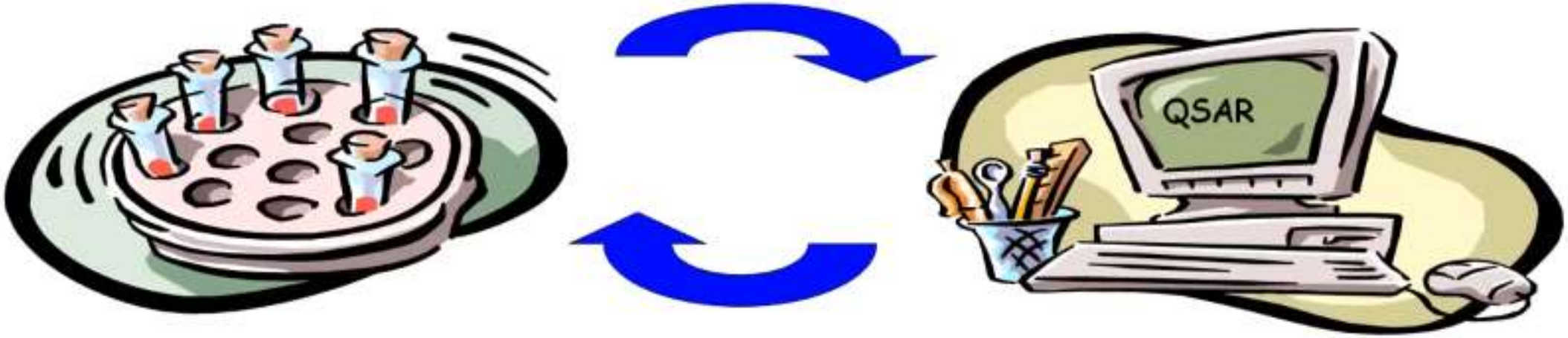
QSAR/QSPR

Compounds + biological activity



New compounds with improved biological activity

Ms. Rina P. Patil

Department of Pharmaceutical Chemistry

Jijamata Education Society'S College of Pharmacy Nandurbar

Drug design Or **Rational drug design:**

It is the inventive process of finding new medications based on the knowledge of a biological target.

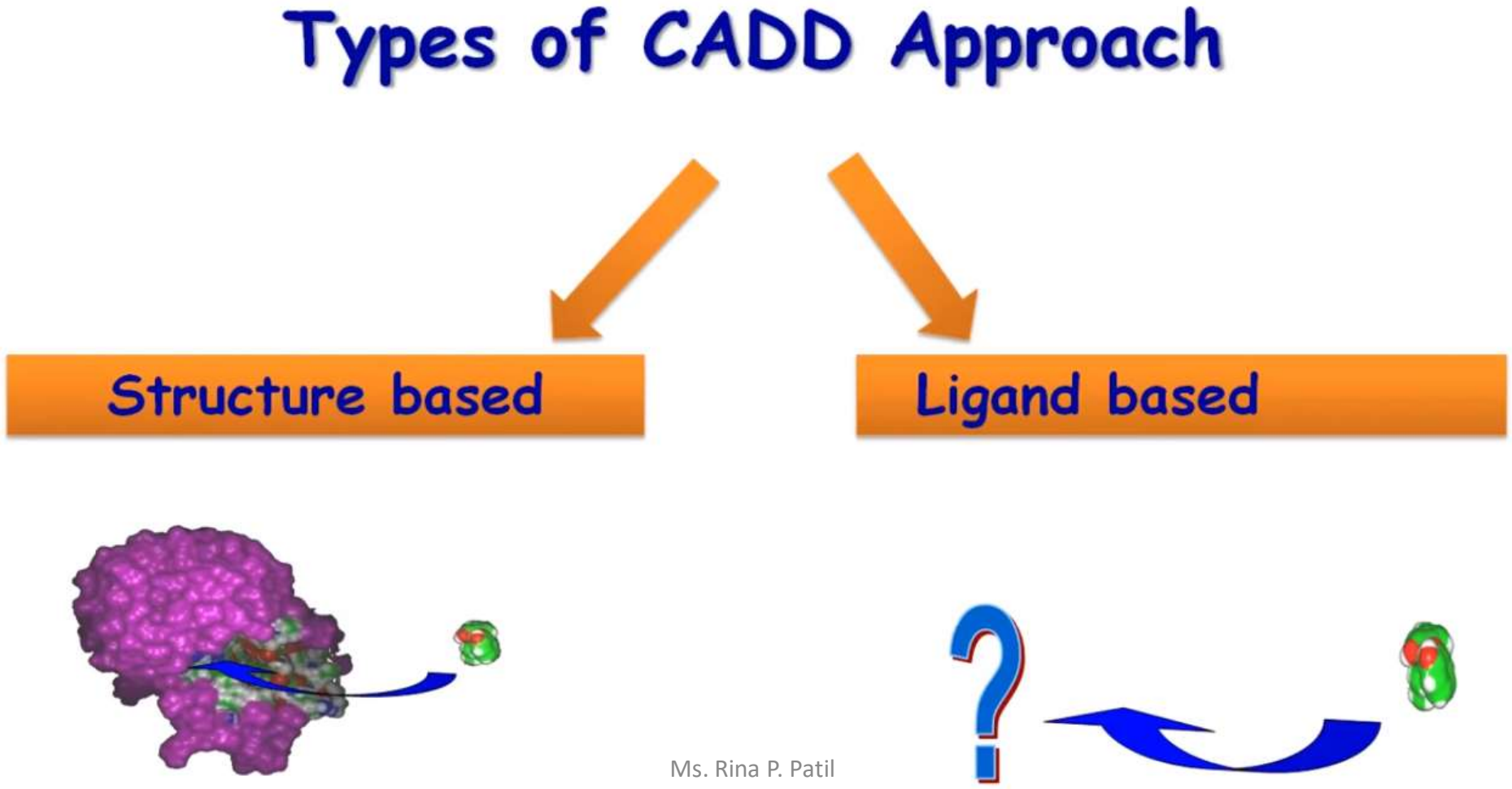
Strategies for design of drugs involved:

- A change in shape such that the new drug had a better fit for its target binding site.
- A change in functional groups or substituents such that the drugs pharmacokinetics or binding site interactions were improved.

Principles of Drug Designing:

- Improving the binding of drugs
- Increasing the selectivity
- Reduces the side effects
- Easy synthesizable
- Arrangement of functional groups and identification of a pharmacophore

Computer-aided drug design uses computational approaches to discover, develop, and analyze **drugs** and similar biologically active molecules.



Structure Based Designing (Hit Generation)

Study the Disease

↓

Identify the Target

↓

Isolate the Target

↓

Establish 3D Structure of Target

↓

Structure Based Approach

↓

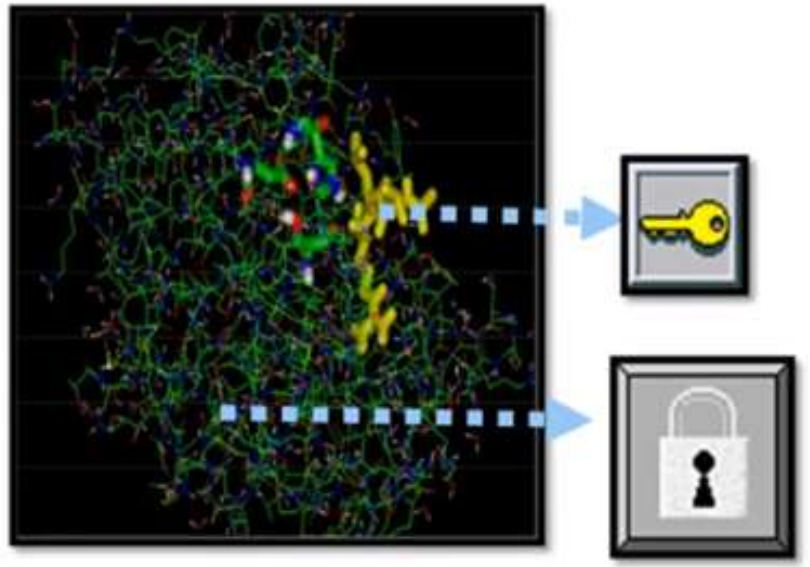
Theoretical Drug

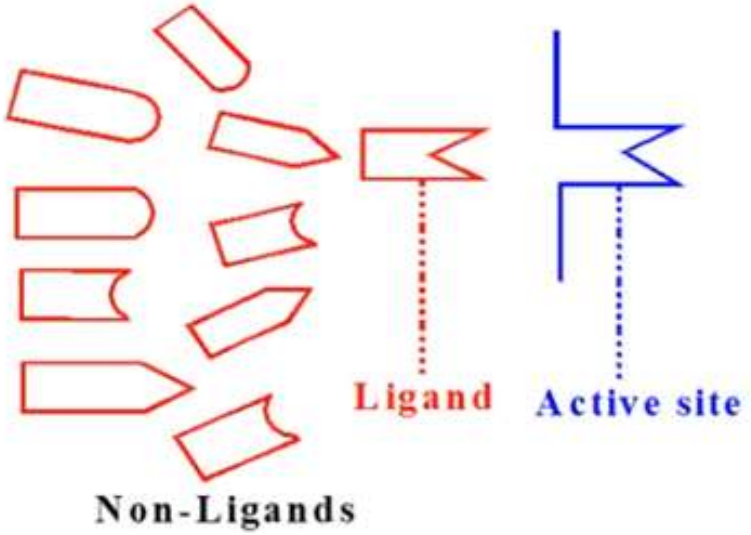
↓

Synthesize the Compounds

↓

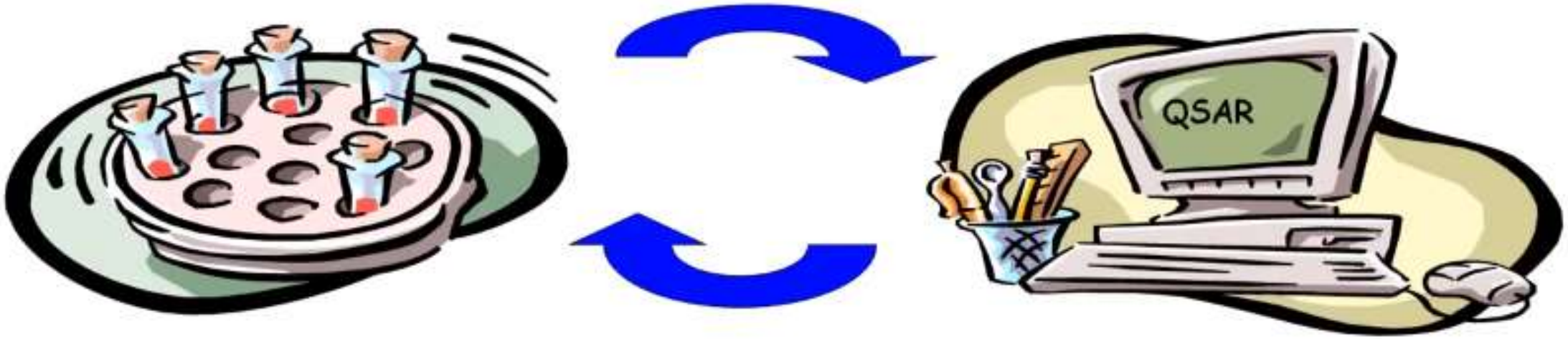
New Chemical





Ligand Based Approach (Hit Optimization)

Compounds + biological activity



New compounds with improved biological activity

What is QSAR?

- QSAR is an integral part of rational drug discovery process.

$$\text{Activity (y)} = f(x_1, x_2, x_3 \ldots\ldots\ldots x_p)$$

- Objectives of QSAR
  - To identify molecular features/properties responsible for variation in desired biological activity.
  - To get direction for the design of novel molecules based on the identified important features/properties governing biological activity.
  - To build a mathematical model which can be used to predict biological activity of designed novel molecules.

QSAR method Requires

- The compounds studied must be **structurally related**, **act at the same target**, and have the **same mechanism of action**.
- It is crucial that the **correct testing procedures** are used.

***In-vitro* tests** carried out on isolated enzymes are relevant for a QSAR study as the activities measured for different inhibitors are related directly to how each compound binds to the active site.

***In-vivo* tests** carried out to measure the physiological effects of enzyme inhibitors are not valid, however, as both pharmacodynamic and pharmacokinetic factors come into play.

This makes it impossible to derive a sensible QSAR equation.

Steps in QSAR

Step 1: Selection of biologically active series (Structure + Biological activity)

Step 2: Calculation of Various Physico-Chemical Descriptors

Step 3: Correlation of Physico-chemical properties with Biological activity by QSAR methods

Step4 : Getting equation

Step 5: Designing of the compounds based upon QSAR equation

Step 6: Predicting the biological activity of the designed compounds

Step7: Synthesis of the compounds

Advantages of QSAR

Advantages of predicting biological activity with quantitative structure-activity relationships modelling include:

- Able to predict activities of a large number of compounds with little to no prior experimental data on activity.
- Can reveal which molecular properties may be worth investigating further.
- Regarded as a “green chemistry” approach since chemical waste is not generated when performing in silico predictions.
- In vivo and in vitro experimentation can be very expensive and time-consuming. QSAR modelling reduces the need for testing on animals and/or on cell cultures and saves time.

Disadvantages of QSAR

- Does not provide an in-depth insight on the mechanism of biological action.
- Some risk of highly inaccurate predictions of pharmacological or biological activity.

Physicochemical Properties/ Parameters

1. Many physical, structural, and chemical properties have been studied by the QSAR approach, but the most common are hydrophobic, electronic, and steric properties.
2. This is because it is possible to quantify these effects. Hydrophobic properties can be easily quantified for complete molecules or for individual substituents.
3. However, it is more difficult to quantify electronic and steric properties for complete molecules, and this is only really feasible for individual substituents.

- **Hydrophobicity of the molecule**
- **Hydrophobicity of substituent's**
- **Electronic properties of substituent's**
- **Steric properties of substituent's**

} **Most common properties studied**

Lipophilic parameters: partition coefficient, π-substitution constant

Polarizability parameters: molar refractivity, parachor

Electronic parameters: Hammet constant, dipole moment.

Steric parameters: Taft's constant.

Miscellaneous parameters: molecular weight, geometric parameters.

Hydrophobicity/ Lipophilicity of the Molecule

1. **The hydrophobic character of a drug is crucial to how easily it crosses cell membranes and may also be important in receptor interactions.**
2. Changing substituents on a drug have significant effects on its hydrophobic character and, hence, its biological activity.
3. Therefore, it is important to predict this quantitatively.

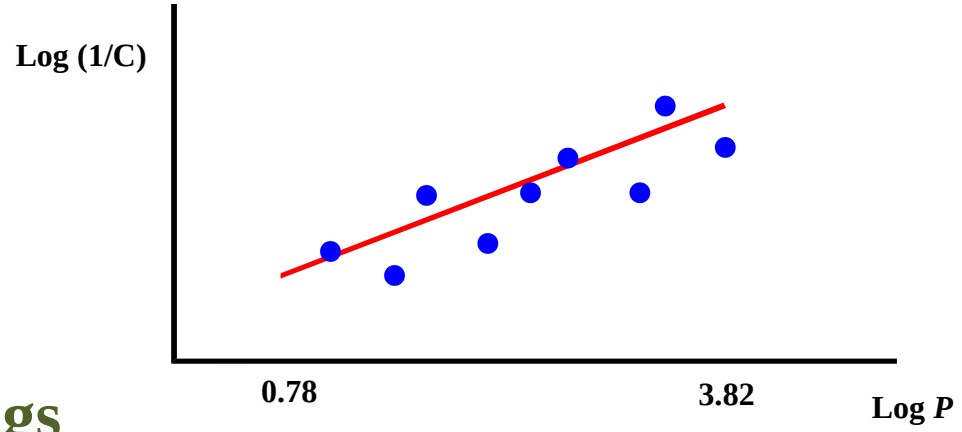
$$\textbf{Partition Coefficient } P = \frac{\textbf{[Drug in octanol]}}{\textbf{[Drug in water]}}$$

High *P* ⇨ **High hydrophobicity**

4. Hydrophobic compounds have a high P value, whereas hydrophilic compounds have a low P value.
5. Varying substituents on the lead compound will produce a series of analogues having different hydrophobicity's and, therefore, different P values.

By plotting these **P values against the biological activity** of these drugs, it is possible to see if there is any relationship between the two properties. The biological activity is normally expressed as **1/C**, where **C is the concentration of drug** required to achieve a defined level of biological activity. Activity of drugs is often related to *P*,

e.g. Binding of drugs to serum albumin (**straight line - limited range of log P**)



Binding increases as log P increases

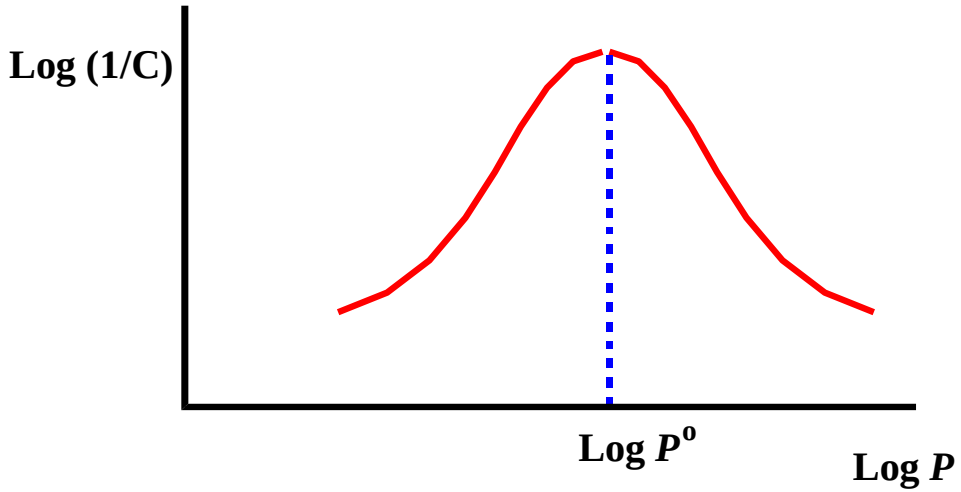
Binding is greater for hydrophobic drugs

log1/C = k1logP + k2

log1/C = 0.75logP + 2.30

r= 0.96 s= 0.159

- If graph is extended to very high logP values, then get a parabolic curve
- Example 2- General anaesthetic activity of ethers, (parabolic curve - larger range of log P values)



$$\log 1/C = -0.22(\log P)^2 + 1.04\log P + k3$$

Optimum value of log P for anaesthetic activity = $\log P^o$

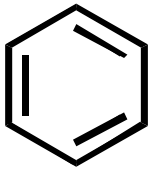
- When *P* small, dominated by log*P* term
- When *P* large, log*P* squared dominates & so activity decreases

π-substituent constant or hydrophobic substituent constant

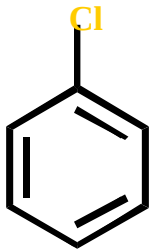
- The π-substituent constant defined by hansch and co-workers by the following equation.

$$\pi_x = \log P_x - \log P_H$$

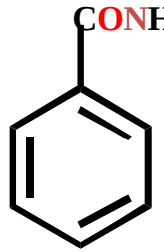
- A positive π value indicates that the π substituent has a higher lipophilicity than hydrogen and the drug favours the organic phase.
- A negative π value indicates that the π substituent has a lower lipophilicity than hydrogen and the drug favours the aqueous phase.
- Example-



Benzene
(Log *P* = 2.13)



Chlorobenzene
(Log *P* = 2.84)

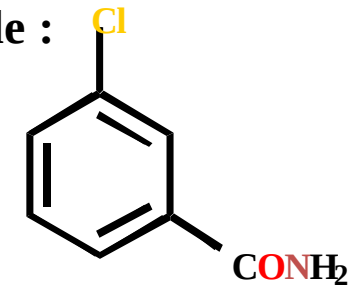


Benzamide
(Log *P* = 0.64)

$\pi_{Cl} = 0.71$, $\pi_{CONH2} = -1.49$

- The value of π is only valid for parent structures
- It is possible to calculate log P using π values

Example :



***meta*-Chlorobenzamide**

$$\text{Log } P_{(\text{theory})} = \log P_{(\text{benzene})} + \pi_{Cl} + \pi_{CONH2}$$

$$= 2.13 + 0.71 - 1.49$$

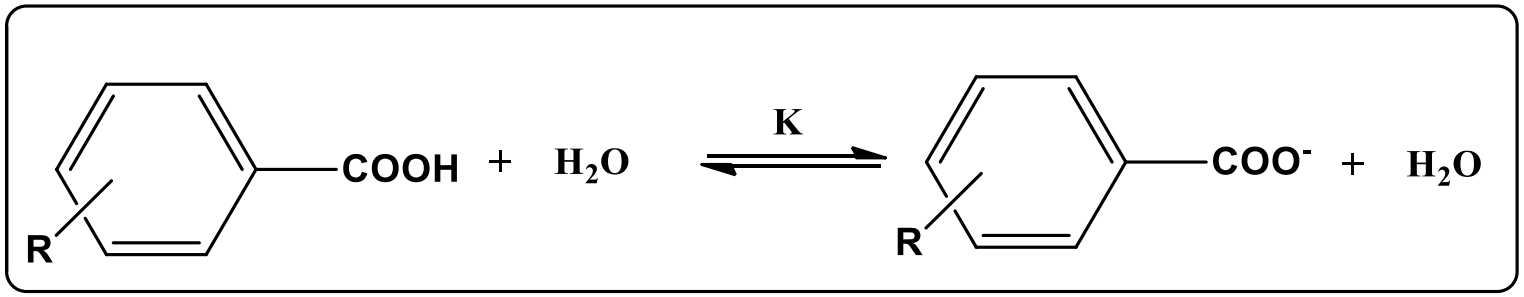
$$= 1.35$$

$$\text{Log } P_{(\text{observed})} = 1.51$$

- A QSAR equation may include both *P* and π.
- *P* measures the importance of a molecule's overall hydrophobicity (relevant to absorption, binding etc)
- π identifies specific regions of the molecule which might interact with hydrophobic regions in the binding site

Electronic Parameters: *The Hammett constant(σ)*

- Electronic effects will have an impact on ionization or polarity. This in turn may have an effect on how easily a drug can pass through cell membrane or how strongly it can interact with a binding site. It is therefore useful to measure the electronic effect of the substitutent.
- The constant (σ) a measure of the e-withdrawing or e-donating influence of substituent's
- It can be measured experimentally and tabulated (e.g. σ for aromatic substituent's is measured by comparing the dissociation constants of substituted benzoic acids with benzoic acid)



$$\log K_{a(\text{substituted acid})} - \log K_{a(\text{unsubstituted acid})} = \log K_{a(RX)} - \log K_{a(RH)}$$

$$= \log [K_{a(RX)} / K_{a(RH)}]$$

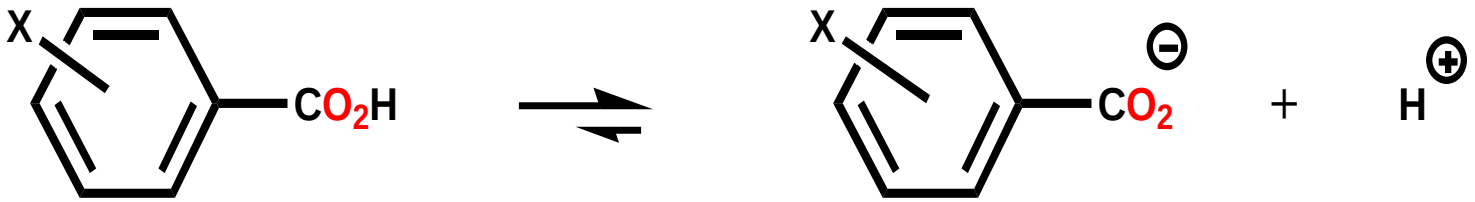
$$= \sigma$$

X=H $\quad K_H$ = **Dissociation constant** = $\dfrac{[PhCO_2^-]}{[PhCO_2H]}$

1. Benzoic acid is a weak acid and only partially ionizes in water. An equilibrium is set up between the ionized and non-ionized forms, where the relative proportion of these species is known as the **equilibrium or dissociation constant *KH*.**
2. Electron-withdrawing groups, such as a nitro group, result in the aromatic ring having a stronger electron-withdrawing and stabilizing influence on the carboxylate anion, and so the equilibrium will shift more to the ionized form. Therefore, the substituted benzoic acid is a stronger acid and has a larger *K X* Value.
3. If the substituent X is an electron-donating group such as an alkyl group, then the aromatic ring is less able to stabilize the carboxylate ion. The equilibrium shifts to the left indicating a weaker acid with a smaller *K X value*

- **X= electron withdrawing group (e.g. NO_2)**

X = electron withdrawing group

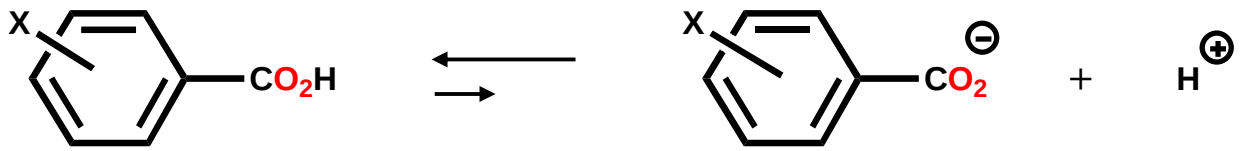


Charge is stabilised by X
Equilibrium shifts to right
$K_X > K_H$

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

Positive value

- **X= electron donating group (e.g. CH_3)**



Charge destabilised
Equilibrium shifts to left
$K_X < K_H$

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

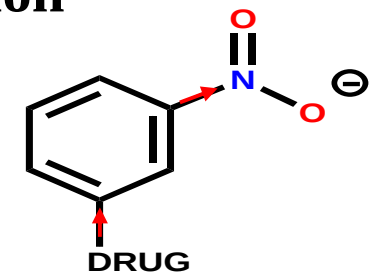
Negative value

- ***σ* value depends on inductive and resonance effects**
- ***σ* value depends on whether the substituent is *meta* or *para***
- ***ortho* values are invalid due to steric factors**

Hammett Substituent Constant (σ)

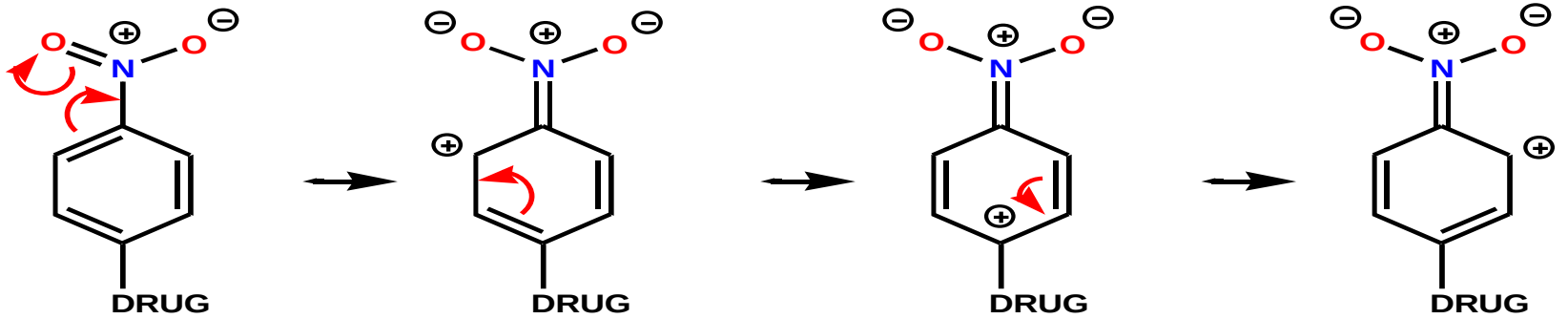
EXAMPLES: $\sigma_p(NO_2) = 0.78$ $\sigma_m(NO_2) = 0.71$

***meta*-Substitution**



e-withdrawing (inductive effect only)

***para*-Substitution**

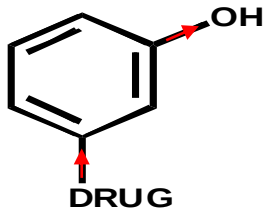


e-withdrawing (inductive + resonance effects)

Hammett Substituent Constant (σ)

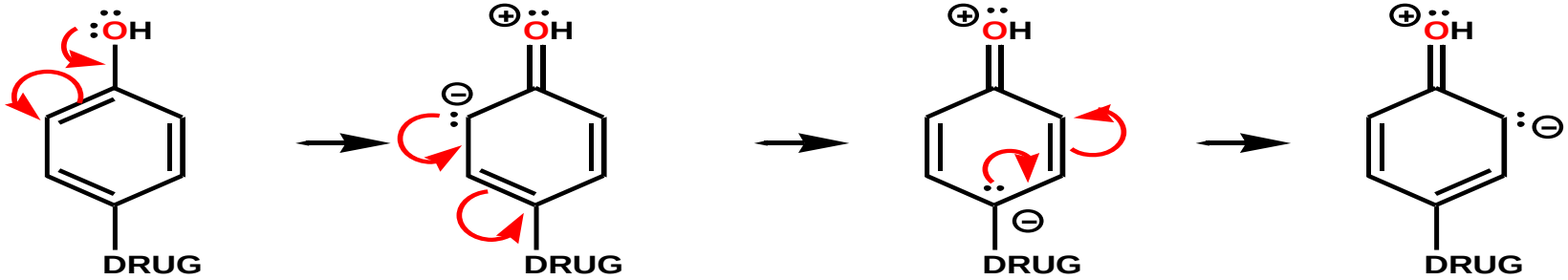
EXAMPLES: $\sigma_m(OH) = 0.12$ $\sigma_p(OH) = -0.37$

***meta*-Substitution**



e-withdrawing (inductive effect only)

***para*-Substitution**



e-donating by resonance more important than inductive effect

Electronic Factors *R* & *F*

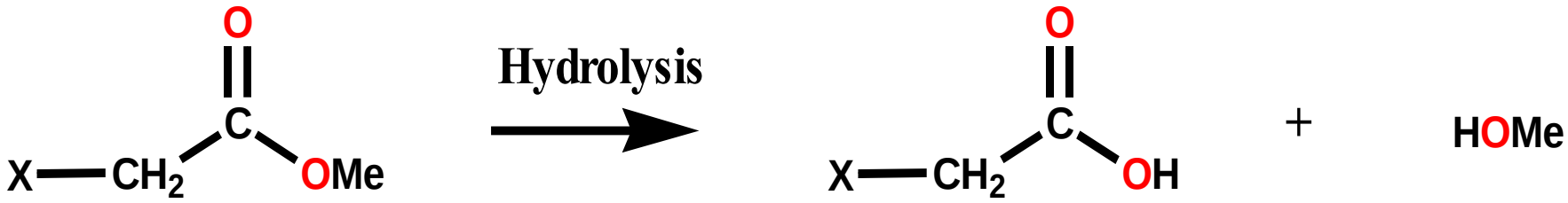
Only used for aromatic substituent's

Only suitable for drugs containing aromatic rings

- ***R*** **- Quantifies a substituent’s resonance effects**
- ***F*** **- Quantifies a substituent’s inductive effects**

Aliphatic electronic substituent's

- Defined by σ1
- Purely inductive effects
- Obtained experimentally by measuring the rates of hydrolyses of aliphatic esters
- Hydrolysis rates measured under basic and acidic conditions



X= electron donating Rate ↓ σ1= -ve

X= electron withdrawing Rate ↑ σ= +ve

Basic conditions: Rate affected by steric + electronic factors
Gives σ_I after correction for steric effect

Acidic conditions: Rate affected by steric factors only (see E_s)

Steric Factors:

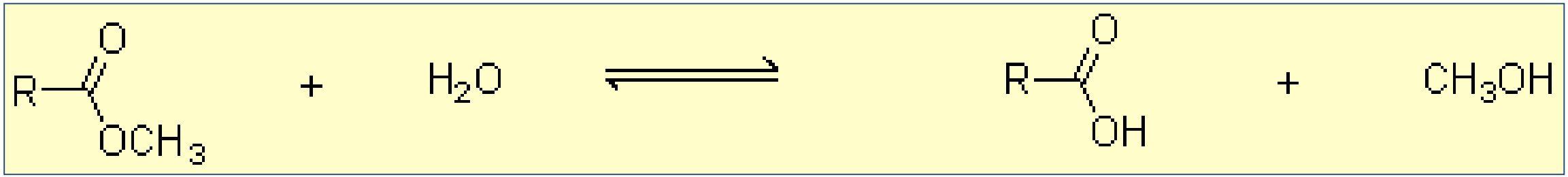
- It is a measure of the bulkiness of the group it represents and it effects on the closeness of contact between the drug and receptor site.
- Much harder to quantitate

Examples are:

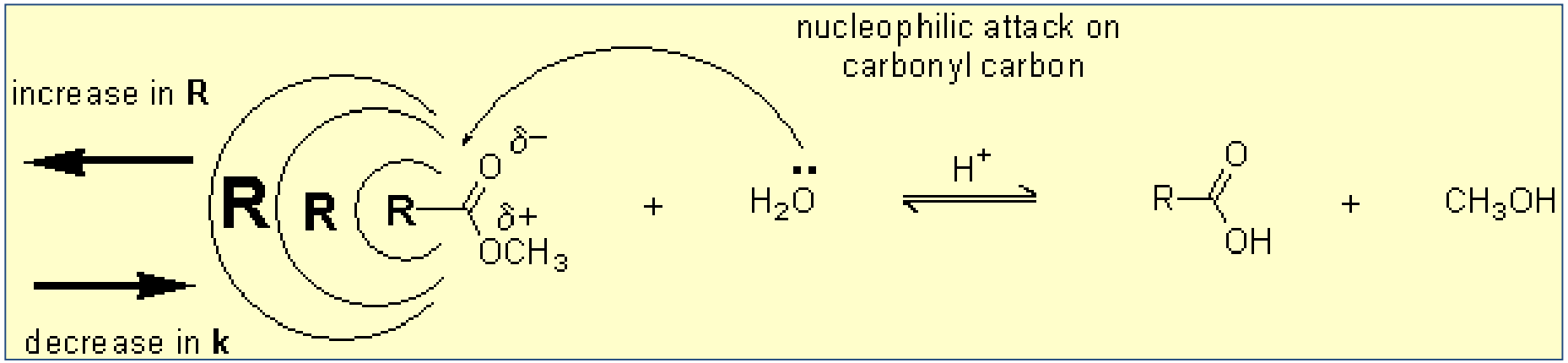
- **Taft's steric factor** (Es) (~1956), an experimental value based on rate constants
- **Molar refractivity (MR)-** measure of the volume occupied by an atom or group- equation includes the MW, density, and the index of refraction--
- **Ver loop steric parameter--**computer program uses bond angles, van der Waals radii, bond lengths

Taft's steric factor (E_s)

- Taft quantified the steric (spatial) effects using the hydrolysis of esters:



- Here, the size of R affects the rate of reaction by *blocking nucleophilic attack by water*.
-



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

k_x represents the rate of hydrolysis of a substituted ester

k_o represents the rate of hydrolysis of the parent ester

Molar Refractivity (MR)

- It is a measure of a substituent's volume

$$MR = \underbrace{\frac{(n^2 - 1)}{(n^2 + 2)}}_{\text{Correction factor for polarisation (n=index of refraction)}} \times \underbrace{\frac{\text{mol. wt.}}{\text{density}}}_{\text{Defines volume}}$$

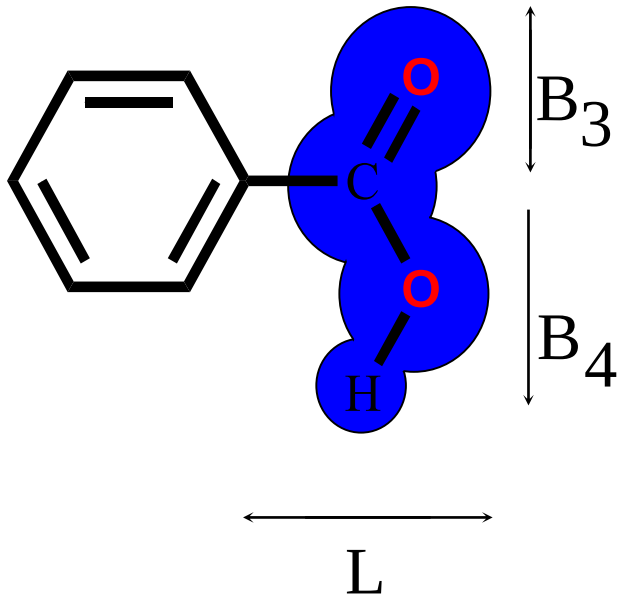
Where n is the index of refraction, MW is the molecular weight, and d is the density.

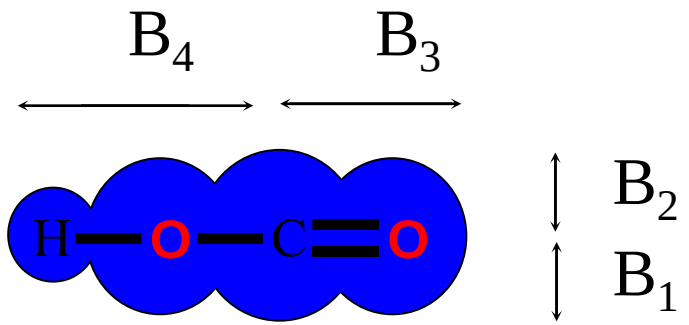
This is particularly significant if the substituent has π electrons or lone pairs of electrons.

Verloop Steric Parameter

- calculated by software (STERIMOL)
- gives dimensions of a substituent
- can be used for any substituent

Example - Carboxylic acid





QSAR Methods:

Quantitative methods

1) Hansch analysis (Linear free energy relationship)

2) Free Wilson analysis (Non-linear free energy relationship)

3) Mixed approach

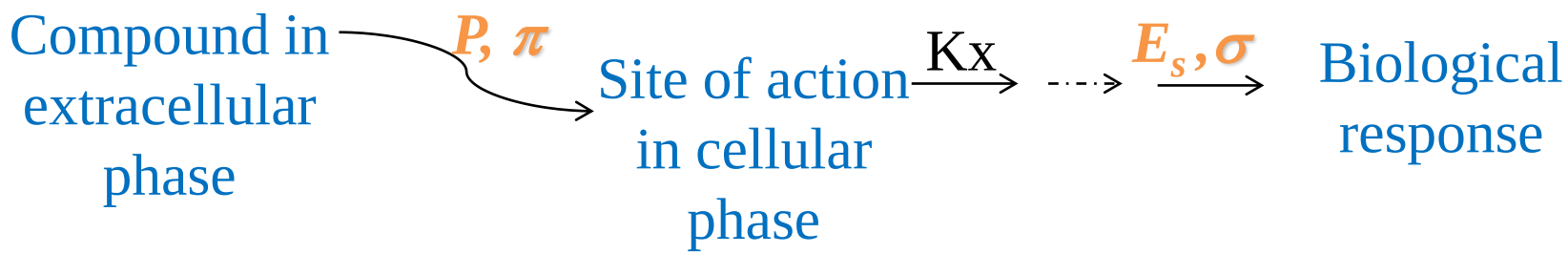
Statistical methods:

1) Regression analysis

2) Principal Component Analysis

3) Partial least squares (PLS) analysis

Hansch Analysis Two stages of drug action was proposed based on



He suggested linear and non- linear dependence of biological activity on different parameters like π,σ and E_S.

Biological Activity = *f* (EL, ST, HPh) + constant

Biological Activity = log1/C , ***C***, drug concentration causes EC50, IC50, etc.

EL (electronic descriptor): σ Hammett constant (σ_m, σ p, σ_p^0, σ_p^+, σ_p^-, R, F)

HPh (hydrophobicity descriptor): π hydrophobic subst. constant, log P octanol/water partition coeff.

ST (steric descriptor): E_s Taft steric constant

Hansch Equation:

- A QSAR equation relating various physicochemical properties to the biological activity of a series of compounds.
- If range of hydrophobicity values is limited to small range then equation will be linear

$$\log 1/C = k_1 \log P + k_2\sigma + k_3E_s + k_4$$

- If log p values spread over large range then equation will be parabolic

$$\log 1/C = -k_1(\log P)^2 + k_2 \log P + k_3\sigma + k_4E_s + k_5$$

- Usually includes log *P*, electronic and steric factors
- Start with simple equations and elaborate as more structures are synthesised
- Typical equation for a wide range of log *P* is parabolic

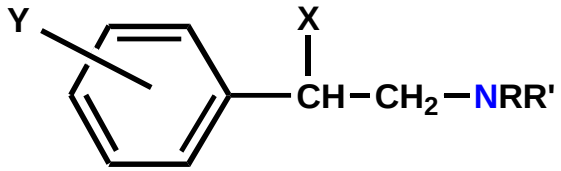
Advantages

- Require no descriptor calculation i.e. utilizes experimentally known substituent constants e.g. π, σ & ρ
- Simple to interpret

Disadvantages

- Applicable only on congeneric series of molecules
- Limited to study of substituent's with known substituent constants
- Does not consider substituent interactions

Example: Adrenergic blocking activity of β-halo-β-arylamines



$$\text{Log}\left(1/C\right) = 1.22\ \pi - 1.59\ \sigma + 7.89$$

Conclusions:

- **Activity increases if π is +ve (i.e. hydrophobic substituents)**
- **Activity increases if σ is negative (i.e. e-donating substituents)**

For Hansch Equation Choosing suitable substituents

Substituents must be chosen to satisfy the following criteria;

- A range of values for each physicochemical property studied
- Values must not be correlated for different properties
- At least 5 structures are required for each parameter studied

Substituent	**H**	**Me**	**Et**	**n-Pr**	**n-Bu**
p	0.00	0.56	1.02	1.50	2.13
MR	0.10	0.56	1.03	1.55	1.96

} Correlated values. Are any differences due to p or MR?

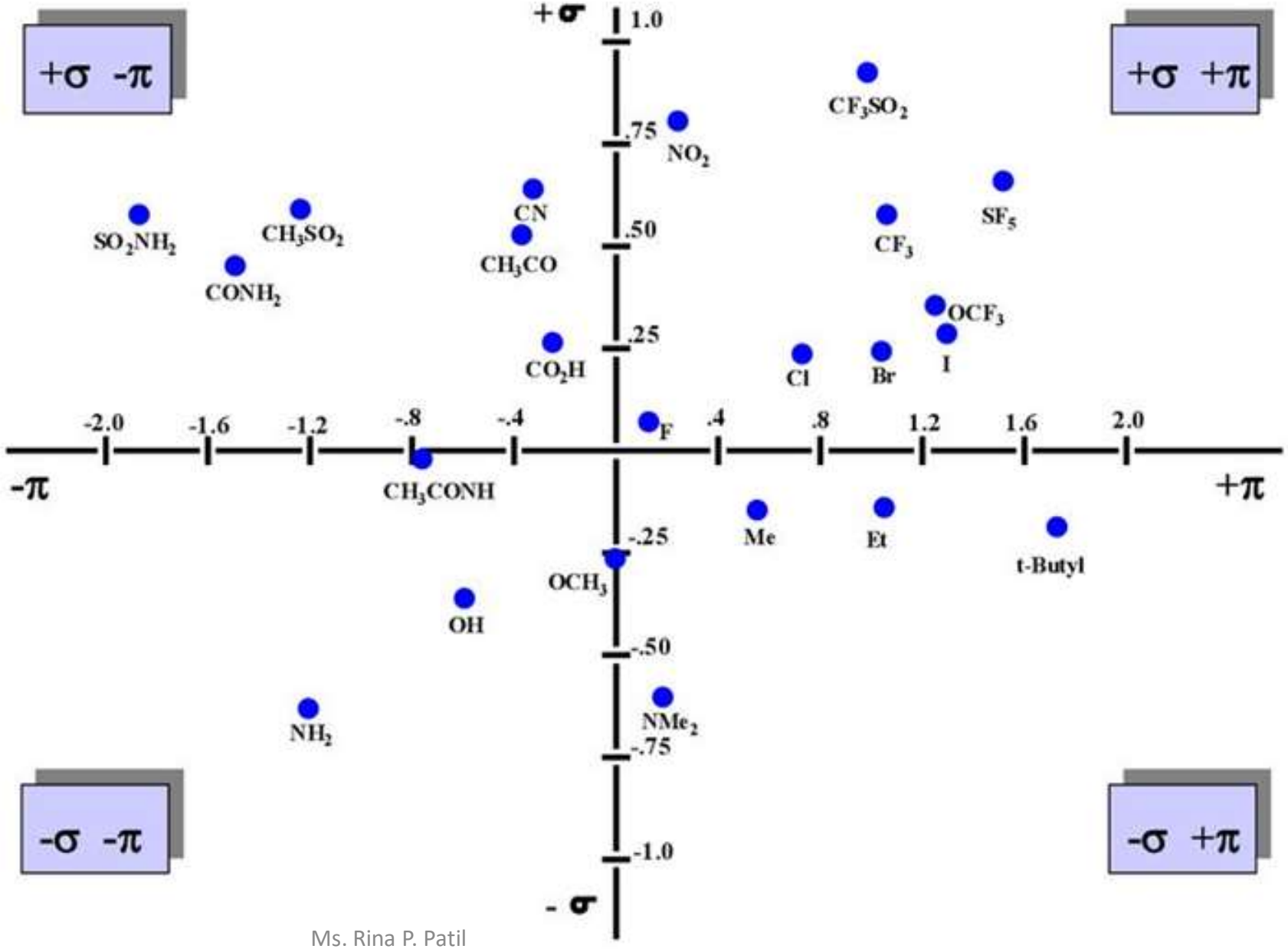
Substituent	**H**	**Me**	**OMe**	**$NHCONH_2$**	**I**	**CN**
p	0.00	0.56	-0.02	-1.30	1.12	-0.57
MR	0.10	0.56	0.79	1.37	1.39	0.63

} No correlation in values Valid for analysing effects of p and MR.

Craig Plot

Craig plot shows values for 2 different physicochemical properties for various substituents

Example:

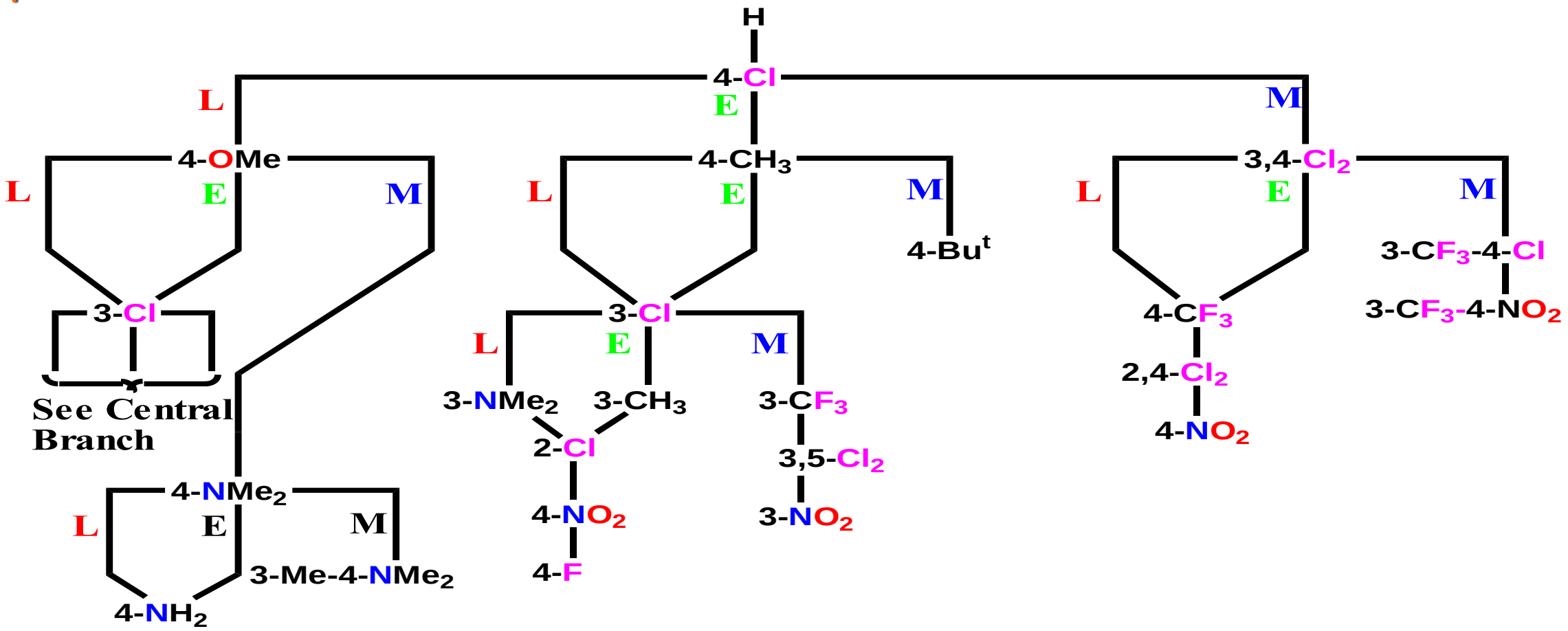


- The plot shows clearly that there is no overall relationship between π and σ. The various substituents are scattered around all four quadrants of the plot.
- It is possible to tell at a glance which substituents have positive π and σ parameters, which substituents have negative π and σ parameters, and which substituents have one positive and one negative parameter.
- It is easy to see which substituents have similar π values. For example, the ethyl, bromo, trifluoromethyl, and trifluoromethylsulfonyl groups are all approximately on the same vertical line on the plot. In theory, these groups could be interchangeable on drugs where the principal factor affecting biological activity is the π factor.
- The Craig plot is useful in planning which substituents should be used in a QSAR study. In order to derive the most accurate equation involving π and σ, analogues should be synthesized with substituents from each quadrant. For example, halogen substituents are useful representatives of substituents with increased hydrophobicity and electron-withdrawing properties (positive π and positive σ), whereas an OH substituent has more hydrophilic and electron-donating properties (negative π and negative σ).
- Once the Hansch equation has been derived, it will show whether π or σ should be negative or positive in order to get good biological activity. Further developments would then concentrate on substituents from the relevant quadrant. For example, if the equation shows that positive π and positive σ values are necessary, then further substituents should only be taken from the top right quadrant.

Topliss Scheme

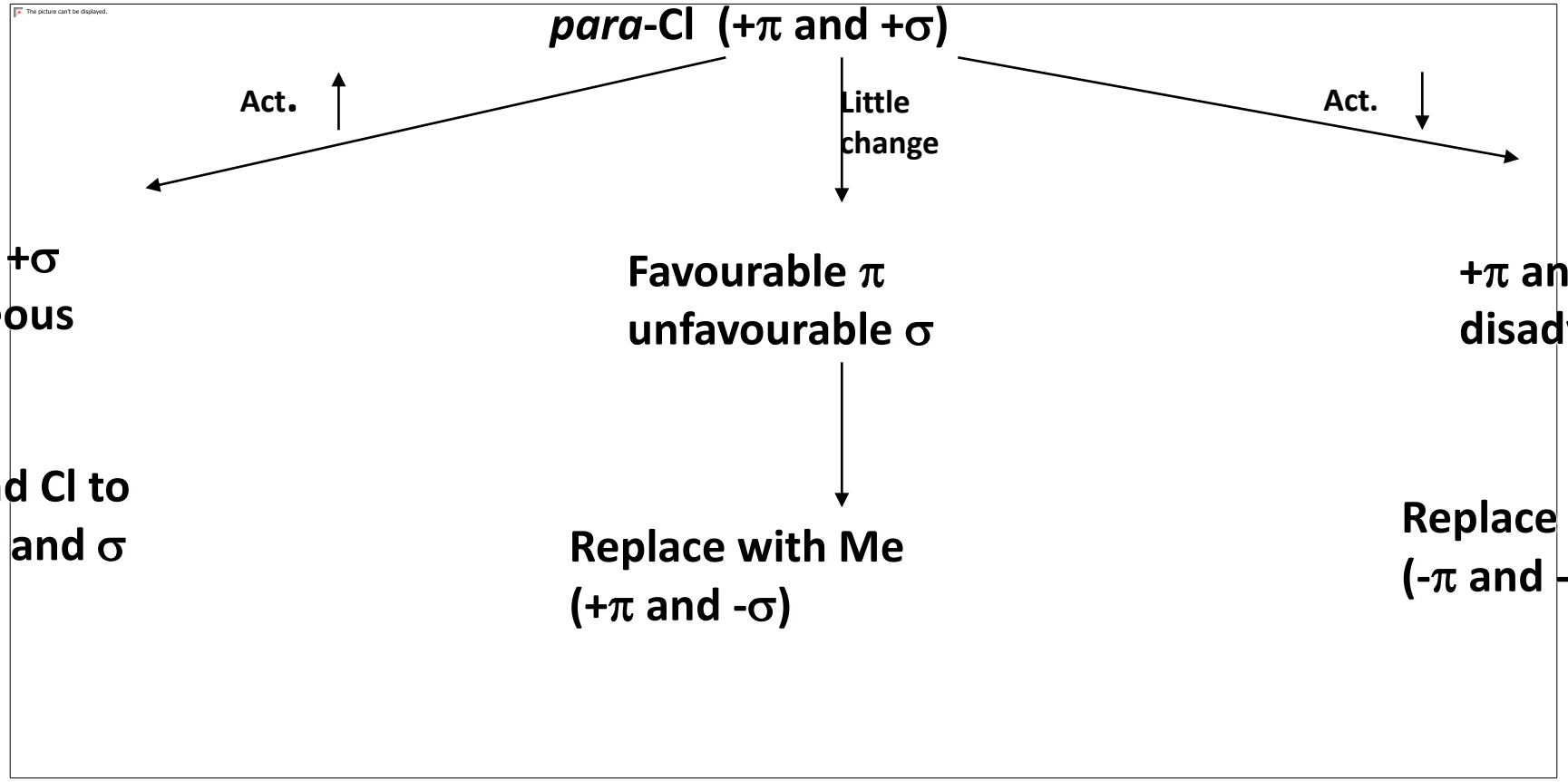
Used to decide which substituent's to use if optimising compounds one by one (where synthesis is complex and slow)

Example: Aromatic substituents



Topliss Scheme

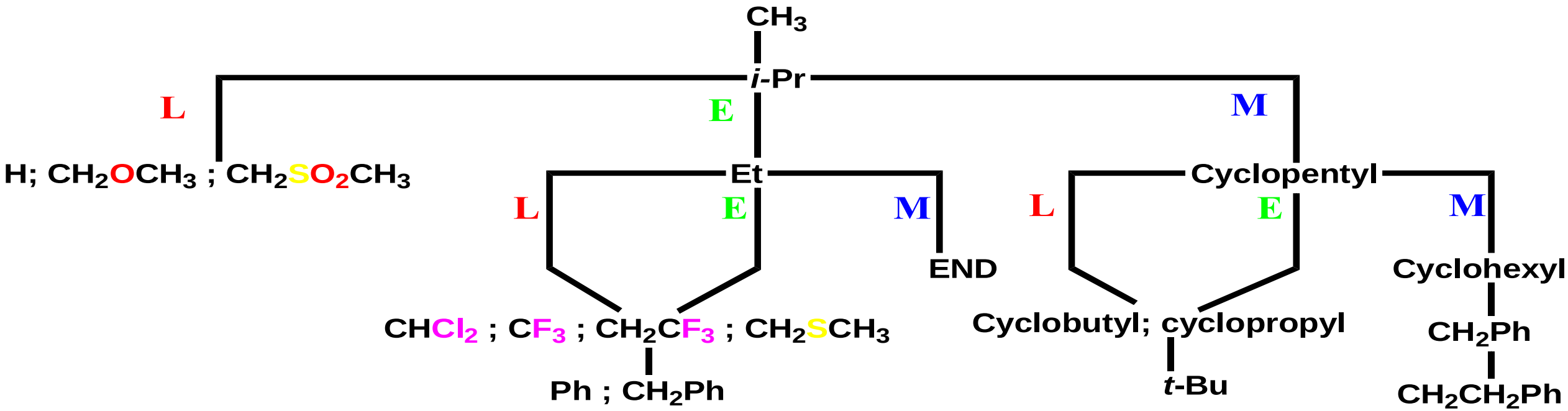
Rationale



Further changes suggested based on arguments of π, σ and steric strain

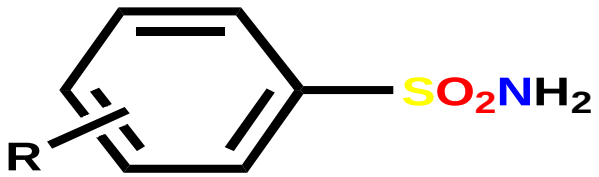
Topliss Scheme

Aliphatic substituents



Topliss Scheme

Example



Order of Synthesis	R	Biological Activity	High Potency
1	H	-	
2	4-Cl	M	
3	3,4-Cl_2	L	
4	4-Br	E	
5	4-NO_2	M	*

M= More Activity
L= Less Activity
E = Equal Activity

Free Wilson analysis

- The biological activity of the parent structure is measured and compared with the activity of analogues bearing different substituents
- Relate biological activity to the presence/absence of a specific functional group at a specific location on the parent molecule
- Basis : The mathematical contribution of chemical (substituent) to structure activity studies

$$\textbf{Activity} = A + G_{ij}X_{ij}$$

- **A** was defined as the average biological activity for the series,
- G_{ij} the contribution to activity of a functional group **i** in the j^{th} position
- and X_{ij} the presence (1.0) or absence (0.0) of the functional group **i** in the j^{th} position

Advantages

- No need for physicochemical constants or tables like π,σ etc
- Useful for structures with unusual substituents
- Useful for quantifying the biological effects of molecular features that cannot be quantified or tabulated by the Hansch method

Disadvantages

- A large number of analogues need to be synthesised to represent each different substituent and each different position of a substituent
- It is difficult to rationalise why specific substituents are good or bad for activity
- The effects of different substituents may not be additive (e.g. intramolecular interactions)

3D-QSAR Notes

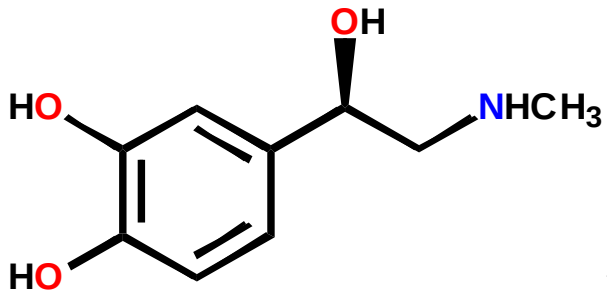
- 3D properties are measured for the molecule as a whole
- Properties are calculated using computer software
- No experimental constants or measurements are involved
- 3D-QSAR assumes that the most important features about a molecule are its overall size and shape, and its electronic properties
- Properties are known as ‘Fields’
  - Steric field - defines the size and shape of the molecule
  - Electrostatic field - defines electron rich/poor regions of molecule
  - Hydrophobic properties are relatively unimportant

Advantages over QSAR

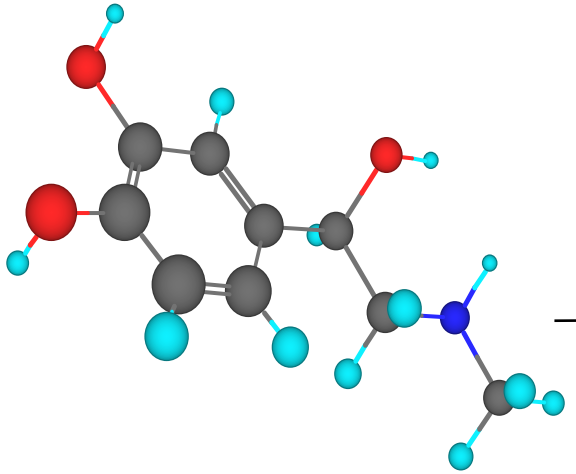
- No reliance on experimental values
- Can be applied to molecules with unusual substituents
- Not restricted to molecules of the same structural class
- Predictive capability

3D-QSAR Method

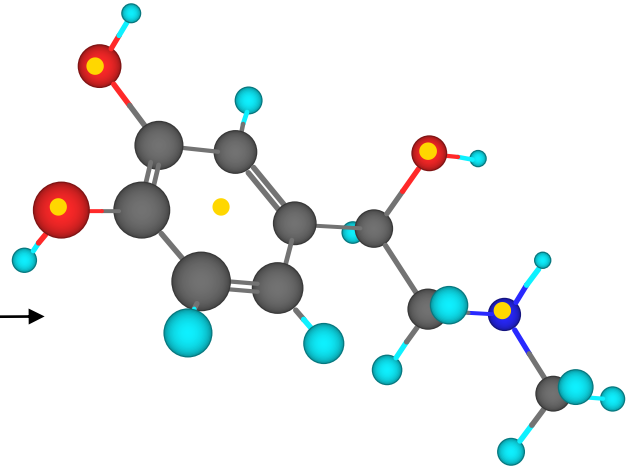
- Comparative molecular field analysis (CoMFA) - Tripos
- Build each molecule using modelling software
- Identify the active conformation for each molecule
- Identify the pharmacophore



Build 3D model



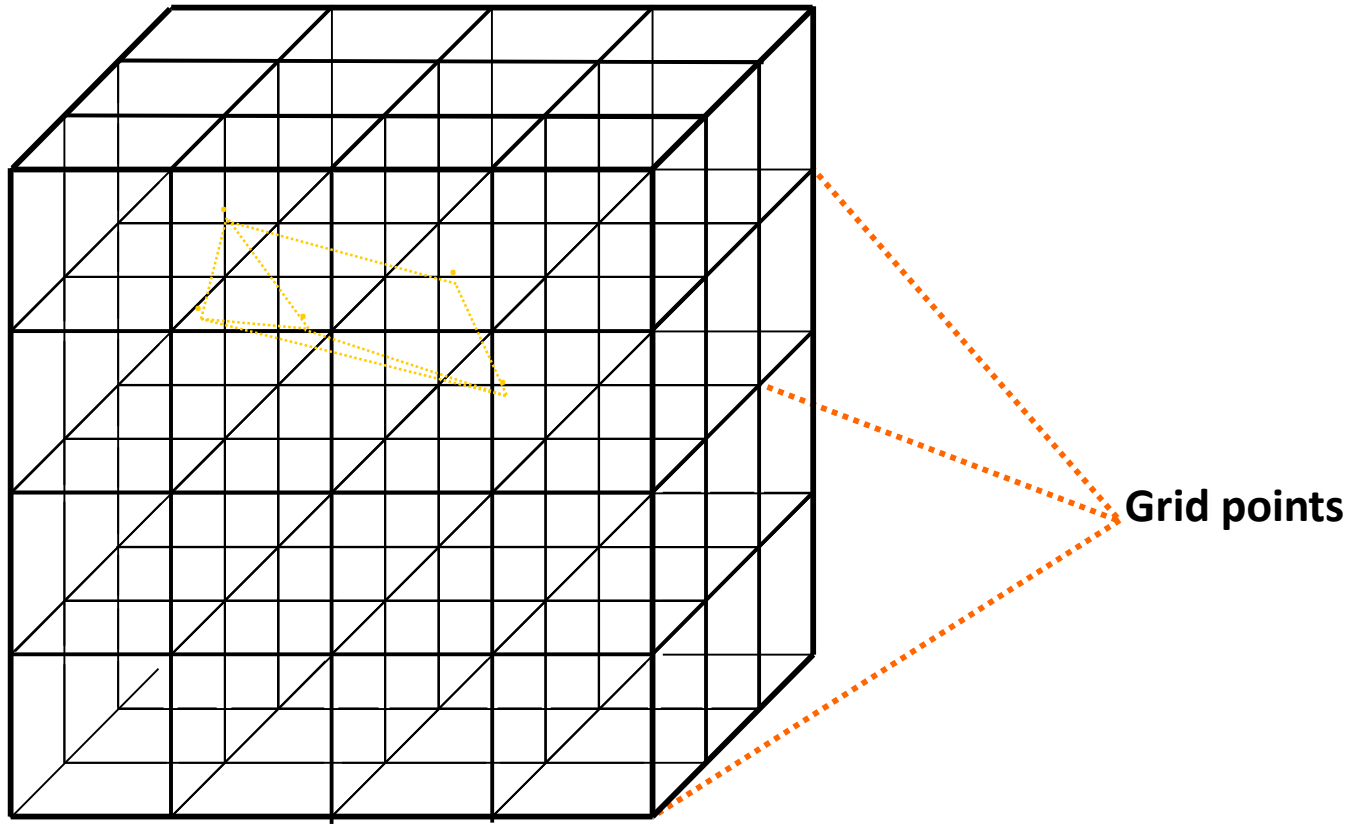
Active conformation



Define pharmacophore

3D-QSAR Method

•Place the pharmacophore into a lattice of grid points

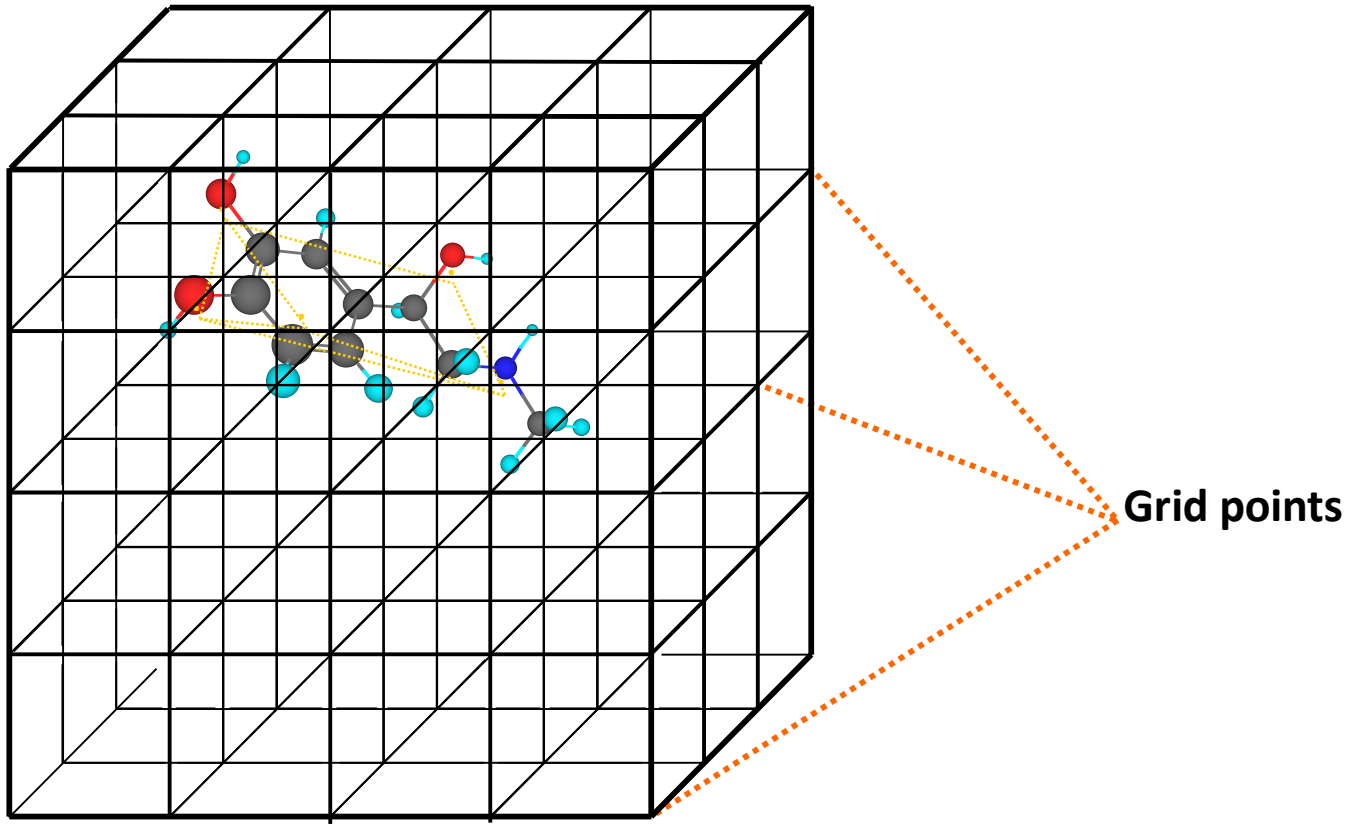


•Each grid point defines a point in space

3D-QSAR

Method

- Position molecule to match the pharmacophore

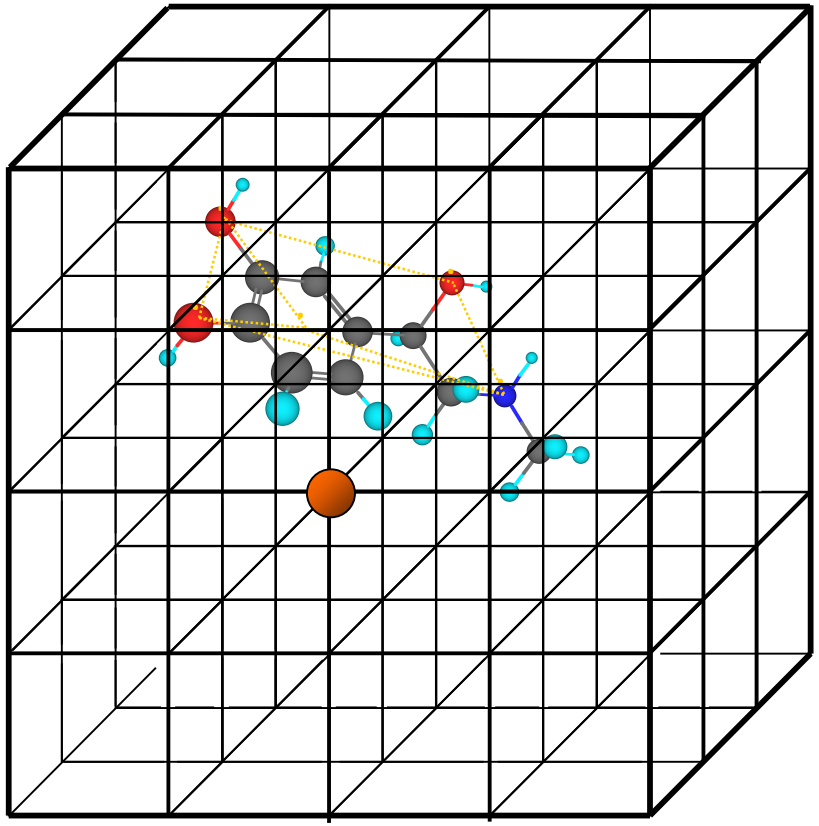


- Each grid point defines a point in space

3D-QSAR

Method

- A probe atom is placed at each grid point in turn



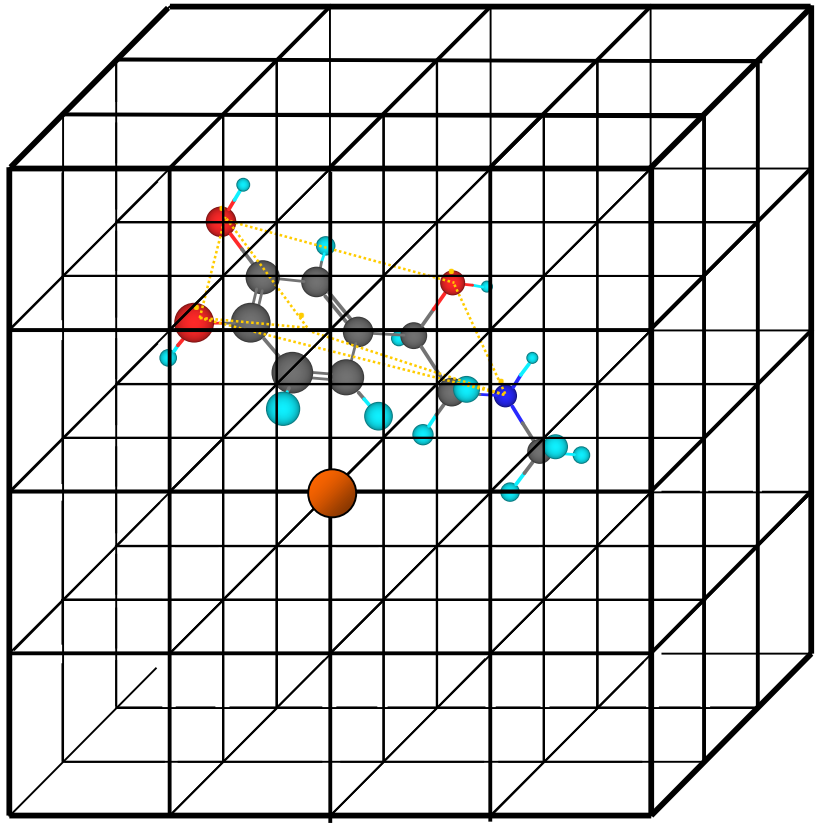
Probe atom

- Probe atom = a proton or sp^3 hybridised carbocation

3D-QSAR

Method

- A probe atom is placed at each grid point in turn



Probe atom

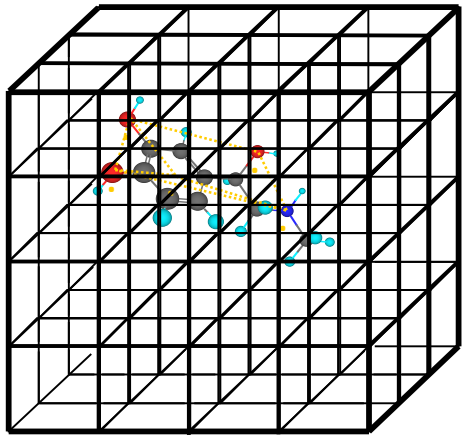
- Measure the steric or electrostatic interaction of the probe atom with the molecule at each grid point

3D-QSAR Method

- The closer the probe atom to the molecule, the higher the steric energy
- Define the shape of the molecule by identifying grid points of equal steric energy (contour line)
- Favourable electrostatic interactions with the positively charged probe indicate molecular regions which are negative in nature
- Unfavourable electrostatic interactions with the positively charged probe indicate molecular regions which are positive in nature
- Define electrostatic fields by identifying grid points of equal energy (contour line)
- Repeat the procedure for each molecule in turn
- Compare the fields of each molecule with their biological activity
- Identify steric and electrostatic fields which are favourable or unfavourable for activity

3D-QSAR

Method



↓ Tabulate fields for each compound at each grid point

Compound	Biological activity	Steric fields (S) at grid points (001-998)					Electrostatic fields (E) at grid points (001-098)				
		S001	S002	S003	S004	S005 etc	E001	E002	E003	E004	E005 etc
1	5.1										
2	6.8										
3	5.3										
4	6.4										
5	6.1										

↓ Partial least squares analysis (PLS)

QSAR equation Activity = aS001 + bS002 +……..mS998 + nE001 +…….+yE998 + z

3D-QSAR

Method

- Define fields using contour maps round a representative molecule