SCIENCEFORECAST
A digital science library

Review Article
Published: 28 Aug, 2019

SF Journal of Pharmaceutical and Analytical Chemistry

# Quantitative Structure Activity Relationship in Drug Design: An Overview

*Kapoor Y[1] and Kumar K[1,2]**

[1]School of Pharmaceutical Sciences, Apeejay Stya University, Sohna-Palwal Road, Sohna, Gurgaon, Haryana, India

[2]School of Pharmacy and Technology Management, SVKM's NMIMS, Hyderabad, Telangana, India

## Abstract

Quantitative Structure-Activity Relationships (QSARs) have been used in developing relationships between biological activities and physicochemical properties of chemical substances in a way to develop a model consistent for activities prediction of new chemical entities. Much of the gained applicability for QSAR concept is a consequence of correlating molecular information with not only biological activities but also with other physicochemical properties, which has therefore been termed as QSPR (quantitative structure-property relationship). Typically, QSAR has been used for drug discovery and development. 3D-QSAR has also emerged as a natural extension to the classical approaches (Free-Wilson and Hansch analysis), which exploits the three-dimensional properties of the ligands to predict their biological activities. This article reviews the processes required for QSAR model development and intend to cover the crucial techniques and concepts relevant for performing QSAR studies.

**Keywords: QSAR; Physicochemical parameters; Classical approaches; Rational approaches; Drug designing; Combinatorial chemistry**

## Introduction

QSAR is an attempt for drug designing by establishing a mathematical relationship in the form of an equation between measurable physicochemical parameters, such as lipophilicity, shape and size of the molecule, biological activity and electron distribution or electronic effects within the molecule, which are believed to influence the drug's activity. The QSAR technique aims to develop correlations between any biological activity and their properties. However, in its general form QSAR has been adapted to cover correlations between structure and chemical reactivity. The properties and activities ideally are connected by some known mathematical function (F). Broadly physicochemical properties can be classified into three general types such as steric, hydrophobic and electronic properties of biologically active molecules, for which a vast range of physicochemical parameters and properties have been defined. The selected parameters should ideally be orthogonal, i.e., has minimal covariance. The function or relationship is usually a mathematical expression derived by statistical and related techniques, e.g., MLR (Multiple Linear Regressions). The parameters describing biological activities are used as dependent variables and physicochemical properties as independent variables.

## Physicochemical Properties

### Lipophilic parameters

Lipophilicity is one of the most studied physicochemical properties. Methods for lipophilicity have been in use for many years. An extensive set of published lipophilicity values for drugs and other compounds has been compiled by Hansch et al [1]. Reliable and inexpensive in silico lipophilicity tools are commonly used in drug discovery.

Partition coefficients (*P*): The partition coefficient is the measure of the lipophilicity of a drug and an indication of its ability to cross the cell membrane. It is defined as the ratio between un-ionized drugs distributed between the organic and aqueous layers at equilibrium. Drugs with high partition-coefficient value can easily permeate through biological membrane. The diffusion of drug molecules across rate-controlling membrane or through the matrix system essentially relies on the partition-coefficient. Drugs having lower partition-coefficient value are not suitable for translating in to oral controlled release formulations and drugs that have higher partition-coefficient are also

OPEN ACCESS

*Correspondence:
Kapil Kumar, School of Pharmacy and Technology Management, SVKM's NMIMS, Hyderabad, Telangana, India.
E-mail: kapil.py@gmail.com

Received Date: 31 Jul 2019
Accepted Date: 21 Aug 2019
Published Date: 28 Aug 2019

Citation: Kapoor Y, Kumar K. Quantitative Structure Activity Relationship in Drug Design: An Overview. SF J Pharm Anal Chem. 2019; 2(2): 1017.

ISSN 2643-8178

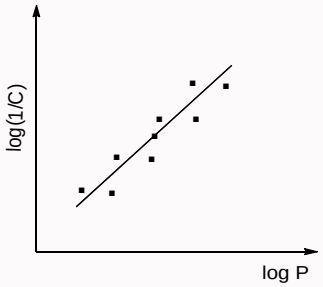



**Figure 1:** Proximity of data points to line of best fit.

poor candidates for oral controlled formulations. A drug in order to reach its site of action has to pass through a number of biological membranes. Consequently, *P* was the obvious parameter to use as a measure of the movement of the drug through these membranes. For the compounds used the nature of relationship obtained depends on the range of *P* values. In case of small range, by using regression analysis the results may be expressed as a straight line equation (equation 1). This equation marks an indication of a linear relationship between the partition coefficient of the drug and its activity.

$$\text{Log}(1/C) = K_1 \log P + K_2 \text{----------(1)}$$

Where,

C is the concentration of compound required to produce a standard response in a given time.

log*P* is the logarithm of the molecules partition coefficient between 1-Octanol and water.

$K_1$ and $K_2$ are constants.

(a) Regression analysis: It is a group of mathematical methods used to obtain mathematical equations relating different sets of data that have been calculated using theoretical considerations or obtained from experimental work. Into a suitable computer program, the data are fed, which produces an equation on execution that represents the line best fit for those data. E.g., assume that an investigation indicated that the relationship between the partition coefficients and activity of a number of related compounds appeared to be linear (Figure 1). Mathematically, these data could be represented in the form of straight line equation (y = mx + c). RA (Regression Analysis) would calculate the values of m and c that gave the line of best fit to the data.

RA does not indicate the spread and accuracy of the data. Consequently, they are normally accompanied by additional data, which at least should include the standard deviation of the observations(s), the regression coefficient(r) and the number of observations used (n).

The value of the 'r' (0-1) is a measure of closeness for the data with the equation, and a value r =1 indicates an ideal match, a value of r>0.9 are usually regarded as representing an acceptable degree of accuracy, provided that they are obtained using a reasonable number of results with a suitable standard deviation. The value of $100r^2$ is a measure of the percentage of the data that can be satisfactorily explained by RA. E.g., r = 0.90 indicates that 81% of the results can be satisfactorily explained by RA using the specified parameters. It indicates that 19% of the data are not satisfactorily explained by these parameters and so indicates that the use of an additional parameter might give a more acceptable account of the results. Assume, e.g., RA using an extra parameter gave a regression coefficient of 0.98 indicating that

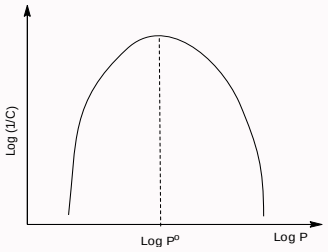



**Figure 2:** Parabolic curve obtained for log(1/C) versus log *P*.

96.04% of the data are now satisfactorily accounted for by the chosen parameters.

Partition coefficient: $\text{Log}(1/C) = K_1 \log P + K_2$

Example:

1. Miscellaneous neutral molecule binding with bovine serum:

$\text{Log}(1/C) = 0.75 \log P + 2.30$

r = 0.92, n = 42, s = 0.087

2. Alcohols toxicity to red spiders:

$\text{Log}(1/C) = 0.69 \log P + 0.16$

r = 0.979, n = 14, s = 0.087

Over larger ranges of P values the graph (Figure 2) of log(1/C) against log*P* has a parabolic form with a maximum value ($\log P^o$) [1(a)].

This maximum value existence implies that for maximum biological activity there is an optimum balance between lipid and aqueous solubility. The drug will be reluctant to enter the membrane below this, whilst will be reluctant to leave the membrane above this. Log$P^o$ represents the optimum partition coefficient for biological activity. This indicates that entities with *P* close to this optimum value are expected to be the most active and worth further investigation. Hansch et al., [2] showed that many of these parabolic relationships could be represented reasonably accurately by equation 2. The values for constants $k_1$, $k_2$ and $k_3$ in equation 2 are usually obtained either by RA or other statistical approaches. E.g., a study involving hypnosis inducement in mice using a series of barbiturates showed that the correlation could be expressed by the equation 3.

$$\text{Log}(1/C) = -K_1(\log P)^2 + K_2 \log P + K_3 \text{-------(2)}$$

$$\text{Log}(1/C) = -0.44(\log P)^2 + 1.5 \log P + 1.93\ (r = 0.969) \text{-------(3)}$$

This equation 3 has a maximum log$P^o$ at about 2.0. Hansch et al., [2] discovered that series of non-specific hypnotic drugs having extensively different types of structure were found to have log*P* values around 2. This indicates that it is not their structure that is the major factor in controlling their activity but the solubility of these different drugs in the membrane. On the basis of these and other partition studies, Hansch [3] suggested in the mid-1960s that any organic compound with a log*P* value of approximately 2 would be rapidly transported into the CNS, provided it have some hypnotic properties, and was not metabolized or eliminated. The fact that thiobarbiturates have log*P* values of about 3.1 implies that these drugs probably have a different site of action from those of the barbiturates. The greater value also suggests that a more lipophilicity of receptor is involved.

Change in log*P* monitoring has been used in drug design. E.g.,

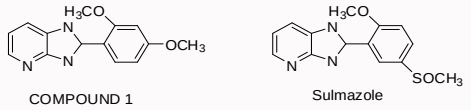



**Figure 3:** Cardiotonic agents.

compound 1 (logP = 2.57) is a cardiotonic agent however, in some patients its use resulted in unnecessary CNS side effects. Alteration of this compound by the replacement of methoxy group with more hydrophilic but approximately same sized hydrophilic methyl sulphone residue provide the cardiotonic drug (sulmazole) with a value log*P* of 1.17 (Figure 3). With the most extensive data base the n-octanol/water system is commonly chosen solvent system. However, if the organic phase is matched to the area of biological activity being studied the more accurate results can be obtained. E.g., for drugs absorbed in the gastro-intestinal tractn-octanol usually gives the most consistent results, whilst solvents with less polarity such as olive oil commonly give more consistent correlations for drugs crossing the BBB (Blood Brain Barrier).

Lipophilic substituents constants ($\pi$): These are also known as hydrophobic substituent's constants. They were defined by Hansch [3] and co-workers by the equation 4. The parameter $\pi$, which is the relative hydrophobicity of a substituent, is defined as

$$\pi = \log P_X - \log P_H \text{---------- (4)}$$

Where,

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

This is a substituent constant denoting the difference in hydrophobicity between a parent compound and a substituted analog and is usually replaced by the more general molecular term the log of the 1-octanol/water partition coefficients, $\log K_{ow}$ or log*P*.

E.g., the value of $\pi$ for the chloro group of chlorobenzene could be calculated from the *P* for chlorobenzene and benzene in the octanol/water system and substituting the appropriate $\pi$ values.

$$\pi_{Cl} = \log P\ (C_6H_5Cl) - \log P\ (C_6H_6)$$

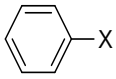
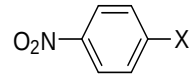
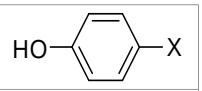
$$\pi_{Cl} = 2.84 - 2.13$$

$$\pi_{Cl} = 0.71$$

Depending on the solvent system used to determine the partition coefficients the values of $\pi$ will vary. However, most $\pi$ values are determined using the n-octanol/water system. A negative $\pi$ value shows that the substituent has a lower lipophilicity than hydrogen. As a result, the concentration of the compound in the aqueous media of biological systems possibly increases. Conversely, a positive value of $\pi$ indicates that a substituent has a higher lipophilicity than hydrogen and so will probably increase the concentration of the compound in the n-octanol layer and by inference its concentration in the lipid material of biological systems [1]. The $\pi$ value for a specific substituent will vary with its structural environment (Table 1).

Consequently, in determining activity relationships average values or the values relevant to the type of structures being investigated may be used. Where several substituents are present, the value of $\pi$ for the compound is the sum of the $\pi$ values of each of the separate substituents equation 5. When dealing with a series of analogues in which only the substituents are different, lipophilic substituents

**Table 1:** Examples of the variations of $\pi$ values with structure.

| X | R-X |  |  |  |
|---|---|---|---|---|
| H | 0.0 | 0.00 | 0.00 | 0.00 |
| $CH_3$ | 0.5 | 0.56 | 0.52 | 0.49 |

constants can be used as an alternative to the partition coefficient. For each of the analogues this usage is based on the assumption that the lipophilic effect of the unchanged part of the structure is similar. Consequently, to the lipophilicity of the molecule the $\pi$ values of substituents indicate the significance of the contribution of that substituent. Furthermore, in determining the lipophilic character of the drug, biological activity-$\pi$ relationships that have low standard deviations and high regression constants are important.

$\pi = \pi$ (substituted 1) + $\pi$ (substituted 2) +……. $\pi$ (substituted n) ---------- (5)

Lipophilic substituent constants can also be used to calculate theoretical partition coefficients using equation for whole molecules. E.g., for 1, 3-dimethylbenzene the calculated partition coefficient would be given by:

$$\pi = \log P_{\text{1,3 di methyl benzene}} - \log P_{\text{benzene}} \text{------- (6)}$$

However, for compounds containing more than one substituent the value of $\pi$ is the sum of the $\pi$ values for each of the substituent groups in equation 5 and so using the $\pi$ value given for the methyl group in methylbenzene given in Table 1, the value of $\pi$ in equation 6 for the 2 methyl groups of 1,3-dimethylbenzene is:

$$\pi = (2.13 + 1.12) - 2.13$$

$$= 3.25 - 2.3$$

$$= 1.12$$

In order to distinguish these calculated values from experimentally determined *P* values, they sometimes are referred to as C log*P*. Provided that the substituents are not satirically crowded they are often in good agreement with the experimentally determined values. However, poorer agreements often result when substituents are located close to each other in the molecule. Strong electron interactions between substituents groups in addition, can result in less accurate theoretical *P* values.

Distribution coefficients: Ionization of many compounds in aqueous solution is not taken into account by *P* values. The extent of this ionization will have a significant effect on the absorption and distribution of these drugs [4]. Consequently, in Hansch and other forms of mathematical analyses of drug behaviour, the lipophilicity of ionisable compound is often represented by the value of their distribution coefficients (D), which is defined as the ratio of the concentrations of the unionized and the ionized compound between an organic solvent and an aqueous medium. E.g., the distribution coefficient of the acid HA is given by:

$$D = [HA]_{organic}/[H^+_{(aq)}/A^-_{(aq)}] \text{----------(7)}$$

Since pH of the aqueous medium is the deciding factor for the ionization of acids and bases it can be shown using equations 8 and 9.

For acids: $\text{Log}(P/D\text{-}1) = \text{pH} - \text{pKa}$--------- (8)

For bases; $\text{Log}(P/D\text{-}1) = \text{pKa} - \text{pH}$----------(9)

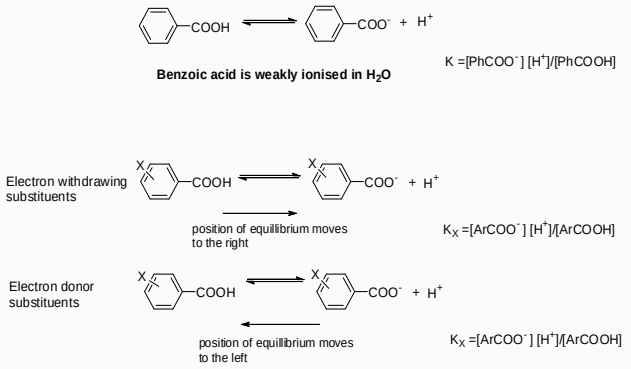



**Figure 4:** Effect of electron withdrawing and donor groups on the position of equilibrium of substituted benzoic acids.

Provided that the pKa and the value of *P* are known for the same solvent system these equations allow the calculation of the effective lipophilicity of a compound at any pH. Log D values usually use distribution coefficients, a large number of which are available from databases.

## Electronic parameters

The electrons distribution in a drug molecule will have a considerable influence on the activity and distribution of a drug. A drug normally has to pass through a number of biological membranes in order to reach its target. Generally, polar and non-polar drugs in their unionized form are usually more readily transported through membranes than polar drugs and drugs in their ionized forms. Furthermore, the electronic distribution in drug structure will control the type of bonds it forms with the target, once it reaches the site of action, which in turn affects its biological activity.

The Hammett constant ($\sigma$): The Hammett equation relates observed changes in equilibrium or rate constants to systematic changes in substituents that govern electron donating/withdrawing ability. The nature of the electron donating and withdrawing groups present in the structure determines the electronic distribution within a molecule. Hammett quantified the effect of substituents on any reaction by defining an empirical electronic substituent parameter ($\sigma$), which is derived from the acidity constants, Ka's of substituted benzoic acids (Figure 4). When an electron withdrawing substituent (-X), such as a nitro group substitutes ring hydrogen, it results in carboxylate anion stabilization and the O-H bond weakening of the carboxyl group. This shifts the equilibrium to the right indicating that the substituted compound is a stronger acid than benzoic acid ($K_x$>K). It also indicates that at equilibrium as anions more of the nitrobenzoic acid will exist, which could hinder its membrane transfer more than that of the less ionized weaker benzoic acid. Conversely, the introduction of an electron donor substituent (-X), such as a methyl group into the ring strengthens the acidic OH group and reduces the carboxylate anion stability. This shifts equilibrium to the left, indicating that the compound is a weaker acid than benzoic acid (K>$K_x$). This in turn indicates that in solution at equilibrium it has fewer anions than benzoic acid and so could cross membranes more easily than benzoic acid. In addition to the effect that changes in the electron distribution have on membranes transfer, they also will affect the binding of these acids to a target site. These observations indicate the possibility of using equilibrium constants to compare the electronic distributions of structurally identical compounds. Hammett used equilibrium constants to study the relationship between the acid strength and structure of aromatic acids.

**Table 2:** Different electronic substitution constants used in QSAR studies.

| Substituent | Hammett constants | | Inductive constants[a] | Taft constants[b] | Swain-Lupton constants[c] | |
|---|---|---|---|---|---|---|
| | $\sigma_m$ | $\sigma_P$ | $\sigma_1$ | $\sigma^*$ | F | R |
| H | 0 | 0 | 0 | 0.49 | 0 | 0 |
| $CH_3$ | -0.07 | -0.17 | -0.05 | 0 | 0.04 | 0.13 |
| OH | 0.12 | -0.37 | 0.25 | - | 0.29 | -0.64 |

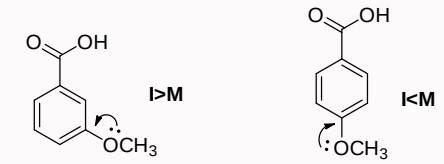



**Figure 5:** The resonance (mesomeric) and inductive effects of m-&p-methoxybenzoic acids.

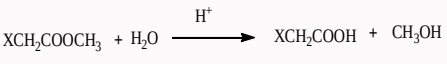



**Figure 5a:** *α*-substituted methyl ethanoates hydrolysis.

In the course of study, calculated constants for a variety of ring substituents (X) of benzoic acid known as Hammett constants or Hammett substitution constants ($\sigma_x$), using this acid as the comparative reference standard as shown in Table 2.

Hammett constants ($\sigma_x$) can be defined as:

$\sigma_x = \log K_X/K$

i.e., $\sigma_x = \log K_x - \log K$

$\sigma_x = pK - pK_x$ [as $pK_a = -\log K_a$]

Since K>>$K_x$ a negative value for $\sigma_x$ indicates that the substituent is acting as an electron donor group. Conversely, as K<$K_x$ a positive value shows that the substituent is acting as an electron withdrawing group. With the position of the substituent in the molecule the value of $\sigma_x$ varies. Usually, by using the subscripts o,m and p this position is indicated. Depending on its position on the ring where a substituent has opposite signs it means that in one case it is acting as an electron withdrawing and as an electron donor group in the other. This is possible because the Hammett constant includes both the mesmeric (resonance) and inductive contributions to the electron distribution. E.g., for the methoxy group of m-methoxybenzoic acid the $\sigma_m$ is 0.12 whilst it is -0.27 for *p*-methoxybenzoic acid (Figure 5). In the former case, the electronic distribution is dominated by the inductive (I or F) contribution whilst it is controlled by the Mesomeric (M) or Resonance effect (R) in the latter case.

Since electron distribution is not the only factor involved, attempt to relate biological activity to the values of Hammett substitution and similar constants have been largely unsuccessful. However, investigation by Metcalf and Fukata was a successful attempt to relate biological activity to structure using Hammett constant into the effectiveness of diethyl aryl phosphates for killing fruit flies. This investigation showed dependency for the activity of these compounds on electronic distribution factors. Their results may be expressed by the relationship in equation 10. This equation shows that, greater the

positive value for σ, greater the biological activity of the analogue. This type of knowledge facilitates one to predict the activities of analogues, testing and synthesizing all the possible analogues.

Log (1/C) = 2.282σ-0.348------- (10)

## Steric factor

The size, shape, and bulk of a drug will influence the ease with which it can approach and interact with a target or binding site. A bulky substituent may act like a hinder or shield for 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 better activity. Steric properties are more difficult to quantify than electronic or hydrophobic properties. Several methods have been opted, of which few are described hereunder.

Taft's steric factor ($E_s$): In order to define steric parameter, Taft in 1956 used the relative rate constants of the acid-catalysed hydrolysis of α-substituted methyl ethanoates (Figure 5a) because it had been shown that the rates of these hydrolyse were almost entirely dependent on steric factors. With the use of methyl ethanoate as standard he defined $E_s$ as;

$E_s = \log K_x - \log K_o$

Where,

$K_o$ represent the hydrolysisrate of parent ester.

$K_x$ represent the hydrolysis rate of substituted ester.

It is assumed that the values for Es obtained for a group using the hydrolysis data are applicable to other structures containing that group.

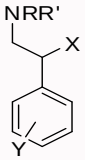
By adding -1.24 to the corresponding methyl-based values the methyl- based $E_s$ values can be converted to H-based values. In a number of investigations Taft steric parameters have been found to be useful. E.g., regression analysis has shown by equation 11 that the antihistaminic effect of a number of related analogues of diphenhydramine was related to their Biological Response (BR), where Esin the most highly substituted ring is the sum of Meta- and ortho-Es values. Regression analysis also showed the relation between biological response and Hammett constant by the relationship in equation 12.

LogBR = 0.440 $E_s$-2.204 (n = 30, s = 0.307, r = 0.886) ---------- (11)

LogBR = 2.814σ-0.233 (n = 30, s = 0.519, r = 0.629) ---------- (12)

The standard deviations (s) comparison for both the equations 11 and 12 shows that for each of the analogues the calculated values for the corresponding Taft $E_s$ values are less scattered than Hammett constants σ values. Furthermore, although both the s and r values for the former equation 12 are reasonable, those for the latter equation are unacceptable. This indicates that the antihistamine activity of these analogues appears to rely majorly on steric than electronic effects. This deduction is supported by the fact that using regression analysis to obtain a relationship involving both the Taft and Hammett constants does not lead to a significant increase in the s and r values. Taft constant encounters the disadvantage that they are bound to be determined by experiment.

Molar Refractivity (MR): It is a measure of both the ease of polarization and volume of a compound. The refractive index term is a measure of the polarisability while the M/ρ term is a measure of the molar volume of the compound in equation (13).




$\log(1/C) = 1.22\pi - 1.59\sigma + 7.89$
$(n = 22, r^2 = 0.841, s = 0.238)$

**Figure 6:** β-halo-arylamines QSAR equation.

MR = $(n^2-1)$ M/ $(n^2+2)$ ρ ---------- (13)

Where,

n is the refractive index.

M is the relative mass.

ρ is the density of the compound.

Verloop steric parameter: Another approach to measure the steric factor involves a computer program called Sterimol, which calculates steric substituent values (Verloop steric parameters) from standard Van der Waals radii, bond length, bond angles, and possible conformations for the substituent. Unlike Es, the Verloop steric parameters can be measured for any substituent [5].

Other steric parameters: A wide variety of other parameters have been used to relate the steric nature of a drug on activity.

Hansch equation: A simple equation can be drawn up in a situation where biological activity is related to only one property. However, for most drugs the biological activity is related to a combination of properties. In such cases, only if other parameters are kept constant, simple equations involving only one parameter are relevant. This is not easy to achieve in reality and equations which relate biological activity to a number of different parameters are more common. These equations are known as Hansch equations and they usually relate biological activity to the most commonly used physicochemical properties (logP, π, σ and a steric factor). If the range of hydrophobicity values is limited to a small range then the equation 14 will be linear. If the log*P* values are spread over a large range, then the equation will be parabolic. The constants $K_1$-$K_5$ are determined by computer software in order to get the best fitting equation.

$\text{Log}(1/C) = K_1 \log P + K_2 \sigma + K_3 E_s + K_4$--------- (14)

$\text{Log}(1/C) = -K_1(\log P)^2 + K_2 \log P + K_3 \sigma + K_4 E_s + K_5$ ---------- (15)

Not all the parameters will necessarily be significant. For e.g.; the adrenergic blocking activity of β-halo-arylamines (Figure 6) was related to π and σ and did not include a steric factor. This equation 15 tells us that biological activity increases if the substituents have a positive π value and a negative σ value. In other words, the substituents should be electron donating and hydrophobic.

## The Craig plot

Although tables of π and σ factors are readily available for a large range of substituents, it is often easier to visualize the relative properties of different substituents by considering a plot where the y axis is the value of the σ factor and the x axis is the value of the π factor. Such a two dimensional plots of one parameter against another is known as a Craig plot as shown in Figure 7.

**Advantages:**

1. The plot shows clearly that there is no overall relationship

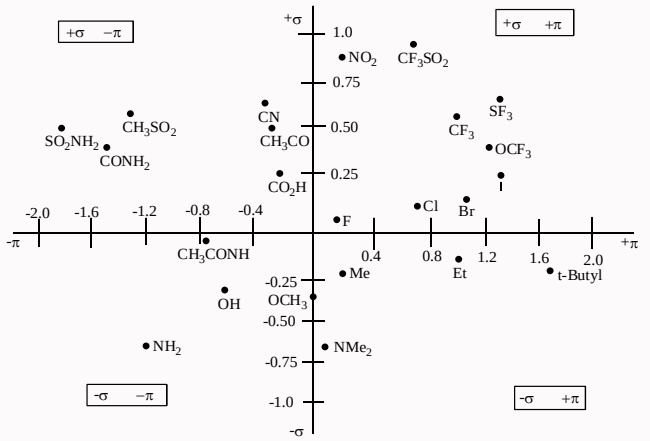



**Figure 7:** Craig plot for the *σ* and *π* factors of *para*- aromatic substituents.

between *σ* and *π*. The various substituents are scattered around all four quadrants of the plot.

2. At a glance it is possible to tell which substituents have positive, negative *σ* and *π* parameters, and which substituents have one positive and one negative parameter.

3. It is easy to observe which substituents have similar *π* values. E.g., the bromo, ethyl, trifluoromethyl and trifluoromethyl sulfonyl groups are all approximately on the same vertical line on the plot. In theory, these groups could be interchangeable on drugs where the principal factor affecting biological activity is the *π* factor. Similarly, groups which form a horizontal line can be identified as being isoelectronic or having similar *σ* values (e.g., Cl, Br, $CO_2H$, I).

4. The Craig plot is useful in planning which substituents should be used in a QSAR study. Analogues should be synthesized with substituents from each quadrant in order to derive the most accurate equation involving *π* and *σ*. E.g., halogen substituents are useful representatives of substituents with increased hydrophobicity and electron-withdrawing properties (+*π*, +*σ*), whereas an OH substituent has more hydrophilic and electron-donating properties (-*π*, -*σ*). Alkyl groups are examples of substituents with +*π*, and -*σ*, whereas acyl groups have -*π*, and +*σ* values.

5. Once the Hansch equation has been derived, it will show whether *π* or *σ* should be negative or positive in order to get good biological activity. Further developments would then concentrate on substituents from the relevant quadrant. E.g., if the equation shows that +*π* and +*σ* values are necessary, then further substituents should only be taken from the top right quadrant.

Drug design involves the development of analogues and prod rug by some chemical modifications from the lead molecule or from a parent compound by modifying the carbon skeletal transformation or by the synthesis of compounds of the same nucleus with various substitutions. Analogues can also be synthesized by changing the position of substitution group. E.g., synthesis of trans-diethylstilboestrol by the modification of oestradiol produced better oestrogenic activity than the latter one.

## Rational Approach to Drug Design

There are many approaches to drug designing in relation with physicochemical parameters and electronic features taken into consideration for designing a drug. These are as follows:

1. Approach with quantum mechanics: This, also called as wave mechanics, comprises the fundamental physical properties of a molecule. These include the properties of protons, neutrons, and electrons, which are explained by quantum mechanics. The basis of drug molecule nature is altered by chemical alterations of the electronic features.

2. Approach with molecular orbital theory: This approach depicts the change in properties that shall be made by the alteration of orbits. Based on this, the electrons present in the molecules are linked with orbital's to change the electronic feature. The molecular orbital approach is the change on electronic charges, evidenced from the investigation of three volatile inhalation aesthetics, and also on molecular conformation, as studied with respect to Acetylcholine, in regard to bond length and angles including torsional angles. These interpretations are carried out by computational methods in respect to SAR.

3. Approach with molecular connectivity: This is based on the structural features of a molecule. All steric and electronic parameters vary according to their configuration. This includes cyclization, unsaturation, presence of heteroatom, skeletal branching, and position in molecules with the aid of numerical indices and series of functional attachments.

4. Approach of linear free-energy: Linear free energy was based on the selection of physicochemical parameters of a molecule with a specific biological activity. But the biological activity may vary in relation to the physicochemical properties of the drug or molecule and does not provide a prompt success, but it may reveal some beneficial features regarding the molecule.

## Discovery of Lead Compounds

The concept of lead discovery envisages two investigational processes i.e.,

1. Exploration of leads: search of new molecules.

2. Exploitation of leads: assessment, chemical modeling, and extension of leads.

### Approaches to lead discovery

The lead identification requires a series of biological evaluation of the lead molecules. Once, after the identification, it can be structurally modified, the potency and the activity are improved.

1. Random screening: The entire synthesized compounds or any chemical constituents obtained from natural products are evaluated in a series for their biologically active components. Thus, random screening may produce unexpected active medicines. Antibiotics, such as streptomycin, tetramycin, and fungal metabolites, such as lovastatin and cyclosporins, were found through this method. This approach needs more manpower, and it is expensive and time-consuming and the success rate is considerably low.

2. Non-random screening: In this method, only compounds that possess similar structural skeletons were evaluated from their particular properties.

3. Pharmacokinetic studies: Biotransformation occurs as the fate by metabolizing enzymes. In order to develop new leads, the metabolites or bio transformed compounds are studied for their properties, and such studies are expected to assess the activity from a comparison with the parent molecule. E.g., the discovery of

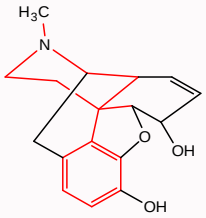



**Figure 8:** Pharmacophore of morphine.

sulphanilamide is reported through the metabolic studies of prontosil.

4. Pharmacodynamic studies: The effect apart from the therapeutic actions, that is, side effects may lead to the finding out of a new molecule with some appreciable structural modifications. E.g., sulphonamide used specifically for the treatment of typhoid, lowered the blood sugar levels drastically. This exerted action led to the finding of aryl sulphonyl thiourea moiety responsible for the lowering of blood glucose level. Amino alkyl derivatives of iminodibenzyl were synthesized as analgesic, sedative and antihistamines that were found to possess antidepressive action. This lead to the synthesis of many tricyclic antidepressants.

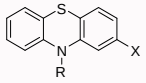
## Pharmacophore identification

Drugs interaction with receptors, known as pharmaco dynamics, is very specific. Therefore, only a small part of the lead compound may be involved in the appropriate receptor interactions. The relevant groups on a molecule that interact with a receptor and are responsible for the activity are collectively known as the pharmacophore. The other atoms in the lead molecule, sometimes referred to as the auxophore, may be extraneous. By determining the auxophoric and pharmacophoric groups on the lead compound, and of the auxophoric groups, which are interfering with lead compound binding and which are not detrimental to binding, it can be known which groups must be excised and which one can retain/modify as needed. One approach in lead modification to help make this determination is to cut away sections of the lead molecule and measure the effect of those modifications on potency. E.g., assume that the addictive analgesics morphine, codeine, and heroin are the lead compounds, and one wants to know which groups are auxophoric and which are pharmacophoric. The morphine family of analgesics binds to the $\mu$ opioid receptors [6] as illustrated in Figure 8. A decrease in potency on removal of a group will suggest that it may have been pharmacophoric, an increase in potency means it was auxophoric and interfering with proper binding, and essentially no change in potency will mean that it is auxophoric but not interfering with binding.

Excision of the dihydrofuran O atom (not in the pharmacophore) gives morphinan [7], the hydroxyanalogue, levorphanol [8], is 3-4 times more potent than morphine but it retains the addictive properties. Removal of half of the cyclohexene ring (also not in the pharmacophore), leaving only methyl substituents gives benzomorphan [9]. This compound shows some separation of analgesic and addictive effects. Pentazocine hydrochloride is less potent than morphine, but has a much lower addiction. Cutting away the methylene group of the cyclohexane fused ring also, has little effect on the analgesic effect in animal's tests.

## Structure modification

Homologation: A homologous series is a group of compounds




**Trimeprazine** (R = $CH_2CH(CH_3)CH_2N(CH_3)_2$)
**Promethazine** (R = $CH_2CH(CH_3)N(CH_3)_2$)
**Promazine** (R = $CH_2CH_2CH_2N(CH_3)_2$)

**Figure 9:**10-aminoalkylphenothiazine analogue.

that differ by a constant unit, generally a -$CH_2$ group. For many series of compounds, lengthening of a saturated C side chain from one (methyl) to 5 to 9 atoms (pentyl to nonyl) produces an increase in pharmacological effects; further lengthening results in a sudden decrease in potency. This phenomenon corresponds to increased lipophilicity of the molecule to permit penetration into cell membranes. In the case of aliphatic amines, another problem is micelle formation, which begins at about $C_{12}$. This effectively removes the compound from potential interaction with the appropriate receptors. According to Richardson [10], who was investigating the hypnotic activity of alcohols, the maximum effect occurred for 1-hexanol to 1-octanol; then the potency declined on chain lengthening until no activity was observed for hexadecanol.

Chain branching: The chain branching lowers the potency of a compound because a branched alkyl chain is less lipophilic than the corresponding straight alkyl chain as a result of larger molar volumes and shape of branched compounds. Major pharmacological changes can occur with chain branching or homologation. Consider the 10-aminoalkylphenothiazines (Figure 9, X=H). When R is $CH_2CH(CH_3)N(CH_3)_2$(promethazine HCl), antispasmodic and antihistaminic activities predominate. However, the straight-chain analogue with R being $CH_2CH_2CH_2N(CH_3)_2$ (promazine) has greatly reduced antispasmodic and antihistaminic activities, but sedative and tranquilizing activities are greatly enhanced. In the case of the branched chain analogue with R = $CH_2CH(CH_3)CH_2N(CH_3)_2$ (trimeprazine) (next larger branched-chain homolog), the tranquilizing activity is reduced and antipruritic (anti-itch) activity increases. This indicates that multiple receptors are involved and branching or homologation can cause the molecule to bind more or less well to the receptors responsible for antispasmodic activity, antihistamine activity, tranquilizing activity or antipruritic activity.

Ring-Chain transformations: Another modification that can be made is the transformation of alkyl substituents into cyclic analogues, which often doesn't affect potency greatly. If the dimethylamino group of the tranquilizer chlorpromazine is substituted by a methylpiperazine ring, antiemetic (prevents nausea and vomiting) activity is greatly enhanced.

Bioisosterism: Bioisosteres are substituents or groups that have chemical/physical similarities, and which produce broadly similar biological properties [11]. Bioisosterism is an important lead modification approach that has been shown to be useful to attenuate toxicity or to modify the activity of a lead, and may have a significant role in the alteration of pharmacokinetics of a lead.

**Classification of Bioisosterism:**

A. Classical isosteres [12,13]

Grimm (1925) [14,15] formulated the hydride displacement law to describe similarities between the groups that have the same number of valence electrons, but may have a different number of atoms.

Erlenmeyer later redefined isosteres as atoms, ions, or molecules in which the peripheral layers of electrons can be considered to be identical. These 2 definitions describe classical isosteres.

B. Nonclassical Bioisosteres [16]

They do not have the same number of atoms and do not fit the steric and electronic rules of the classical isosteres, but do produce similar biological activity.

## Prodrug Approach

A pro-drug is a chemically modified inert precursor of the drug that on biotransformation liberates the pharmacologically active parent compound. A pro-drug is also called as pro-agent, bio-reversible derivative or latentiated drug. The design of pro-drug approach is also called as drug latentiation.

### Ideal properties

1. It should not have intrinsic pharmacologic activity.

2. Drug and the carrier linkage must be cleared *in vivo*.

3. The metabolic fragments, apart from the active drug, should be nontoxic.

4. It should rapidly transform, chemically or enzymatically, into the active form where desired.

### Classification

Depending on the constitution, lipophilicity, and method of bio-activation, prodrugs are classified into two categories.

1. Carrier-linked pro-drug/simple pro-drug: They are generally esters/amides. Carrier-linked pro-drugs are the ones where the active drug is covalently linked to an inert carrier / transport moiety. Such prodrug modifies the lipophilicity due to the attached carrier. The active drug is released by hydrolytic cleavage, either chemically/ enzymatically.

2. Bio-precursors: They are inert molecules obtained by a chemical modification of the active drugs, but do not contain a carrier. E.g., NSAID, sulindac, is inactive as sulphoxide and must be reduced metabolically to active sulphide.

**Classification of pro-drugs according to the functional group:**

1. Carboxylic acid and alcohols: Pro-drugs of carboxylic acid and alcohol functionalities can be prepared by conversion to esters. The esters can be easily hydrolyzed by esterase enzymes (e.g., lipase, ester hydrolase, cholesterol esterase, acetyl cholinesterase, and carboxy peptidase) present in plasma and other tissues to give active drug.

2. Amines: Due to high stability and lack of amidase enzyme necessary for hydrolysis, the conversion of amines to amide as a pro-drug is not been used for most of the drugs. A more common approach adopted is to use Mannich bases as pro-drug form of amines. Hetacillin is a prodrug form of ampicillin in which amide N and $\alpha$ amino functionalities have been allowed to react with acetone to give a Mannich base (imidazolidine ring system). This leads to decrease in the basicity and increase in the lipophilicity and absorption (Figure 10).

3. Azo linkage: Pro-drugs of amines are occasionally prepared by incorporating them in to an azo linkage. By the action of azo reductase, the amino compounds are released *in vivo*. Prontosil drug is inactive *in vitro*, but it is converted to sulphanilamide by azo reductase enzymesas shown in Figure 10.

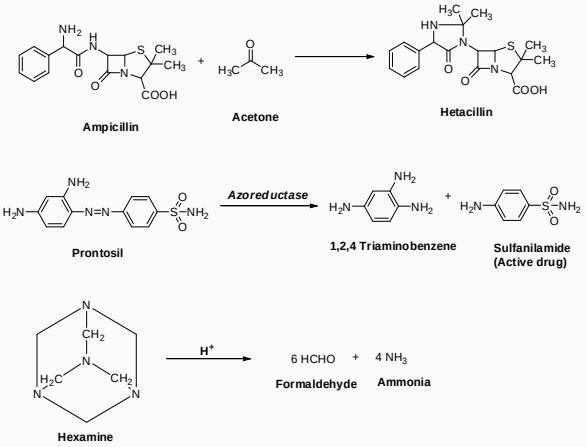
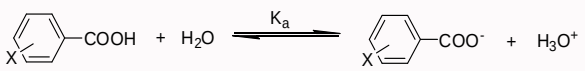



**Figure 10:** Activation of certain prodrug.




**Scheme 1:** Ionization of substituted benzoates.

4. Carbonyl moiety: Conversion of carbonyl functionalities, such as aldehyde and ketone, to pro-drug has not been found wide clinical use. They are converted into derivatives in which the $sp^2$ carbonyl C is converted as $sp^3$ hybridized C attached to hetero atoms. These pro-drugs are re-converted to carbonyl compounds by hydrolysis. E.g., hexamine releases formaldehyde in the urine (acidic pH), which acts as an antibacterial agent (Figure 10).

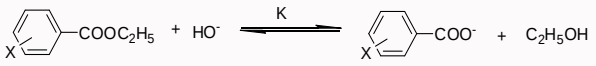
## Physicochemical Alterations

### Electronic effects: The Hammett equation

Hammett's postulate was that the electronic effects (both inductive and resonance effects) of a set of substituents should be similar for different organic reactions, by choosing benzoic acid as the standard system. In Scheme 1, it seems reasonable that as X becomes electron withdrawing, the equilibrium constant ($K_a$) should increase (the reaction should be favored to the right) because X is inductively pulling electron density from the carboxylic acid group, making it more acidic; it also is stabilizing the negative charge on the carboxylate group of the product. Conversely, when X is electron donating, the equilibrium constant should decrease.

A similar relationship should exist for a rate constant k, an electron-withdrawing substituent would stabilize a negative change in the transition state, thereby lowering the activation energy, and increasing the rate, and an electron donating group would destabilize the transition state, decreasing the rate. Scheme 2 was chosen by Hammett as the standard system to determine electronic effects of substituents on the rate constant of a reaction.

If $K_a$ is measured from scheme 1 and k from scheme 2 for a series of substituents X, and the data are expressed in a double logarithm plot, then a straight line can be drawn through most of the data points. This is known as a linear free-energy relationship. When X is meta/ para substituent, then virtually all of the points fall on the straight line; the ortho-substituent points are badly scattered. The straight line can be expressed as equation 16. The electronic parameter $\sigma$, depends on




**Scheme 2:** Saponification of substituted ethyl benzoates.

the electronic properties and position of the substituent on the ring and, therefore, is called the substituent constant. The more electron withdrawing a substituent, the more positive its $\sigma$ value (relative to H, which is set at 0.0): conversely, the more electron donating, the more negative its $\sigma$ value. A large $\rho$ value, either positive/negative, indicates great sensitivity to substituent effects. Reactions that are favoured by electron donation in the transition state (such as reactions that proceed *via* Carbocation intermediates) have negative $\rho$; reactions that are aided by electron withdrawal (such as reactions that proceed *via* carbanion intermediates) have positive $\rho$ values.

$$\text{Log K} = \rho \log K + C \text{ ---------- (16)}$$

### Lipophilicity effects

Hansch believed that, just as Hammett equation relates the electronic effects of substituents to reaction rates, there should be a linear free-energy relationship between lipophilicity and biological activity. As a measure of lipophilicity, Hansch proposed the partition coefficient, P, a measure of the solubility in 1-octanol versus water, and P was determined by equation 17 ($\alpha$ is the degree of dissociation of the compound in water calculated from ionization constant). A model for a drug traversing through the body to its site of action, the relative potency of the drug, expressed as log1/C, where C is the concentration of the drug that produces some standard biological effect, was related by Hansch to its lipophilicity by the parabolic equation 18.

$$P = [\text{compound}]_{\text{octanol}} / [\text{compound}]_{\text{aqueous}} (1-\alpha) \text{ --------- (17)}$$

$$\text{Log1/C} = -K (\log P)^2 + K' (\log P) + K'' \text{ ----------(18)}$$

On the basis of equation 17, it is apparent that if a compound is more soluble in water than in 1-octanol, $P<1$, and, therefore, log$P$ is negative. Conversely, a molecule more soluble in 1-octanol has a $P>1$, and the log$P$ is positive. Therefore, the more positive the log$P$, the more lipophilicity it is. The larger the value of $P$, the more there will be an interaction of the drug with the lipid phase (membranes). As $P$ approaches infinity, micelles will form and the drug interaction will become so great that the drug will not be able to cross the aqueous phase, and it will localize in the first lipophilic phase with which it comes into contact. As $P$ approaches 0, the drug will be too water soluble that it will not be capable of crossing the lipid phase and will localize in the aqueous phase. Somewhere between $P=0$ and $P=\infty$, there will be a value of P such that drugs having this value will be least hindered in their journey through macromolecules to their site of action. This value is called log$P_0$, the logarithm of the optimum partition coefficient for biological activity. This random walk analysis supports the parabolic relationship (equation 8) between potency (log1/C) and log$P$. An increase in the alkyl chain length increases the lipophilicity of the molecule; the log$P_0$ generally occurs in the range of 5-9 Carbon atoms. Hansch [17] found that a number of series of nonspecific hypnotics had similar log$P_0$ values, approximately 2, and they suggested that this is the value of log$P_0$ needed for penetration into the CNS, i.e., for crossing the blood-brain barrier. If a hypnotic agent has a log$P$ considerably different from 2, then its activity probably is derived from mechanisms other than just lipid transport. If a lead compound has modest CNS activity and has a log$P$ value of 0, it would be reasonable to synthesize an analogue with a higher log$P$.

### Effects of ionization on lipophilicity

What if the drug attempting to discover binds at an ionized site in the receptor, so ionization of the drug favours binding to the receptor, but ionization of the drug also blocks its ability to cross various membranes prior to reaching the receptor. However, because equilibrium is established between the neutral and ionized form of a molecule/group that depends on the pH of the medium and the $pK_a$ of the ionizable group it is possible to design a compound that is neutral when it needs to cross membranes, but ionized when it finally reaches the target receptor. When the pH of the medium equals the $pK_a$ of the molecule, half of the molecules are in the neutral form and half in the ionized form. The ones that are neutral may be able to cross membranes, but once on the other side, the equilibrium with the ionized form is re-established (the equilibrium mixture will again depend on the pH on the other side of the membrane), so there are now ionized molecules on the other side of the membrane that can interact with the target receptor. The ionized molecules that didn't cross the membrane also re-establish equilibrium and become a mixture of ionized and neutral molecules, so more neutral molecules can get across the membrane. If the equilibrium could be re-established indefinitely, eventually all of the molecules would cross the membranes and bind to the target receptor. To adjust the ionization equilibrium of the lead compound, it is necessary to add electron donating/electron withdrawing groups to vary the $pK_a$ of the molecule. Electron-donating groups will increase the $pK_a$, making acids less ionisable and bases more ionisable; the opposite holds for electron withdrawing groups.

## Molecular Modeling

In order to utilize molecular modeling for drug design a variety of approaches can be taken into account; when the structure of the target receptor is known, direct design approaches are used and when the receptor structure is not known, indirect design approaches are used. The basic premise in the utilization of molecular graphics is that the better the complementary fit of the drug to the receptor, the more potent the drug will be. This is the lock-and-key hypothesis of Fischer [18] in which the receptor is the lock into which the key (i.e., the drug) fits. The structure of the receptor (either X-ray crystal structure/NMR solution structure) should be known in order to apply this concept most effectively; different drug analogous can then be docked into the receptor. Docking is a molecular graphics term for the computer-assisted movement of a terminal-displayed molecule into its receptor.

The most effective use of molecular modeling is when a high-resolution crystal structure (or NMR solution structure) of a receptor with a ligand bound is available. Molecular graphics visualization of the electron density map of this complex may reveal empty pockets in the complex that could be filled by appropriate modification of a lead compound. An important example of structure-based drug design is the discovery of zanamivir, an antiviral agent used against influenza A and B infections [19]. Sialic acid residues on receptors at the host cell surface binds to the hemagglutinin at the surface of the virus. The virus then replicates in the nucleus after it enters the cell. The progeny virus particles escape the cell and stick to the sialic acid residues on the cell surface as well as to each other. Neuraminidase (also known as sialidase) is a key viral surface enzyme that catalyses the cleavage of terminal sialic acid residues from the cell surface, which releases the virus particles to spread into the respiratory tract and infect new cells. The important feature of this enzyme that made it an attractive

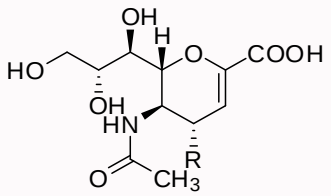



**Figure 11:** Zanamivir analogue.

target for drug design is that its active site is lined with amino acids that are invariant in neuraminidases of all known strains of influenza A and B. Therefore, inhibition of this enzyme should be effective against all strains of influenza A and B. Random screening didn't produce any potent inhibitors of the enzyme, although a nonselective neuraminidase inhibitor was identified. The breakthrough came when the crystal structures of the influenza a neuraminidase with inhibitors bound were obtained [20-23]. The active site of the enzyme (Figure 11, R=OH) bound was probed computationally using Good ford's GRID program. Predictions by GRID of energetically favorable substitutions suggested replacement of the 4-hydroxyl group of Figure 11(R=OH) by an amino group (R=$NH_2$), which when protonated would form a favorable electrostatic interaction.

Kuntz et al., [24] reported an algorithm called DOCK that was designed to fit small molecules into their macromolecular receptors for lead discovery. This shape matching method, which was originally restricted to rigid ligands (receptor-bound molecules) and receptors, was modified for flexible ligands where a ligand is approximated as a small set of rigid fragments. Ideally, a high-resolution structure (X-ray crystal structure or NMR spectral structure) [25] of the receptor with a ligand bound should be available. The ligand is removed from the binding site in the graphic display, and then DOCK fills the binding site with sets of overlapping spheres, where a set of sphere centres serves as the negative image of the binding site. When a crystal structure of a receptor is available, but without a ligand bound, DOCK characterizes the entire surface of the receptor with regard to grooves that could potentially form target binding sites, which are filled with the overlapping spheres. Next DOCK matches X-ray or computer-derived structures of putative ligands to the image of the receptor on the basis of a comparison of internal distances. Then the program searches 3D databases [such as Cambridge Structural Database (CSD), Fine Chemicals Directory (FCD) and Available Chemicals Directory (ACD)] of small molecules and ranks each candidate on the basis of the best orientations.

### *De novo* design

Use of docking programs to design new lead structures that fit a particular target site is known as *de novo* design [26-28]. Normally two strategies are followed, the first is to use a template structure and the second is the use of component fragments of structure to construct a new molecule. Both these approaches involve fitting fragments of structure into the target site. Consequently, if the structure of the target site is reasonably well established both approaches can then only be followed. The *de novo* design methods treat all the structures they use as being rigid, i.e., bonds that can exhibit a degree of free rotation are locked into one conformation. Consequently, *de novo* programs don't usually take into account the flexibility of molecules and target sites. In other words, the software doesn't take into account that ligands and the target site can exist in more than one conformation so that those structures used in the program are not necessarily those they assume in real life. In addition, in fully automated procedures the design is normally limited by the extent of the library of fragments, conformations and links held in the software's data base. For these and other reasons, *de novo* design is normally used to find new leads.

Another program for *de novo* molecular design, Grow Mol (called Alle Grow in latest version) approaches the problem of receptor binding from a different direction. Instead of docking known molecules into the binding site, it generates molecules with steric and chemical complementarily to the 3D structure of the receptor binding site by evaluating each new atom according to its chemical complementarily to the nearby receptor atoms. In the modified method for docking flexible ligands into a receptor described above, an *X*-ray structure of the receptor is not necessary to characterize the shape of the receptor binding site. Rather, the receptor binding site can be deduced from the shapes of active ligands. This technique, which is useful for identification of the pharmacophore geometry, is called receptor mapping. A variety of receptor mapping techniques has been described. An approach termed steric mapping uses molecular graphics to combine the volumes of compounds known to bind to the desired receptor. A major improvement in the use of molecular modelling came when high-throughput crystallography was coupled with combinatorial chemistry approaches. Because structure based drug design usually involves targets whose structures are already known, but which have different ligand bound; only the part of the structure where the ligand binds needs to be resolved. Software such as AutoSolve (Astex Technology) analyzes and interprets electron density data automatically without the need for an expert crystallographer, so 100s of receptor complex crystals can be analysed in just a few days.

## Combinatorial Chemistry

In order to produce the large numbers of compounds required for high-throughput screening (techniques used by rapid, efficient, drug testing systems are collectively known as HTS, which even when extremely small amounts of test substance are available give accurate results) combinatorial chemistry was developed. It allows the simultaneous synthesis of a large number of the possible compounds that could be formed from a number of building blocks. The products of such a process are known as a combinatorial library. Libraries may be a collection of individual compounds/mixtures of compounds. Screening the components of a library for activity using HTS enables the development team to select suitable compounds for a more detailed investigation by combinatorial chemistry. Example: Consider the reaction of a set of 3 compounds ($A_1$-$A_3$) with a set of 3 building blocks ($B_1$-$B_3$). In combinatorial synthesis [29], $A_1$ would simultaneously undergo separate reactions with compounds $B_1$, $B_2$ and $B_3$, respectively. At the same time compounds $A_2$ and $A_3$ would also be undergoing reactions with compounds $B_1$, $B_2$ and $B_3$. These simultaneous reactions would produce a library of 9 products. If this process is repeated by reaction of these 9 products with 3 new building blocks ($C_1$-$C_3$), a combinatorial library of 27 new products would be obtained.

### The design of combinatorial syntheses

First strategy (linear synthesis): The building blocks are successively added to the preceding structure so that it grows in only one direction. It relies on the medicinal chemist finding suitable protecting groups so that the reactions are selective. This design approach is useful if the product is a polymer formed from a small

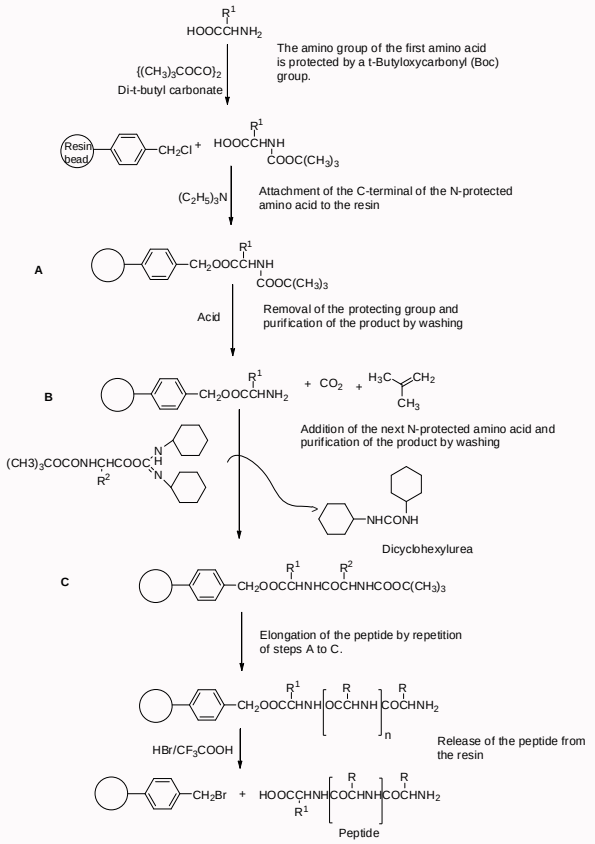



**Figure 12:** Outline of the Merrifield peptide synthesis where R is any amino acid side chain.

number of monomeric units.

Second strategy: the synthesis can proceed in different directions from an initial building block known as a template provided that the template has either the necessary functional groups/they can be generated during the course of synthesis.

## The solid support method (Merrifields, 1969)

For the product of each stage of the synthesis this method used polystyrene-divinyl benzene resin beads as a solid support. Each bead had a large number of monochlorinated methyl side chains. The C-terminal of the first amino acid in the peptide chain was attached to the bead by an $S_N2$ displacement reaction of these chloro groups by a suitable amino acid. The large number of chlorinated side chains on the bead meant that one bead acts as the solid support for the formation of a large number of peptide molecules of the same type. To the growing peptide chain using the reaction sequence additional amino acids were added as shown in Figure 12 [30].

To control the position of amino acid coupling this sequence uses protecting groups such as t-Boc (butyloxycarbonyl). The N-protected amino acids were converted to a more active acylated derivative of DCC (Di-Cyclohexyl Carbodiimide) to form the amide peptide link, which to link the new amino acid residue to the growing peptide reacted with an unprotected amino group. At the end of the synthesis using a mixture of trifluro ethanoic acid and hydrogen bromide the peptide was detached from the bead.

General methods in solid support combinatorial chemistry, Parallel synthesis: In a separate reaction vessel in parallel the

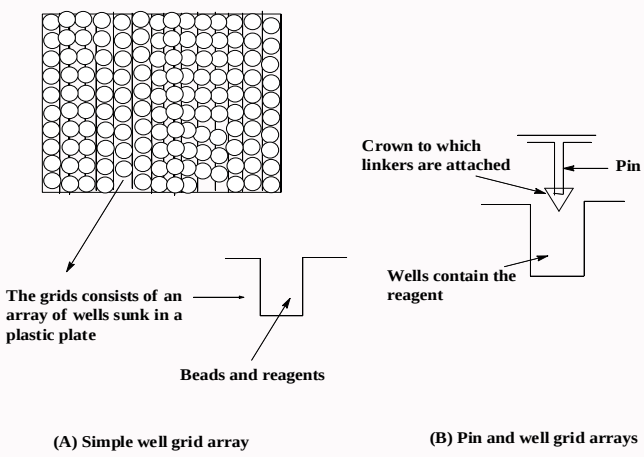



**Figure 13:** Examples of the arrays in combinatorial chemical analysis.

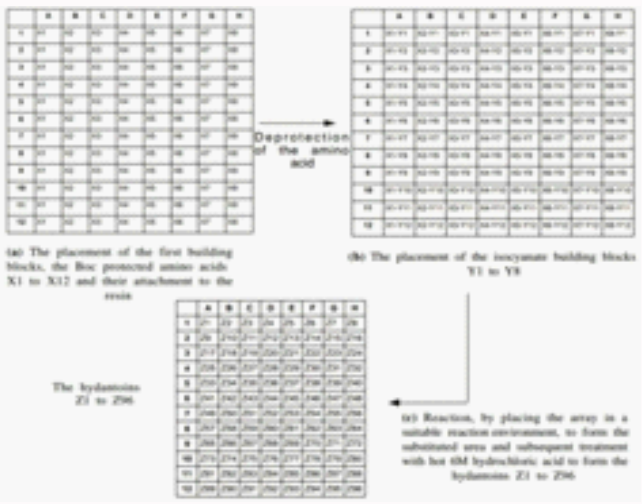



**Figure 14:** The pattern of well loading for the formation of a combinatorial library of 96 hydantoins.

compounds are prepared. The array of individual reaction vessels often takes the form of either a grid of plastic rods called pins or a grid of wells in a plastic plate as illustrated in Figure 13.

Ex., Preparation of a combinatorial library of hydrations by the reaction of isocyanates with amino acids using a 96-well array. Eight N-protected amino acids ($X_1$, $X_2$…$X_8$) are placed in the well array so that only one type of amino acid occupies a row, i.e., row A will only contain amino acid $X_1$, row B will only contain amino acid $X_2$, and so on (Figure 14). To each well beads are added and the array placed in a reaction environment that will join the X compound to the linker of the bead. The amino acids are deprotected by hydrogenolysis and 12 isocyanates ($Y_1$, $Y_2$…$Y_8$) added to the wells so that compound $Y_1$ is only added to row 1, compound is only added to row 2, and so on. To the substituted ureas the isocyanates are allowed to react. Each well is treated with 6M HCl and the whole array heated to simultaneously form the hydantoins and release them from the resin. Hence, it is possible to simultaneously synthesise a total of 96 different hydantoins (Z1-Z96).

Furka's mix and split technique: The Furka method produces the library of compounds on resin beads. These beads are divided into a number of equally sized portions corresponding to the number of

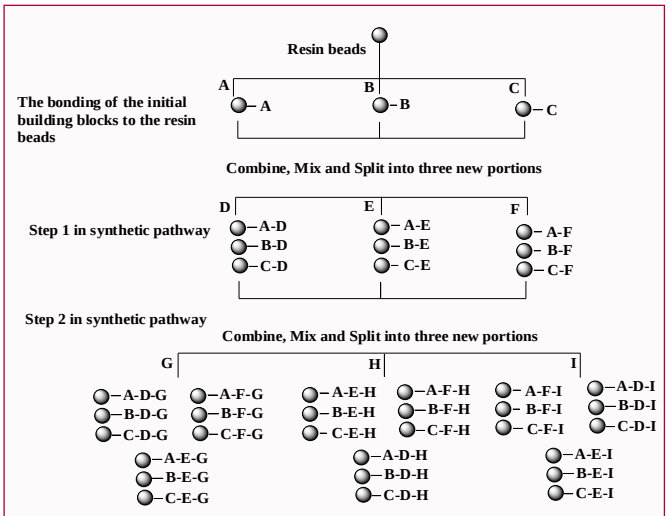



**Figure 15:** An example of the Furka approach to combinatorial libraries using a two-step synthesis involving three building blocks at each stage.

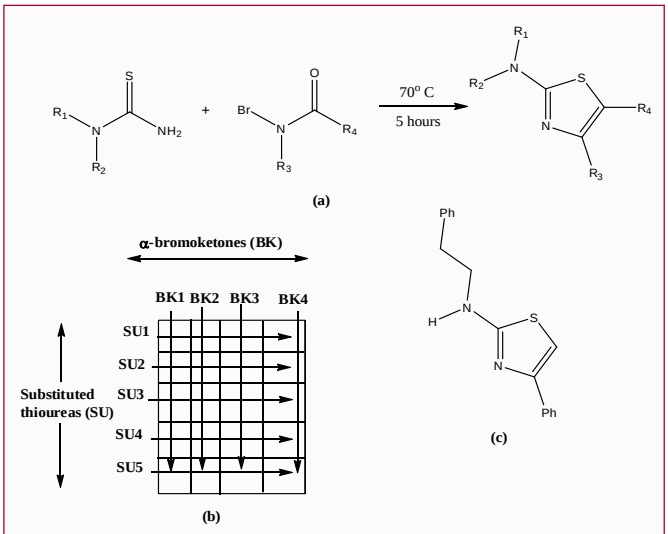



**Figure 16:** (a) The Hantzsch synthesis of 2-aminothiozoles; (b) the reaction grid; (c) Fanetizole.

initial building blocks. Each of the starting compounds is attached to its own group of beads using the appropriate chemical reaction (Figure 15). All the portions of beads are now mixed and separated into the number of equal portions corresponding to the number of different starting compounds being used for the first stage of synthesis. A different reactant building block is added to each portion and the reaction is carried out by putting the mixtures of resin beads and reactants in a suitable reaction vessel. After reaction, all the beads are mixed before separating them into the number of equal portions corresponding to the number of building blocks being used in the second stage of synthesis. This process of mix and split is continued until the required library is synthesized [31].

### Encoding methods

Sequential chemical tagging: This uses specific compounds (tags) as a code for the individual steps in the synthesis. These tag compounds are sequentially attached in the form of a polymer-like molecule to the same bead as the library compound at each step in the synthesis, usually by the use of a branched linker. One branch is used for the library synthesis and the other for the encoding. At the end of the synthesis both the library compound and the tag compound are liberated from the bead.

### Combinatorial synthesis in solution

Solid phase combinatorial synthesis encounters number of disadvantages:

1. All the libraries have a common functional group at the position corresponding to the one used to link the initial building block to the linker/bead.

2. By using the linear approach syntheses are usually carried out.

3. If multistep syntheses are needed it requires especially modified reactions with high yields (>98%).

4. To attach the building block to and remove the product from the support it requires additional synthesis steps.

Parallel synthesis in solution: It is used to prepare libraries of single compounds. E.g. Bailey (1996) by means of Hantzsch synthesis produced a library of 20, 2-aminothioazoles. They used a 5×4 grid of glass vials (Figure 16). Five different substituted thioureas, one per row, were treated with 4 different α-bromoketones. Each of the bromoketones, only one per row, was added to a separate row. After reaction the products were isolated before being characterised by high resolution NMR and MS. One of the compounds synthesized by this method was the anti-inflammatory fanetizole [32].

## Conclusion

In drug design and medicinal chemistry QSAR has proved to enhance our understanding of fundamental processes and phenomena. Much of knowledge and discussion has been generated by the concept of hydrophobicity and its calculation as well as spawned a mini-industry. QSAR has refined our thinking on selectivity at cellular level as well as molecular level. QSAR studies in the pharmacokinetic arena have established different hydrophobic requirements for renal/ non-renal clearance, whereas the optimum hydrophobicity for CNS penetration has been determined by Hansch et al. Over the last few decades QSAR has matured in terms of the descriptors, methods of analysis, models, and choice of compounds and substituents. QSAR modeling remains one of the most important instruments of computer-aided drug design in the modern age of medicinal chemistry. Based on our interest in organic and medicinal chemistry [33,34], discussed methodologies in this review article will afford validated QSAR models, which should continue to facilitate and enrich the experimental process of drug discovery and development.

## References

1. Hansch C, Leo A, Hoekman D. Exploring QSAR. Fundamentals and Applications in Chemistry and Biology, Volume 1. Hydrophobic, Electronic and Steric Constants, Volume 2. American Chemical Society.1995.

2. Hansch C, Kim D, Leo AJ, Novellino E, Silipo C, Vittoria A. Toward a quantitative comparative toxicology of organic compounds. Crit Rev Toxicol. 1989; 19: 185-226.

3. Hansch C, Steward AR, Anderson SM, Bentley DL. Parabolic dependence of drug action upon lipophilic character as revealed by a study of hypnotics. J Med Chem. 1968; 11: 1-11.

4. Manallack DT. The $pK_a$ Distribution of Drugs: Application to Drug Discovery. Perspect Medicin Chem. 2007; 1: 25-38.

5. Tipker J, Verloop A. Use of STERIMOL, MTD, and MTD* Steric Parameters in Quantitative Structure-Activity Relationships. American

Chemical Society. 1984; 255: 279-296.

6. Pathan H, Williams J. Basic opioid pharmacology: an update. Br J Pain. 2012; 6: 11-16.
7. Grewe R. Das Problem der Morphin-Synthese. Naturwissenschaften. 1946; 33: 333-336.
8. Schnider O, Grussner A. Synthese von Oxy-morphinanen. Helv Chem Acta. 1949; 32: 821-828.
9. May EL, Murphy JG. Structures related to Morphine III. Synthesis of an analog of N-Methylmorphinan. J Org Chem. 1955; 20: 257-263.
10. Richardson BW. Physiological research on alcohols. Med Times Gaz. 1869; 18: 703-706.
11. Thornber CW. Isosterism and molecular modification in drug design. Chem Soc Rev. 1979; 8: 563-580.
12. Burger A. Relationships of chemical structure and biological activity in drug design. Wiley. 1970.
13. Korolkovas A. Essentials of Molecular Pharmacology: Background for Drug Design. Wiley: New York.1970.
14. Grimm HG. ÜberBau und Grösse der Nichtmetallhydride. Elektrochem. 1925; 31: 474-480.
15. Grimm HG. Allgemeinesüber die verschiedenen Bindungsarten. Elektrochem.1928; 34: 430-437.
16. Lipinski CA. Bioisosterism in Drug Design. Annu Rep Med Chem.1986; 21: 283-291.
17. Hansch C, Steward AR, Anderson SM, Bentley DL. Parabolic dependence of drug action upon lipophilic character as revealed by a study of hypnotics. J Med Chem. 1968; 11: 1-11.
18. Fischer E. Einfluss der Configuration auf die Wirkung der Enzyme. Ber Deut Chem Ges. 1894; 27: 2985-2993.
19. Itzstein MV, Wu WY, Kok GB, Pegg MS, Dyason JC, Jin B, et al. Rational design of potent sialidase-based inhibitors of influenza virus replication. Nature. 1993; 363: 418-423.
20. Colman PM, Varghese JN, Laver WG. Structure of the catalytic and antigenic sites in influenza virus neuraminidase. Nature. 1983; 303: 41-44.
21. Varghese JN, Laver WG, Colman PM. Structure of the influenza virus glycoprotein antigen neuraminidase at 2.9 Å resolution. Nature. 1983; 303: 35-40.
22. Varghese JN, Colman PM. Three-dimensional structure of the neuraminidase of influenza virus A/Tokyo/3/67 at 2.2 A resolution. J Mol Biol. 1991; 221: 473-486.
23. Varghese JN, McKimm-Breschkin JL, Caldwell JB, Kortt AA, Colman PM. The structure of the complex between influenza virus neuraminidase and sialic acid, the viral receptor. Proteins. 1992; 14: 327-332.
24. Kuntz ID, Blaney JM, Oatley SJ, Langridge R, Ferrin TE. A geometric approach to macromolecule-ligand interactions. J Mol Biol. 1982; 161: 269-288.
25. Leach AR, Harren J. Structure-based Drug Discovery. Springer: Berlin; 2007.
26. Schneider G, Fechner U. Computer-based *de novo* design of drug-like molecules. Nat Rev Drug Discov. 2005; 4: 649-663.
27. Dey F, Caflisch AJ. Fragment-based *de novo* ligand design by multi objective evolutionary optimization. Chem Inf Model. 2008; 48: 679-690.
28. Nicolaou CA, Apostolakis J, Pattichis CS. De Novo Drug Design Using Multi objective Evolutionary Graphs. J Chem Inf Model. 2009; 49: 295-307.
29. Szostak JW. Introduction: Combinatorial Chemistry. Chem Rev. 1997; 97: 347-348.
30. Merrifield RB. Solid-phase peptide synthesis. Adv Enzy mol Relat Areas Mol Biol. 1969; 32: 221-296.
31. Furka A, Bennett WD. Combinatorial libraries by portioning and mixing. Comb Chem High Throughput Screen. 1999; 2:105-122.
32. Bailey N, Dean AW, Judd DB, Middlemiss D, Storer R, Watson SP. A convenient procedure for the solution phase preparation of 2-aminothiazole combinatorial libraries. Bioorg Med Chem Lett. 1996; 6: 1409-1414.
33. (a) Kumar K, Mudshinge SR, Goyal S, Gangar M, Nair VA. A catalyst free, one pot approach for the synthesis of quinoxaline derivatives via oxidative cyclisations of 1,2-diamines and phenacyl bromides. Tetrahedron Letters. 2015; 56: 1266-1271.
33. (b) Sharma R, Kumar K, Chouhan M, Grover V, Nair VA. Lithium hydroxide mediated synthesis of 3,4-disubstituted pyrroles. RSC Advances. 2013; 3: 14521-14527.
33. (c) Kumar K, More SS, Goyal S, Gangar M, Khatik G, Rawal R, Nair VA. A convenient synthesis of 4-alkyl-3-benzoylpyrroles from $\alpha,\beta$-unsaturated ketones and tosylmethylisocyanide. Tetrahedron Letters. 2016; 57: 2315-2319.
33. (d) Kumar K, Siddique J, Gangar M, Goyal S, Rawal RK, NairVA. $ZrCl_4$ catalysed diastereoselective synthesis of spirocarbocyclic oxindoles *via* [4+2] cycloaddition. Chemistry Select. 2016; 1: 2409-2412.
33. (e) Kaur R, Manjal SK, Rawal RK, Kumar K. Recent synthetic and medicinal perspectives of tryptanthrin. Bioorg Med Chem. 2017; 25: 4533-4552.
33. (f) Kaur R, Choudhry S, Kumar K, Gupta MK, Rawal RK. Recent synthetic and medicinal perspectives of dihydropyrimidinones: A review. Eur J Med Chem. 2017; 132: 108-134.
34. (a) Talwan P, Chaudhary S, Kumar K, Rawal RK. Chemical and medicinal versatility of substituted 1,4-dihydropyridines. Current Bioactive Compounds. 2017;13:109-120.
34. (b) Kaur R, Kumar K. One-pot synthesis of [4-(tert-butyl)-1H-pyrrol-3-yl] (phenyl)methanone from tosylmethyl isocyanide and carbonyl compound. Chemistry Heterocyclic Compounds. 2018; 54: 700-702.
34. (c) Kapoor Y, Dubey S, Kar S, Kumar K. CAR T Cell: A Novel Treatment Regime for Cancer. SF J Pharm Anal Chem. 2018; 1: 1008.
34. (d) Kapoor Y, Kumar K. Medicinal and Chemical Perspectives of Nitric Oxide: An Overview. SF J Pharm Anal Chem. 2019; 2: 1015.
34. (e) Kaur R, Kumar K. A Journey Towards FeCl3Catalysed Synthesis of Multisubstituted Pyrrole. J Med Chem Sci. 2019; 2: 110-117.