IV-B.Pharm (8th semester)

BPH-803/BPL-803 Medicinal Chemistry-III theory

UNIT-5

QSAR AND DRUG DESIGN

B) INTRODUCTION OF QSAR

INTRODUCTION

Structure-activity relationship (SAR) and quantitative structure-activity relationship (QSAR) study, collectively referred to as (Q)SAR, are theoretical models that can be used to predict the physicochemical, biological, and environmental fate properties of molecules.

The aim of QSAR techniques is to develop correlations between any biological property form of activity, frequently *biological* activity, and their properties, usually, physicochemical properties of a set of molecules, in particular, substituent properties. However, in its most general form, QSAR has been adapted to cover correlations independent of actual physicochemical properties. QSAR started with similar correlations between chemical reactivity and structure. Ideally, the activities and properties are connected by some known mathematical function, F:

Biological activity = F (physicochemical properties)

Biological activity can be any measure of, such as C, K_i, IC_{50}, ED_{50}, and K_m.

Physicochemical properties can be broadly classified into three general types such as electronic, steric, and hydrophobic property of biologically active molecules, for which an enormous range of properties and physicochemical parameters have been defined. Ideally, the parameters selected should be orthogonal, that is, have minimal covariance. The relationship or function is usually (but not always) a mathematical expression derived by statistical and related techniques, for example, multiple linear regression (MLR). The parameters describing physicochemical properties are used as independent variables and the biological activities are dependent variables. In some cases, a function cannot be found, and this reflects the multivariate, nonlinear nature of biological and physical properties. Usage of such data may be possible with neural networks to deduce essential data for biological activities and then using them for prediction.

Usually, some data are used to generate a relationship (the training set), while another set of data is reserved as a test set on which predictions using the rule are made. In this manner, a model can be tested for validity. The complete range of techniques used to derive functional relationships between the data is collectively known as chemometrics.

HISTORICAL DEVELOPMENT OF QSAR

In early 1868, Crum-Brown and Fraser published an equation, which is considered to be the first general equation of QSAR. Richet discovered that the toxicity of organic compounds (aldehyde, alcohol, ethers, ketone, etc.) inversely follow their water solubility. Later, Meyer and Overton have worked independently and observed the linear relationships between lipophilicity (oil-water partition coefficients) and narcotic activities. Further, Fuhner has given the first evidence of additivity of groups contributing to biological activity values in a homologous series of compounds for their narcotic activity. Ferguson postulated that the thermodynamic principles could be related to drug activities.

After 1950, QSAR methodology has progressed remarkably. During this year, Bruice, Kharasch, and Winzler reported group contributions to biological activity values in a series of thyroid hormone analogues; this may be considered as a first Free-Wilson-type analysis. Zahradnik has tried to apply the concept of the Hammett equation. However, significant progress has occurred in the field of QSAR in 1960s through the research work of Bocek and Kopecky, Hansch, Free-Wilson and Fujita. Among these, only Bocek-Kopecky method has become unsuccessful due to the involvement of a number of parameters. Among the different methods, Hansch, Free-Wilson, and modified Free-Wilson approaches are the widely practiced ones for modelling the biological response. Also, QSAR is one of the approaches attracting increasing levels of interest in the pharmaceutical industry as a productive and cost-effective technology in the search for novel lead compounds.

Hansch Analysis

QSAR based on Hammett's relationship utilize electronic properties as the descriptors of structures. Difficulties were encountered when investigators attempted to apply Hammett-type relationships to biological systems, indicating that other structural descriptors were necessary. In 1962, Hansch et al entered the scenario with the numerical information on lipophilicity, electronic, and steric effect on the model development. The general form of Hansch equation is as follows:

$$\text{Log BA} = a \log p + b\,\sigma + c\,\text{Es} + \text{constant (linear)}$$
$$\text{Log BA} = a \log p + b\,(\log p)^2 + c\,\sigma + d\,\text{Es} + \text{constant (nonlinear)}$$

Hansch model correlates biological activity with physicochemical properties. The coefficients (*a*, *b*, *c*, *d*, and constant) are determined by multiple regression analysis.

Free-Wilson Analysis

It is also known as the additivity model or *de novo* approach. This method is based on the assumption that the introduction of a particular substituent at a particular molecular position always contributes in the same way to the biological potency of the whole molecule, as expressed by the equation:

$$\text{Log BA} = \text{contribution of unsubstituted parent compound} + \text{contribution of corresponding substituents}$$
$$\text{Log BA} = \mu + \sum a_i a_j$$

where a_i = number of positions at which substitution occurs

a_j = number of substituents at that position

μ = overall average.

The equation is solved by MLR using the presence (1) or absence (0) of the different substituents as independent parameters, while the measured activity serves as dependent variable.

Mixed Approach

Kubinyi has presented the combination of Hansch and Free-Wilson approach as mixed approach.

$$\text{Log BA} = k_1\pi + k_2\sigma + k_3 Es + k \text{ (Hansch analysis)}$$
$$\text{Log BA} = \mu + \sum a_i a_j \text{ (Free-Wilson approach)}$$

So, the mixed approach can be written as

$$\text{Log BA} = \sum a_i a_j + \sum k_i \phi_j + k$$

Where $\sum (a_i a_j)$ is the Free-Wilson part for the substituents

$\phi_j = \sigma, \pi$, and Es contribution of the parent skeleton.

Among the above-mentioned approaches, Hansch approach became the most popular approach in QSAR. The high-dimensional QSAR analyses (3D, 4D, and 5D) are developed to avoid pitfalls of classical method and to create the hypothetical drug receptor model.

Advantages of QSAR

- It gives quantifying the relationship between structure and activity with their physiochemical property basis.
- Possible to make predictions of designed compounds before the chemical synthesis of novel analogues.
- It may help to understand the interactions between functional group of designed molecules and their activity of target enzyme or protein.

Disadvantages of QSAR

- Due to biological data experimental error it may give false correlations.
- If training set of molecule is less, the data may not reflect the complete property and it cannot be used to predict the most active compounds.
- In some 3D QSAR study ligands binding receptor or protein may not be available in that case the common approach result may not represent the reality.
- Cannot expect that the QSAR works all the time give successful applications.

A model perfectly predicts that the training data may not be good or even useless for prediction. The problem of QSAR is to find coefficients $C_0, C_1, \ldots, C_n$ such that

$$\textit{Biological activity} = C_0 + (C_1 \times P_1) + \cdots + (C_n \times P_n)$$

and the prediction error is minimized for a list of given compounds.

Partial least squares (PLSs) is a technique used for computation of the coefficients of structural descriptors.

BASIC REQUIREMENTS FOR QSAR ANALYSIS

Some basic requirements are very essential for best model development. They are the following:

- All analogues belong to a congeneric series (classical QSAR studies) exerting the same mechanism of action. This is a series of compounds with a similar basic structure, but with varying substituents. Noncongeneric series are widely used for higher dimensional (3D and 4D) studies.

- Also, the set of compounds with same mechanism of action is essential.
- Biological response should be distributed over a wide range.
- Observed biological activity should be in specific units (concentration in molar units or IC_{50} or percentage inhibition).
- A simple rule is that the total number of compounds in the training set divided by the number of variables in the final model should be greater than approximately five or six.

This will assure that a data set will not be 'over predicted' and that the model will have a better chance to retain the predictive value.

Steps Involved in QSAR Studies

The QSAR methodology enables the development of mathematical models, which can be used to predict the biological activity of newly designed compounds. There are three steps involved in this procedure; the first step is the creation of a database in which calculation of various physicochemical and structural parameters of a congeneric series takes place followed by regression analyses leading to model development between biological activities versus derived physiochemical descriptors. The third step involves the validation of the models and prediction of the biological activity of the designed compounds.

Statistical Methods Used in QSAR Analysis

Statistical methods are an essential component of QSAR work. They help to build models, estimate a model's predictive abilities, validate an already existing model, and find the relationships and co-relationship among the variables and the activities. Data analysis methods are used to recombine data into forms and groups and observations into hierarchies.

Regression Methods

It is a mathematical procedure, which co-relates dependent (X) variable with the independent (Y) variables. There can be different forms of regression analysis:

Simple linear regression analysis: An independent variable is correlated with a dependent variable and produces a linear one-term equation. It is useful for discovering some of the most important descriptors.

MLR analysis: More than one independent variable is correlated with a dependent variable and a single multiterm equation is formed. The number of variables should be one-fifth of the molecules in a series, that is, for each five molecules in the series one can have one variable.

Stepwise linear regression analysis: This is useful when the number of independent variables is very high and is thus correlated in a stepwise manner with the dependent variable producing a multiterm linear equation.

Partial Least Square (PLS)

Hundreds or even thousands of independent variables (X-block) can be correlated with one or several dependent variables (Y-block). PLS is used when X data contain co-linearities or when N is less than 5M, where N is the number of compounds and M is the number of independent variables. Often perfect correlations are obtained in PLS analysis, due to the usually large number of X variables and cross-validation

procedure must be used to select the model that is having the highest predictive values. Several PLS are performed in which one or several objects are eliminated from the data set. It is the method of choice in 3D QSAR method.

Genetic Function Approximation (GFA)

It provides multiple models that are created by evolving random initial models using a genetic algorithm. Models are improved by performing a cross over operation to recombine better sorting models. This method is used when dealing with a large numbers of descriptors.

Genetic Partial Least Squares (G/PLSs)

This method combines the best of GFA and PLS. Each generation has a PLS applied to it instead of MLR and so each model can have more terms in it without fear of overfilling. G/PLS retains the ease of interpretations of GFA by back transforming the PLS component to the original variable.

Principal Component Analysis (PCA)

PCA is a data reduction method, using mathematical techniques to identify the pattern in a data matrix. The main element of this approach consists of the construction of a small set of new orthogonal, that is, uncorrelated variables derived from a linear combination of the original variables.

Statistical Measures Commonly Used in Regression Analysis

Correlation coefficient (r)/Square of the correlation coefficient (r^2): The correlation coefficient '*r*' and square of the correlation coefficient (r^2) are measures of the quality of the fit of the model. It is computed using the following equation

$$r = \sqrt{1 - \Sigma\Delta^2/SSY}$$
$$r^2 = 1 - \Sigma\Delta^2/SSY$$

Where, $SSY = \Sigma (Y_{obs} - Y_{mean})$

$\Sigma\Delta^2 = \Sigma (Y_{obs} - Y_{cal})^2$

Where SSY is the overall variance, that is, $S = \Sigma (Y_{obs} - Y_{mean})$ Y_{obs} is observed biological activities

Y_{mean} is mean of biological activities value

Y_{cal} is calculated biological activity used in the equation.

A high value of correlation coefficient (r) indicates the statistical significance of the regression equation and thereby the participating substituent constants. The squared correlation r^2 is a measure of the explained variance, most often presented as a percentage value, for example, $r = 0.8$, then $r^2 = 0.664$ or 66.4% as the variance accounted by regression parameters.

Standard error of the estimate (S): This is a measure of how well the function derived by the QSAR analysis predicts the observed biological activity. Its value considers the number of objects n and the number of variable k. Therefore, S depends not only on the quality of fit, but also on the number of degrees of freedom. The smaller the value of S the better is the QSAR.

$DF = n - k - 1$

$S = \sqrt{\Sigma (Y_{obs} - Y_{cal})^2}/n - k - 1$

***F*-value:** It is a measure of the statistical significance of the regression model, the influence of the number of variables included in the model is even larger than the standard deviation.

$$F\text{-value} = r^2 (n - k - 1)/k(1 - r^2)$$

Predicted sum of squares (PRESS): The sum of the overall compounds of the square difference between the actual and the predicted value of dependent variables.

$$P = \Sigma (Y_{obs} - Y_{pred})^2$$

Cross-validation r^2 (q^2): Cross-validation is an approach for assessing the predictive value of a model. The cross validation $r^2(q^2)$ is generated during a validation procedure. It is calculated using the formula

$$q^2 = 1.0 - \Sigma (Y_{pred} - Y_{obs})^2 / \Sigma (Y_{obs} - Y_{mean})^2$$

Where Y_{pred} is a predicted value; Y_{obs} is an actual value or experimental value; Y_{mean} is the best estimate of the mean of all values that might be predicted.

A cross-validated r^2 is usually smaller than the overall r^2 for a QSAR equation. It is used as a diagnostic tool to evaluate the predictive power of an equation. Cross-validation proceeds by omitting one or more rows of input data, re-deriving the model, and predicting the target property values of the omitted rows. The re-derivation and predicting cycle continues until all the target property values have been predicted at least once. The root mean square error of all the target predictions, the predictive sum of squares (PRESS) is the basis for evaluating the model.

Outliers: An outlier is defined as a structure with a residual greater than two times the standard deviation.

Bootstrapping: Bootstrapping is another technique for model validation. It is based on simulating a large number of data sets sampled from the original data set that are of the same size as the original. The same data can be sampled more than once. The statistical analysis is performed on each of the simulating data sets. The component model with consistent results is then chosen as the final model.

MODEL DEVELOPMENT PROCEDURES

Classical or 2D QSAR Analysis

2D descriptors are usually developed by using the atoms and connective information of the molecule, but 3D coordinates and individual conformations are not considered. In 2D QSAR, physicochemical parameters such as hydrophobic (π), steric (molar refractivity or MR), hydrogen acceptor (HA), hydrogen donor (HD), and electronic (field effect or F, resonance or R, Hammett's constant or σ) are normally used. In addition to these parameters, *de novo* constants or indicator variables with 0 or 1 values denoting the absence or presence of certain features (*cis*/*trans* ring atom and bridge atom or chain, different test model, etc) are also used to adequately parameterize the compounds. In all this, many topological indices are also considered as parameters for analysis.

Drug distribution and binding processes are equilibrium processes governed by the corresponding free energy differences, $K = e^{-\Delta G/RT} = e^{-\Delta H - T\Delta S)/RT}$, such relationships should use logarithmic scale. For these the biological inhibitory values, that is, IC_{50} or ED_{50} or LD_{50} or *Ki* must be converted into logarithmic form, such as log ($1/IC_{50}$) or log ($1/ED_{50}$) or log ($1/LD_{50}$) or log ($1/Ki$) values to obtain appropriate activity parameters for the QSAR study. The logarithmic scales also ensure a normal distribution for the experimental

error of biological tests, a requirement for regression-type statistical analyses. In some cases, the activity percentage (%) values (A) are converted to Log {A/(100 – A)} as a binding equilibrium constant, which physicochemically is more meaningful than A alone for QSAR analysis.

3D-QSAR Analysis

Three-dimensional quantitative structure-activity relationships (3D-QSARs) are quantitative models that relate the biological activity of small molecules with their properties calculated in 3D space (Fig 4.1). Hence, 3D properties of a molecule are considered rather than that of the individual substituents. The 3D structures are usually generated from 2D or 2D with configurational information or 3D-structure database or X-ray crystallographic analysis or 2D NMR study. This structure is optimized to refine the geometry based on the size of the molecule such as molecular mechanics (large systems; thousands of atoms) or semi-empirical (medium size systems; hundreds of atoms) or *ab initio* (small systems; tens of atoms), in order to obtain one lowest energy structure per molecule. There are many 3D-QSAR techniques used for various purposes. A few of them are the following:

- Comparative molecular field analysis (CoMFA)
- Comparative molecular similarity indices analysis (CoMSIA)
- Molecular shape analysis (MSA)
- The distance geometry approach
- The binding site model approach
- COMPASS, the hypothetical active lattice method
- The molecular similarity approach
- Genetically evolved receptor models

Among all the 3D-QSAR techniques, CoMFA is the most widely used technique and has shown unprecedented accuracy in prediction. Some of these approaches to QSAR are based on the statistical analysis of the 3D interaction fields. These are generated by measuring over a regular 3D grid the interaction energy between a small probe atom or a group and the ligands. Initially, the 3D structures of the training set of compounds are aligned based on common molecular features, so as to occupy the same volume of space. The interaction energies of the small probe, usually, a methyl group and a proton, is measured with each of the training set compounds at each grid co-ordinates in space. The interaction energy at each grid point in space becomes a descriptor in a QSAR analysis. It results in a data table containing several hundreds or even thousands of descriptors for the analysis.

Comparative Molecular Field Analysis (CoMFA)

CoMFA is a 3D-QSAR technique employing both interactive graphics and statistical techniques for correlating the shapes and the biological properties of the molecules. It was proposed and developed by R.D. Cramer in 1988. The principle underlying CoMFA is that differences in a target property related to differences in the shapes of the noncovalent fields of tested molecules. The molecular shape of tested moelcule field into a QSAR table and the magnitude of steric (Lennard-Jones) and electrostatic (Coulombic) fields are sampled at regular intervals throughout a defined region of rigid box.

To do so, bioactive conformation of each compound is chosen and they are superimposed in a manner defined by the supposed mode of interaction with the target receptor. Further, CoMFA compares the steric

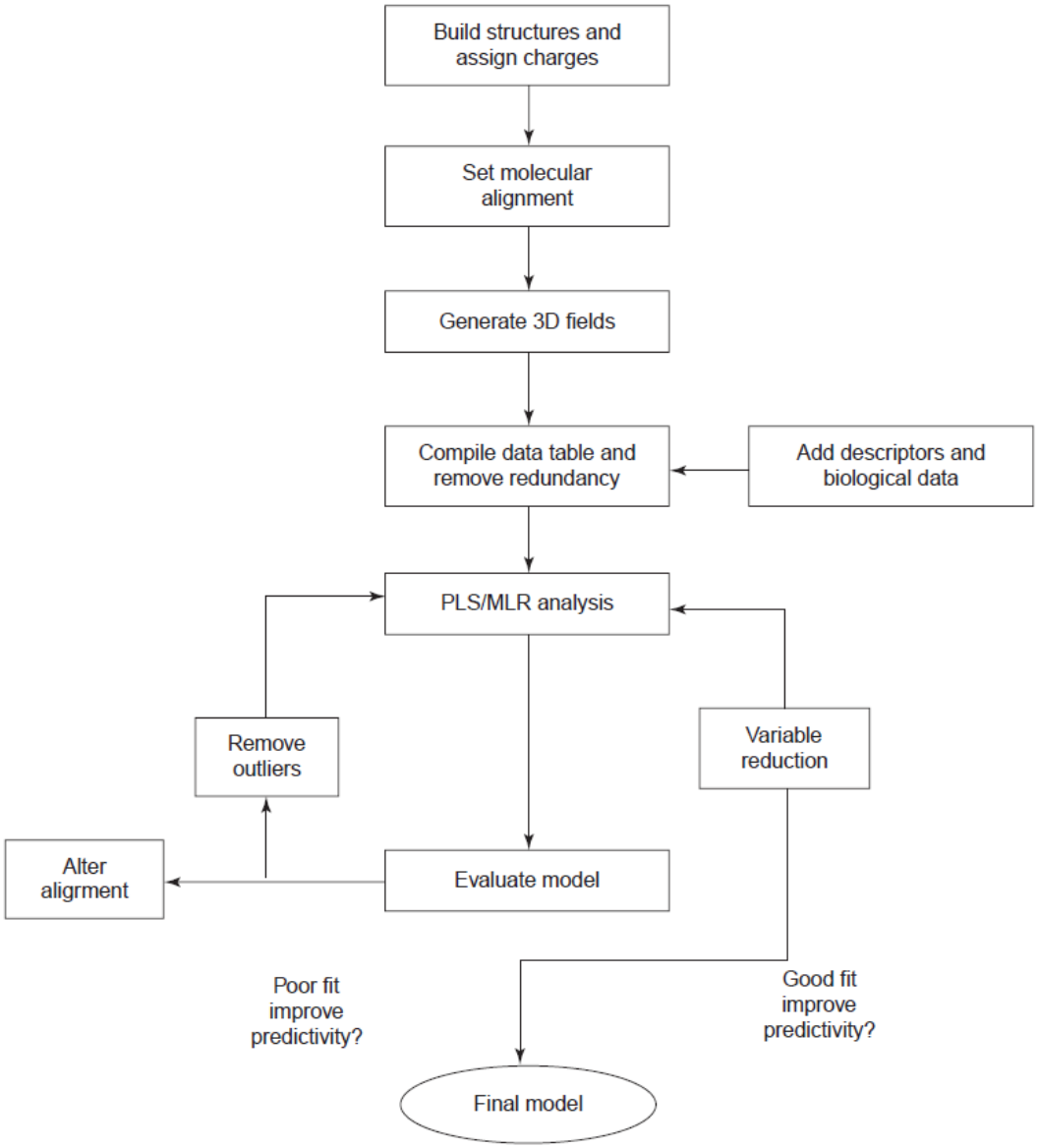


Figure 4.1 Systematic methodology for 3D-QSAR analysis.

and the electrostatic fields calculated around the molecules with various probe groups in three dimensions and extract the important features related to the biological activity. With this information, CoMFA tries to identify the quantitative influence of the specific chemical features of the molecules on their potencies. In 3D space the contour plots result showing that biological activity important regions of designed or active molecules. Advantages of CoMFA technique include the prediction of activity of new compounds and representation of QSAR models in the form of contour maps.

ELECTRONIC PARAMETERS

There are many approaches to drug designing in relation with physiochemical 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 orbitals to change the electronic feature. The molecular orbital approach is the change on electronic charges, evidenced from the investigation of three volatile inhalation anaesthetics, and also on molecular conformation, as studied with respect to acetylcholine, in regard to bond lengths and angles including torsional angles. These interpretations are carried out by computational methods in respect to structure activity relationship (SAR).
3. **Approach with molecular connectivity:** This is based on the structural features of a molecule. All seteric and electronic parameters varies according to their configuration. These includes cyclization, unsaturation, presence of heteroatom, skeletal branching, and position in molecules with the aid of numerical indices and the series of functional attachments.
4. **Approach of linear free-energy:** Linear free energy approach was based on the selection of physiochemical parameters of a molecule with a specific biological activity. But the biological activity may vary in relation to the physiochemical properties of the drug or molecule and does not provide a prompt success, but it may reveal some beneficial features regarding the molecule.

A) LEAD DISCOVERY

INTRODUCTION

The discovery or design of new drugs requires a design process along with the synthetic techniques, methods of administration, development of tests, and further procedures to establish pharmacodynamics and their toxicological assessment. The diversity in synthetic chemistry is ever increasing as a result of the increase in the number of chemical compounds, which requires a sequence of screening modules with appropriate time consumption and affidavits little success. These consequences made it necessary to find out the possibilities for the development of new, logical, and scientific approaches in the discovery of new molecules or drugs and came to be known as drug designing.

Drug design is an integrated advancing discipline, in which a biologically active molecule is produced by chemical synthesis followed by an evaluation of its activity and toxicological studies with the limitation of trial-and-error screening. In the broader sense, it implies the random evaluation of congeners produced from the lead molecule either by implementing tailor-made techniques or by applying the basic concepts of physicochemical properties to produce an able molelcule.

DESIGN OF LEAD AND LEAD DISCOVERY

The concept of lead discovery envisages two investigational processes. They are:

1. **Exploration of leads:** search of new molecules.
2. **Exploitation of leads:** assessment, chemical modelling, and extension of leads.

The drug discoveries without a lead are quiet few in number. The most prominent examples include penicillium and librium. Librium is the first benzodiazepine tranquillizer. A series of quinazoline-3-oxides were synthesized by Leo Sternbach at Roche in a new tranquilizer development programme. None of the molecules produced reliable pharmacological activity. One of the molecules that had been missed and identified while a laboratory clean up was happening, found to posses probable activity and chemically it was benzodiazepine-4-oxide. It produced chloromethyl quinazoline-3-oxide with methylamine. This lead identification was further exploited to develop potent analogues such as diazepam, which was found to be 10 times more potent than the lead.

Approaches to Lead Discovery

The lead identifications require 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.

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, tetramycins, 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.

Nonrandom Screening

In this method, only compounds that possess similar structural skeletons were evaluated from their particular properties.

Pharmacokinetic Studies

Biotransformation occurs as the fate by metabolizing enzymes. In order to develop new leads, the metabolites or biotransformed compounds are studied for their properties, and such studies are expected to asses the activity from a comparison with the parent molecule. For example, the discovery of sulphanilamide is reported through the metabolic studies of prontosil.

Pharmacodynamic Studies

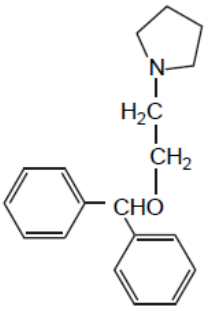
The effects apart from the therapeutic actions, that is, side effects may lead to the finding out of a new molecule with some appreciable structural modification. For example, sulphonamide used specifically for the treatment of typhoid, lowered the blood sugar levels drastically. This exerted action 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 was found to posses antidepressive action. This lead to the synthesis of many tricyclic antidepressants.

BIO-ISOSTERES

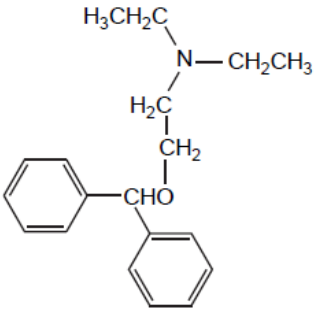
INTRODUCTION

Isosterism is of vital importance to the medicinal chemists because the biological characteristics of isosteres appear to be similar; more frequently than physical or chemical characteristics. Keeping in view the numerous advantageous application of isosterism in resolving biological problems effectively, Friedman proposed the following definition of bioisosterism—the phenomenon by which compounds usually fit the broadest definition of isosteres and possess the same type of biological activity.

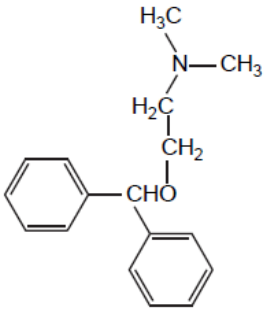
For instance, among antihistamines it is always preferable to have small compact substituents on the terminal nitrogen.



(A) Pyrrolidone analogue



(B) Diethyl analogue



(C) Dimethyl analogue

In the above-mentioned three structural analogues, it has been observed that 'A' possesses twice the activity of 'C', whereas it showed an activity that is many times greater than that of the open-chain diethylamino analogue.

It has been duly observed that it is more or less difficult to correlate the biological properties *vis-à-vis* physicochemical properties inherited by specific individual atoms, functional groups, or entire molecule by virtue of the glaring and established fact that a host of physical and chemical parameters are involved simultaneously, and are, therefore, extremely difficult to quantify them justifiably.

Besides simpler relationships, for example, isosterism invariably do not delay across the several varieties of biological systems that are often encountered with medicinal agents. In other words, a specific isosteric replacement in one particular biological system may either work or fail in response to the other. Thus, bioisosteres were further explained as follows: bioisosteres are (functional) groups or molecules that have chemical and physical similarities producing broadly similar biological properties.

CLASSIFICATION

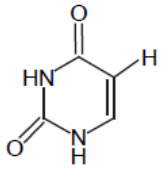
Bioisosteres are classified into the following two types:

1. Classical bioisosteres
2. Nonclassical bioisosteres

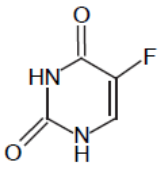
Classical bioisosteres: Classical bioisosteres have similarities in shape and electronic configuration of atoms, groups, and molecules, which they replace. Actual applications of bioisosteres in the successful design of a specific given molecule interacting with particular receptor is one glaring example, and very often either fails or negates the biological characteristics in another environment. Therefore, it is pertinent to state at this juncture that the logical use of biological replacement (classical or nonclassical) in the design of a new target drug molecules is solely and significantly depend on the specific biological system under critical investigation. Hence, there are no predetermined, well-established, predictable hard and fast guidelines, or laid generalized rules that may be useful to a medicinal chemist to affect biosteric replacement gainfully towards improved biological activity. Various classical bioisosteres with their appropriate examples are listed as follows:

i. Monovalent atoms and groups

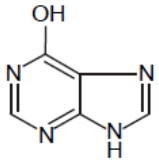
F, H
OH, NH
F, OH, NH, or CH_3 for H
SH, OH
Cl, Br, and CF_3



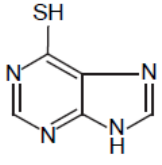
Uracil



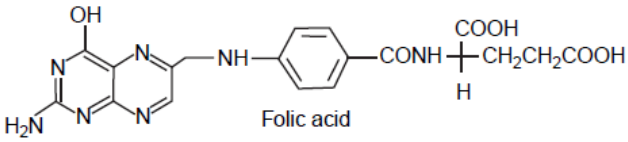
5-Flurouracil



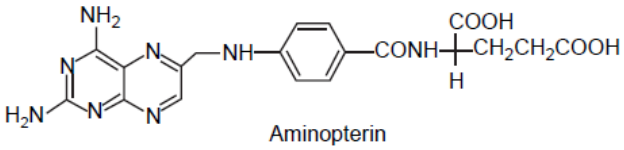
Hypoxanthine



6-Mercaptopurine



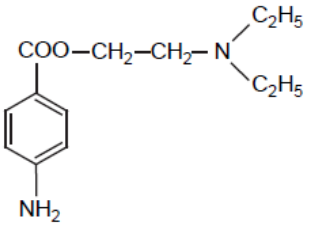
Folic acid



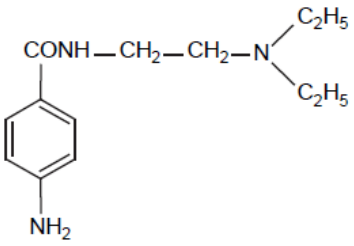
Aminopterin

ii. Divalent atoms and groups

–C=S,–C=O, –C=NH, –C=C–



Procaine



Procainamide

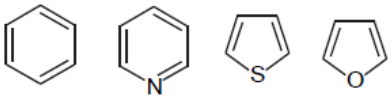
iii. Trivalent atoms and groups

–CH =, –N =
–P =, –As =

iv. Tetravalent atoms and groups

$=N^{+}=$, =C=, $=P^{+}=$, $=As^{+}=$

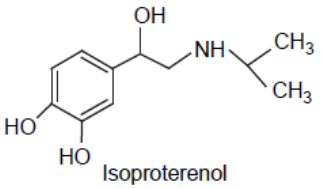
v. Ring equivalents



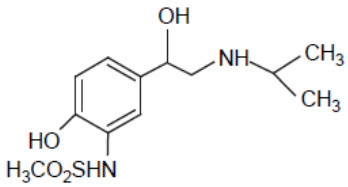
Nonclassical bioisosteres: They do not obey the steric and electronic definition of classical isosteres. Also, they do not have the same number of atoms as replacement. Although, many of these functional moieties practically just behave as one, they have one of the following characteristic features, such as

- electronic properties
- physicochemical properties
- spatial arrangements
- functional moiety critical for biological activity

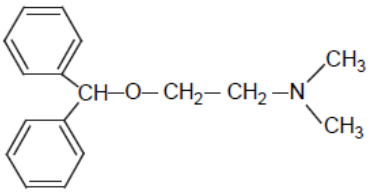
1. Exchangeable groups



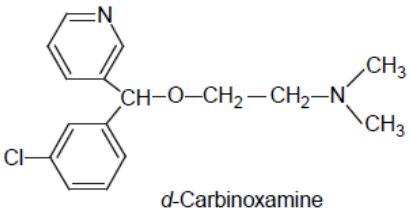
Isoproterenol



Soterenol

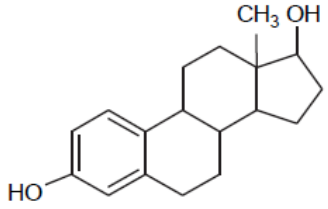


Diphenhydramine

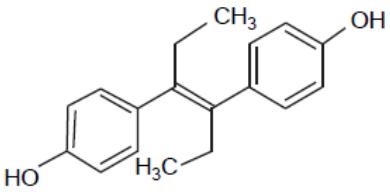


d-Carbinoxamine

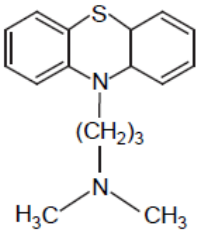
2. Cyclic versus noncyclic structure



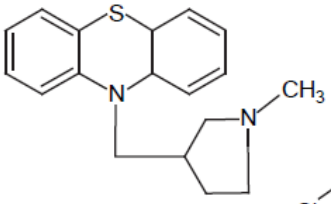
Estradiol



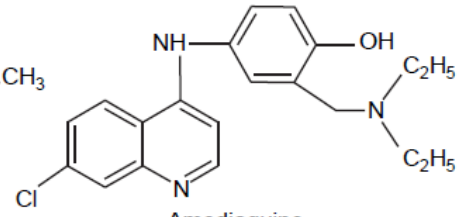
Diethylstilbestrol (*trans*)



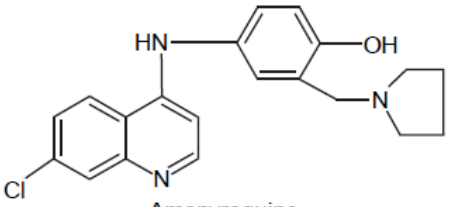
Promazine



Methdilazine



Amodiaquine



Amopyroquine

References:

1. V.Alagarsamy; Textbooks of Medicinal Chemistry volume 1; Elsevier; ISBN: 978-81-312-2189-1.