Chapter 16: Drug Design

Introduction to Drug Design

Various approaches used in drug design.

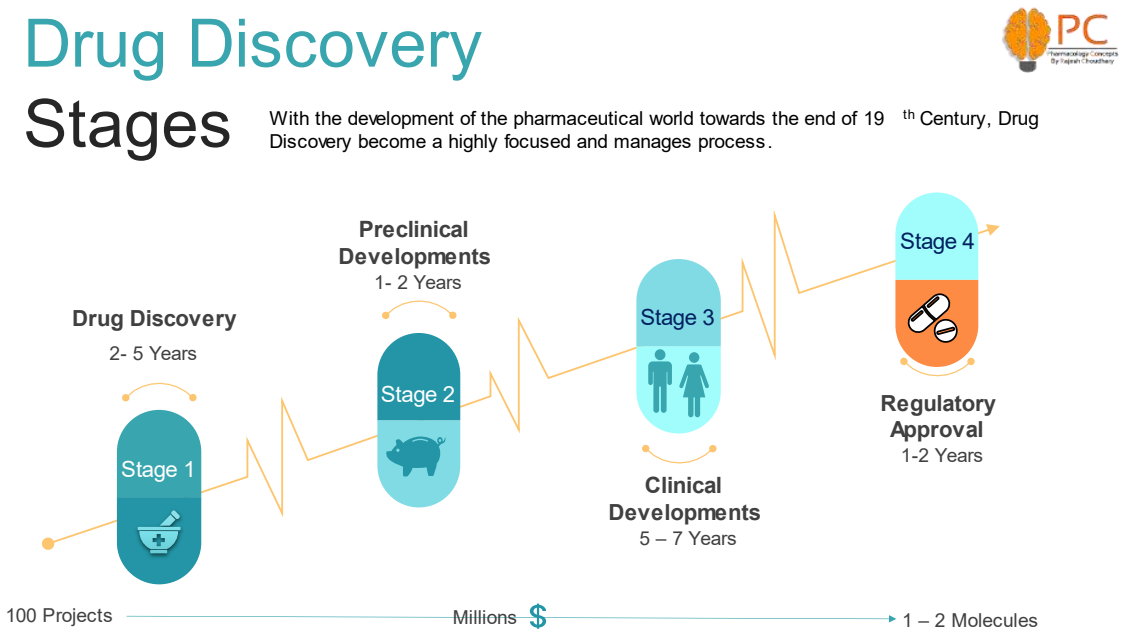
Physicochemical parameters used in quantitative structure activity relationship (QSAR) such as partition coefficient, Hammet's electronic parameter, Tafts steric parameter and Hansch analysis.

Pharmacophore modelling and docking techniques.

Combinatorial Chemistry: Concept and applications of combinatorial chemistry: solid phase and solution phase synthesis.

16.1. INTRODUCTION

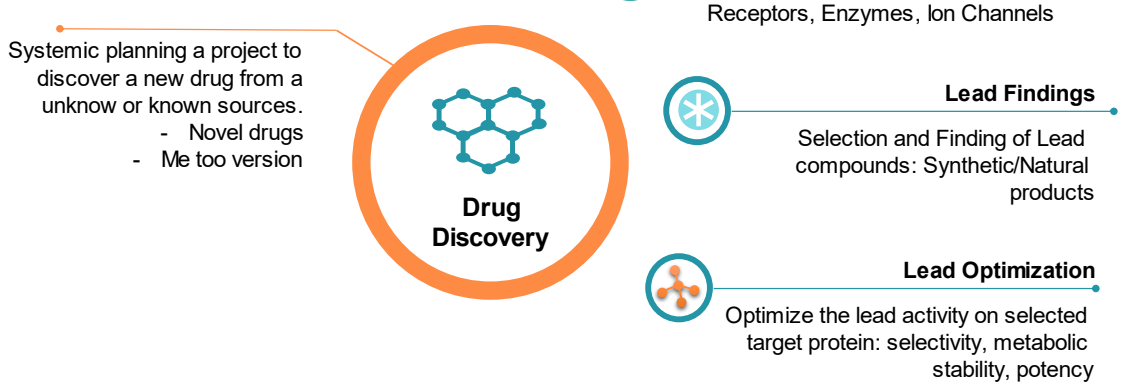
- Drug Design is the important part of the drug discovery. It is a systematic approach to finding, selection, optimization of drug molecules on the basic of molecular interactions (Strero-structural basis) between drug and target proteins and or its physico-chemical properties involved.
- The drug discovery and development are very time and resources consuming process. Due to the high research & development (R&D) costs and extensive clinical testing, drug discovery and development has become an expensive process (**Avg cost millions of USD and 10-15 years)**



Video lecture: https://youtu.be/xukCiYeywWI

Drug Discovery
Stage 1 (Drug Design)

"To discover a effective medicine, select a proper target & lead"



- Drug design or rational drug design is the innovative process of finding new drug molecules based on the knowledge of the biological target.
- The drug is mostly an organic molecule which activates or inhibits the function of target proteins that in turn results in a therapeutic effect.
- Frequently the drug design is based on a computer modeling technique known as computer-aided drug design (CADD).
- In CADD attempts are made to find a ligand that will interact favorably with a receptor or target protein that represents the target site.
- Binding of ligand to the receptor may include hydrophobic, electrostatic, and hydrogen-bonding
- The approach used in CADD is dependent upon the amount of information that is available about the ligand and receptor. Ideally, one would have 3-dimensional structural information for the receptor and the ligand-receptor complex from X-ray diffraction or NMR.
- Based on the information that is available, one can apply either ligand-based or structure based drug design methods.
- Regulatory agencies as well as pharmaceutical industry are actively involved in development of computational tools that will improve effectiveness and efficiency of drug discovery and development process, decrease use of animals, and increase predictability.
- **There are two major types or approaches to drug design.**
    - Ligand based drug design (Indirect drug design)

- Structure based drug design (Direct drug design)

Ligand based drug design (Indirect drug design): Ligand based drug design is based on the knowledge of other molecules that bind to the biological target of interest so as to derive a **pharmacophore** which will bind to the target.

(A) Q SAR

(B) Analog drug design

(C) Combinatorial chemistry

(D) Natural Products as a lead, etc

Structure based drug design (Direct drug design): Structure based drug design is based on the knowledge of the three dimensional structure of the biological target. Using the structure of the biological target, candidate drugs that are predicted to bind with high affinity and selectivity to the target may be designed.

Pharmacophore: is a group of atoms in the molecule of a drug responsible for the drug's action

16.2. RATIONAL APPROACHES & CONCEPTS

Drug design seeks to explore:

- ✓ Effects of biological compounds on the basis of molecular interaction
- ✓ Explore the various process involve in drug discovery
- ✓ Explore drug-protein interaction to elicit biological response
- ✓ Probable relationship between biological activity and chemical structure.

Concept of LEAD

Lead is the prototype bioactive molecule subjected to drug design and drug discovery and needs to exploration and exploitation.

- ✓ **Exploration of Lead:** The search for new lead.
- ✓ **Exploitation of Lead:** Requires assessment, improvement and extension of lead.

Concept of Analogue

- ✓ Chemical Derivatives or structural analogue having similar structure with little medication

Concepts of Prodrug

- ✓ Prodrug is the appropriative derivatives which converts active metabolite invo

16.3. VARIOUS APPROACHES USED IN DRUG DESIGN

- The various approaches used in drug design (**Ligand based or Structural based**) include the following.

  1) Drug discovery via Random screening of synthetic compounds or chemicals and natural products by bioassay procedures.

  2) Drug discovery via metabolic studies

  3) Drug discovery via Novel compounds preparation based on the known structures of biologically active, natural substances of plant and animal origin, i.e., lead skeleton.

  4) Drug discovery via Preparation of structural analogs of lead with increasing biological activity and Application of bio isosteric principle.

I. Ligand Based Drug Design

- **Ligand based drug design** or **indirect drug design** is an approach used in the absence of the receptor 3D information and it based on knowledge of molecules that bind to the biological target of interest.
- A model of the biological target may be built based on the knowledge of
  - what binds to it ????
  - Hydrophobic and hydrophillic groups
- 3D quantitative structure activity relationships (3D QSAR) and **pharmacophore** modeling are the most important and widely used tools in ligand based drug design.
  - *****Pharmacophore: is a group of atoms in the molecule of a drug responsible for the drug's action***
- They can provide predictive models suitable for lead identification and optimization
- **Basic Approaches:**

  (A) Q SAR

  (B) Analog drug design

  (C) Combinatorial chemistry

  (D) Natural Products as a lead, etc

A) QSAR(Quantitative Structure Activity Relationships)

- QSAR is mathematical or statistical approaches to define the relationship between biological activity (experimental data) of a molecular system and its geometrical, physical, electronic, and chemical properties

Activity = function (property 1 , property 2…..)

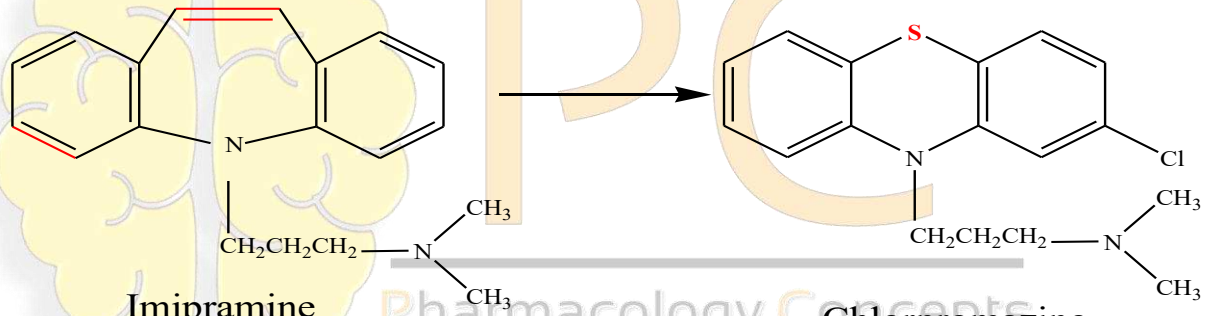
Activity = function (xi)

xi- descriptor

Property- geometry, steric, or steric etc

B) Analogue Drug Design

- Analog design is most fruitful in the study of pharmacologically active molecules that are structurally specific: their biological activity depends on the nature and the details of their chemical structure.



Imipramine Chlorpromazine

- Hence, a minor modification of the molecule may result in a profound change in the pharmacological response (increase, diminish, completely destroy, or alter the nature of the response).
- Lead compounds are frequently identified as endogenous participants (hormones, neurotransmitters, second messengers, or enzyme cofactors) in the body's biochemistry and physiology.
- A lead may result from routine, random biological screening of natural products or of synthetic molecules that were created for purposes other than for use as drugs.
- In analog design, molecular modification of the lead compound can involve one or more of the following strategies:
    - Bioisosteric replacement
    - Design of rigid analogs

- Homologation of alkyl chain(s) or alteration of chain branching, design of aromatic ring-position isomers, alteration of ring size, and substitution of an aromatic ring for a saturated one
- Alteration of stereochemistry, or design of geometri isomers (or) stereoisomers
- Design of fragments of the lead molecule that contain the pharmacophoric group (bond disconnection)
- Alteration of interatomic distances within the pharmacophoric group or in other parts of the molecule

- Bioisosteric replacement strategy has been fruitful in design of psychoactive agents, by use of the antidepressant dibenzazepine derivative **Imipramine** as the lead. The structural similarity between imipramine and the phenothiazine antipsychotics [typified by **chlorpromazine** is apparent.

C) Combinatorial chemistry

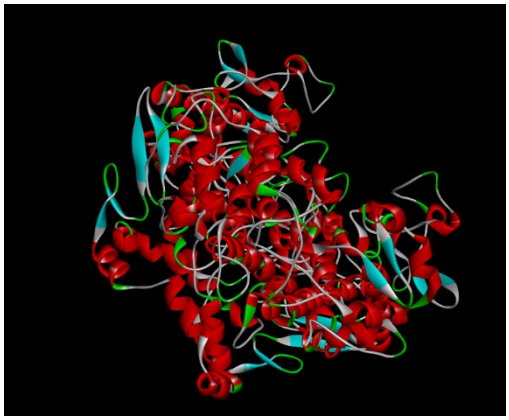
- Combinatorial chemistry comprises chemical synthetic methods that make possible to prepare large number (tens to thousands or even millions) of compounds in a single process.
- These compound libraries can be made as mixtures, sets of individual compounds or chemical structures generated in computer

D) Natural Product as Lead

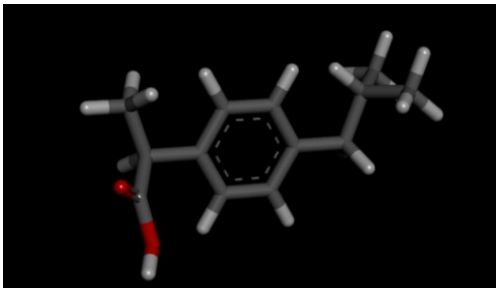
- A total no. of 520 new pharmaceuticals approved between 1983 and 1994, among which 39% were derived from natural products, the proportion of antibacterials and anticancer agents of which was over 60%.
- Between 1990 and 2000, a total of 41 drugs derived from natural products were launched on the market by major pharmaceutical companies including azithromycin, orlistat, paclitaxel, sirolimus (rapamycin), Synercid, tacrolimus, and topotecan

II. Structure Based Drug Design

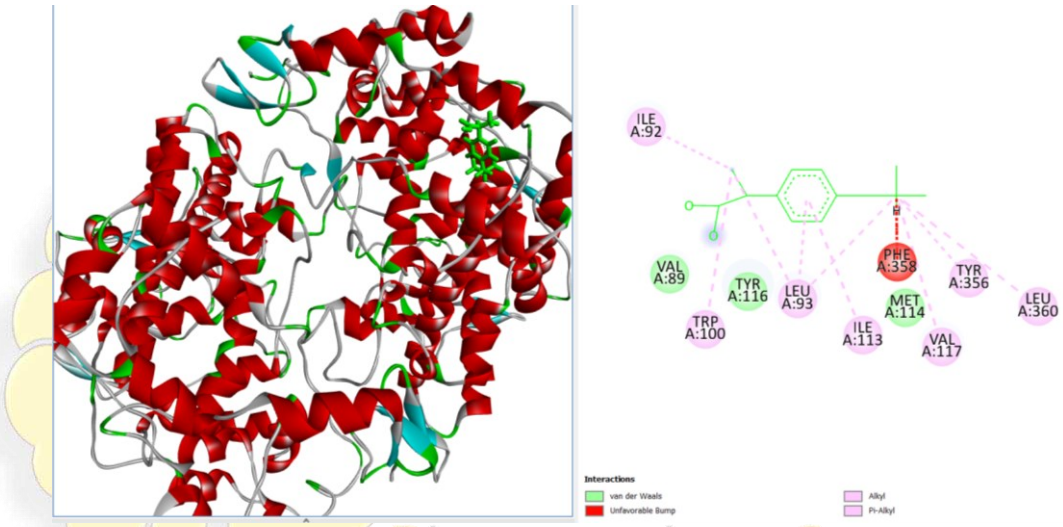
- Structure-based drug design (or direct drug design) relies on knowledge of the three-dimensional structure of the biological target obtained through methods such as x-ray crystallography or NMR spectroscopy



Structure of COX-2

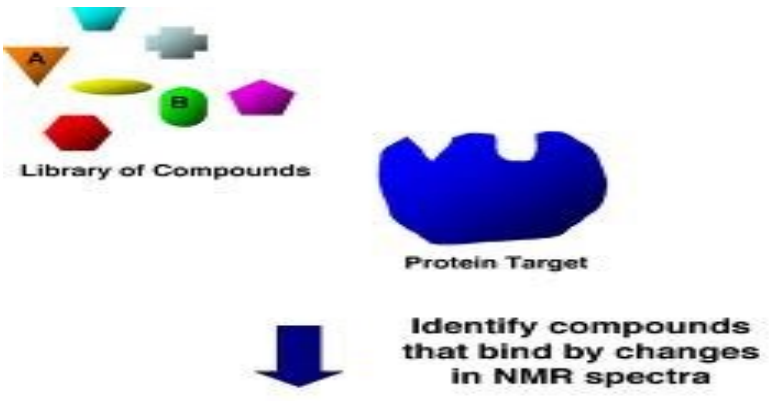


Ibuprofen



COX-2- Ibuprofen Interaction

- The ultimate goal of structure-based drug design is a simple robust process that starts with high resolution crystal structure of a validated biological macromolecular target and reliably generates an easily synthesized, high-affinity small molecule with desirable pharmacological properties.

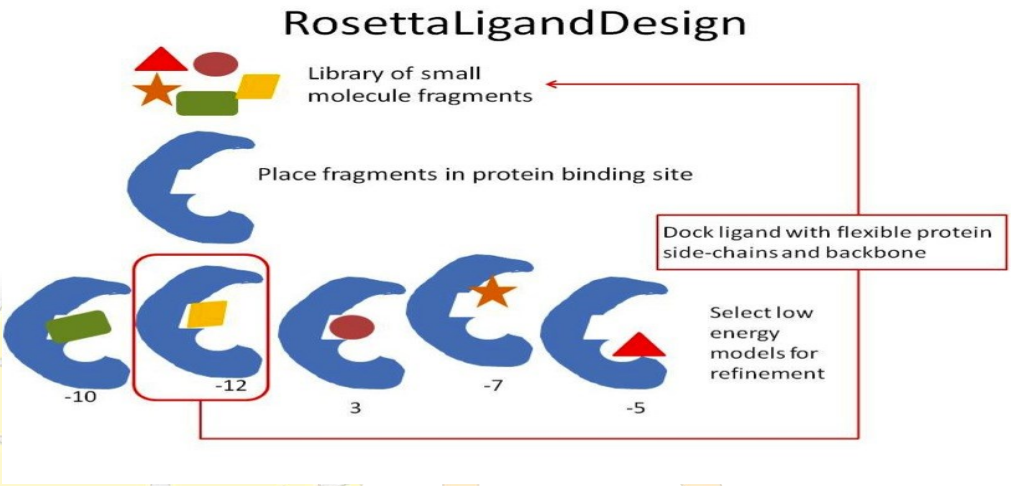


- The prediction of affinity based on the docking score, developed by Böhm to develop a general-purposed empirical scoring function in order to describe the binding energy. The following "Master Equation" was derived:

$$\Delta G_{bind} = -RT \ln K_d$$

$$K_d = \frac{[Receptor][Acceptor]}{[Complex]}$$

$$\Delta G_{bind} = \Delta G_{desolvation} + \Delta G_{motion} + \Delta G_{configuration} + \Delta G_{interaction}$$



(A) Computerised Drug Design:

1. hit identification: using virtual screening (structure- or ligand-based design)

2. hit-to-lead optimization: affinity and selectivity (structure-based design, QSAR, etc.)

3. **lead optimization:** optimization of other pharmaceutical properties while maintaining affinity

(B) Rational Drug Design:

- A drug target is a key molecule involved in a particular metabolic or signaling pathway that is specific to a disease condition or pathology or to the infectivity or survival of a microbial pathogen.
- The first unequivocal example of the application of structure-based drug design leading to an approved drug is the **carbonic anhydrase inhibitor dorzolamide**, which was approved in 1995.
- Another important case study in rational drug design is imatinib, a tyrosine kinase inhibitor.

QSAR (QUANTITATIVE STRUCTURE ACTIVITY RELATIONSHIPS)

- QSAR is a computational modeling method for revealing relationships between structural properties of chemical compounds and biological activities
- Hansch, (1964)- Structural properties of a chemical influence its biological activity and similar compounds behave similarly.
- So QSAR is mathematical or statistical approaches to define the relationship between biological activity (experimental data) of a molecular system and its geometrical, physical, electronic, and chemical properties

Activity = function (property 1 , property 2…..)

Activity = function (xi)

xi- descriptor

Property (xi)- size, shape, no. of H-bond, electrostatic etc

- **Advantages**:
  - a) we can predict the activity of unknown or new compound
  - b) Short list of active compounds
  - estimation of physical/chemical property of unknown compounds based on knowledge of known compounds
- In molecular docking the geometrical structure of both the ligand and the target protein must be known. But the Quantitative Structure-Activity Relationships (QSAR) is a method which can be applied regardless of whether the structure is known or not.
- QSAR explore how a given protein interacts with some tested compounds. As an example, it may be known from previous experiments that the protein under investigation shows signs of activity against one group of compounds, but not against another group. In terms of the lock and key metaphor, we do not know what the lock looks like, but we do know which keys work, and which do not.
- In order to build a QSAR model for deciding why some compounds show sign of activity and others do not, a set of descriptors are chosen. These are assumed to influence whether a given compound will succeed or fail in binding to a given target. The parameters such as molecular weight, molecular volume, electrical and thermodynamical properties are used as descriptors.
- QSAR derive models which describe the structural dependence of biological activities either by physicochemical parameters (**Hansch analysis**), by indicator variables encoding different structural features (**Free Wilson analysis**), or by three-dimensional molecular property profiles of the compounds (**comparative molecular field analysis, CoMFA**).

QSAR based on **Hammett's relationship** utilize electronic properties as the descriptors of structures.

Descriptor/ Properties/ Features:

- ✓ Physico-chemical properties
- ✓ Electronic
- ✓ Steric
- ✓ Lipophilicity
- ✓ Hydrogen bonding
- ✓ Shape
- ✓ Charge
- ✓ Polarizability

Activity = function (Descriptors)

Molecular Descriptor Used in QSAR

A) Hydrophobic Parameter

- ➢ Partition coefficient; log P
- ➢ Hansch's Substitution Constant; π
- ➢ Hydrophobic fragmental constant; f, f'
- ➢ Distribution Cofficient; log D
- ➢ Solubility parameter; log S

B) Electronic Parameters

- ➢ Hammett costant; σ, σ+, σ-
- ➢ Taft's inductive (polar) constant; σ*
- ➢ Swain and Lupton field parameter
- ➢ Ionization constant; pKa, ΔpKa

C) Steric Parameter

- ➢ Taft's steric parameter; Es
- ➢ Molar volume; MV
- ➢ Van der waals radius and volume
- ➢ Molar refractivity; MR

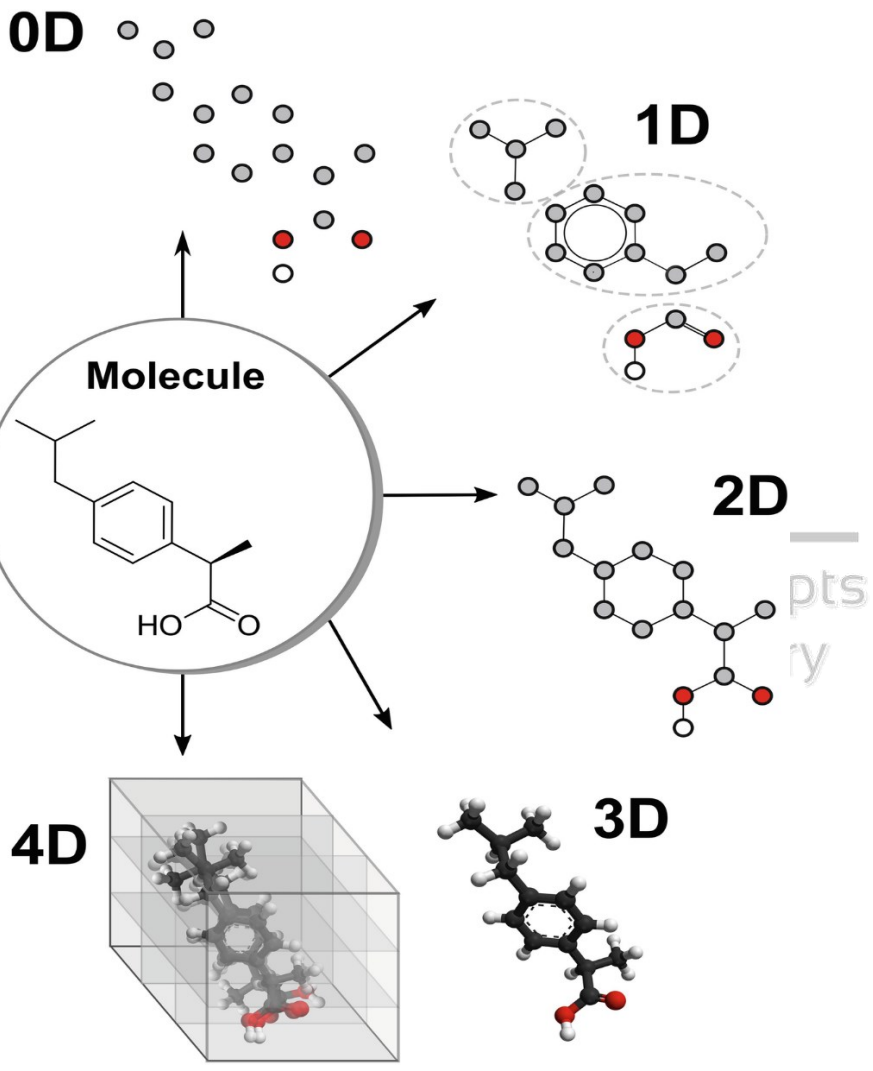
D) Quantum Chemical Descriptor

- ➢ Atomic net charge; Qσ, Q π
- ➢ Super Delocalizability
- ➢ Energy of highest occupied molecular orbital; EHOMO
- ➢ Energy of lowest unoccupied molecular orbital; ELOMO

E) Spatial Descriptor

- Jurs descriptor
- Shadow indices
- Radius of Gyration
- Principle moment of inertia

Descriptors based on the dimensionality of their molecular representation



A) 0-Dimension descriptors

Atom count, bond counts, molecular weight, sum of atomic properties, chemical Formula

Examples: Molecular weight, average molecular weight number of: atoms, hydrogen atoms carbon atoms, hetero-atoms, non-hydrogen atoms, double bonds, triple bonds, aromatic bonds, rotatable bonds, rings, 3 or 4 or 5 or 6 – membered ring,

B) 1-Dimension descriptors: Fragments counts

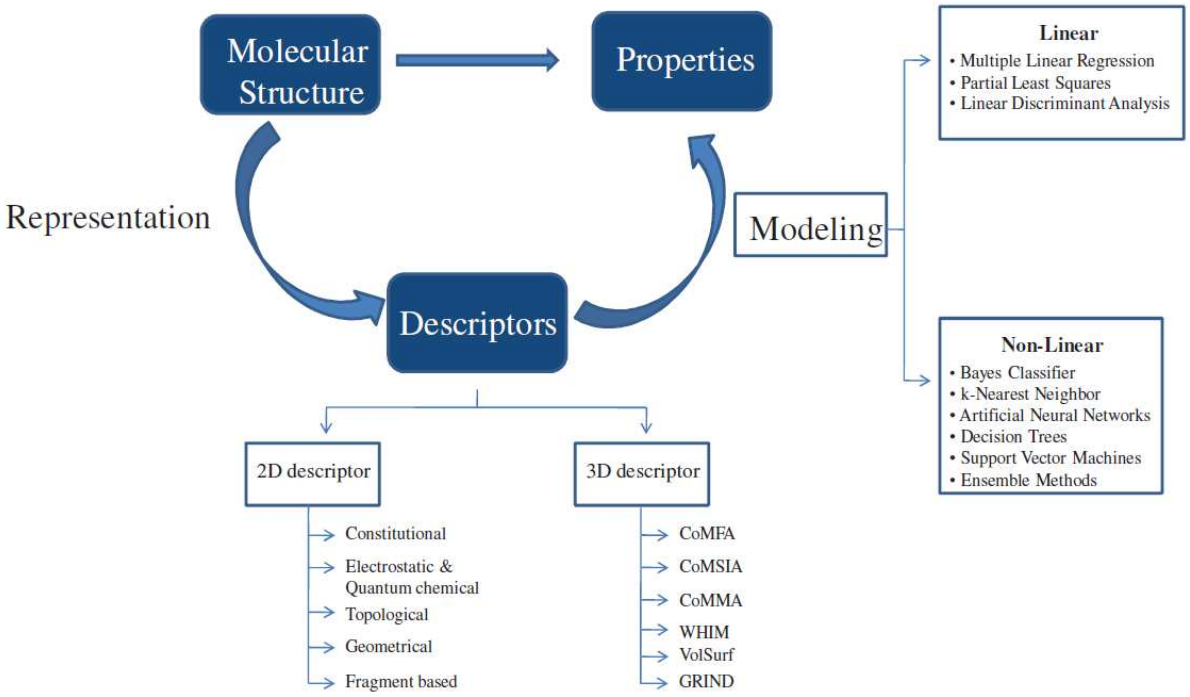
Examples: Number of: primary C, secondary C, tertiary C, quaternary C, secondary carbon in ring, tertiary carbon in ring, quaternary carbon in ring, unsubstituted aromatic carbon, substituted carbon, number of H-bond donar atoms, number of H-bond acceptor atoms, unsaturation index, hydrophilic factor, molecular refractivity

C) 2-Dimension descriptors: Topological descriptors

Example: Zagreb index, Wiener index, Balaban J index, connectivity indices, chi (x), kappa (K) shape indices

D) 3-Dimension descriptors: Geometrical descriptors

Example: Radius of gyration, E-state topological parameters, 3D Wiener index, 3D Balaban index, Surfaces (Van der Waals Surface Area / Solvent-accessible Surface Area / Polar surface area / lsotropic surface area-Non-specific interactions with the solvent / Hydrophobic — Hydrophilic surface area)/ Volumes (Solvent-Excluded Volume) Geometrical / Electrostatic / Charged partial surface areas / HOMO and LUMO Potential energy (Energy differences / Active conformation — global minimum conformation / Binding energies / lonization potential



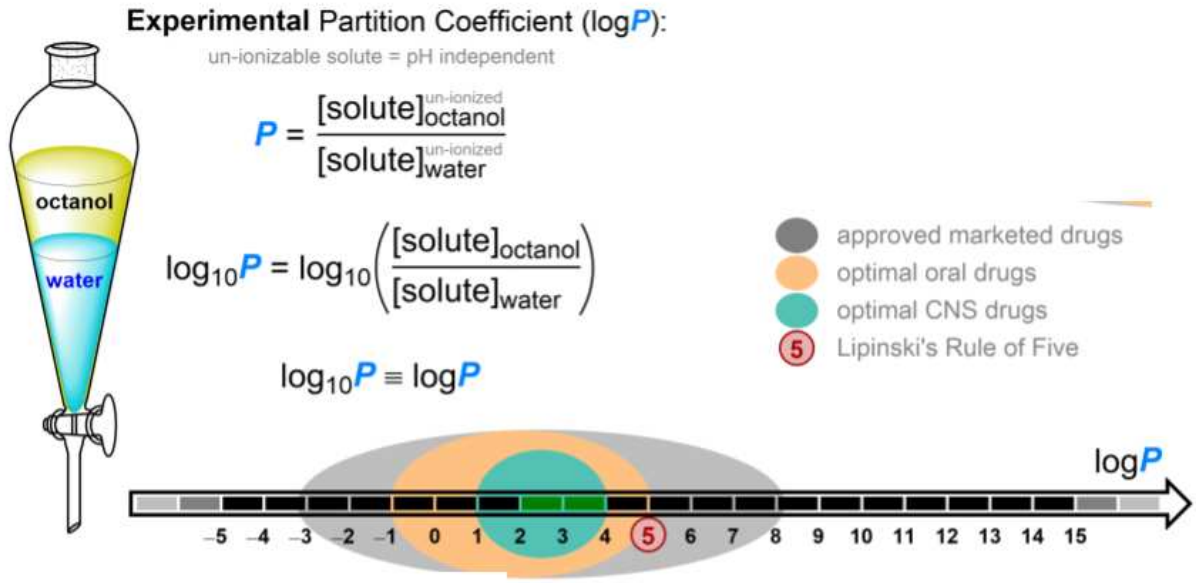
PHYSICOCHEMICAL PARAMETERS USED QSAR

Partition coefficient, Hammet's electronic parameter, Tafts steric parameter and Hansch analysis.

Partition Coefficient (Log P)

- Log P, or octanol-water partition coefficient, is a measure of hydrophilicity and hydrophobicity of the compounds.
- In the physical sciences, a partition coefficient (P) or distribution coefficient (D) is the ratio of concentrations of a compound in a mixture of two immiscible solvents at equilibrium.
- This ratio is therefore a comparison of the solubilities of the solute in these two liquids. The partition coefficient generally refers to the concentration ratio of un-ionized species of compound, whereas the distribution coefficient refers to the concentration ratio of all species of the compound (ionized plus un-ionized).
- In Pharmaceutical/Chemical Sciences, both phases are usually solvents. Most commonly polar solvent- water and non-polar solvent- octanol.
- Partition coefficients are useful in estimating the distribution of drugs within the body.
- **Hydrophobic drugs** with high octanol-water partition coefficients are mainly distributed to hydrophobic areas such as **lipid bilayers of cells**. Conversely, **hydrophilic drugs** (low

octanol/water partition coefficients) are found primarily in **aqueous regions such as blood serum**



- A more polar, hydrophilic compound will have a lower log P (the value can even be negative), and prefer to "reside" in the aqueous phase. (Log P < 0)
- More non-polar, hydrophobic compounds will have a higher log P, and will partition into an organic phase. Typical values range from -3 (polar) to 7 (non-polar). (Log > 0)
- Log P also depends on substituents hydrophobic (hydrophobicity constant; π>0) or hydrophilic (π<0).

π = Log P (substituted compound) – Log P (Parent compound)

$\pi = \text{Log } (R_X/R_H)$

Hydrophobic (π>0)		Hydrpophili	
CH_3	0.56	NO2	-0.28
$C(CH_3)_3$	1.98	OH	-0.67
C6H5	1.96	COOH	-0.32
C6H11	2.51	NH2	-1.23
CF3	0.88	CHO	-0.65

Component	log $P_{O/W}$	T (°C)
Acetamide	−1.16	25
Methanol	−0.81	19
Formic acid	−0.41	25
Diethyl ether	0.83	20
p-Dichlorobenzene	3.37	25
Hexamethylbenzene	4.61	25

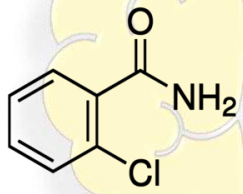
In QSAR, we can calculate the log P value by following equation

Log P(compound) = Log P (ring) + π (group 1) + π (group 2)..

- Hansch's Substitution Constant/ Hydrophobicity Constant; π

Example :

Q. find out the Log P of m-chlorobenzamide;



Log P (benzene)= 2.13,

π(cl) = 0.71,

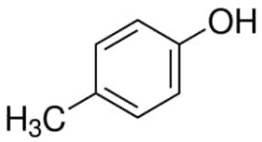
π(CONH2) = -1.49

ANS

Log P(m-chlorobenzamide) = Log P (benzene) + π (Cl) + π (CONH2)

= 2.13 + 0.71 – 1.49

= 1.35



Q. find out the Log P of p-cresol

Log P (benzene)= 2.13,

π(OH) = -0.67,

π(CH3) = 0.52

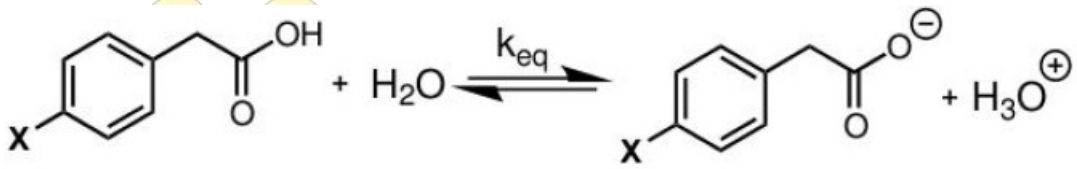
Log P(m-cresol) = 2.13-0.67+0.52 = **1.98**

Hammet's Electronic Parameter

- The Hammett Plot is a type of Linear Free-Energy Relationship (LFER) analysis designed to model the electronic effect of substituents on aromatic systems (in the para and meta positions only).
- This equation was developed and published by Louis Plack Hammett in 1937
- Information gathered can be used to probe the mechanism of the reaction and can be applied in the optimization of reaction conditions.
- Hammett equation accounts for how field, inductive, and resonance effects influence reaction rates

Relationship of σ to new reactions

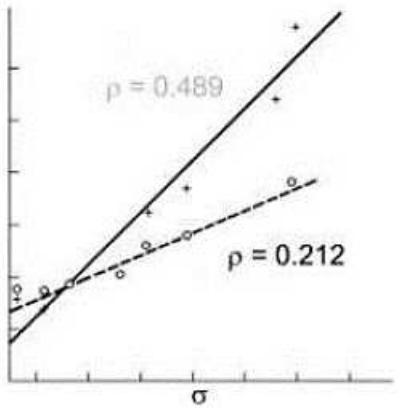
Consider similar reaction



Equation:

$$\text{Log } (K_X/K_H) = \rho\ \sigma$$

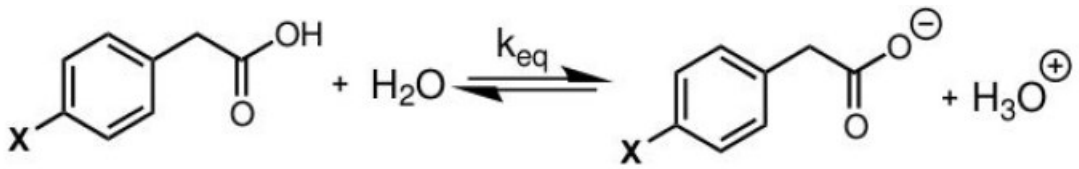
- Relating the equilibrium constant, Kx, for a given equilibrium reaction with substituent X and the reference K_H constant when X is a hydrogen atom to the **substituent constant σ**
- **ρ** is a correction factor to compare a new reaction to the original



- The reaction constant, or sensitivity constant, ρ, describes the susceptibility of the reaction to substituents, compared to the ionization of benzoic acid. It is equivalent to the slope of the Hammett plot.

- ρ>1, the reaction is more sensitive to substituents than benzoic acid and negative charge is built during the reaction (or positive charge is lost).
- 0<ρ<1, the reaction is less sensitive to substituents than benzoic acid and negative charge is built (or positive charge is lost).
- ρ=0, no sensitivity to substituents, and no charge is built or lost.
- ρ<0, the reaction builds positive charge (or loses negative charge).

Dissociation Constant (Ka) of substituted benzoic acid (K x 10^5, at 25 C)



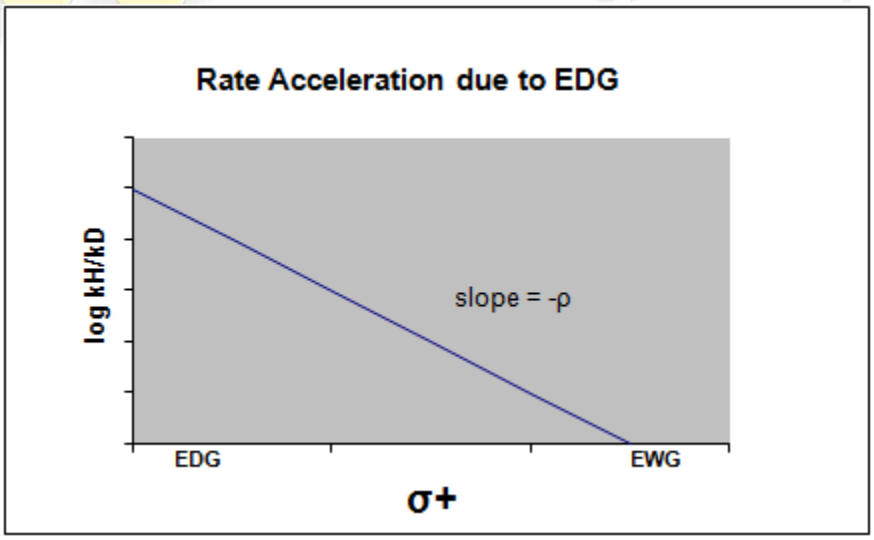
Ka of Substituents

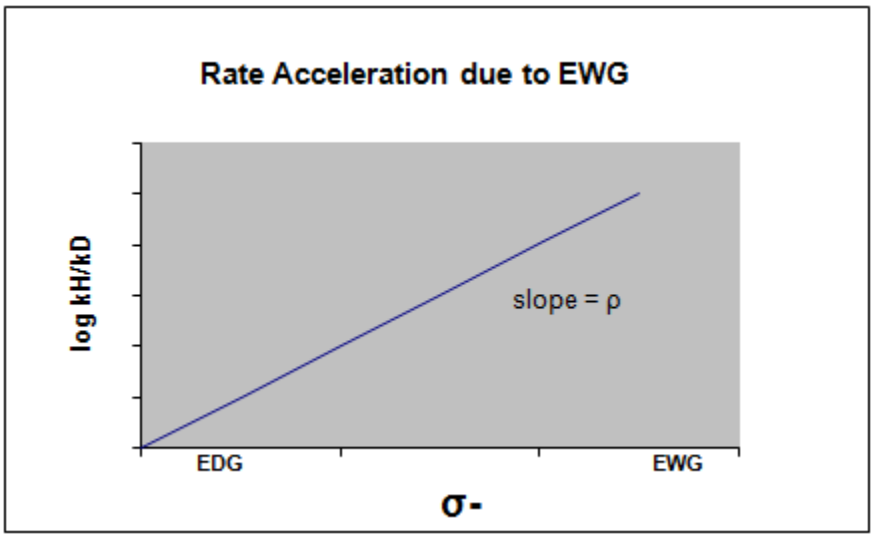
R	H	CH_3	OCH_3	F	Cl	NO_2
ortho	6.27	12.3	8.06	54.1	11.4	671
meta	6.27	5.35	8.17	13.6	14.8	32.1
para	6.27	4.24	3,38	7.22	10.5	37.0

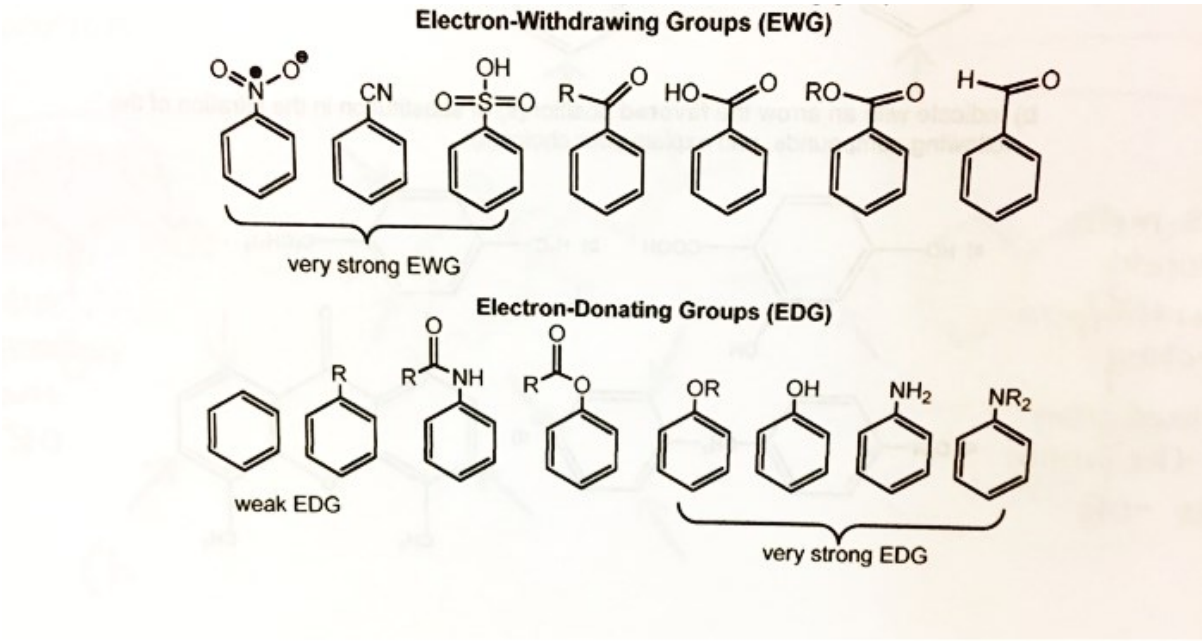
Acidity and σ values

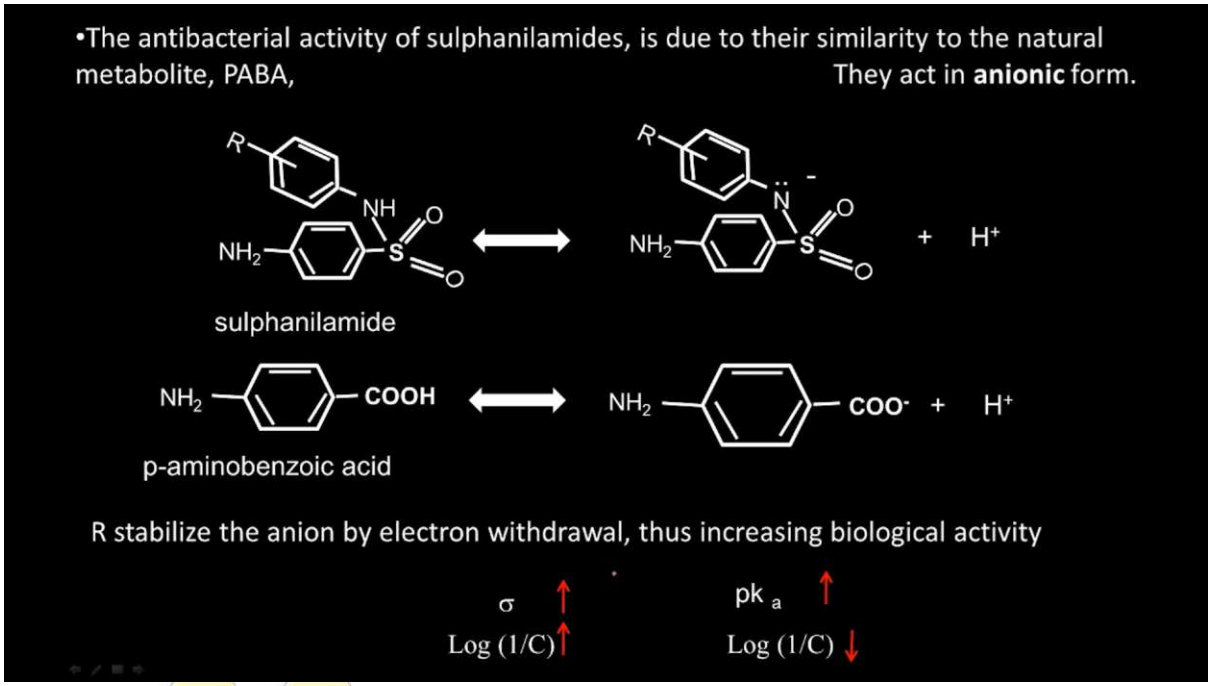
X	OMe	CH_3	H	Cl	NO_2
pK_a	4.47	4.36	4.2	3.99	3.44
	-0.27	-0.17	0	0.23	0.71

electron-donating ⟷ electron-withdrawing




Rate Acceleration due to EWG
log kH/kD
slope = ρ
EDG
EWG
σ-


Electron-Withdrawing Groups (EWG)
very strong EWG
Electron-Donating Groups (EDG)
weak EDG
very strong EDG



Taft's Steric Parameter (Es)

The Taft equation is a linear free energy relationship (LFER) used in physical organic chemistry in the study of reaction mechanisms and in the development of quantitative structure–activity relationships for organic compounds. It was developed by Robert W. Taft in 1952 as a modification to the Hammett equation

While the Hammett equation accounts for how field, inductive, and resonance effects influence reaction rates, the Taft equation also describes the **steric effects** of a substituent. The Taft equation is written as:

$RCOOCH_3 + H_2O \rightleftharpoons RCOOH + CH_3OH$

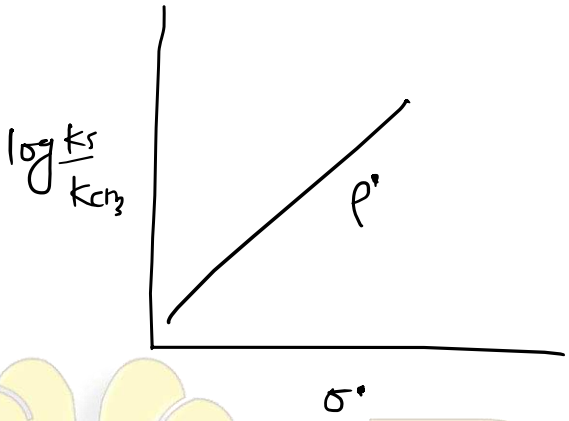
$$Es = Log (K_{RCOOCH_3}/K_{CH_3COOCH_3})$$

$$Log (K_S/K_{CH_3}) = \rho^* \sigma^* + \delta Es$$

- **$Log (K_S/K_{CH_3})$** is ratio of the rate of the substituted reaction compared to the reference reaction,
- **ρ^*** is the sensitivity factor for the reaction to polar effects,
- **σ^*** is the polar substituent constant that describes the field and inductive effects of the substituent,
- **δ** is the sensitivity factor for the reaction to steric effects,
- **Es** is the steric substituent constant.

- the polar sensitivity factor ρ* for Taft plots will describe the susceptibility of a reaction series to polar effects. When the steric effects of substituents do not significantly influence the reaction rate the Taft equation simplifies to a form of the Hammett equation

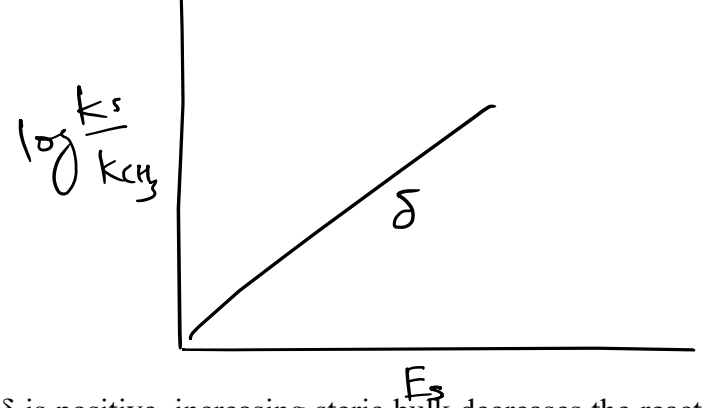
$$\textbf{Log } (K_S/K_{CH3}) = \rho^* \sigma^*$$



  - If ρ* > 1, the reaction accumulates negative charge in the transition state and is accelerated by electron withdrawing groups.
  - If 1 > ρ* > 0, negative charge is built up and the reaction is mildly sensitive to polar effects.
  - If ρ* = 0, the reaction is not influenced by polar effects.
  - If 0 > ρ* > −1, positive charge is built up and the reaction is mildly sensitive to polar effects.
  - If −1 > ρ*, the reaction accumulates positive charge and is accelerated by electron donating groups.
- Similar to the polar sensitivity factor, the steric sensitivity factor δ for a new reaction series will describe to what magnitude the reaction rate is influenced by steric effects. When a reaction series is not significantly influenced by polar effects, the Taft equation reduces to:

$$\textbf{Log } (K_S/K_{CH3}) = \delta\ Es$$

- A plot of the ratio of the rates versus the Es value for the substituent will give a straight line with a slope equal to δ.

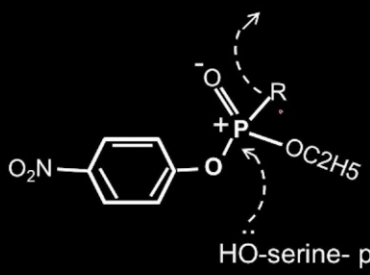


- If δ is positive, increasing steric bulk decreases the reaction rate and steric effects are greater in the transition state.
- If δ is negative, increasing steric bulk increases the reaction rate and steric effects are lessened in the transition state

Constants used in the Taft equation		
Substituent	**E_s**	**σ***
–H	1.24	0.49
$-CH_3$	0	0
$-CH_2CH_3$	–0.07	–0.1
$-CH(CH_3)_2$	–0.47	–0.19
$-C(CH_3)_3$	–1.54	–0.3
$-CH_2Ph$	–0.38	0.22
–Ph	–2.55	0.6

steric

hydrolysis of inhibitors of acetylcholine esterase



HO-serine- protein

Organophosphates must be hydrolysed to be active and their biological activity is:

$$\text{Log } (1/C) = 2.58\ E_s + 7.94$$

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:

$$\textbf{Log BA} = \textbf{a log p} + \textbf{b } \sigma + \textbf{c Es} + \textbf{constant (linear)}$$

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

  - Partition coefficient; log P
  - Hammett constant; σ
  - Taft's steric parameter; Es

- 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

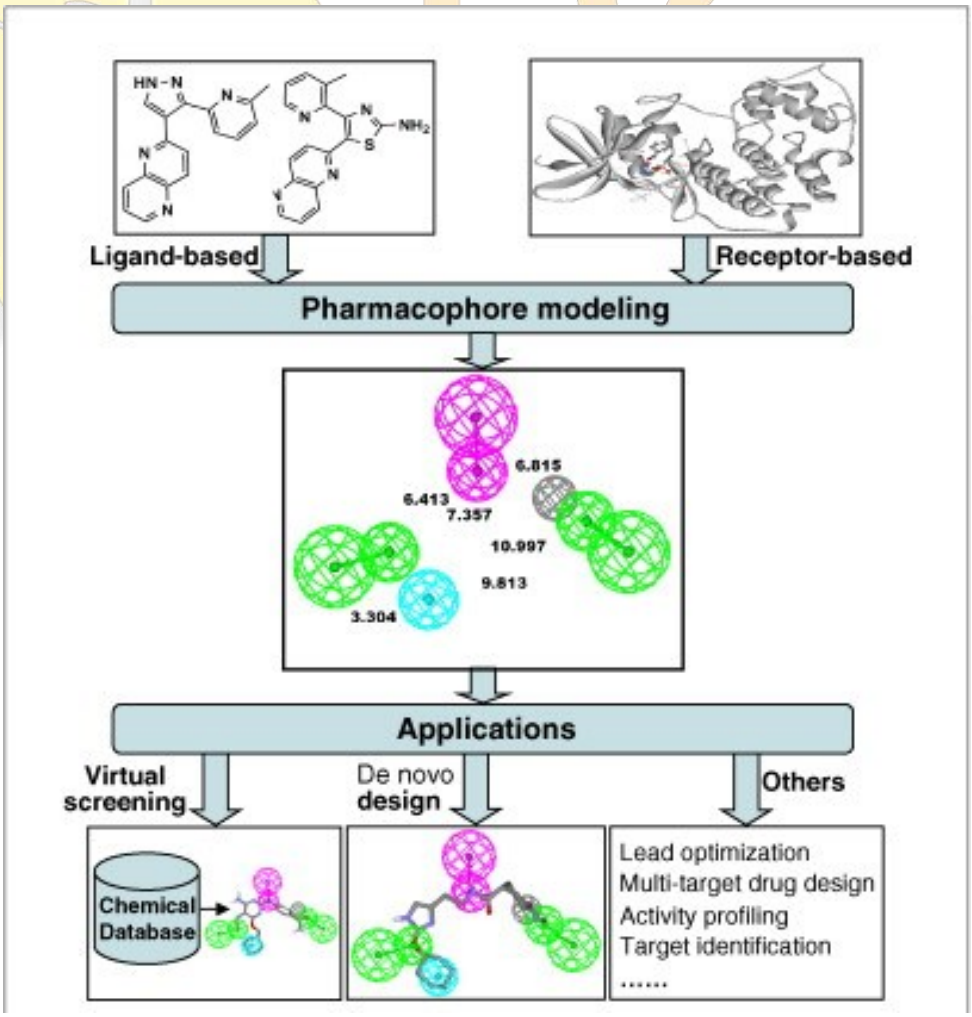
$$\textbf{Log BA} = \mu + \Sigma \textbf{ ai aj}$$

  - where ai = number of positions at which substitution occurs
  - aj = 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.

PHARMACOPHORE MODELLING

- The pharmacophore concept was introduced by **Paul Ehrlich** in the early 1900s. Then, the term pharmacophore was coined by **Schueler** in his 1960 book *Chemobiodynamics and Drug Design*, and was defined as "a molecular framework that carries (phoros) the essential features responsible for a drug's (pharmacon) biological activity."
- In those year's pharmacophore was understood as chemical or functional groups on a molecule that are responsible for the biological activity.
- In 1997, IUPAC (International Union of Pure and Applied Chemistry) defined pharmacophore as the sum of steric and electronic properties that are required for the interaction of a molecule with a target and thus provide the biological activity
- a pharmacophore does not represent a real molecule or a set of chemical groups, but is an abstract concept; "A pharmacophore is the pattern of features of a molecule that is responsible for a biological effect," which captures the essential notion that a pharmacophore is built from features rather than defined chemical groups.



- Each atom or group of a compound that shows features associated with molecular recognition can be converted into a pharmacophore pattern. Molecular pharmacophore patterns can be hydrogen bond donors (HBD), hydrogen bond acceptors (HBA), positive features, negative features, aromatic rings, hydrophobic features and their combinations.
- A pharmacophore model includes several patterns arranged in a particular 3D (three dimensional) pattern. Each pattern is depicted by a typical sphere containing radius that determines the deviation tolerance from the exact position. There are also various other displaying ways. These patterns can be displayed as a single pattern or their combination

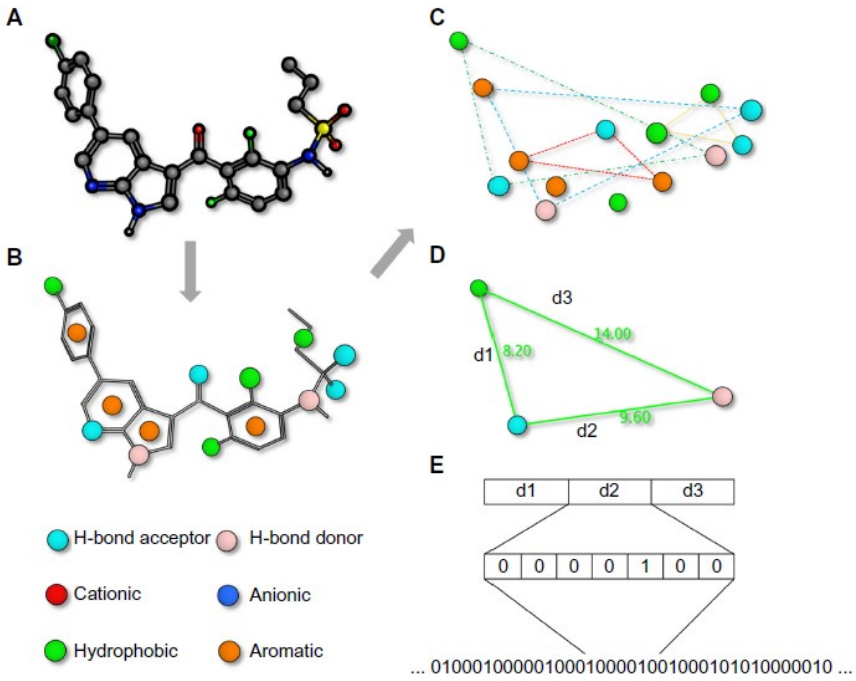


Figure 1 Pharmacophore fingerprints.
Notes: A pharmacophore fingerprint is the representation of a small molecule ligand (**A**) annotated with molecular interaction features (**B**) into a string. Typically, every possible three- (or four-) point combination of molecular interaction features (**C**), with different distances between the features, calculated either through space or by the number of bond lengths (**D**), is calculated and the frequency of occurrence is stored in a string (**E**). Such strings are useful for the easy comparison of similarity between multiple molecules.

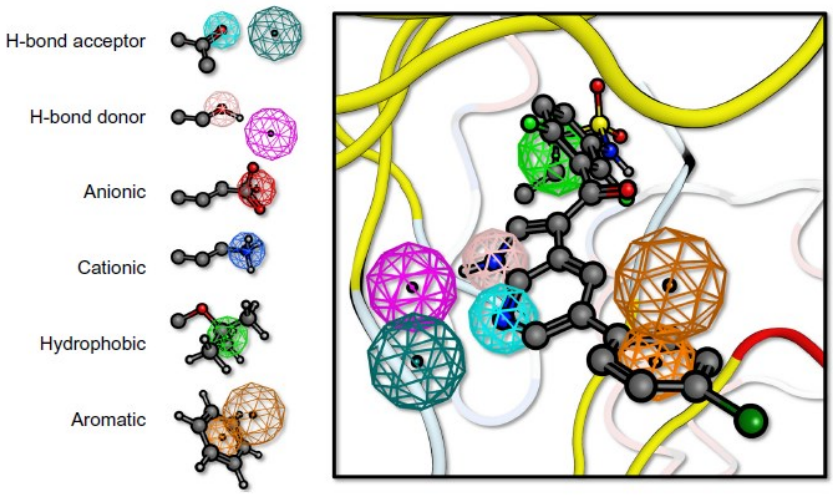


Figure 2 Pharmacophore query.
Notes: A pharmacophore query is comprised of different features. The features represent molecular recognition motifs such as hydrogen bond acceptors or donors, anionic, cationic, hydrophobic, and aromatic groups. The radius of the sphere determines the strictness of the geometric constraint. For features where the correct orientation of the interaction is important such as hydrogen bonds and the aromatic plane, a second feature can be used indicating the vector of the interaction (or the normal of the plane). A pharmacophore query can combine any of these features, with different radii and logic operations such as "AND," "OR," and "NOT." On the left a hypothetical pharmacophore query for BRAF kinase is given.

Approaches of Pharmacophore Modeling

- There are two principal approaches of pharmacophore modeling that are used in the drug discovery process: Ligand-based pharmacophore modeling and structure-based pharmacophore modeling.
- In the ligand-based pharmacophore modeling approach, novel ligands are designed by using a set of active ligands available. This approach is employed if the target structure is not available. In a similar manner, the structure-based pharmacophore approach is employed when the structure of the target protein is available.

Ligand Based Drug Design

- In the ligand-based pharmacophore modeling, first active ligands are identified by using the literature available or database search. The data set is split into a training set and test set.
- Then, feature analysis of the training set ligands is done. The common features are detected through the alignment of the active ligands.
- The next step is pharmacophore model generation and ranking of the generated models. Finally, pharmacophore model validation is performed and the best pharmacophore model is selected depending on the results obtained.

Structural Based Pharmacophore Modeling

- In the structure-based pharmacophore modeling, selection and preparation of target protein structure is the first step.

- The second step is binding site prediction. Then, complemental chemical features of the binding site amino acids and their layouts are identified by analyzing it carefully.
- After this, the pharmacophore features, which should be optimized by the adjusted tools in the programs employed, are generated.
- Finally, crucial pharmacophore features responsible for the activity are selected.
- LigandScout , MOE, Pocket v2 and Snooker are among the commonly used software for structure-based pharmacophore modeling. Similarly, there are various software and servers used in pharmacophore modeling.
- The commonly employed programs and servers are summarized in the alphabetical order (Table 1).

Table 1: Programs and servers used in pharmacophore modeling. Program/Server Brief Description

CATALYST-HipHop	CATALYST is now part of the BIOVIA Discovery Studio. It consists of algorithms used in pharmacophore generation: HipHop and HypoGen. HipHop gives the alignment of active ligands against a specific target and finds the three dimensional arrangements of common features by overlapping various structures.
CATALYST-HypoGen	It generates hypotheses that are able to estimate the activity of molecules quantitatively by using biological analysis data. Thus, it allows the correlation of the structural and activity data for pharmacophore modeling.
GALAHAD	The program uses modified genetic algorithm and fixes certain shortcomings of the GASP program and thus increases its performance. It increases the computational speed by using prebuilt structures as a starting point.
GASP	GASP is available in the SYBYL package. It uses genetic algorithm for the detection of pharmacophores. Unlike the other pharmacophore determinations, conformational search is carried out instantly in the GASP process and is an integral part of the program. A single low energy structure and random spinings are applied to examine conformational changes before superimposing on each input compound.
LigandScout	Though it is possible to perform both structure-based and ligandbased phamacophore modeling with LigandScout, it is among the first programs specialized in structure-based pharmacophore modeling.

	Especially, if the structure of the target protein is present in its ligand bound state, LigandScout is widely used.
MOE	MOE is able to perform ligand-based and structure-based pharmacophore modeling. Model building is performed by the pairwise alignment of the active ligands. It is recommended to decrease the magnitude of the training set by grouping similar molecules.
PharmaGist	It is a freely accessible server used in ligand-based pharmacophore generation. This web server detects pharmacophores via multiple flexible alignments of the input molecules.
Pharmer	It is a pharmacophore method that makes searching based on the width and complexity of the query instead of the molecular library screened. It is a very fast method and its source code is available under an open-source license.
PharmMapper	It is a freely accessible web server used for the identification of potential targets for the input ligands. It calculates pharmacophores by using semi-rigid pharmacophore mapping.
PHASE	It is provided by Schrödinger package. It is a convenient approach used in drug discovery with or without its receptor structure. It creates a hypothesis from one or more ligands, protein-ligand complexes and apo proteins. It has a special algorithm designed for use in optimization of lead compounds and virtual screening.

Application

- Pharmacophore modeling is employed in virtual screening, fishing drug targets, ligand profiling, docking, and ADMET prediction.
- New perspectives are also expected for various applications of pharmacophore modeling in the future due to the simplicity and versatility of the concept. In this way, besides the applications explained here, it may have applications in polypharmacology, drug repurposing and side effect prediction

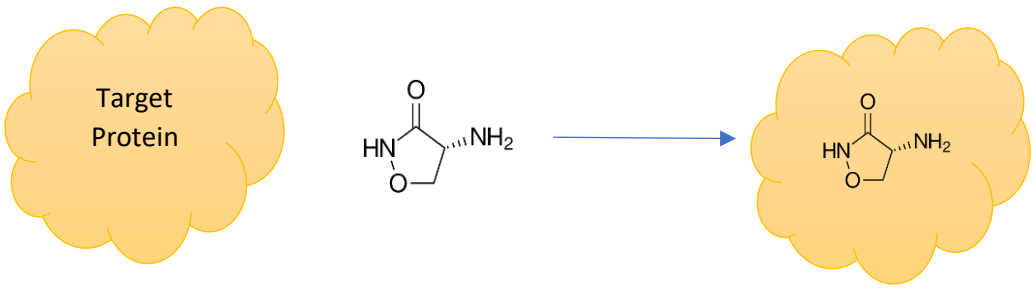
Source:

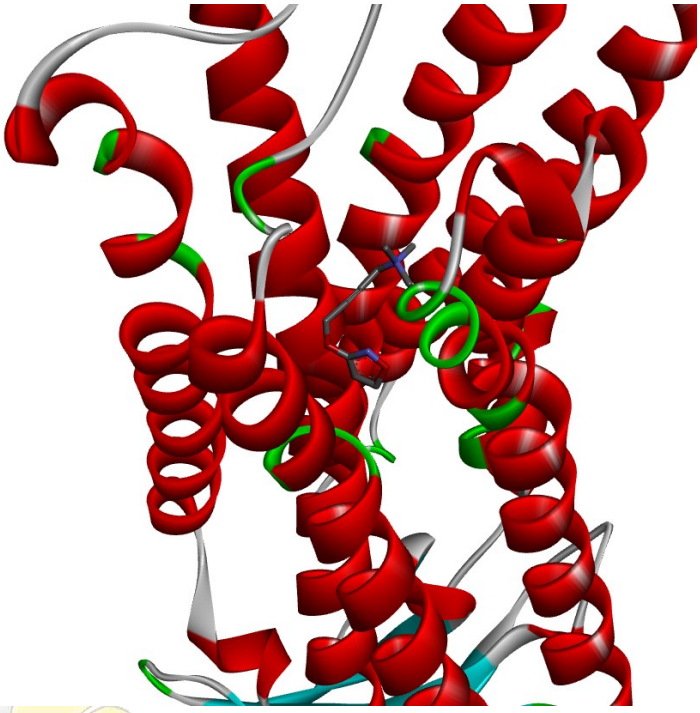
1. Yang et al., 2010, Drug Discovery Today, 15(11), 444-450. Doi: https://doi.org/10.1016/j.drudis.2010.03.0134
2. Muhammed and Akı-Yalçın, JOTCSA. 2021; 8(3): 749-762. https://doi.org/ 10.18596/jotcsa.927426.

3. Qing et al. *Journal of Receptor, Ligand and Channel Research*. 2014;7:81-92 https://doi.org/10.2147/JRLCR.S46843

MOLECULAR DOCKING

- Molecular docking is a computational method to identify the architecture of compounds generated by two or more distinct molecules.
- Docking is widely used to anticipate the interaction between ligand and target protein in terms of affinity and activity.
- Docking plays a critical role in rational drug design. Considering the biological and pharmacological importance of docking studies, much effort has been made to improve the algorithms for docking prediction.
- Docking is a mathematical technique that anticipates the preferable orientation of one molecule (may be drug, which has ligand) relative to another (may be target protein, which has binding site) when they are linked together to create a stable complex.
- Using scoring functions (binding energy), it is possible to estimate the strength of the connection or binding affinity across two compounds based on their preferential orientation.
- Signal transduction is dependent on the interactions of physiologically significant substances such as proteins, nucleic acids, carbohydrates, and lipids
- The goal of docking studies is to optimize the shape of both the ligand and protein, as well as the relative orientation of the protein and ligand, to reduce the total system's free energy





Types of Docking

Rigid docking

Assuming the compounds are inflexible, we are seeking a rearrangement of one of the compounds in three-dimensional space that results in the best match to the other compounds in parameters of a scoring system. The ligand's conformation can be formed with or without receptor binding activity.

Flexible docking

In conjunction with transformation, we evaluate molecular flexibility to identify confirmations for the receptor and ligand molecules as they exist in the complex

Theories

Lock & Key Theory:

Emil Fischer created a concept termed the "lock-andkey model" in 1890, as seen in figure, to describe how biological processes operate

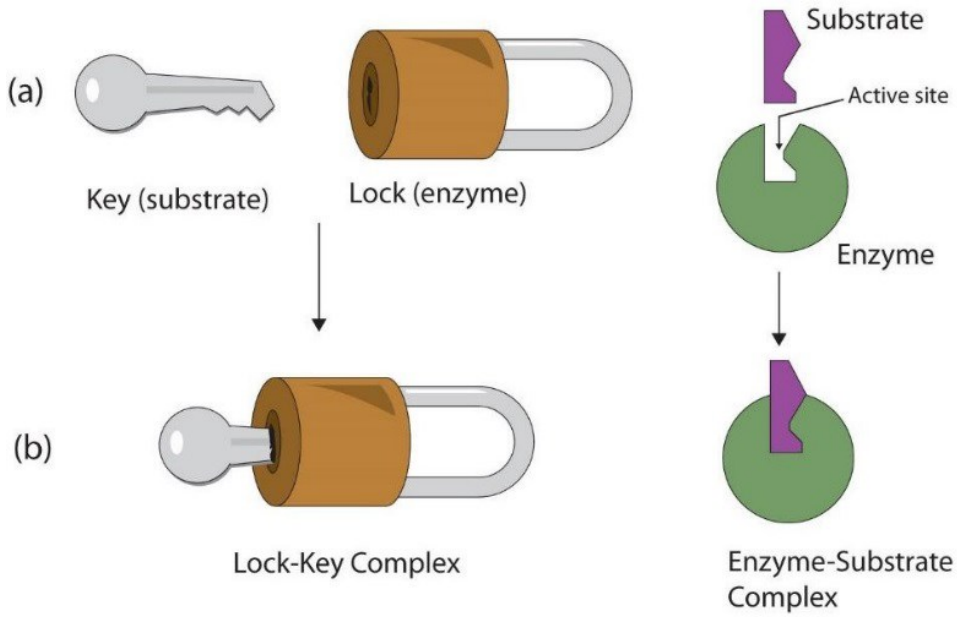


Fig. : The lock and key theory

The induced-fit theory

Daniel Koshland proposed the "induced fit theory" in 1958. The fundamental concept is that throughout the character recognition, both the ligand and target, as seen in figure, adapt to one another by modest conformational changes until an ideal match is reached.

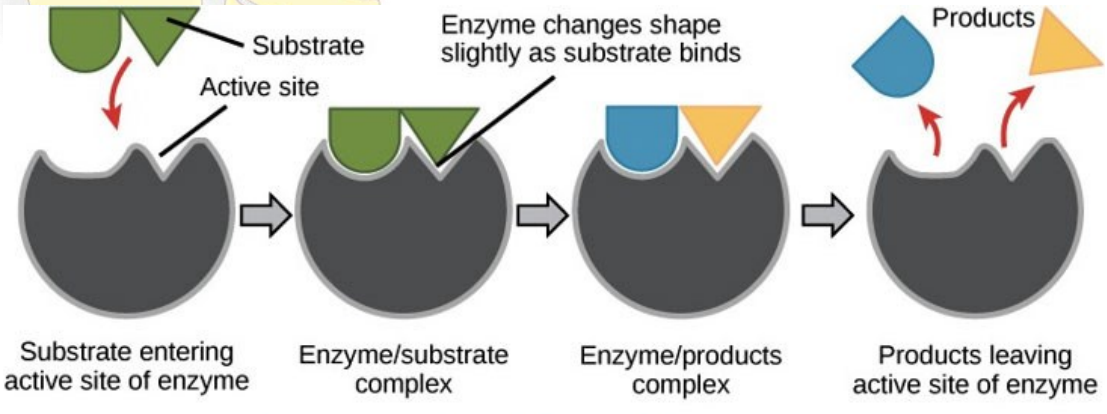


Fig. 5: Induced fit model

Molecular Docking Approaches

Monte carlo approach

It creates a randomized conformation, translations, and rotation of a ligand in an active site. It assigns an initial configuration value. Then it develops and scores a new configuration. It determines if the new configuration is kept using the Metropolis criterion. (Metropolis criterion- If a new approach outperforms the prior one, it is approved instantly.

Matching approach

This strategy emphasizes idleness, the optimal location of the ligand atom in the site is determined, resulting in a ligand-receptor arrangement that might also need improvement.

Ligand fit approach

Ligand fit is a word that refers to a quick and precise methodology for docking small molecules ligands into protein active sites while taking shape complementarity into account

Point complimentarily approach

These techniques are focused on comparing the shapes and/or chemical properties of different molecules. Blind Docking: This technique was developed to identify potential peptide ligand binding sites and mechanisms of action by scanning the full interface of target molecules

Software

There is several other software are available for docking such as Discovery studio, ArgusLab, Schrödinger, Hammerhead, ICM, MCDock, GOLD, GemDock, Glide, Flex-X, Autodock and Yucca

Application

- Drug Design
- Hit identification
- Lead optimization
- Ligand Preparation
- Selection of target site
- Receptor preparation

Source: Raval K, Ganatra T. Basics, types and applications of molecular docking: A review. *IP Int J Comprehensive Adv Pharmacol* 2022;7(1):12-16. DOI: https://doi.org/10.18231/j.ijcaap.2022.003

****