a range of substituents

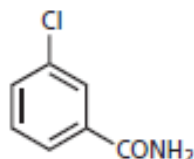| Group	CH$_3$	t-Bu	OH	OCH$_3$	CF$_3$	Cl	Br	F
π (aliphatic substituents)	0.50	1.68	−1.16	0.47	1.07	0.39	0.60	−0.17
π (aromatic substituents)	0.52	1.68	−0.67	−0.02	1.16	0.71	0.86	0.14

Benzene
(log P = 2.13)

Chlorobenzene
(log P = 2.84)
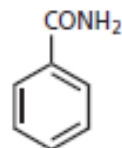
Benzamide
(log P = 0.64)

meta-Chlorobenzamide

Values for log P.

$$\log P_{(chlorobenzamide)} = \log P_{(benzene)} + \pi_{Cl} + \pi_{CONH_2}$$
$$= 2.13 + 0.17 + (-1.49)$$
$$= 1.35$$

Electronic effects

The electronic effects of various substituent's 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 interact with a binding site.

As far as substituents on an aromatic ring are concerned, the measure used is known as the **Hammett substituent constant (σ). This is a measure of the electron withdrawing** or electron-donating ability of a substituent, and has been determined by measuring the dissociation of a series of substituted benzoic acids compared with the dissociation of benzoic acid itself.

Benzoic acid is a weak acid and only partially ionizes in water. An equilibrium is set up between the ionized and non-ionized forms, where the relative proportion of these species is known as the **equilibrium or dissociation constant K_H *(the subscript H signifi es that** there are no substituents on the aromatic ring).*

$$K_H = \frac{[PhCO_2^-]}{[PhCO_2H]}$$

Navneet F. Chauhan, Ph.D.
Saraswati Institute of Pharmaceutical Sciences
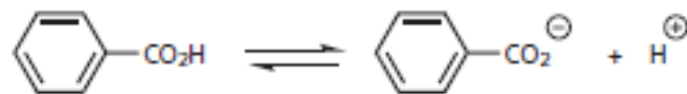
Ionization of benzoic acid in water.

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, and so the equilibrium will shift more to the ionized form. Therefore, the substituted benzoic acid is a stronger acid and has a larger K_X value (X represents the substituent on the aromatic ring
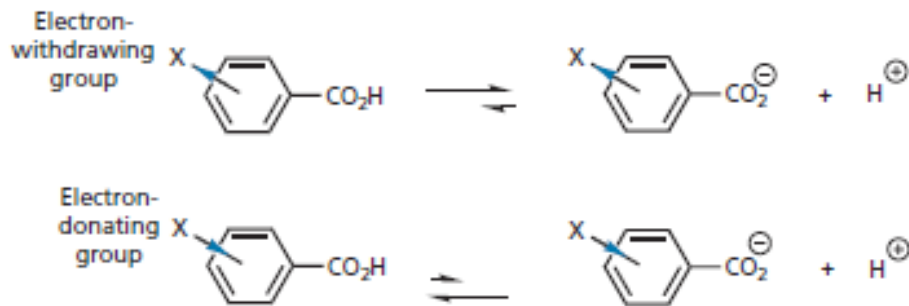
Position of equilibrium dependent on substituent group X.

If the substituent X is an electron-donating group such as an alkyl group, then the aromatic ring is less able to stabilize the carboxylate ion. The equilibrium shifts to the left indicating a weaker acid with a smaller K_X value

Navneet F. Chauhan, Ph.D.
Saraswati Institute of Pharmaceutical Sciences

Th e Hammett substituent constant (σ X) for a particular substituent (X) is defi ned by the following equation:

$$\sigma_X = \log \frac{K_X}{K_H} = \log K_X - \log K_H$$

Benzoic acids containing electron-withdrawing substituents will have larger K X values than benzoic acid itself (K_H) and, therefore, the value of ex. for an electron withdrawing substituent will be positive. Substituents such as Cl, CN, or CF3 have positive σ values.

Benzoic acids containing electron-donating substituents will have smaller KX values than benzoic acid itself and, hence, the value of σ X for an electron-donating substituent will be negative.

Substituents such as Me, Et, and t -Bu have negative values of σ. The Hammett substituent constant for H is zero.

The Hammett substituent constant takes into account both resonance and inductive eff ects. Therefore, the value of σ for a particular substituent will depend on whether the substituent is meta or para .

Navneet F. Chauhan, Ph.D.
Saraswati Institute of Pharmaceutical Sciences

Steric factors

The bulk, size, and shape of a drug will influence how easily it can approach and interact with a binding site.

Hydrolysis of an aliphatic ester.
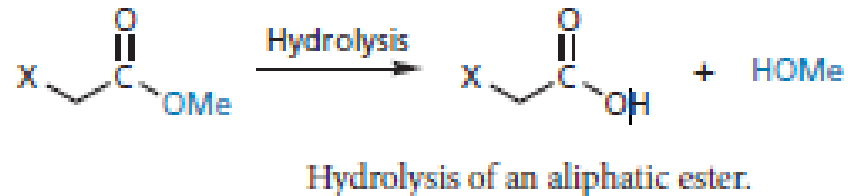
A bulky substituent may act like a shield and hinder the ideal interaction between a drug and its binding site.

Alternatively, a bulky substituent may help to orientate a drug properly for maximum binding and increase activity.

Steric properties are more difficult to quantify than hydrophobic or electronic properties.

Navneet F. Chauhan, Ph.D.
Saraswati Institute of Pharmaceutical Sciences

Taft's Steric factor (*Es*)

The value for Es can be obtained by comparing the rates of hydrolysis of substituted aliphatic esters against a standard ester under acidic conditions. Thus,

$$E_s = \log k_x - \log k_o$$

Where kx represents the rate of hydrolysis of an aliphatic ester bearing the substituent X and ko represents the rate of hydrolysis of the reference ester.

Note that the reference ester is X = Me. Substituents such as H and F, which are smaller than a methyl group, result in a faster rate of hydrolysis (k x > k o), making Es positive.

Substituents which are larger than methyl reduce the rate of hydrolysis (k x < k o), making Es negative.

A disadvantage of Es values is that they are a measure of an intramolecular steric effect, whereas drugs interact with target binding sites in an intermolecular manner.

Molar refractivity

Another measure of the steric factor is provided by a parameter known as molar refractivity (MR). This is a measure of the volume occupied by an atom or a group of atoms. The MR is obtained from the following equation:

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

where n is the index of refraction, MW is the molecular weight, and d is the density.

The term MW/d defi nes a volume and the ($n^2 - 1$)/($n^2 + 2$) term provides a correction
factor by defining how easily the substituent can be polarized.

This is particularly significant if the substituent has π electrons or lone pairs of electrons.

Verloop steric parameter

Another approach to measuring the steric factor involves a computer program called Sterimol, which calculates steric substituent values (Verloop Steric Parameters) from standard bond angles, van der Waals radii, bond lengths, and possible conformations for the substituent.

Unlike Es, the Verloop steric parameters can be measured for any substituent.

For example, the Verloop steric parameters for a carboxylic acid group are demonstrated.

L is the length of the substituent and B1-B4 are the radii of the group in different dimensions
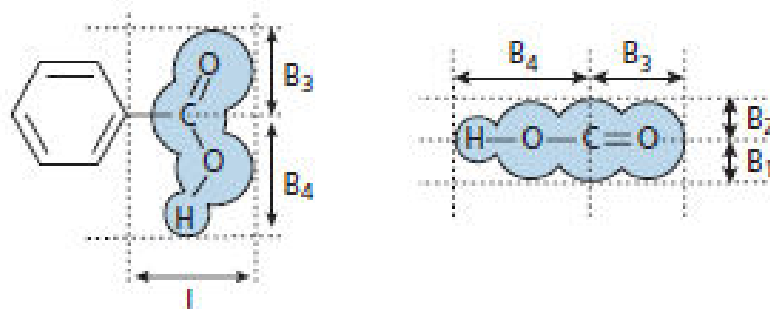
Verloop parameters for a carboxylic acid group.

Navneet F. Chauhan, Ph.D.
Saraswati Institute of Pharmaceutical Sciences

QUANTATIVE MODEL

Navneet F. Chauhan, Ph.D.
Saraswati Institute of Pharmaceutical Sciences

LFER (LINEAR FREE ENERGY APPROACH)/EXTRA THERMODYNAMIC APPROACH

WHAT IS HANSCH ANALYSIS ?

Quantitative relationship between

⬇

Biological activity of a series of compounds and their physicochemical substituent &

⬇

Global parameters representing hydrophobic, electronic, steric and other effects calcuated by using multiple regression correlation methodology

Navneet F. Chauhan, Ph.D.
Saraswati Institute of Pharmaceutical Sciences

IMPORTANCE OF LIPOPHILICTY

Drugs diffusion into lipid bilayer structure which is depends on molecular Structure of drug

Increasing the lipophilicity of a lead compound results in an increase in biological activity at a targeted site

‣ 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.

- cytoplasm is made of fatty acids and membrane is made of glycolipids and phospholipids, they have two ends –

(i) Lipophilicity or Hydrophobic end:- steroids and hydrocarbons,

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

- Increasing hydrophobicity must give positive results but 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.

ASSUMPTIONS IN HANSCH ANALYSIS:-

1. Conformational changes takes place in target site can be ignored,
2. Metabolism doesn't interferes in it,
3. Linear free energy terms are relevant to receptor's affinity and additive in nature,
4. Relationship between potency and lipophilicity is linear or parabolic,
5. 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.

Navneet F. Chauhan, Ph.D.
Saraswati Institute of Pharmaceutical Sciences

LFER

The most commonly used physichochemical parameters for Hansch analysis are log P, π, σ and stearic parameters as practically all parameters used in Hansch analysis are linear free energy related i.e. derived from equilibrium constant so it is known as "linear free energy approach" or "extra thermodynamic approach"

Navneet F. Chauhan, Ph.D.
Saraswati Institute of Pharmaceutical Sciences

LINEAR HANSCH MODEL

- The correlation of biological activity with physicochemical properties is often termed an Extrathermodynamic relationship.

- Because it follows the line of **Hammett and Taft equations** that correlate thermodynamic and related parameters.

- 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.

- Hansch 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 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'.

- They are found to be additive in nature independently. Therefore, it has given rise to the underlying linear Hansch equation also called extrathermodynamic approach.

$$\log 1/C = a \log P + b \sigma + c ES + d \quad (1)$$

Where,

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

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

ES = Taft constant (for steric effects),

σ = Hammett substitution constant

a,b,c,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.

APPLICATIONS

❑ To predict the activity of an as yet unsynthesised compound.

❑ To give an indication of the importance of influence of parameters on mechanism by which drug acts.

❑ Applied to various problems in order to correlate the biological activity with chemical structure.

❑ Hansch analysis serves as guide in future testing and synthesis of new compound and to play roles of hydrophobic, electronic and stearic factors in drug receptor interaction

ADVANTAGES OF HANSCH ANALYSIS

A) Use of descriptors (π, σ, E_s etc.) from small organic molecules may be applied to biological systems.

B) Predictions are quantitative and may be evaluated statistically.

C) Quick and easy.

D) Potential extrapolation: conclusions reached *may* be extended to chemical substituent's not included in the original analysis.

DISADVANTAGES OF HANSCH ANALYSIS

A) Descriptors required for substituent's being studied.

B) Large number of compounds required (training set for which physicochemical parameters and biological activity is available).

C) Limitations associated with using small molecule descriptors, such as steric factors, on biological systems (i.e. descriptors from physical chemistry).

D) Partial protontation of drugs at physiological conditions (can be included in mathematical model).

E) Extrapolations beyond the values of descriptors used in the study are limited.

F) Correlation between physical descriptors. For example, the hydrophobicity will have some correlation with the size and, thus, the Taft steric term.

Free Wilson Analysis

🍀 The Free-Wilson approach is truly a structure-activity based methodology in which the biological activity of the parent structure is measured and compared with the activity of analogues bearing different substituents

Parent structure

Activity of analogues bearing different substituent's

Biological activity measured and compared with

Navneet F. Chauhan, Ph.D.
Saraswati Institute of Pharmaceutical Sciences
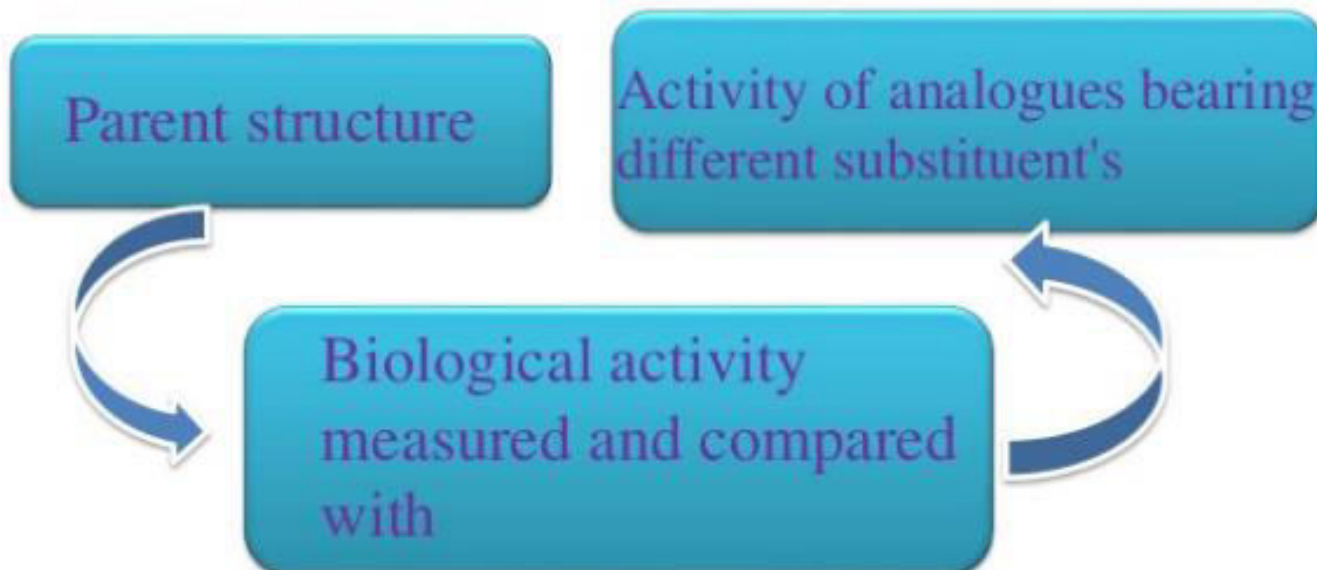
In 1964, Free and Wilson derived a mathematical model that describes the presence and absence of certain structural features, i.e. those groups that are represented by the values 1 or 0 that correlate the resulting structural matrix with biological activity values

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

- Free-Wilson analysis is a regression technique using the presence or absence of substituents or groups as the only molecule descriptors in correlations with biological activity

- It is the only numerical method which directly relates structural features with biological properties, in contrast to Hansch analysis, where physicochemical properties are correlated with biological activity values

It is represented by equation

$$BA = \Sigma a_i x_i + \mu$$

Where,

BA is the biological activity,

μ is the activity contribution of reference compound,

a i is the biological activity group contributions of the substituents X1, X2,…Xi in the different positions P of compound,

x i denotes the presence (x i = 1) or absence (x i = 0) of particular structural fragment

Drawbacks of Free Wilson analysis

At least two different positions of substitution must be chemically modified

Predictions can only be made for new combinations of substituent's already included in the analysis

Single point determinations obscure the statistical results

mostly centered on the large number of parameters and subsequent loss of the statistical degree of freedom.