Name of Unit	Introduction to Drug Design/Combinatorial Chemistry
Subject /Course name	Medicinal Chemistry-III
Subject/Course ID	BP-601T
Class: B.Pharm. Semester	VI
Module no.	1
Course coordinator	Ms. Noel
Mobile No.	6280907651
Email id	Noelchem00@gmail.com

Learning Outcome of Module-5

LO	Learning Outcome (LO)	Course Outcome Code
LO1	To Understand the importance of drug design	BP 601.6
LO2	To Understand the different techniques of drug design	BP 601.6
LO3	To Understand the concept of combinatorial chemistry	BP 601.6

Content Table

Topics
• Various approaches used in drug design. • Physicochemical parameters used in quantitative structure activity relationship • (QSAR) such as partition coefficient, Hammet's electronic parameter, Tafts • Steric parameter and Hansch analysis. • Pharmacophore modeling and docking techniques. • Concept and applications of combinatorial chemistry: solid phase and solution • Phase synthesis.

INTRODUCTION TO DRUG DESIGN

Drug design involves the development of analogues and prodrugs 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. For example, synthesis of *trans*-diethylstilbosterol by the modification of oestradiol produced better oestrogenic activity than the latter one.

The term prodrug implies an appropriate derivative of a drug. The prodrugs are discovered by the screening of active metabolites after in vivo biotransformation that are formed from parent compounds. For example, phenylbutazone is transformed into oxyphenbutazone by hydroxylation reaction mediated by phase I (nonsynthetic) metabolic reaction.

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 assess 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.

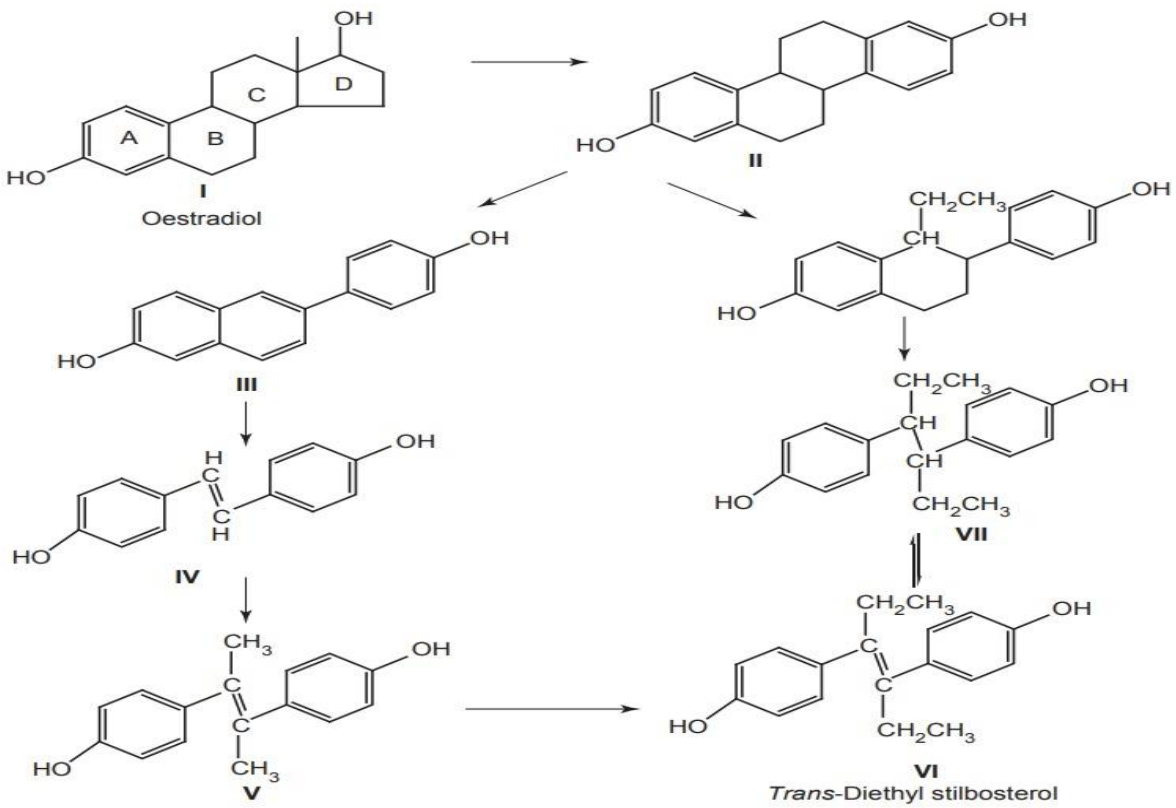
Drug Design through Disjunction

Disjunction comes into the systemic formulation of analogues of a prototype agent, generally, towards structurally simpler products, which may be carried as quasi-replicas or partials of the prototype agent. This is employed by various methods, that is

- Unjoining of certain bonds.
- Substitution of aromatic cyclic system for saturated bonds.
- Elimination of the size of hydrocarbon portion of the parent molecule.
- Decyclization of any ring system.
- Cyclization of hydrocarbon chains.

For example, oestrogenic study of oestradiol through drug design by disjunction produced successful molecules of *trans*-diethylstilbosterol.

The disjunction of various steps in the design from II to III and to IV has not successively produced a reliable molecule, but it has succeeded in the total elimination of ring B and C in estradiol (I). By plotting the response curve, the maximal activity in the series was attributed to *trans*-diethylstilbosterol (D_1) and the possibility of reductions in activity depends on the distance between two hydroxyl groups.



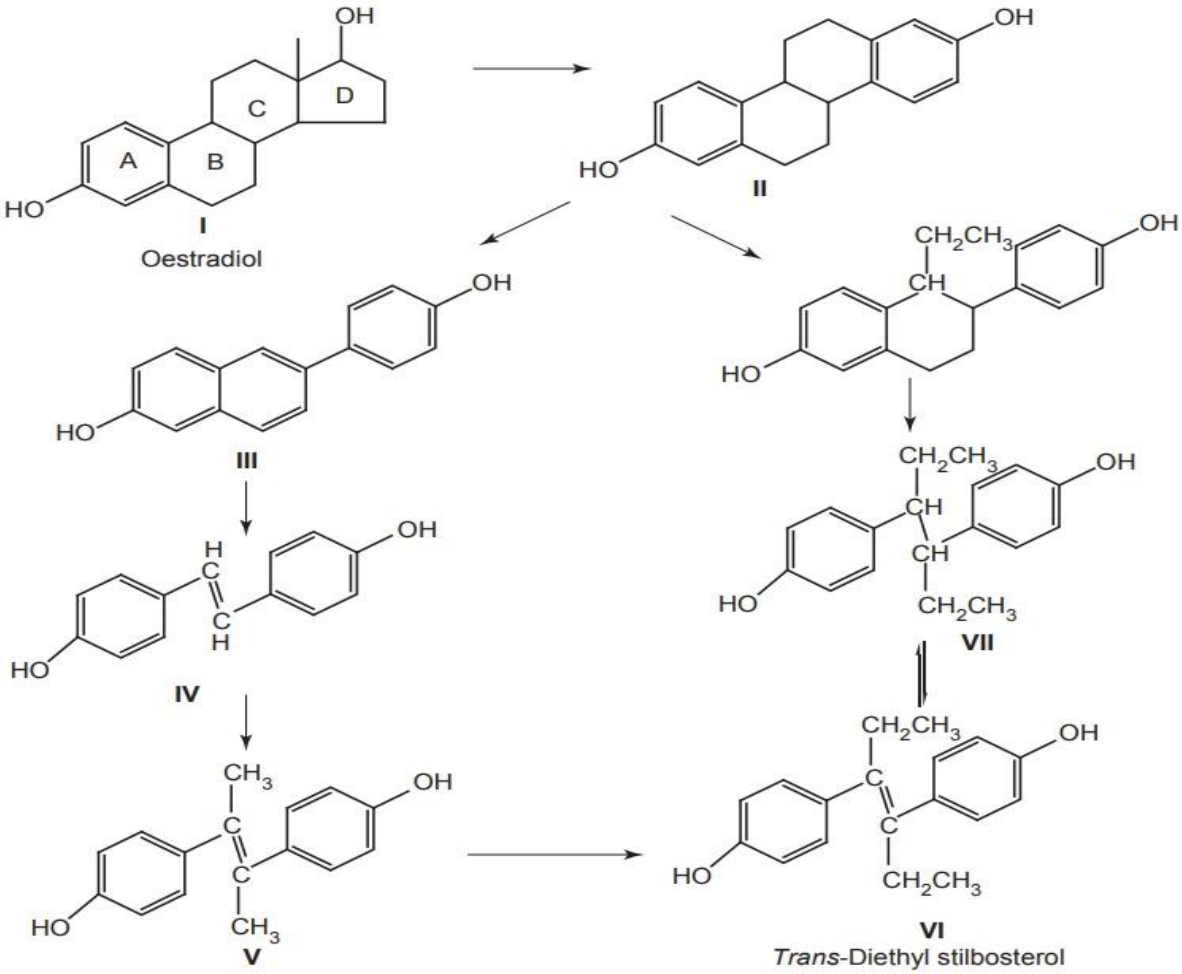
Drug Design through Conjunction

In this method, a systemic formulation of analogues of a prototype agent is employed. In general, principally mixed moieties are derived by conjunction of two pharmacophoreic molecules. An example for this is ganglionic blocking agents and its development is based on the principle of mixed moieties. Acetylcholine is a neurotransmitter, which acts as a parasympathetic muscarnic stimulant and produces appreciable changes in ganglionic functions; whereas, hexamethonium is a ganglionic blocker, and posses only a slight action at postganglionic parasympathetic endings and produces a high degree of ganglionic blockade. The evaluation of Muscarnic moiety on being studied in relation with a particular bisquartenary type of structure, for example, hexamethonium, promptly suggests the following proposed design, thus, embodying the ganglionic moiety and Muscarnic moiety into a single molecule. It is, however, pertinent to mention here that the internitrogen distance essentially constitute an important factor in many of the series of *bis* quarternary salts possessing ganglionic blocking activity. It is worthwhile when the distance is more or less the same as that in hexamethonium. However, the actual synthesis and pharmacological evaluation of conjunctioned hexamethyl analogue reveals the presence of both a weak Muscranic stimulant and possessing of a good ganglionic blocking action.

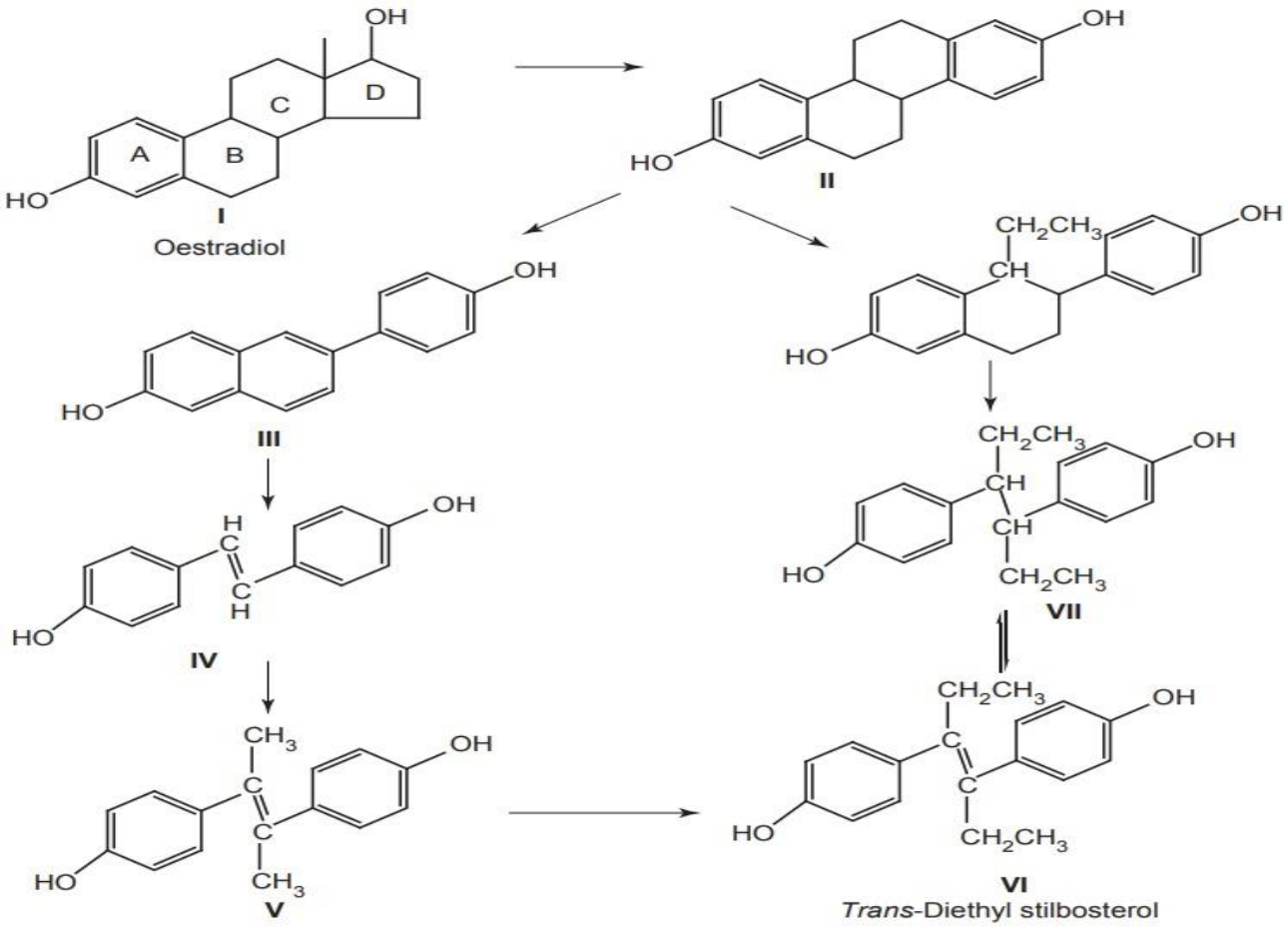
MOLECULAR HYBRIDIZATION IN DRUG DESIGN

Molecular hybridization essentially embodies the synthesis of strategically designed new breeds of bioactive agents from two or more compounds having different characteristic features with the aid of covalent-bond synthesis. In 1886, Necki exploited the beneficial properties of phenols and carboxylic acids possessing potent antibacterial characteristic feature into the design of newer drug molecules with better and improved pharmacological activities by simple esterification.

Example 1. A molecule of streptomycin and a molecule of isoniazid by means of a strong double bond between 'C' and 'N' forms a hybridized molecule through the elimination of a molecule of water. This hybridized molecule exhibits significant potentiated antibacterial and tuberculostatic activity.



2. Hybridization of acetyl salicylic acid (antipyretic) and quinine (antimalarial) to lose a molecule of water to form a hybridized molecule for a potent antimalarial drug with substantial antipyretic and analgesic activity.



APPROACH TO DRUG DESIGN

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.

Quantitative Structure Activity Relationship

Quantitative structure-activity relationship (QSAR) study, collectively referred to as QSAR, 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 chemo metrics.

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:

Log BA = contribution of unsubstituted parent compound + contribution of corresponding substituents

$$\text{Log BA} = \mu + \Sigma_i\, a_i\, a_j$$

where ai = 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.

$$\text{Log BA} = k_1 \pi + k_2 \sigma + k_3 Es + k \text{ (Hansch analysis)}$$

$$\text{Log BA} = \mu + \Sigma a_i a_j \text{ (Free-Wilson approach)}$$

So, the mixed approach can be written as

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

Where $\Sigma (a_i a_j)$ 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

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

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_{30} 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

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

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

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 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

$$q^2 = 1.0 - \Sigma (Ypred - Yobs)^2 / \Sigma (Yobs - Ymean)^2$$

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,

$$K = e^{-\Delta G/RT} = e^{-\Delta H - T\Delta S)/RT},$$

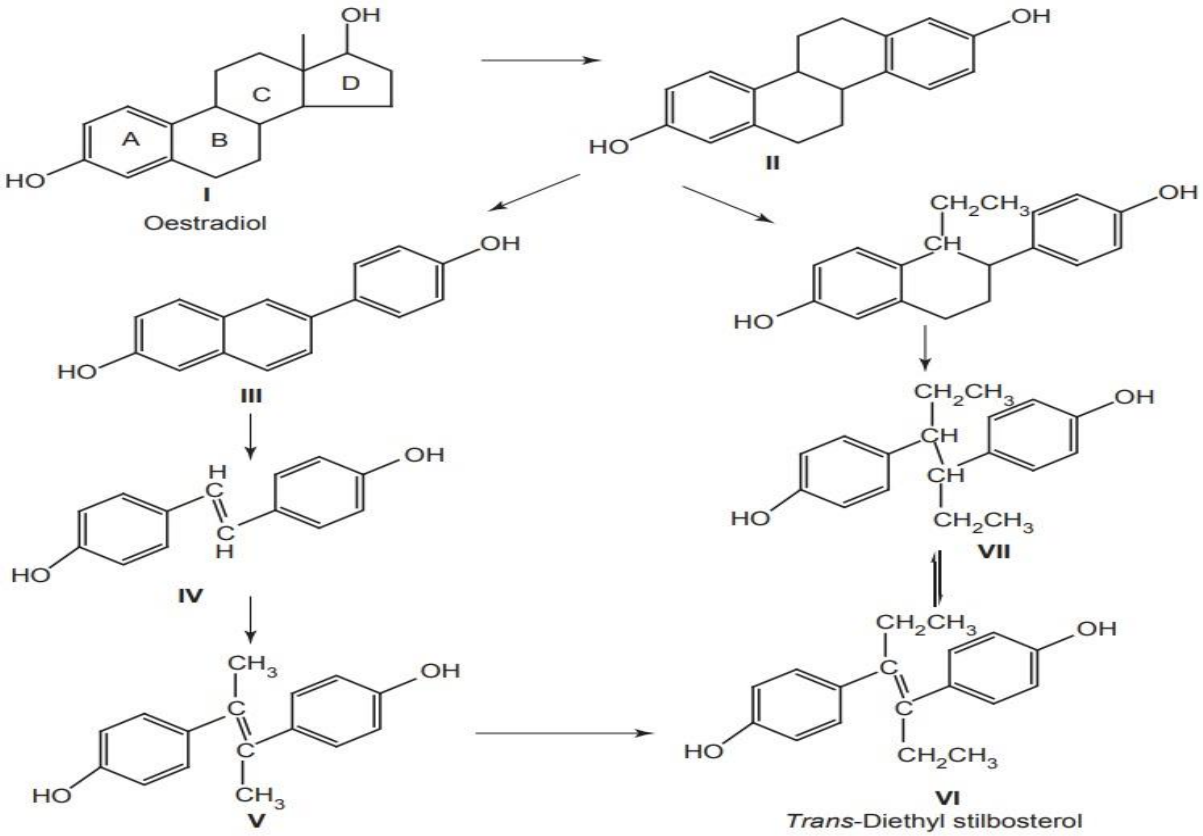
such relationships should use logarithmic scale. For these the biological inhibitory values, that is, IC50 or ED50 or LD50 or *Ki* must be converted into logarithmic form, such as log ($1/IC_{50}$) or log ($1/ED_{50}$) or log ($1/LD_{50}$) or log ($1/K_i$) values to obtain appropriate activity parameters 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 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.

There are many important aspects that need to be considered for developing a good CoMFA model. They include the following factors:

- Biological data, selection of compounds, and series design, generation of 3D structure of ligand molecules.
- Conformational analysis of each molecule.
- Establishment of bioactive conformation of each molecule, binding mode and superimposition of the molecules.

- Position of lattice points, choice of force fields and calculation of interaction energies.
- Statistical analysis of the data and selection of the 3D QSAR model.
- Display of results in contour plots and interpretation of them, design and forecasting the activity of unknown compounds.

COMPARATIVE MOLECULAR SIMILARITY INDICES ANALYSIS (COMSIA)

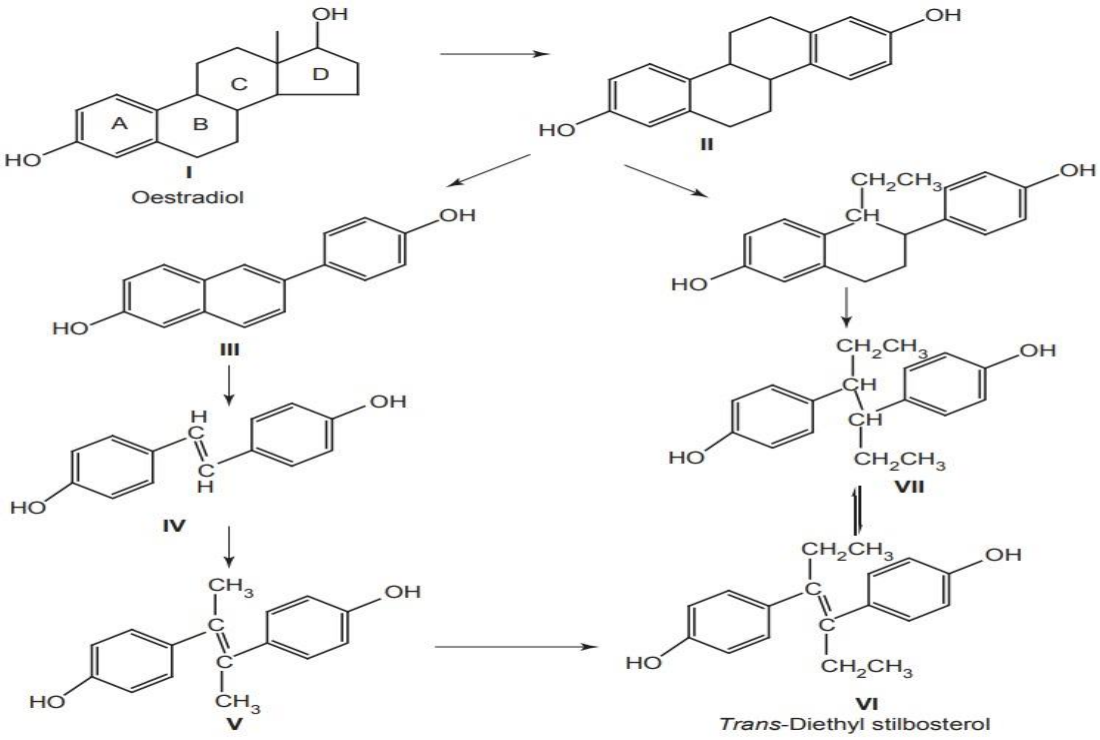
The general methodology and crucial variables for CoMSIA are same as for CoMFA. The primary difference between them is that in case of CoMFA, the contribution due to dispersion forces between molecules are described by Lennard-Jones potential and electrostatic properties are characterized by Coulomb-type potential while in CoMSIA a special Gaussian function is considered for calculation of interaction energies. CoMSIA avoids some of the inherent deficiencies arising from the functional form of the LennardJones and Coulomb potentials used in the original version of CoMFA. Both the potentials are very steep, close to the Vander Waal's surface and produce singularities at the atomic positions. As a consequence, the potential energy expressed at the grid points in the proximity of the surface changes dramatically. To avoid unacceptably large energy values, the potential evaluations are normally restricted to the regions outside the molecules and require the definition of some arbitrarily determined cutoff values. Due to the differences in the slope of the Lennard-Jones and Coulomb potentials, these cut-off values are exceeded at different distances from the molecules, requiring further arbitrary scaling of the two fields in a simultaneous evaluation, which can involve the loss of information about one of the fields. To overcome such problems, CoMSIA evaluates molecular similarity in space. Furthermore, in addition to the steric and electrostatic fields, CoMSIA defines explicit hydrophobic and hydrogen bond donor and acceptor descriptor fields, which are not available with standard CoMFA.

COMBINTORIAL CHEMISTRY

Introduction

Combinatorial chemistry is a technique through which large numbers of structurally distinct molecules may be synthesized at a time and submitted for high throughput screening (HTS) assay. Combinatorial chemistry is one of the recent methodologies developed by researchers in the pharmaceutical industry to reduce the time and costs associated with producing successful and competitive new drugs. By accelerating the process of biologically active compounds, this method is having a profound effect on all the branches of chemistry, especially on drug discovery. Through the rapidly evolving technology of combinatorial chemistry, it is now possible to produce compound libraries to screen for novel bioactivities. This powerful new technology has begun to help pharmaceutical companies to find novel drug candidates quickly, save significant money in preclinical development costs, and ultimately change their fundamental approach to drug discovery.

The aim of this chapter is to provide a basic introduction to the field of combinatorial chemistry describing the development of major techniques and applications.

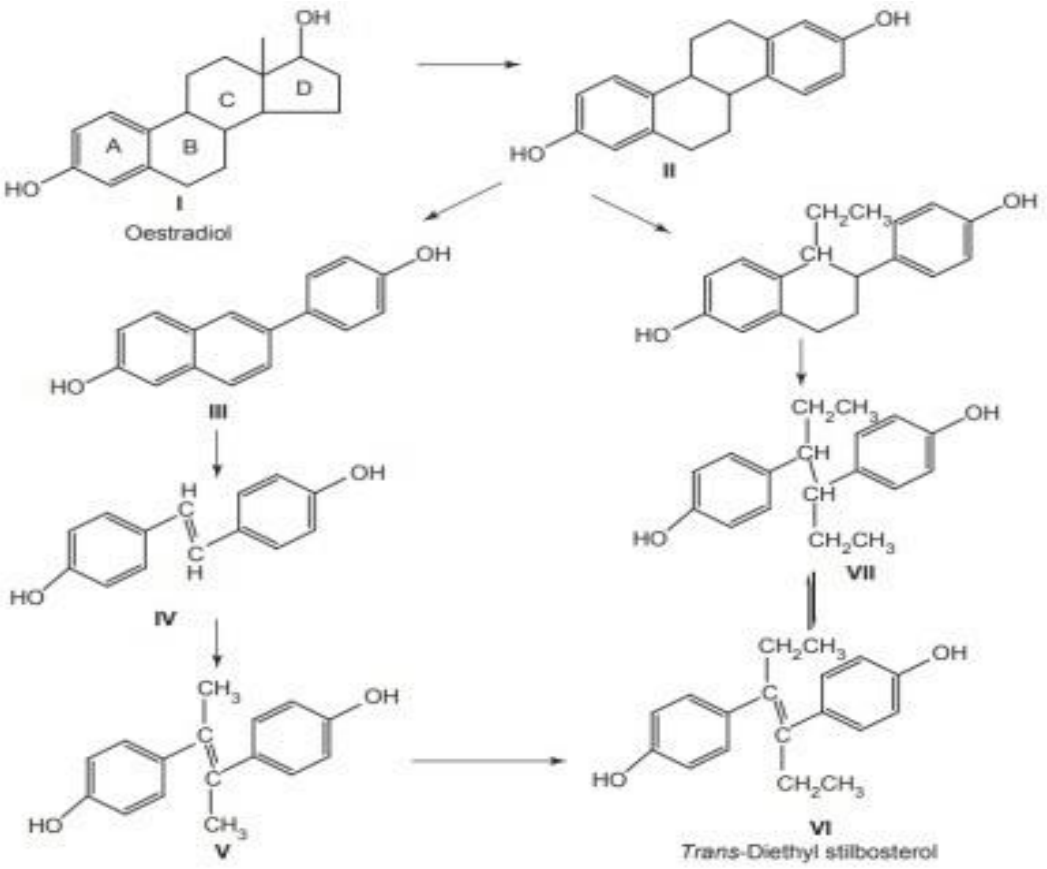


A + B → AB

Orthodox synthesis

Principles of Combinatorial Chemistry

The key of combinatorial chemistry is that a large range of analogues are synthesized using the same reaction conditions and the same reaction vessels. In this way, the organic chemist can synthesize hundreds or thousands of compounds at one time instead of preparing only a few by a traditional methodology. For example, compound A would have been reacted with compound B to give product AB, which would have been isolated after reaction, work up, and purification.



In contrast to this approach, combinatorial chemistry offers the potential to make every combination of a compound A1 to An with compound B1 to Bn.

The range of combinatorial techniques is highly diverse, and these products could be made individually in a parallel or in mixtures, using either solution or solid-phase techniques. Whatever be the technique used, the common denominator is that productivity has been amplified beyond the levels that have been routine for the last hundred years.

Applications of combintorial chemistry

The synthesis of small molecules for non- proteinase targets the synthesis of proteinase inhibitors, often peptidic in nature, is close to the roots of combinatorial chemistry which originated in solid-phase peptide synthesis. Combinatorial chemistry has now been extended to the synthesis of a wide variety of molecular structures. This in turn has lead to its use in tackling targets across the spectrum of medicinal chemistry.

Combinatorial routes to oligosaccharides a number of groups are developing new methods to allow the construction of libraries of oligosaccharides. An interesting solution-phase approach was presented by Wong who designed a building block with four orthogonal protection groups.

Multi-component reactions

Multi-component reactions (MCR) allow the assembly of several building blocks in a single chemical step into a single product. The efficiency of this approach has attracted several groups building on the pioneering work of Ugi, amongst others. Ugi reactions remain some of the most popular methods for the synthesis of libraries of heterocyclic ring systems by MCR. Rhône-Poulenc recently reported a solution-phase synthesis of diketopiperazines using a Ugi/de-Boc/cyclization sequence.

COMBINATORIAL COMPOUND LIBRARIES

The origin of combinatorial chemistry lies in the use of solid supports for peptide synthesis. By coupling the growing peptide to a solid support, such as a polystyrene bead, it is possible to use excess reagents and so ensure that the reaction proceeds to completion. Any excess reagent is simply washed away. In the original applications of solid-phase chemistry to peptide synthesis the goal was generally the synthesis of a single molecular target. A key breakthrough was the recognition that this methodology could be used to generate large numbers of molecules using a scheme known as *split–mix* technique. This technique starts with a set of reagents (which we may also refer to as monomers), each of which is coupled to the solid support. These are then mixed together and divided into equal-sized aliquots for reaction with the second reagent. The products from this reaction are reacted with the third reagent, and so on. If the number of reagents at each step are n_1, n_2, n_3, etc., then the total number of molecules produced is the product is $n_1n_2n_3$. The size of the library, thus, increases exponentially with the number of reagents—hence the use of the term ‘combinatorial’.

The original split-mix method is capable of producing extremely large libraries, but it does suffer from some drawbacks. A particular limitation is that due to the various mixing stages the identity of the product on each bead is unknown (except for the final reagent). It is important to note, however, that each bead contains just one discrete chemical entity. In the recent years, many progress has subsequently been made in the technology for automated synthesis and purification since the first reports were published. These developments have enabled many of the limitations of the early combinatorial techniques to be overcome, making automated synthesis methods a common place in both industrial and academic laboratories.

COMBINTORIAL SYNTHESIS ON A SOLID PHASE

In 1963, Merrifield pioneered the solid phase synthesis (SPS) work, which earned him a nobel prize. Merrifield's SPS concept was first applied for a developed biopolymer, recently it has spread in every field where organic synthesis is involved. Nowadays, many academic laboratories and pharmaceutical companies focused on the development of the technologies and chemistry suitable for SPS. This resulted in the impressive outbreak of combinatorial chemistry, which profoundly changed the approach to new drugs, new catalyst, or new natural discovery.

The utilization of solid support for the organic synthesis relies on three interconnected requirements.

These are as follows:

1. A cross-linked, insoluble polymeric material should be inert to the condition of synthesis.
2. The linking substrate (linker) to the solid phase that permits selective cleavage of some or all the products from the solid support during synthesis for analysis of the extent of reaction (s) and ultimately to give the final product of interest.
3. The chemical protection strategy must allow selective protection and deprotection of reactive groups.

ADVANTAGES AND DISADVANTAGES OF SOLID SUPPORT REAGENTS

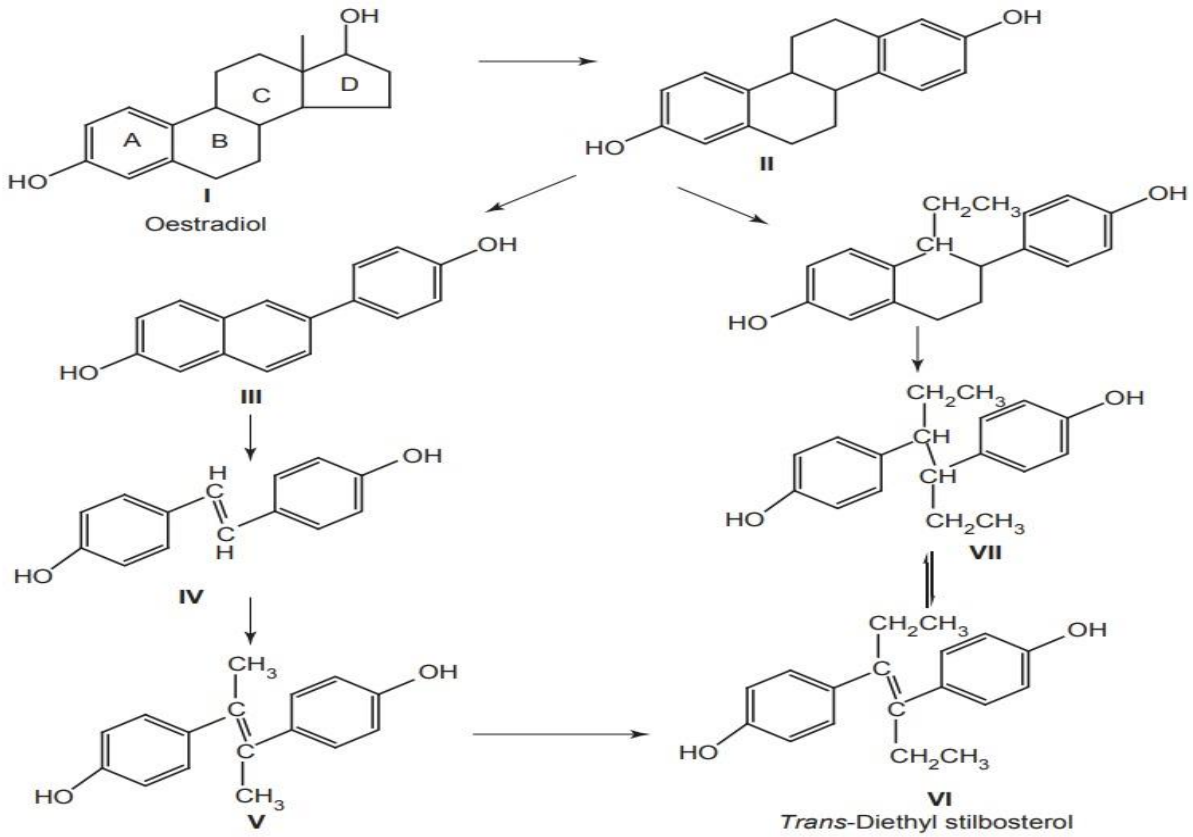
Advantages

- Solid-supported reagents are easily removed from reactions by filtration.
- Excess reagents can be used to drive reactions to completion without introducing difficulties in purification.

- Recycling of recovered reagents is economical, environmentally-sound, and efficient.
- Ease of handling is especially important when dealing with expensive or time-intensive catalysts, which can be incorporated into flow reactors and automated processes.
- Finely tune chemical properties by altering choice of support and its preparation.
- Toxic, explosive, and noxious reagents are often more safely handled when contained on solid support.
- Reagents on solid-support react differently, mostly more selectively, than their unbound counterparts.

Disadvantages

- Some reagents may not interact well with solid support.
- Ability to recycle reagents on solid support is not assured.
- Reactions may run more slowly due to diffusional constraints.
- Polymeric support materials can be very expensive to prepare.
- Stability of the support material can be poor under harsh reaction conditions.
- Side reactions with the polymer support itself may occur.



Resins for SPS

In solid phase synthesis, resin supports for SPS include spherical beads of lightly cross-linked gel type polystyrene (GPS) (1%–2% divinylbenzene) and poly(styrene-oxyethylene) graft copolymers, which are functionalized to allow attachment of linkers and substrate molecules. Each of these materials has advantages and disadvantages, depending on the particular application. There are several types of resins available for different type of reactions.

QUESTIONS

Very Short Answer Type Questions (2 Marks)

Q1 Who developed the Dock 4.0 programme of software for DOCKING?

Q2 Who developed the GLIDE programme of software for DOCKING?

Q3 Define QSAR?

Short Answer Type Questions (5 Marks)

Q1 what do you mean by quantitative structure activity relationship.

Q2 what is auto dock?

Q3 what is Docking?

Long Answer Type Questions (10 Marks)

Q.1 What do you mean by ligand preparation for binding mode preparation for binding mode assessment?

Q.2 Explain Monte Carlo method in Docking.

Q.3 What is glide program in Docking?

Q.4 What is induced fit fit method in Docking ?

Q.5 Explain the application of molecular Docking

Q.6 Write a short note on combinatorial chemistry.