LEARNING MATERIAL

COURSE: B.PHARMACY, 6th Sem, Medicinal Chemistry, BP -601T

Module 05 Introduction to Drug Design. Combinatorial chemistry

Ms. Baljeet Kaur

Assistant Professor

Department of Pharmaceutical Chemistry

ASBASJSM College of Pharmacy, Bela (Ropar)

Objectives:

1. Understand the importance of Drug Design and different technique for drug designing.
2.Know the importance of SAR ofdrugs.

Learning Outcomes:

1.Student will learn about the different techniques of drug designing.
2. Student will learn about the Relation of drug and its Activity.

Quantitative Structure Activity Relationships (QSAR)

CONTENTS

- INTRODUCTION
- HYDROPHOBICITY OF MOLECULE
- HYDROPHOBICITY OF SUBSTITUENTS
- ELECTRONIC EFFECT
- STERIC EFFECT
- HANSCH EQUATION
- CRAIG PLOT
- REFERENCES

INTRODUCTION

Drug designing....

Principles of drug designing

- ♠Improving the binding of drugs
- ♠Increasing the selectivity
- ♠Reduce side effects
- ♠Easy synthesisable
- ♠Arrangement functional groups and identification of a pharmacophore

Drug designing...

Based on lead molecule

♠Traditional
♠Lead compound
♠ Analogue molecules designing new molecule
Eg; salicylic acid and aspirin

Based on target structure

¬By identifying the structure of drug target
¬Designing by denovo drug designing

Based on both leading compound and drug target

- Combination of both methods

QSAR

QSAR approach attempts to identify and quantify the physicochemical properties of a drug and to see whether any of these properties has an effect on the drug's biological activity by using a mathematical equation

PHYSICOCHEMICAL PROPERTIES

- Hydrophobicity of the molecule
- Hydrophobicity of substituents
- Electronic properties of substituents
- Steric properties of substituents

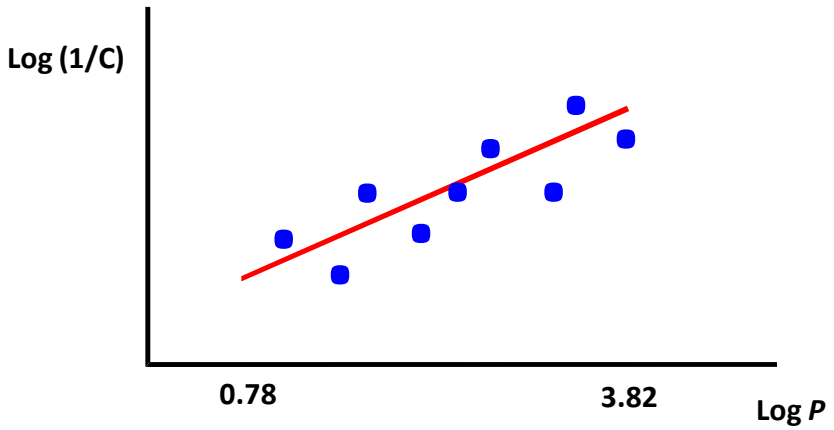
- A range of compounds is synthesized in order to vary one physicochemical property and to test it affects the bioactivity.

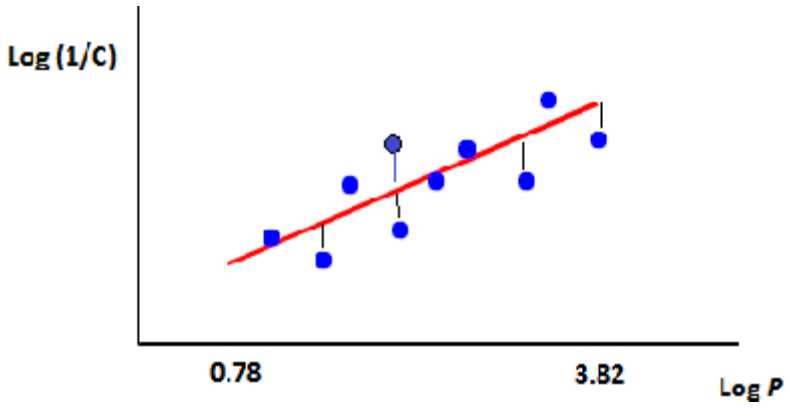
- A graph is then drawn to plot the biological activity on the y axis versus the physicochemical feature on the x axis.

- It is necessary to draw the best possible line through the data points on the graph. This done by procedure known as ***linear regression analysis by the least square method.***

θ If we draw a line through a set of data points will be scattered on either side of the line. The best line will be the one closest to the data points.

θ To measure how close the data points are , vertical lines are drawn from each point.





HYDROPHOBICITY

Hydrophobic character of a drug is crucial to how easily it crosses the cell membrane and may also important in receptor interactions.

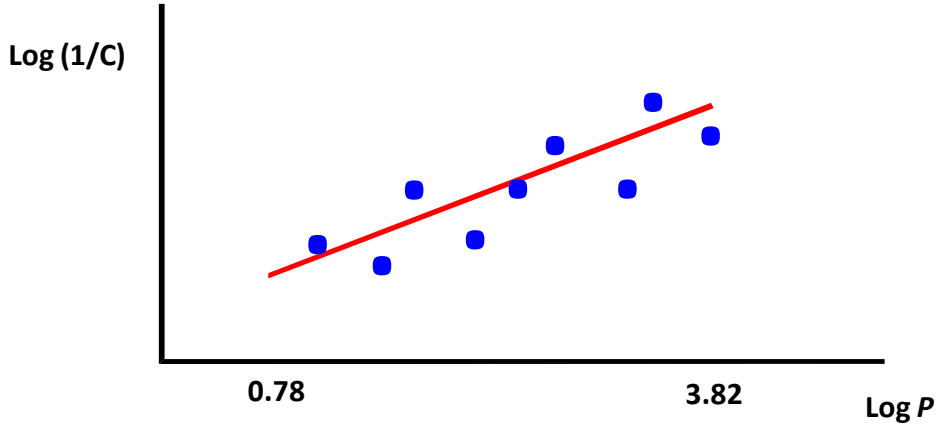
Hydrophobicity of a drug is measured experimentally by testing the drugs relative distribution in octanol water mixture. This relative distribution is known as partition coefficient.

$$\text{Partition Coefficient } P = \frac{\text{conc. Drug in in octanol]}}{\text{[Conc.of drug in water]}}$$

- Activity of drugs is often related to *P*

Biological activity log(1/c) = K1 log P + K2

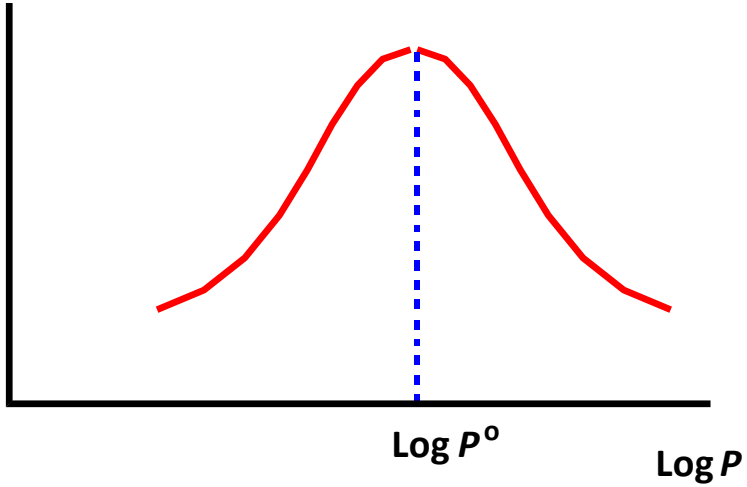
Eg: binding of a drug to serum albumin determined by hydrophobicity and study of 42 compounds.
(straight line - limited range of log *P*)



log(1/c) = 0.075 log P + 230
(n= 42, r= 0.960 s= 0.159)

If the partition coefficient is the only factor influencing biological activity, the parabolic curve can expressed by the equation

$$\log(1/c) = -K1 (\log P)2 + K2 \log P + k3$$



Few drugs where activity is related to log P factor alone.

- QSAR equations are only applicable to compounds in the same structural class (e.g. ethers)

- However, log *P*o is similar for anaesthetics of different structural classes

THE SUBSTITUENT HYDROPHOBICITY CONSTANT (π)

Partition coefficient can be calculated by knowing the contribution that various substituents, is known as substituent hydrophobicity constant(π)

- A measure of a substituent's hydrophobicity relative to hydrogen
- Partition coefficient is measured experimently for a standard compound such as benzene with or without a substituent (X).
- The hydrophobicity constant (π x) for sustituent X.

The equation is

$$\pi x = \log Px - \log PH$$

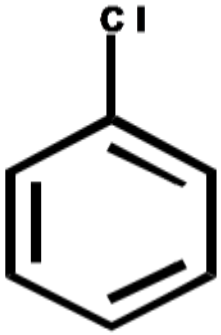
A possitive π value shows that the substituent is more hydrophobic than hydrogen

A negative value indicates that the substituent is less hydrophobic.

The π value is charecteristic for sustituent.

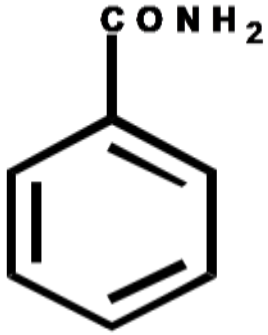
Example:

Benzene
(Log P = 2.13)



Chlorobenzene
(Log P = 2.84)

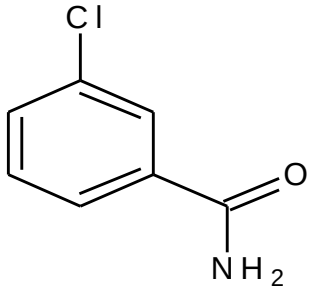
$\pi Cl = 0.71$



Benzamide
(Log P = 0.64)

$\pi CONH_2 = -1.49$

THE SUBSTITUENT HYDROPHOBICITY CONSTANT (π)



meta chlorobenzamide

Log *P*(theory) = log *P*(benzene) + $\pi X\lambda + \pi XONH_2$

= 2.13 + 0.71 – 1.49

= 1.35

Λογ Π(οβσερωεδ) = 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

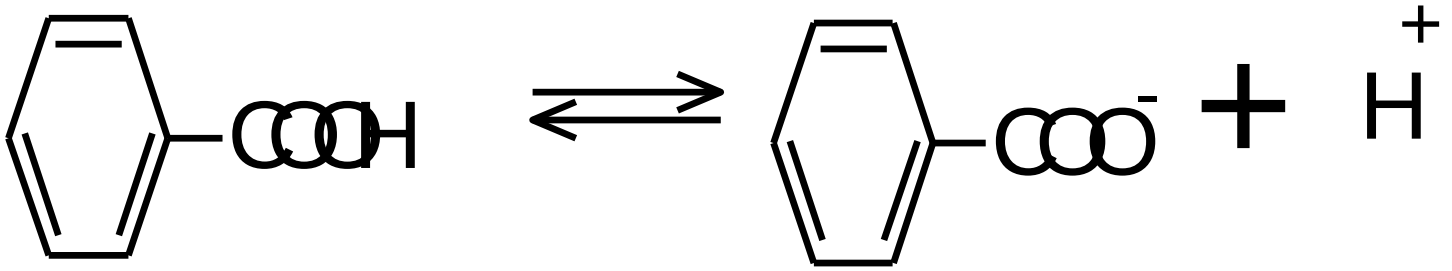
ELECTRONIC EFFECT

The electronic effect of various sustituents will clearly have an effect on drug ionisation and polarity.

Have an effect on how easily a drug can pass through the cell membrane or how strongly it can interact with a binding site.

Hammet substituent constant(σ) this is a measure of electron with-drawing or electron-donating ability of a substituents on an aromatic ring.

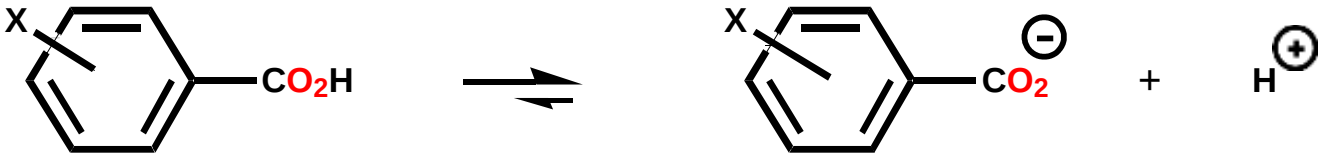
o for aromatic substituents is measured by comparing the dissociation constants of substituted benzoic acids with benzoic acid



$$K_H = \text{Dissociation constant} = \frac{[PhCO_2^-]}{[PhCO_2H]}$$

X= electron withdrawing group (e.g. NO2,)

X = electron withdrawing group

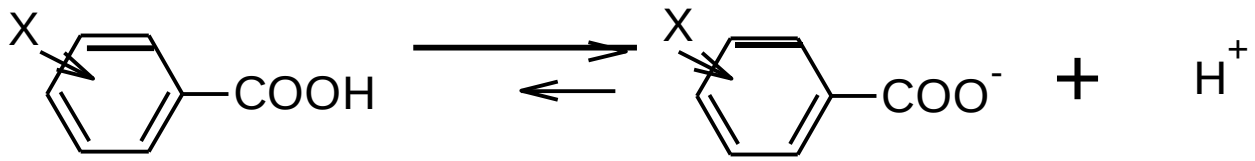


Charge is stabilised by X
Equilibrium shifts to right
KX > KH

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

Positive value

X= electron donating group (e.g. CH3)



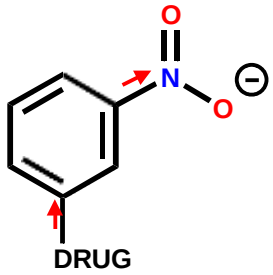
Charge destabilised
Equilibrium shifts to left
KX < KH

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

Negative value

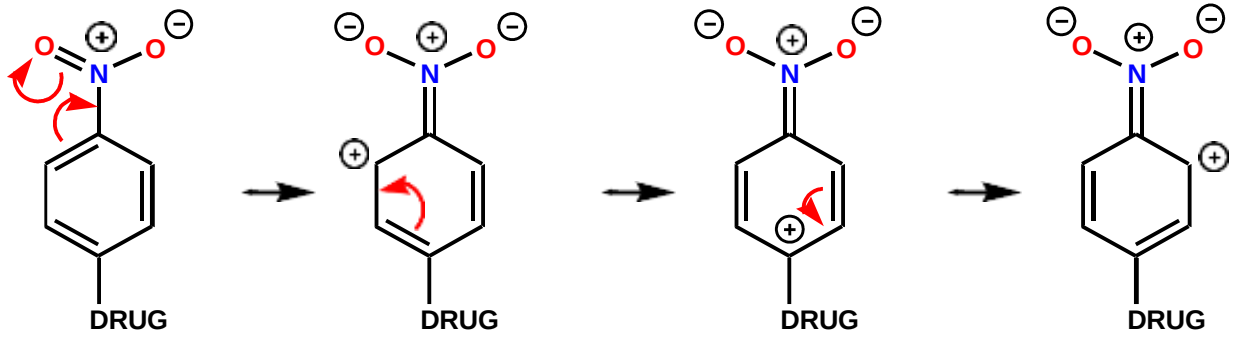
EXAMPLES: σ **p (NO2)** $= 0.78$ σ **m (NO2)** $= 0.71$

***meta*-Substitution**



e-withdrawing (inductive effect only)

***para*-Substitution**



e-withdrawing (inductive + resonance effects)

- value depends on inductive and resonance effects
- value depends on whether the substituent is *meta* or *para*

ortho values are invalid due to steric factors

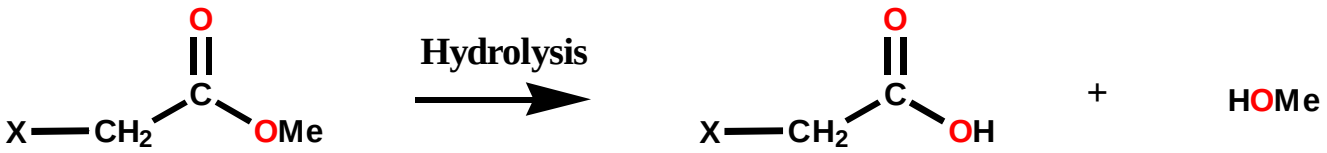
Electronic Factors *R* & *F*

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

The constants σ,R and F can only be used for aromatic substituents

Aliphatic electronic substituents

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



X= electron donating Rate ↓ σI = -ve

X= electron withdrawing Rate ↑ σI = +ve

Basic conditions: Rate affected by steric + electronic factors

Gives σI after correction for steric effect

Acidic conditions: Rate affected by steric factors only (see *E*s)

STERIC FACTORS

The bulk, size and shape of a drug will influence how easily it can approach and interact with binding site.

A bulky substituents may act like a shield and hinder the ideal interaction between a drug and its binding site.

Bulky substituent may help to orient a drug properly for maximum binding and increase activity.

Taft's Steric Factor (*E*s)

- Measured by comparing the rates of hydrolysis of substituted aliphatic esters against a standard ester under acidic conditions

$$Es = \log kx - \log ko$$

*k*x represents the rate of hydrolysis of a substituted ester
*k*o represents the rate of hydrolysis of the parent ester

- Limited to substituents which interact sterically with the tetrahedral transition state for the reaction
- Not by resonance or hydrogen bonding

Disadvantages

ES value measures intramolecular steric effect but drugs interact with target binding site in intermolecular process (i.e. a drug

Molar Refractivity (*MR*)

this is a measure of a substituent's volume

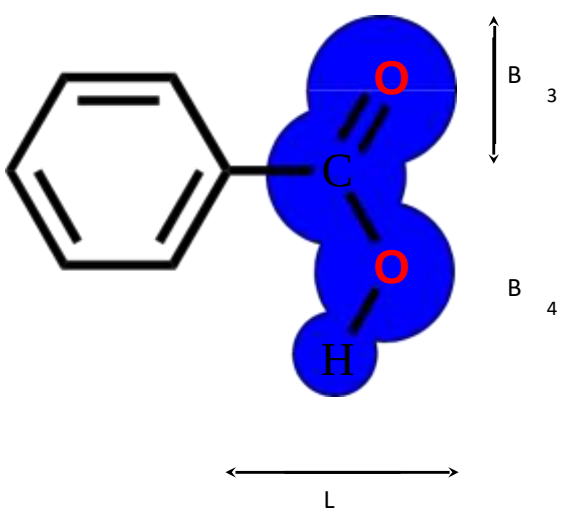
$$MR = \underbrace{\frac{(n^2 - 1)}{(n^2 - 2)}}_{\text{Correction factor for polarisation (n=index of refraction)}} \times \underbrace{\frac{\text{mol. wt.}}{\text{density}}}_{\text{Defines volume}}$$

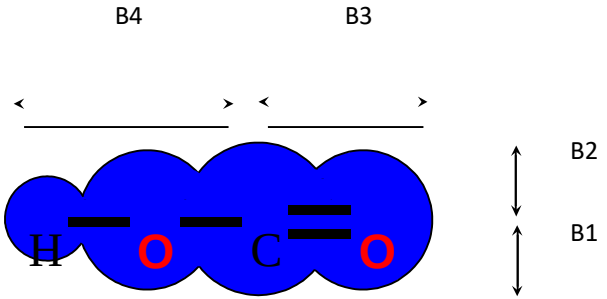
This is perticularly significant if the substituent has π electrons or lone pair of electrons

Verloop Steric Parameter

- calculated by software (STERIMOL)
- gives dimensions of a substituent from the standard bond angle ,van der Waals radii, bond length and possible conformations for the substituents
- can be used for any substituent

Example - Carboxylic acid





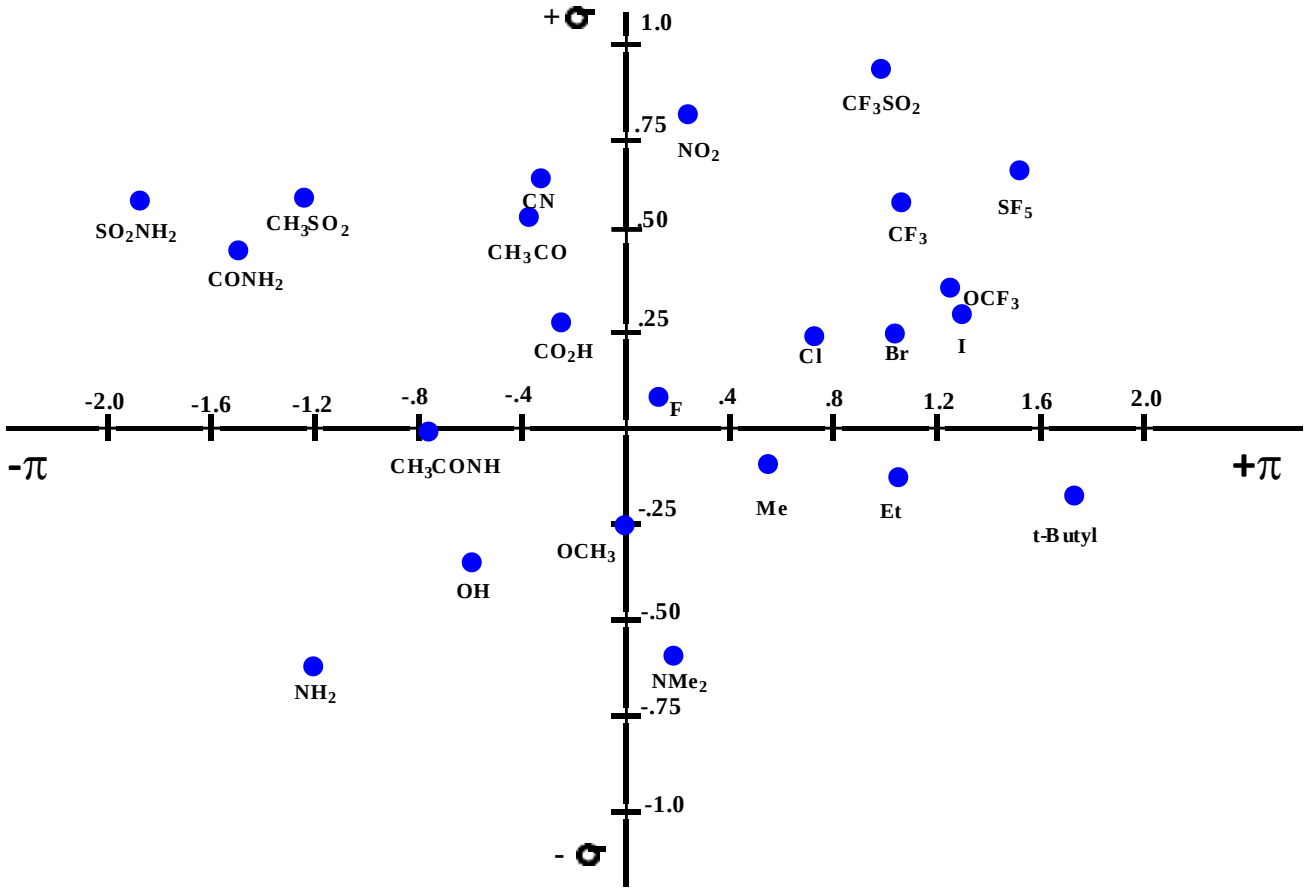
HANSCH EQUATION

- A QSAR equation relating various physicochemical properties to the biological activity of a series of compounds
- 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

$$\text{Log}\left(1/C\right) = -k_1(\log P)^2 + k_2 \log P + k_3\sigma + k_4 E_s + k_5$$

Craig Plot

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



- Allows an easy identification of suitable substituents for a QSAR analysis which includes both relevant properties

- Choose a substituent from each quadrant to ensure orthogonality

- Choose substituents with a range of values for each property

REFERENCES

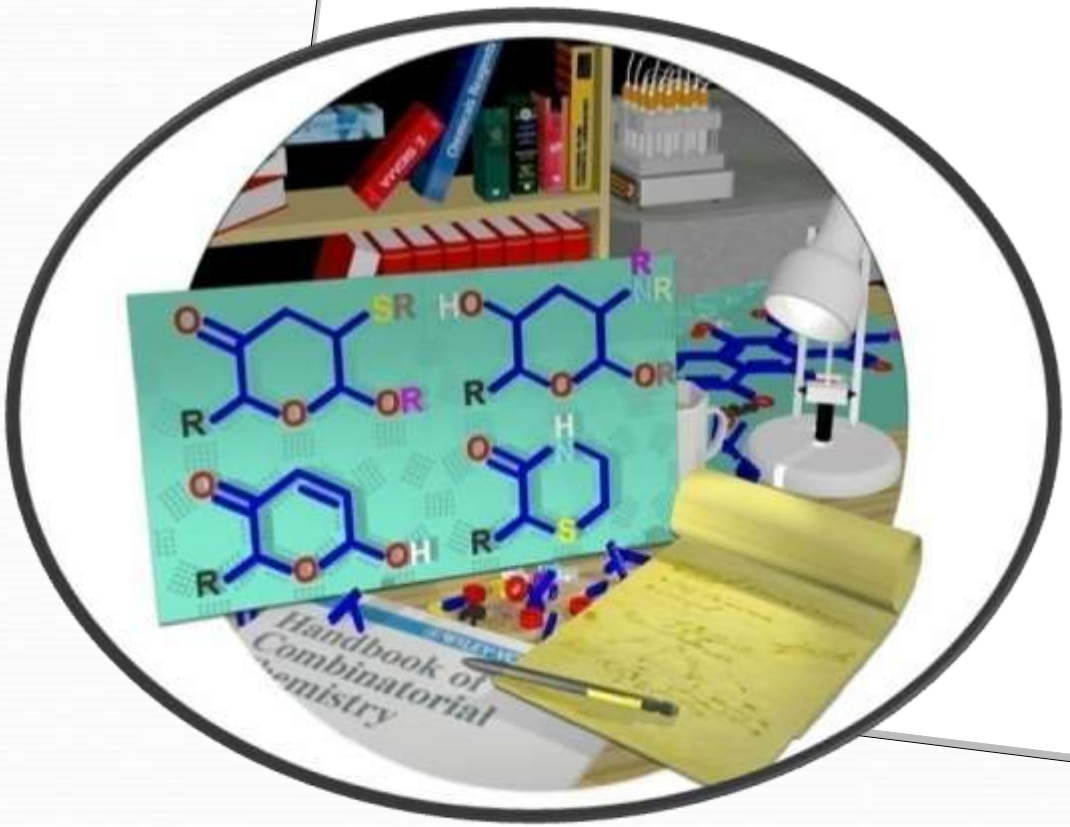
1. An introduction to medicinal chemistry by Graham L Patric

   3rd edition pagee no:271-298

2. Foye : Principles of medicinal chemistry

3. Burgers medicinal chemistry

Combinatorial Chemistry



Introduction:

•Combinatorial Chemistry is a new method developed by academics and researchers to reduce the time and cost of producing effective, marketable and competitive new drugs.

•Scientists use Combinatorial Chemistry to create large numbers of molecules that can be detected efficiently.

•This technique has captured the attention of many areas such as Pharmaceutical chemistry, Biotechnology and Agro chemistry.

Definition:

•Combinatorial chemistry is a technique by which large numbers of different but structurally similar molecules are produced rapidly and submitted for pharmacological assay.

•This technique uses the same reaction conditions with the same reaction vessels to produce a large range of analogues.

•Technique invented in the late 1980s and early 1990s to enable tasks to be applied to many molecules simultaneously


STRATEGIES
Conventional
Combinatorial
– One molecule at a time
– Make → Purity → Test
– hudreds of molecules a month
– Slower lead generation
– High risk of failure
– Many molecules at a time
– Make → Test → Purity
– Thousands of molecules a month
– Faster leads generation
– Low risk of failure
Synergy
LEAD IDENTIFICATION

Application:

- Applications of combinatorial chemistry are very wide Scientists use combinatorial chemistry to create large populations of molecules that can be screened efficiently.
- By producing larger, more diverse compound libraries, companies increase the probability that they will find novel compounds of significant therapeutic and commercial value.
- Provides a stimulus for robot-controlled and immobilization strategies that allow high-thrughput and multiple parallel approaches to drug discovery

Advantages:

Fast

Combinatorial approach can give rise to million of compound in same time as it will take to produce one compound by traditional method of synthesis .

Economical

A negative result of mixture saves the effort of synthesis, purification & identification of each compound

Easy

Isolation purification & identification of active molecule from combinatorial library is relatively easy.

Drug Discovery

Mixed Combinatorial synthesis produces chemical pool. Probability of finding a molecule in a random screening process is *proportional* to the number of molecules subjected to the screening process

Drug Optimization

Parallel synthesis produces analogues with slight differences which is required for lead optimization

Disadvantages:

Efficiency is highly affected by compound's size, solubility and function group.

Compounds produced tend to be Achiral of Racemic

Combinatorial Chemistry within drug design (1)

Combinatorial Chemistry within drug design

Impact at lead discovery

- traditionally lead drugs were found from
    - natural products
    - synthetic custom crafted organic molecules made in small numbers
    - analogues of known actives (analogue me-toos)

- High Throughput screening (HTS) requires large numbers of compounds to fuel the discovery process

- As an alternative to traditional synthesis many compounds rapidly constructed was needed

Tools:

1. Solid Phase Techniques
    - 2.1. Advantages
    - 2.2. Requirements
    - 2.3. Examples of Solid Supports (2 slides)
    - 2.4. Anchor or linker
        1. Merrifield resin for peptide synthesis (chloromethyl group)
        2. Wang resin (2 slides)
        3. Rink resin (2 slides)
        4. Dihydropyran resin (2 slides)
2. Parallel Synthesis
    - 3.1. Houghton's Tea Bag Procedure
    - 3.2. Automated parallel synthesis (2 slides)
    - 3.3. Automated parallel synthesis of all 27 tripeptides from 3 amino acids (2 slides)
3. Mixed Combinatorial Synthesis
4. Solution phase synthesis

1. SOLID PHASE TECHNIQUES

- Reactants are bound to a polymeric surface and modified whilst still attached. Final product is released at the end of the synthesis

Advantages

- Specific reactants can be bound to specific beads
- Beads can be mixed and reacted in the same reaction vessel
- Products formed are distinctive for each bead and physically distinct
- Excess reagents can be used to drive reactions to completion
- Excess reagents and by products are easily removed
- Reaction intermediates are attached to bead and do not need to be isolated and purified
- Individual beads can be separated to isolate individual products
- Polymeric support can be regenerated and re-used after cleaving the product
- Automation is possible

1. SOLID PHASE TECHNIQUES

Requirements

- A resin bead or a functionalised surface to act as a solid support
- An anchor or linker
- A bond linking the substrate to the linker. The bond must be stable to the reaction conditions used in the synthesis
- A means of cleaving the product from the linker at the end
- Protecting groups for functional groups not involved in the synthesis

1. SOLID PHASE TECHNIQUES

Examples of Solid Supports

- Partially cross-linked polystyrene beads hydrophobic in nature

  causes problems in peptide synthesis due to peptide folding

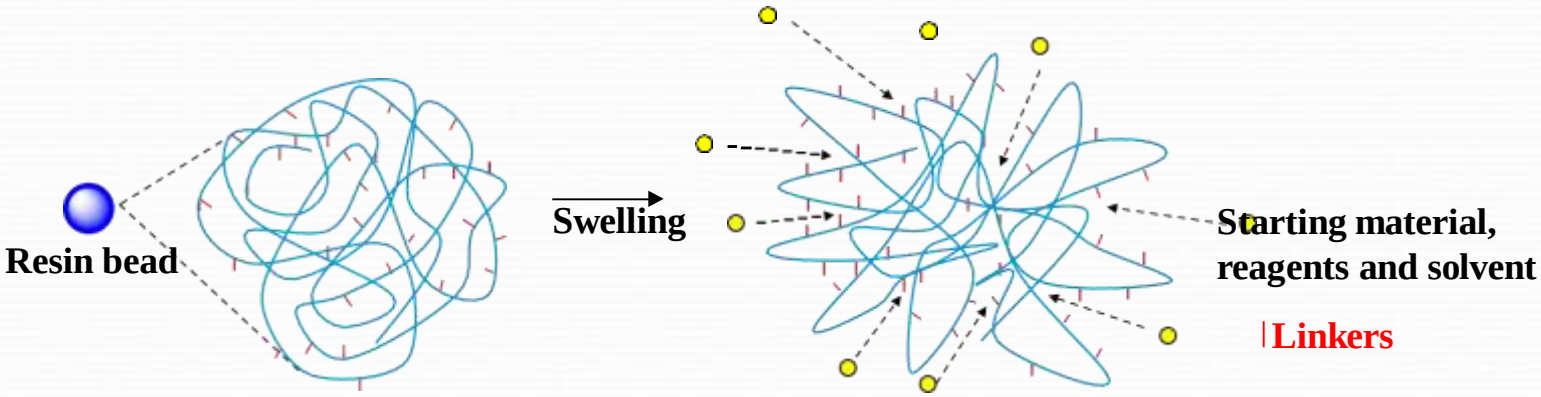
- Sheppard's polyamide resin - more polar

- Tentagel resin - similar environment to ether or THF

- Beads, pins and functionalised glass surfaces

1. SOLID PHASE TECHNIQUES

- Beads must be able to swell in the solvent used, and remain stable
- Most reactions occur in the bead interior



1. SOLID PHASE TECHNIQUES

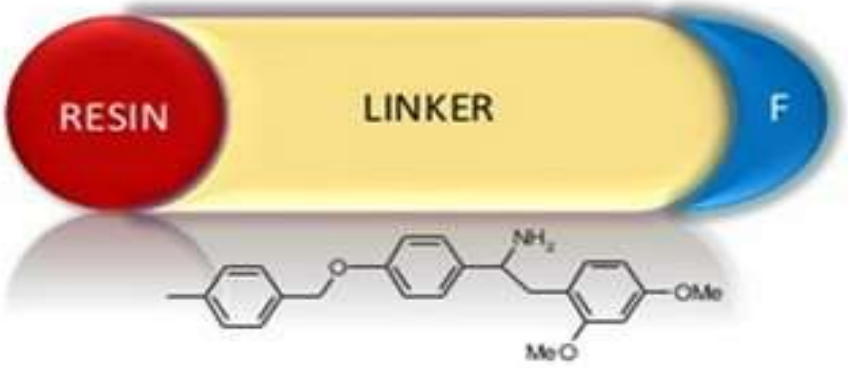
Anchor or linker

- A molecular moiety which is covalently attached to the solid support, and which contains a reactive functional group
- Allows attachment of the first reactant
- The link must be stable to the reaction conditions in the synthesis but easily cleaved to release the final compound
- Different linkers are available depending on the functional group to be attached and the desired functional group on the product
- Resins are named to define the linker e.g.

Merrifield, Wang, Rink

WANG RESIN: linker suitable for attachment & release of carboxylic acids.

MERRIFIELD RESIN : linker suitable for peptide products.



RINK RESIN : Linker suitable for attachment & release of carboxamide.

Solid phase synthesis: protecting groups

- A few protecting groups used in solid phase synthesis.

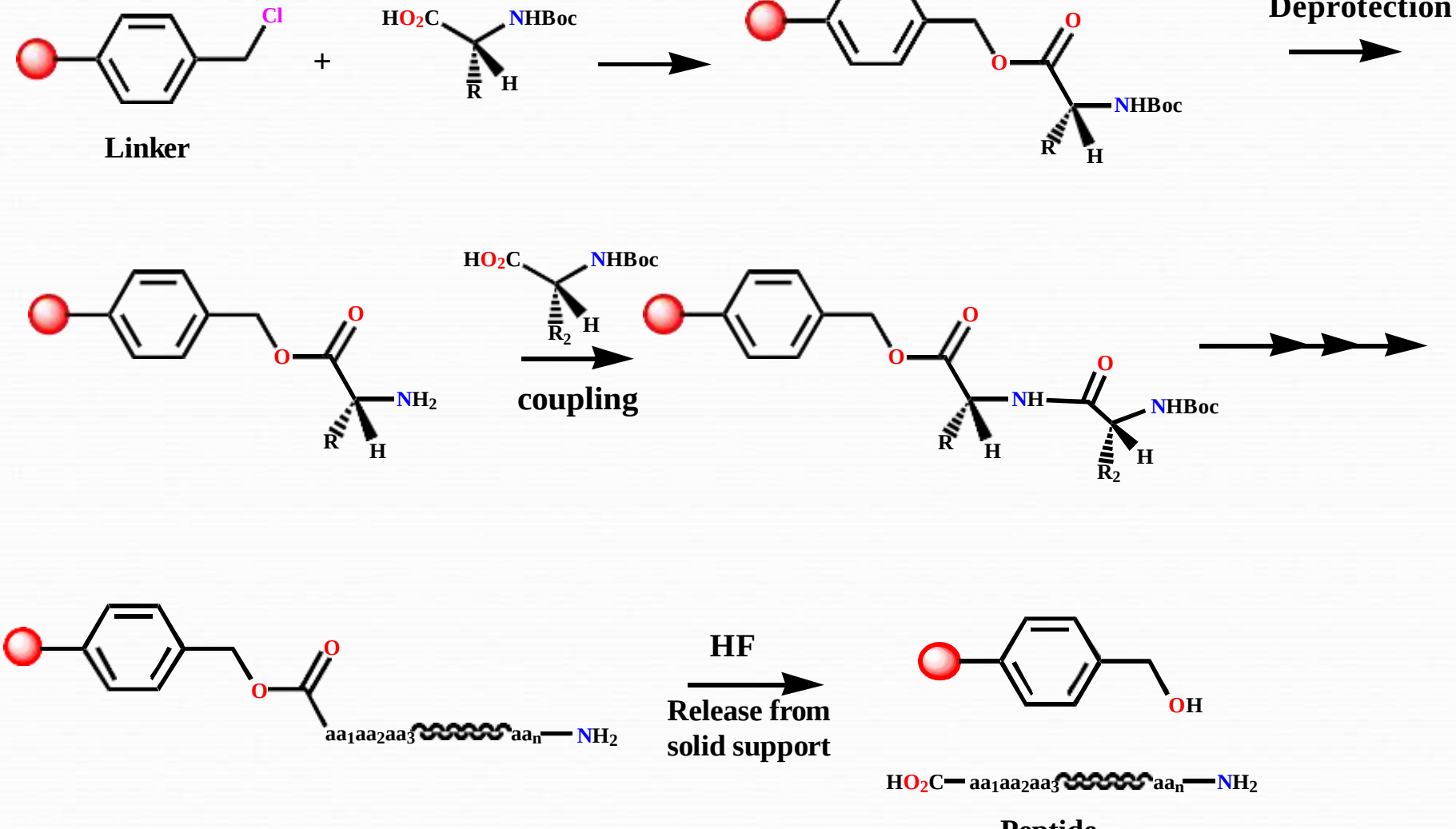
- **For amines.**
  - Boc (t-butoxycarbonyl)
  - Fmoc (9-fluorenylmetoxy carbonyl)
- Tmsec (2 [trimethylsilyl] ethoxycarbonyl)

- **For carboxylic acids.**
  - Tert Bu ester(t-butyl ester)
  - Fm ester(9-fluronyl methyl ester)
  - Tmse ester(2 [trimethylsilyl] ethyl)

17

ield resin for peptide synthesis (c hloromethyl group)

= resin bead



equipment for Solid Phase Peptide Synthesis

4.Solution phase synthesis

- Unlike one bead - one compound synthesis ,solution phase synthesis often lead to mixture of products in one pool.

- Most of the org reaction occurs in solution phase . For this reason there has been much interest in solution phase synthesis.

- The main problem here is the difficulty of removing unwanted impurities at each step in synthesis.

Acid chloride based set:-

A1 + (B1,B2,B3.....B10) ⟶ mixture1 containing all possible A1-B compounds

A2 + (B1,B2,B3.....B10) ⟶ mixture2 containing all possible A2-B compounds

A10+(B1,B2,B3.....B10) ⟶ mixture10 containing all possible A10-B compounds

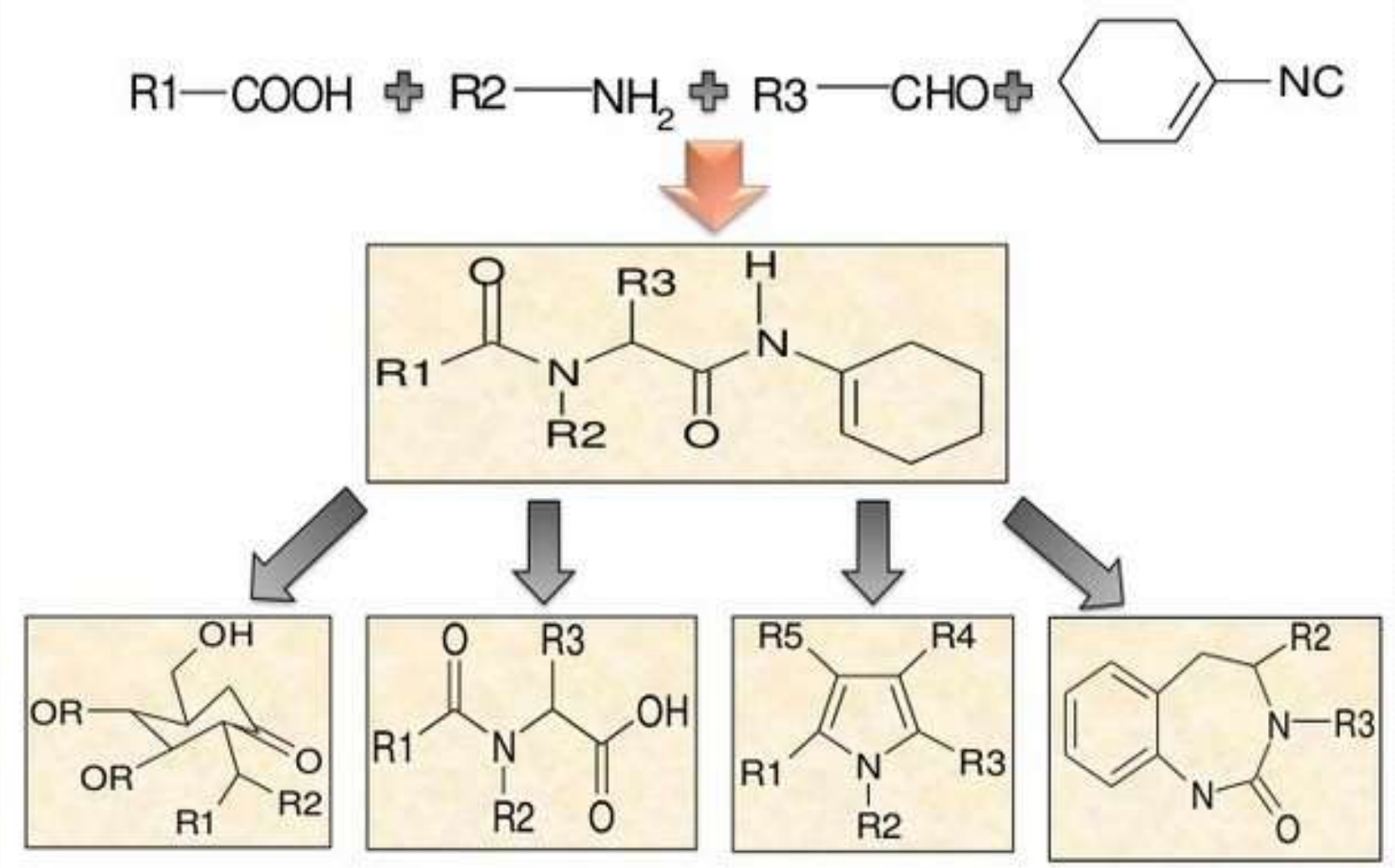
Amine based set:-

B1 + (A1,A2,A3.....A10) ⟶ mixture1 containing all possible B1-A compounds

B2 + (A1,A2,A3.....A10) ⟶ mixture2 containing all possible B2-A compounds

B10+(A1,A2,A3.....A10) ⟶ mixture10 containing all possible B10-Acompounds

4 component UGI reaction:



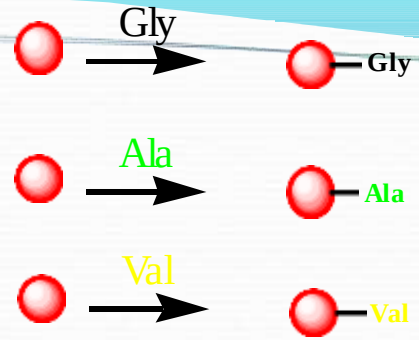
5. Identification of structures from mixed combinatorial synthesis

5.1 Recursive Deconvolution:

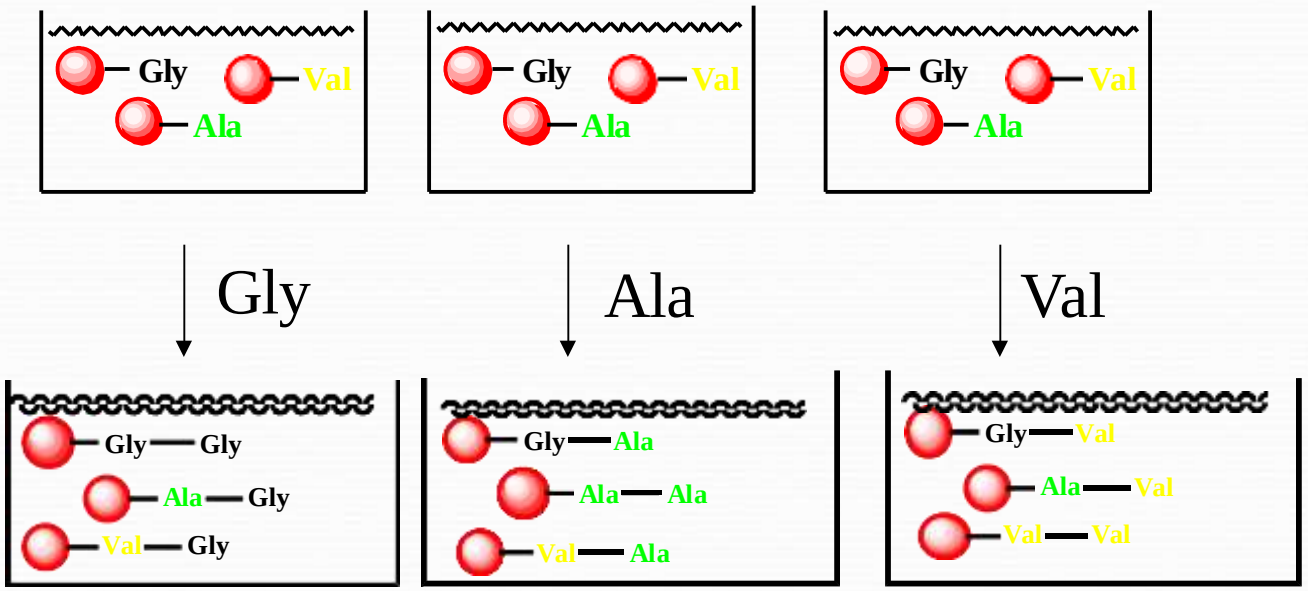
- Method of identifying the active component in a mixture
- Quicker than separately synthesising all possible components
- Need to retain samples before each mix and split stage

Example:

Consider all 27 tripeptides synthesised by the mix and split strategy from glycine, alanine and valine

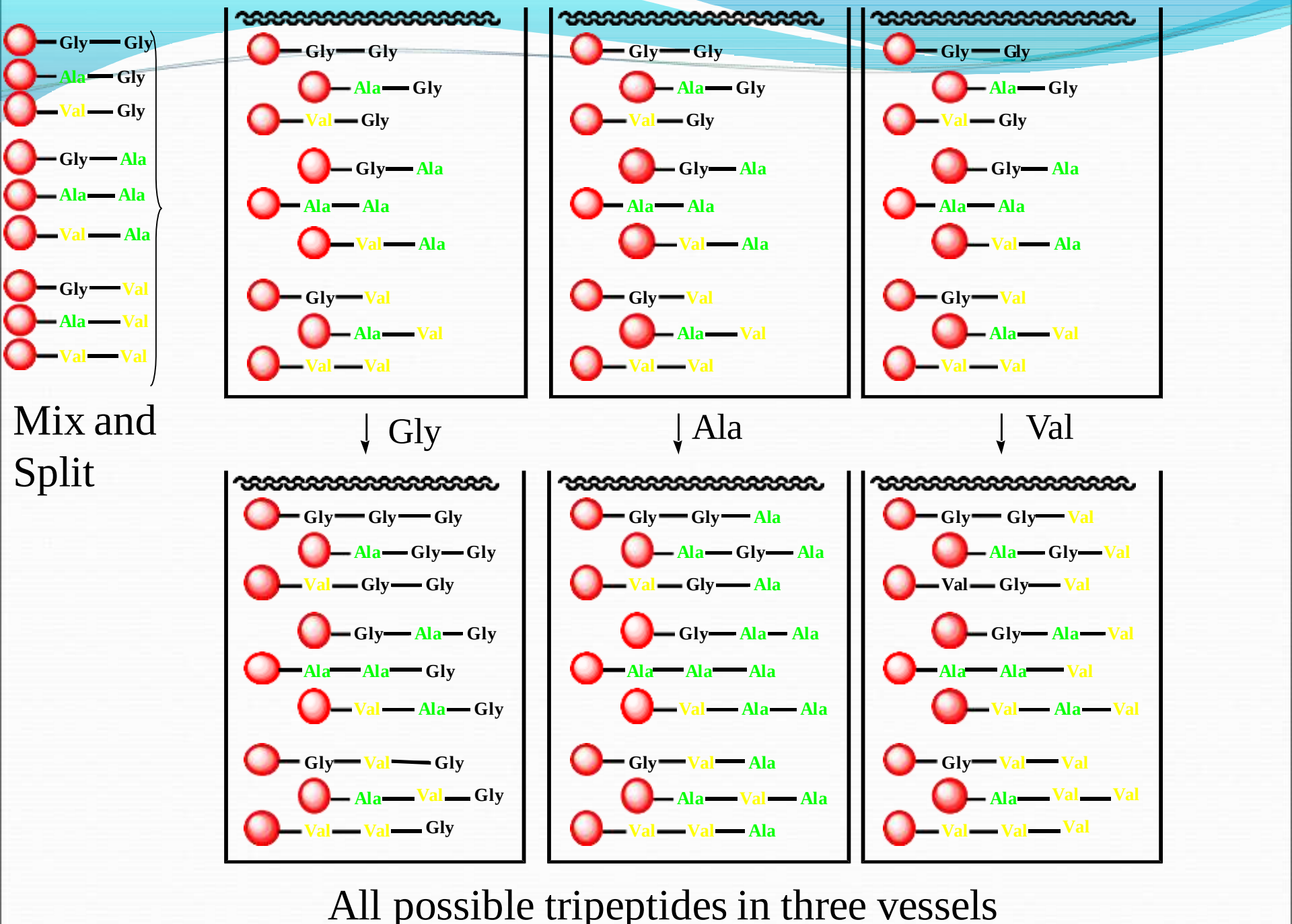


Mix and Split



All possible di peptides in three vessels

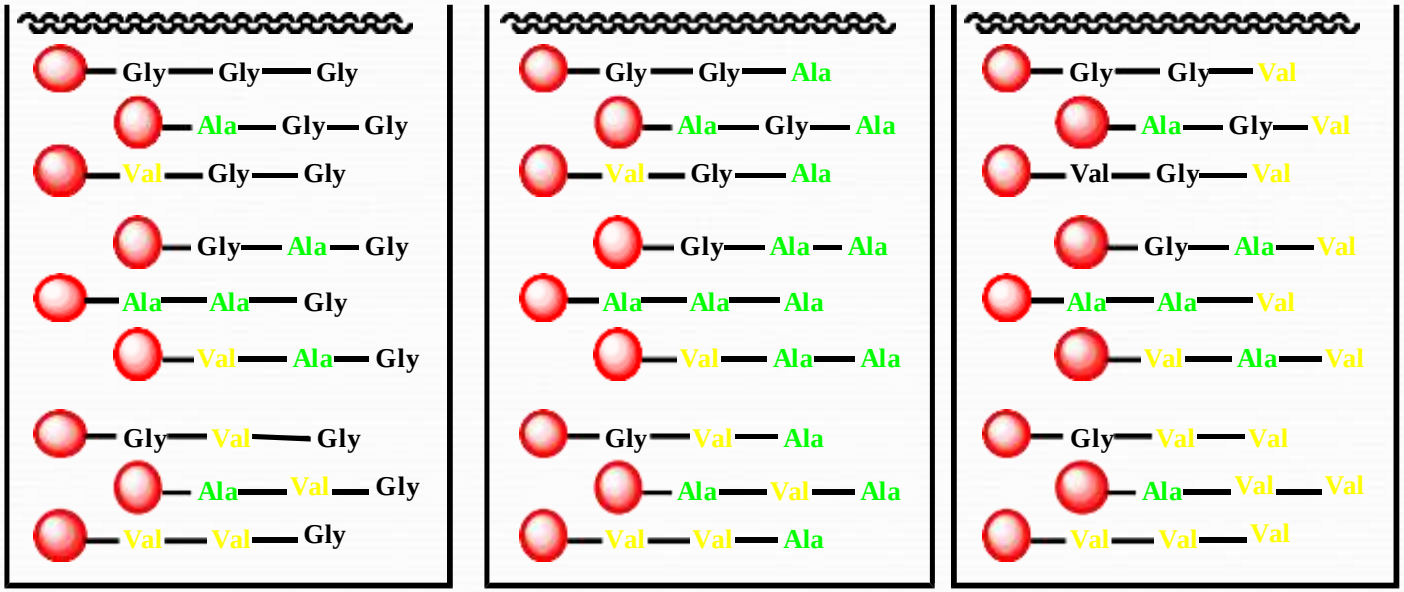
Retain a sample from each vessel



All possible tripeptides in three vessels

5. Identification of structures from mixed combinatorial synthesis

Recursive Deconvolution



Mixture
Inactive

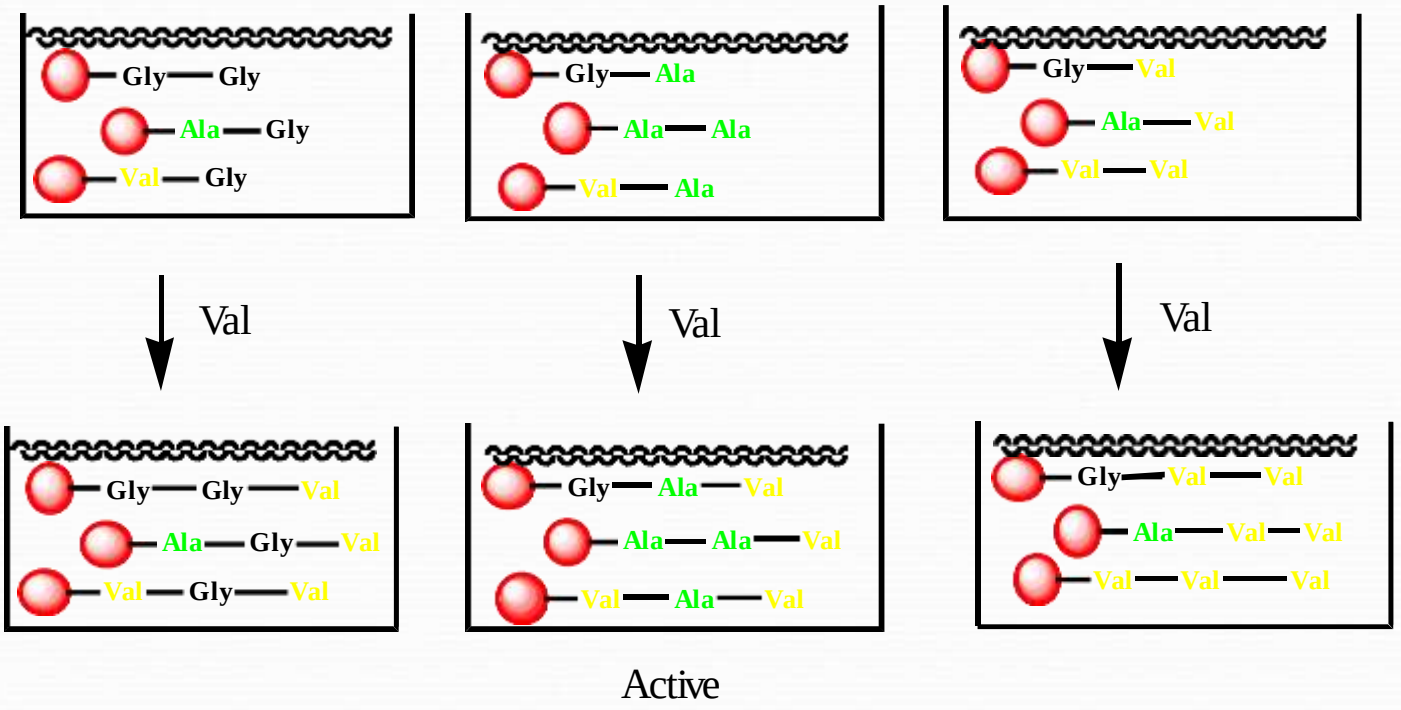
Mixture
Inactive

Mixture
Active

- 9 Possible tripeptides in active mixture
- All end in valine
- Add valine to the three retained dipeptide mixtures

5. Identification of structures from mixed combinatorial synthesis

Recursive Deconvolution

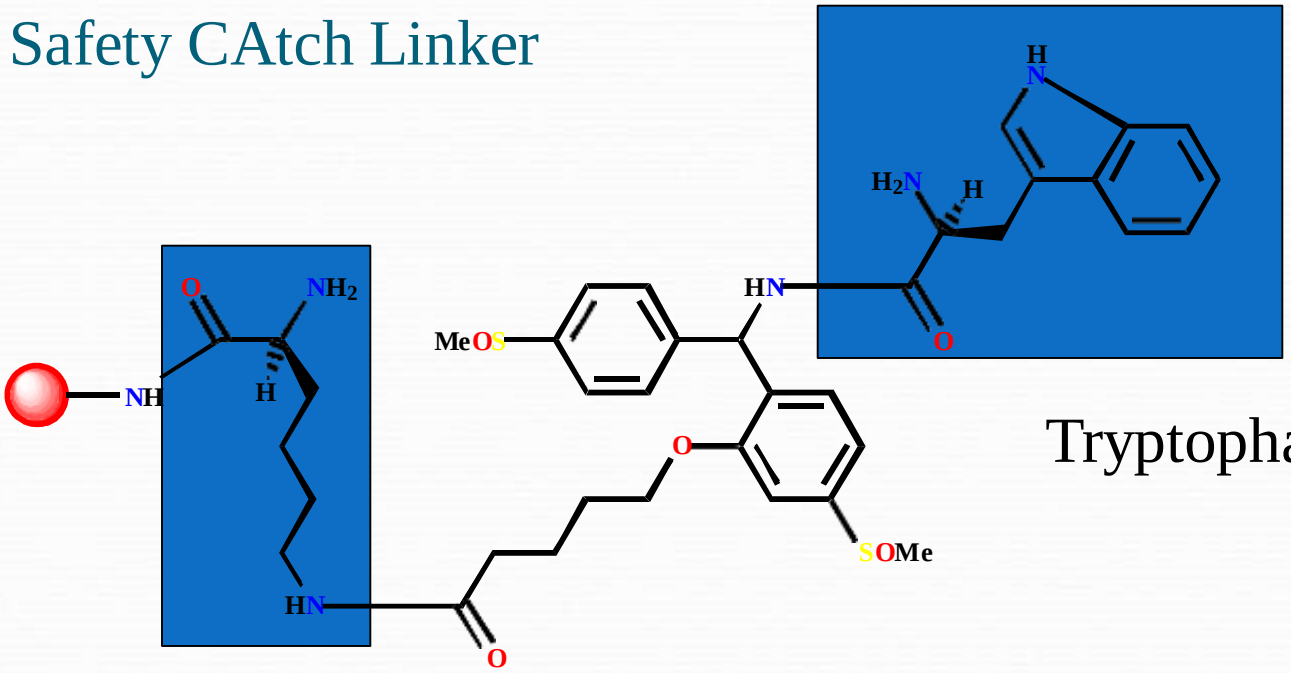


- Active component narrowed down to one of three possible tripeptides
- Synthesise each tripeptide and test

5. Identification of structures from mixed combinatorial synthesis

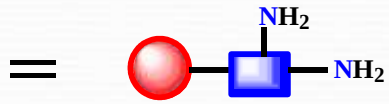
5.2 Tagging:

SCAL = Safety CAtch Linker



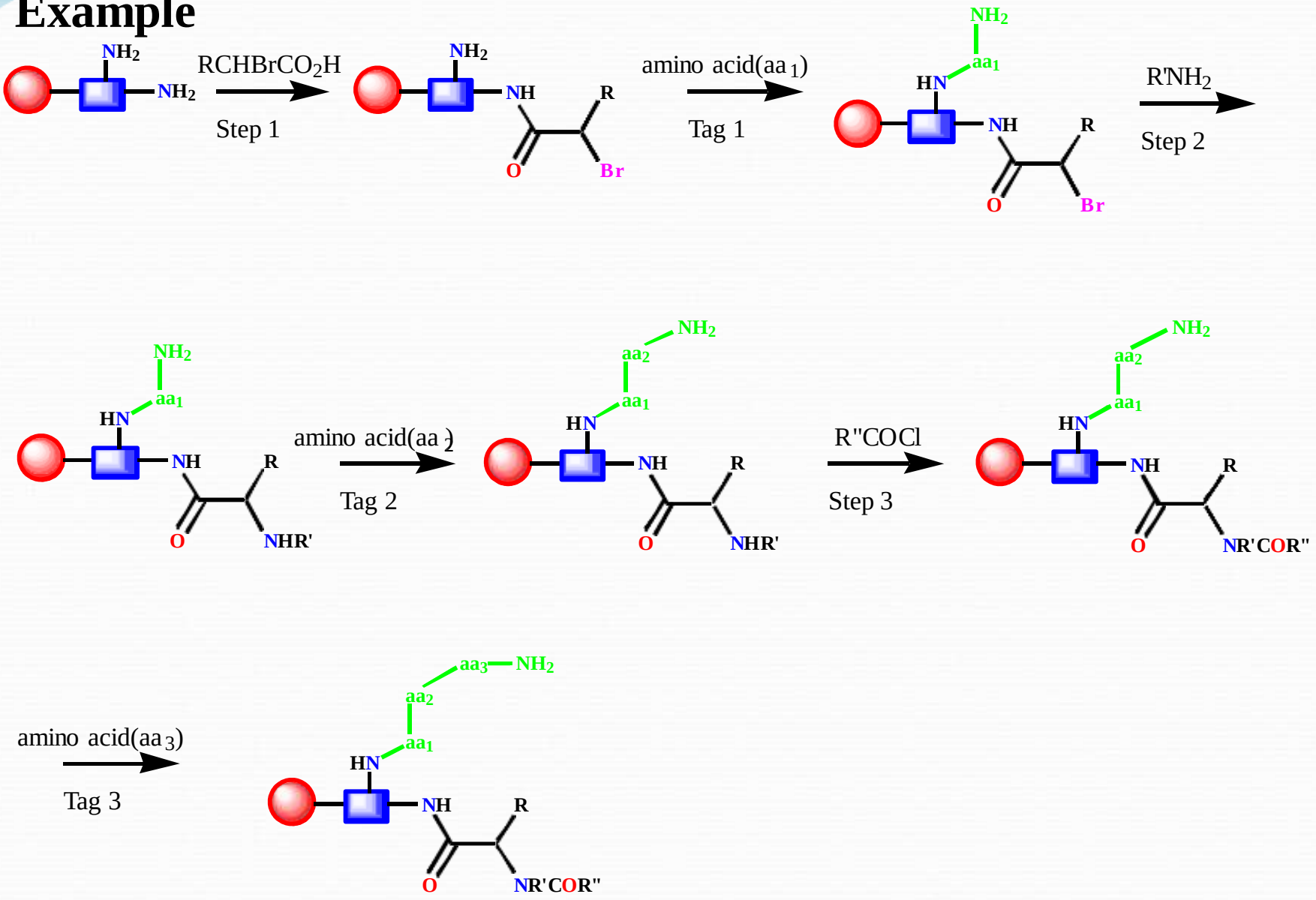
Tryptophan

Lysine



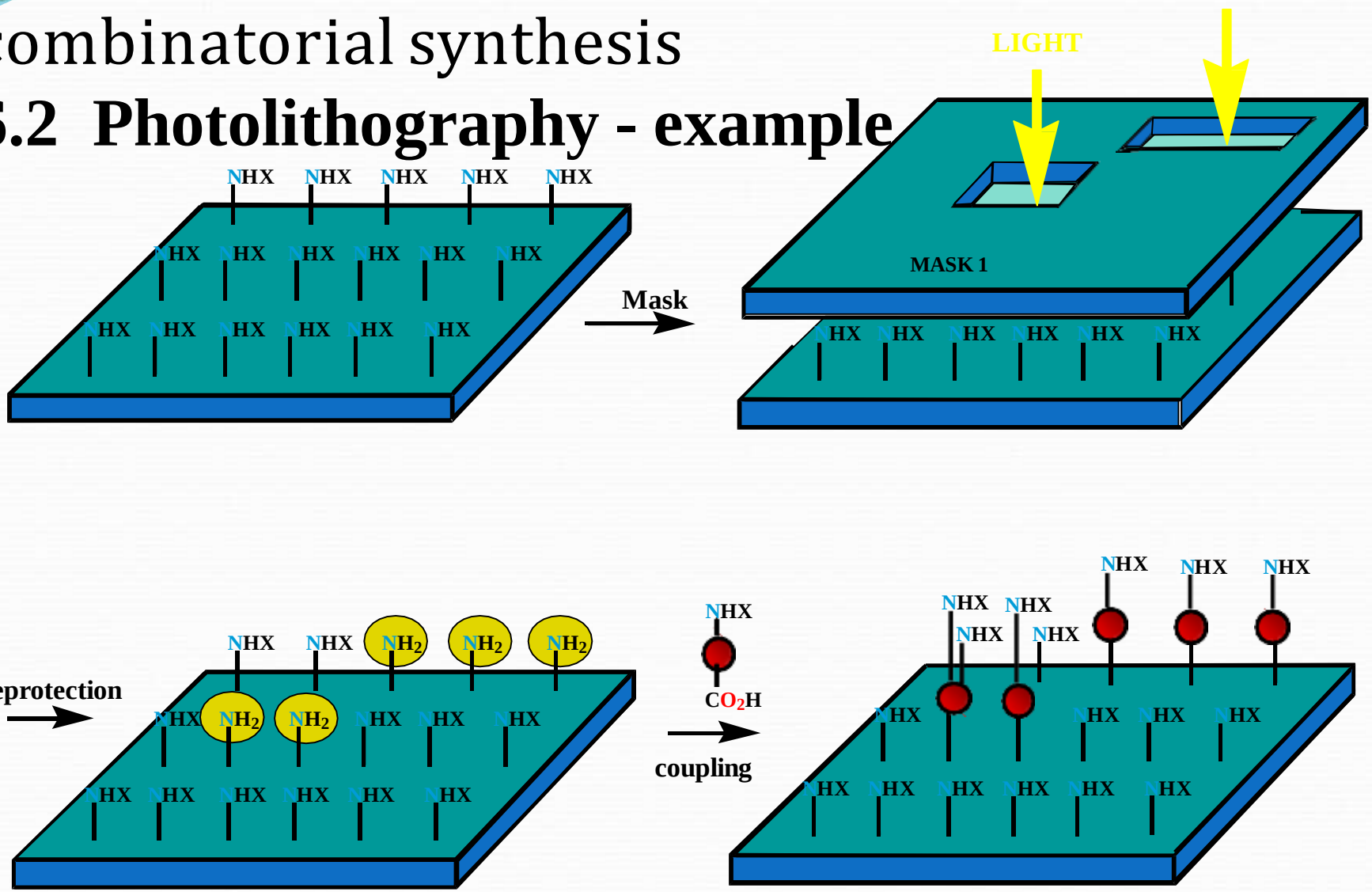
5.2 Tagging

Example



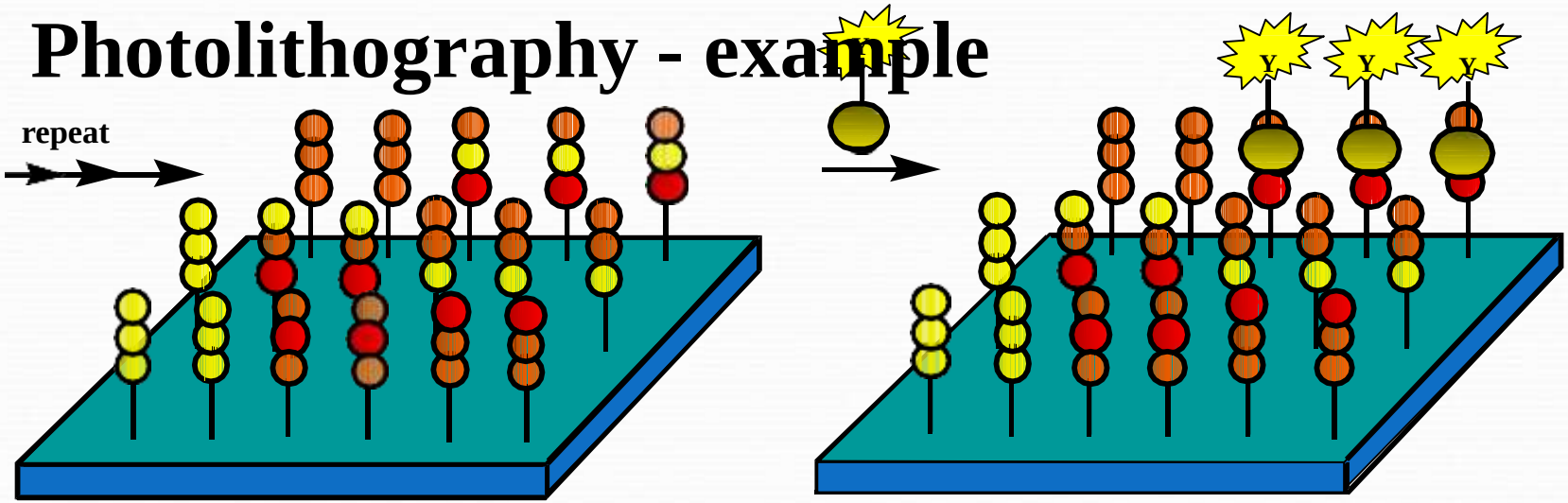
6. Identification of structures from combinatorial synthesis

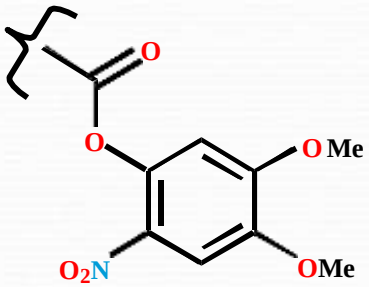
6.2 Photolithography - example



6. Identification of structures from combinatorial synthesis

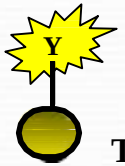
6.2 Photolithography - example





amino acids

X= Nitroveratryloxycarbonyl



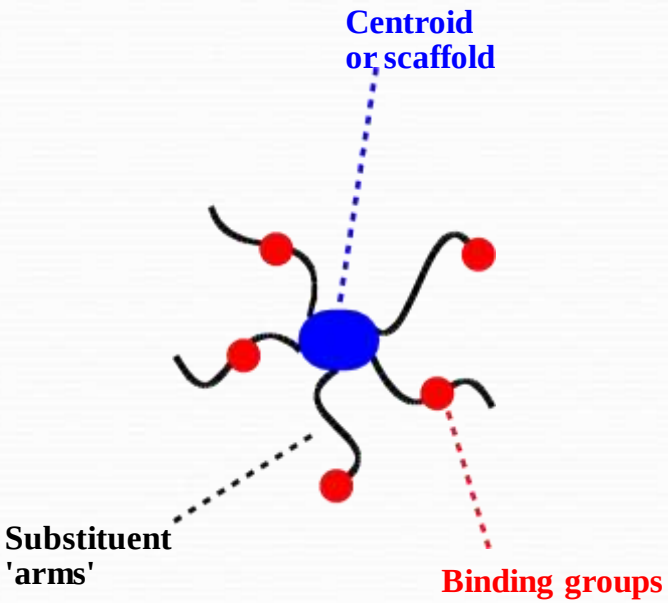
fluorescent tag

Target receptor

7. Planning a Combinatorial Synthesis

7.1 Aims

- To generate a large number of compounds
- To generate a diverse range of compounds
- Increase chances of finding a lead compound to fit a binding site
- Synthesis based on producing a molecular core or scaffold with functionality attached



7. Planning a Combinatorial Syntheses

7.1 Aims

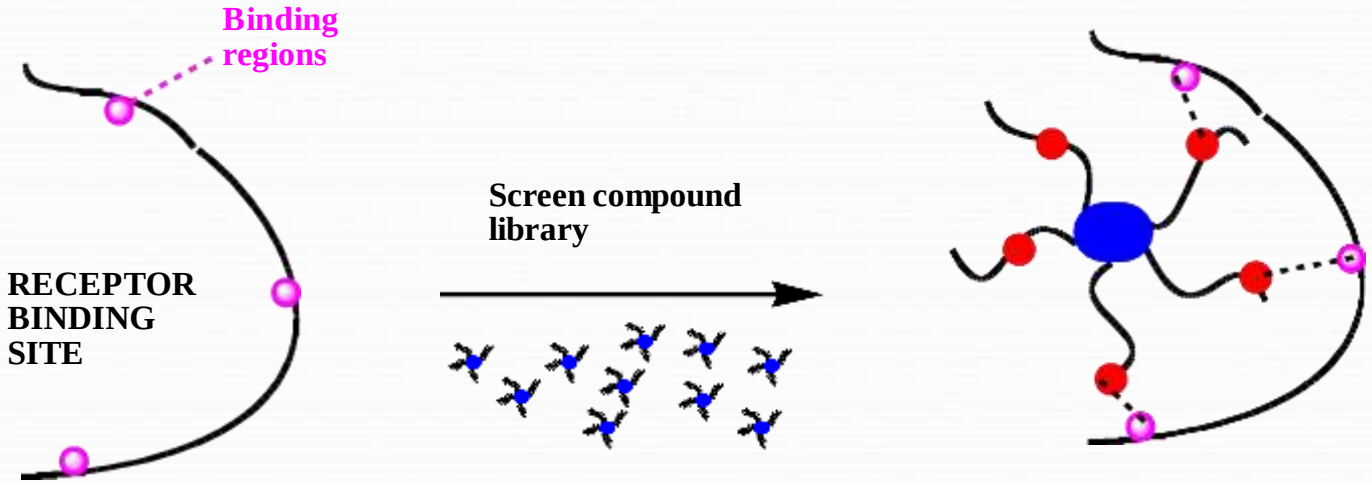
Target molecules should obey Lipinski's 'Rule of Five' for oral activity

- a molecular weight less than 500
- a calculated log *P* value less than +5
- no more than 5 H-bond donating groups
- no more than 10 H-bond accepting groups

7. Planning a Combinatorial Syntheses

7.2 Scaffolds

- 'Spider' scaffolds preferable for exploring conformational space
- Allows variation of functional groups around whole molecule to increase chances of finding suitable binding interactions



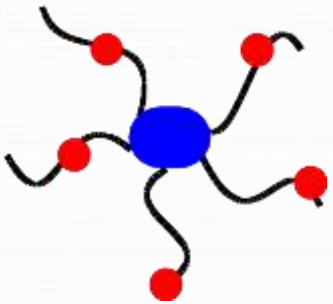
Molecular weight of scaffold should be low to allow variation of functionality, without getting products with a MWt > 500

7. Planning a Combinatorial Syntheses

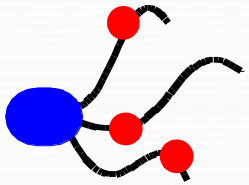
2. Scaffolds

Tadpole scaffolds

- variation restricted to a specific region round the molecule
- less chance of favourable interactions with a binding site



'Spider' Scaffold with 'dispersed' substituents

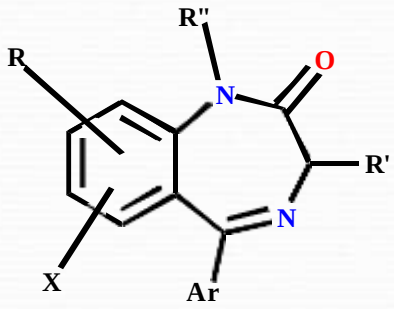


'Tadpole' scaffold with 'restricted' substituents

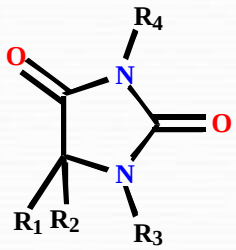
Privileged scaffolds

- scaffolds which are common in medicinal chemistry and which are associated with a diverse range of activities
- benzodiazepines, hydantoins, benzenesulphonamide etc

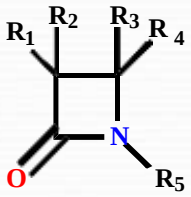
7. Planning a Combinatorial Syntheses
7.2 Scaffolds - examples



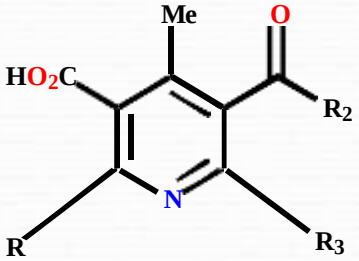
Benzodiazepines



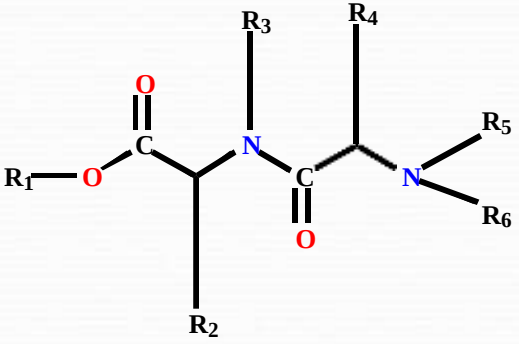
Hydantoins



β-Lactams



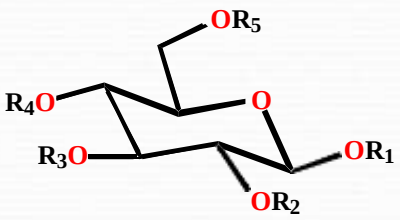
Pyridines



Dipeptides

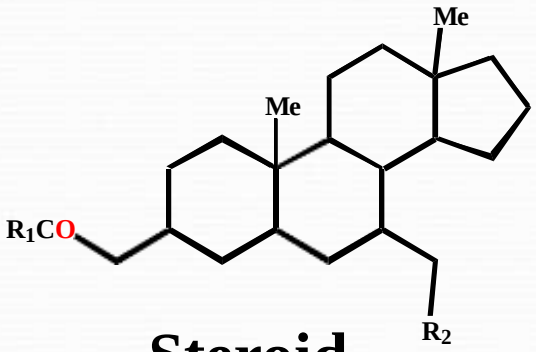
- Good scaffolds
- Spider like
- Low molecular weight
- Variety of synthetic routes available

7. Planning a Combinatorial Syntheses
7.2 Scaffolds - poor examples



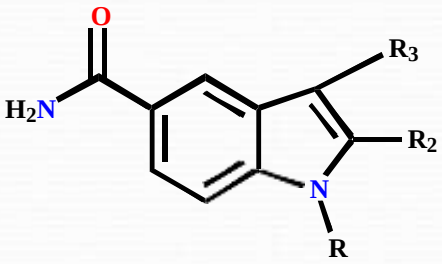
Glucose

Spider like and small molecular weight - good points
But multiple OH groups
Difficult to vary R_1-R_5 independently



Steroid

M.Wt. relatively high
Restricts no. of functional groups to keep MWt.< 500
Relatively few positions where substituents easily added



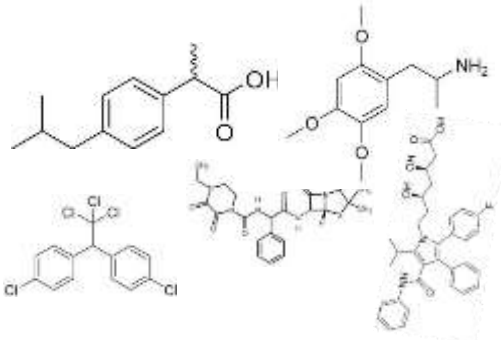
Indole

Tadpole like scaffold
Restricted region of variability

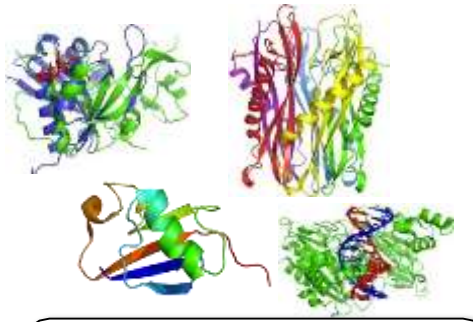
Thank You and Questions

VIRTUAL SCREENING IN DRUG DISCOVERY

Challenge in Drug Discovery



Chemical Space
$> 10^{20}$



Biological Space
10^4 to 10^5

Computational Screening, Experimental Screening

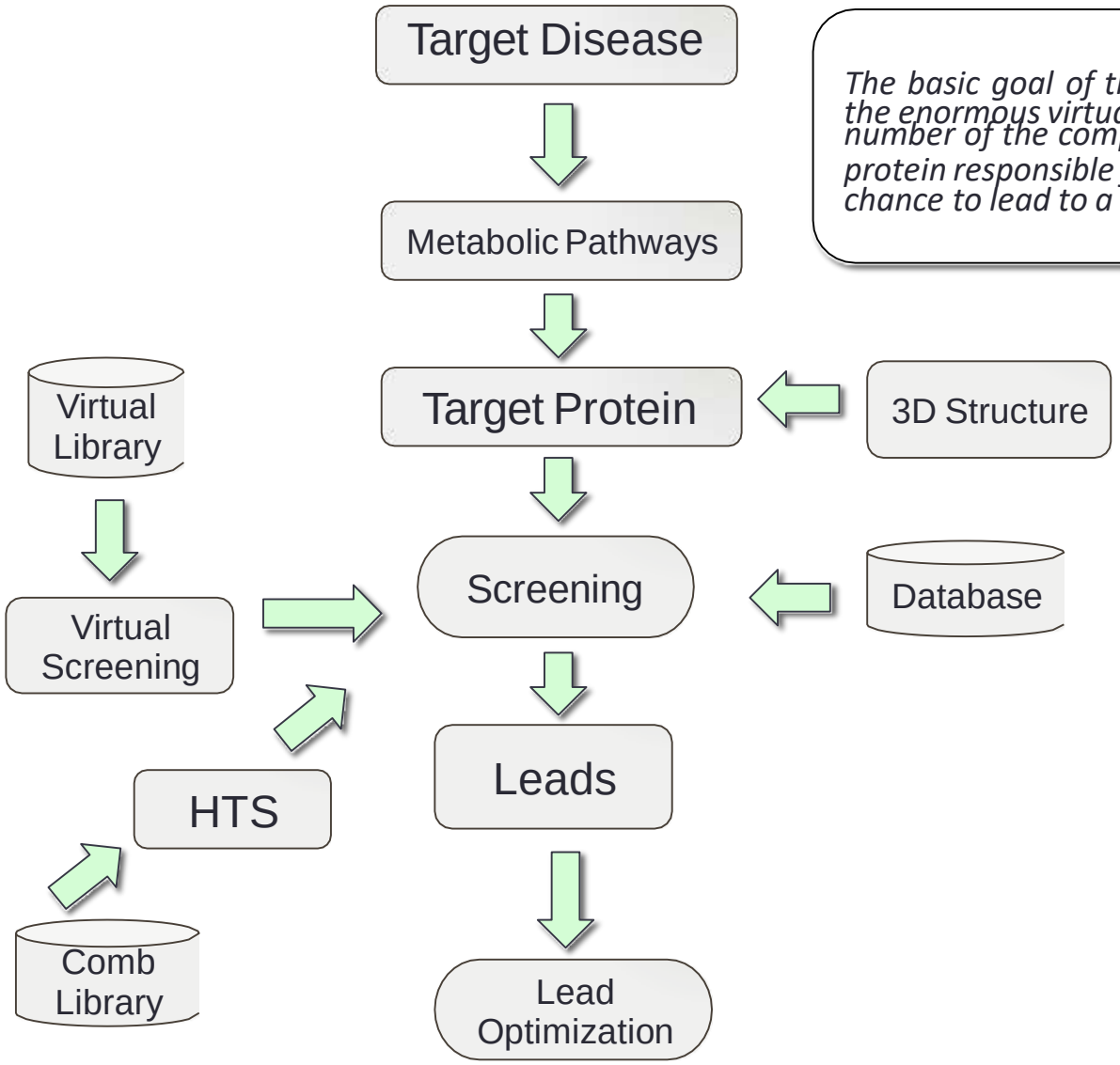
"Available" Compounds

Validated Targets

Choosing the right molecule

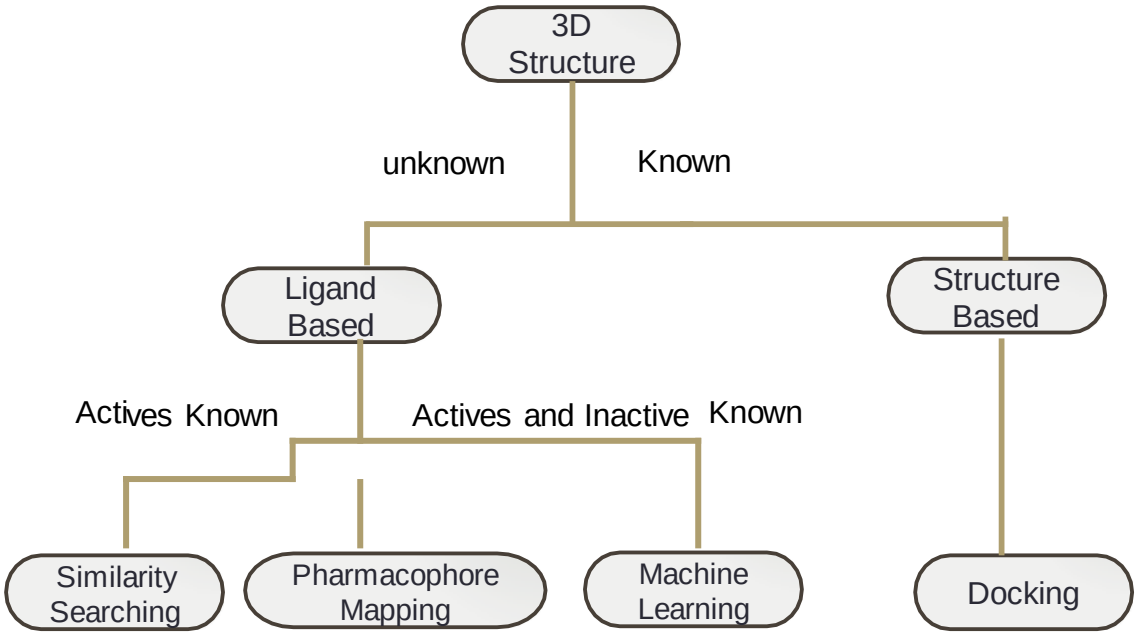
- Goal: to find a lead compound that can be optimized to give a drug candidate
  - Optimization: using chemical synthesis to modify the lead molecule in order to improve its chances of being a successful drug.
- The challenge: chemical space is vast
  - Estimates vary
    - Reymond et al. suggest there are ~1 billion compounds with up to 13 heavy atoms
    - There are ~65 million known compounds (example UniChem, PubChem)
    - A typical pharmaceutical compound collection contains ~1-5 million compounds
- High throughput screening allows large (up to 1 million) numbers of compounds to be tested
  - But very small proportion of “available” compounds
  - Large scale screening is expensive
  - Not all targets are suitable for HTS

Screening Schema in Drug Discovery



The basic goal of the virtual screening is the reduc4on of the enormous virtual chemical space, to a manageable number of the compounds that would inhibit a target protein responsible for disease and also have highest chance to lead to a drug candidate.

Virtual Screening

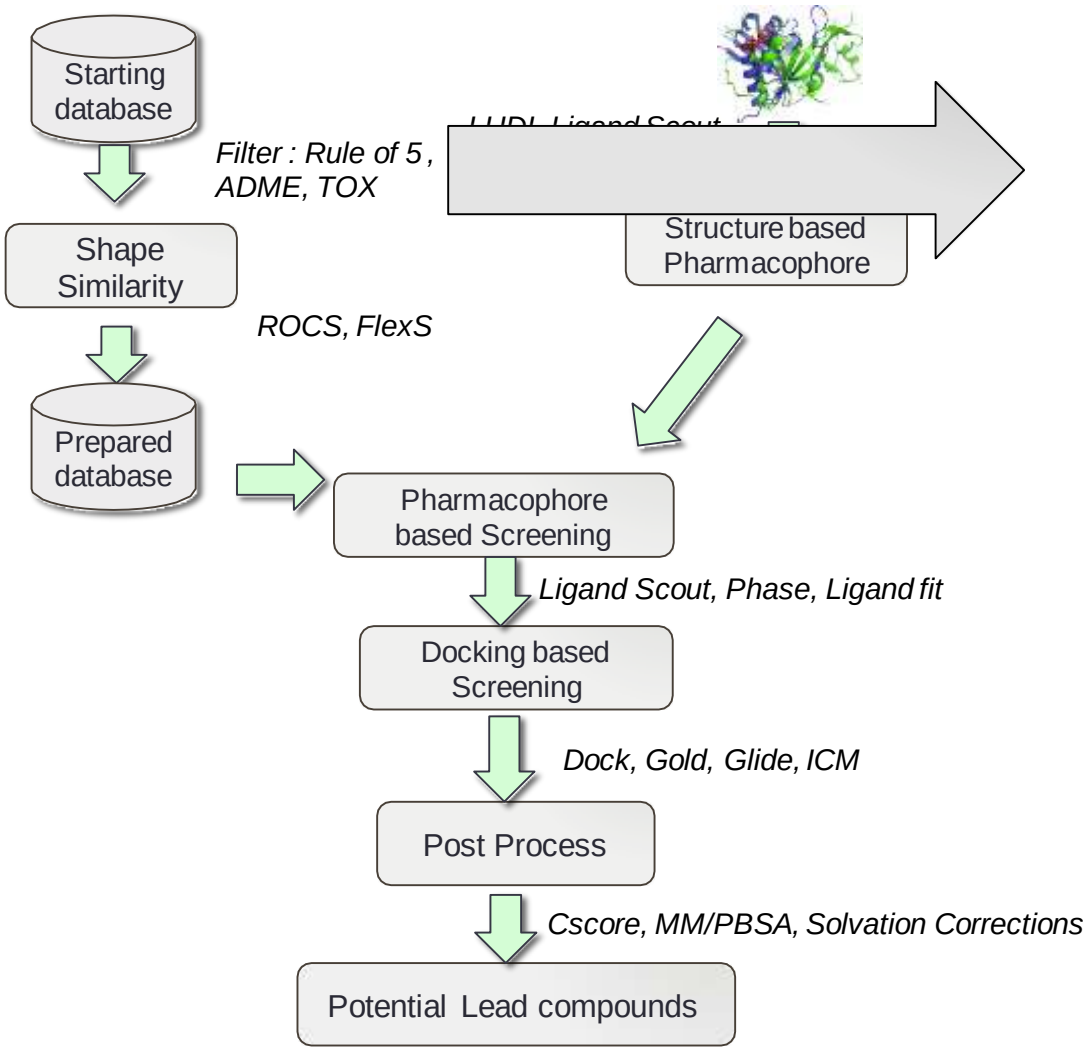


Depending upon structural and Bioactvity data available :

- *One or more actives molecule known perform similarity searching.*
- *Several active known try to identify a common 3D pharmacophore and then do 3D database search.*
- *Reasonable number of active and inactive known train a machine learning model.*
- *3D structure of protein known use protein ligand docking.*

Hybrid Virtual Screening

Mostly, people in pharmaceutical industry does not follow a specific route they follow a hybrid of methods as discussed in previous slide.



Cleaning Molecules

- Remove isotopes, salts and mixtures
- Protonation and normalization
- Remove duplicates and invalid structures

Filtering Molecules

- User defined or other filter
- Remove problematic moieties using PAINS, Frequent Hitters etc.
- PhyChem property descriptor calculation and filtration
- Apply protonation at pH 7.4

Drug Like Properties

Drug-like properties are an integral element of drug discovery projects. Properties of interest to discovery scientists include the following:

- Structural properties

  Hydrogen Bonding, Polar Surface area , Lipophilicity, Shape , Molecular Weight, Reactivity, pk_a

- Physicochemical Properties

  Solubility, Permeability, Chemical Stability

- Biochemical Properties

  Metabolism(Phase 1 and 2) , Protein and tissue binding, transport

- Pharmacokinetics(PK) and toxicity

  Clearance, Half-life, Bioavailability, Drug-Drug Interaction,LD_{50}

Leadlike & Druglike

- **Leadlike**
  - Molecular weight (MW) = 200–350 (optimization might add 100–200)
  - clogP <1.0–3.0 (optimization might increase by 1–2 log units)
  - Single charge present (secondary or tertiary amine preferred)
  -Importantly, exclude chemically reactive functional groups ,'promiscuous inhibitors', 'frequent hitters' and warheads
  - Non-substrate peptides are suitable.

- **Druglike**
  -Importantly, exclude chemically reactive functional groups ,'promiscuous inhibitors', 'frequent hitters' and warheads
  - MW < 500
  - cloP < 5
  - H-bond donors < 5
  - Sum of N and O (H-bond acceptors) < 10
  - Polar surface area < 140 A^2
  - Number of rotatable bonds <= 10

Filtering molecules using structural properties

Basic Washing –

- Removing Salts & Unwanted Elements

  Filter out cationic atoms: Ca2+, Na+, etc.

  Filter out metals:

  Sc,Ti,V,Cr,Mn,Fe,Co,Ni,Cu,Zn,Y,Zr,Nb,Mo,Tc,Ru,Rh,Pd,Ag,Cd

  Often the salt “filter” = keeping the largest molecule in the sdf entry.
- ALLOWED_ELEMENTS H, C, N, O, F, P, S, Cl, Br, I
- Check proper Atom Types by adding hydrogen and checks if O, N, C valences are correct.
- Check formal charge

Filter out Reactives (false positives for proteins)

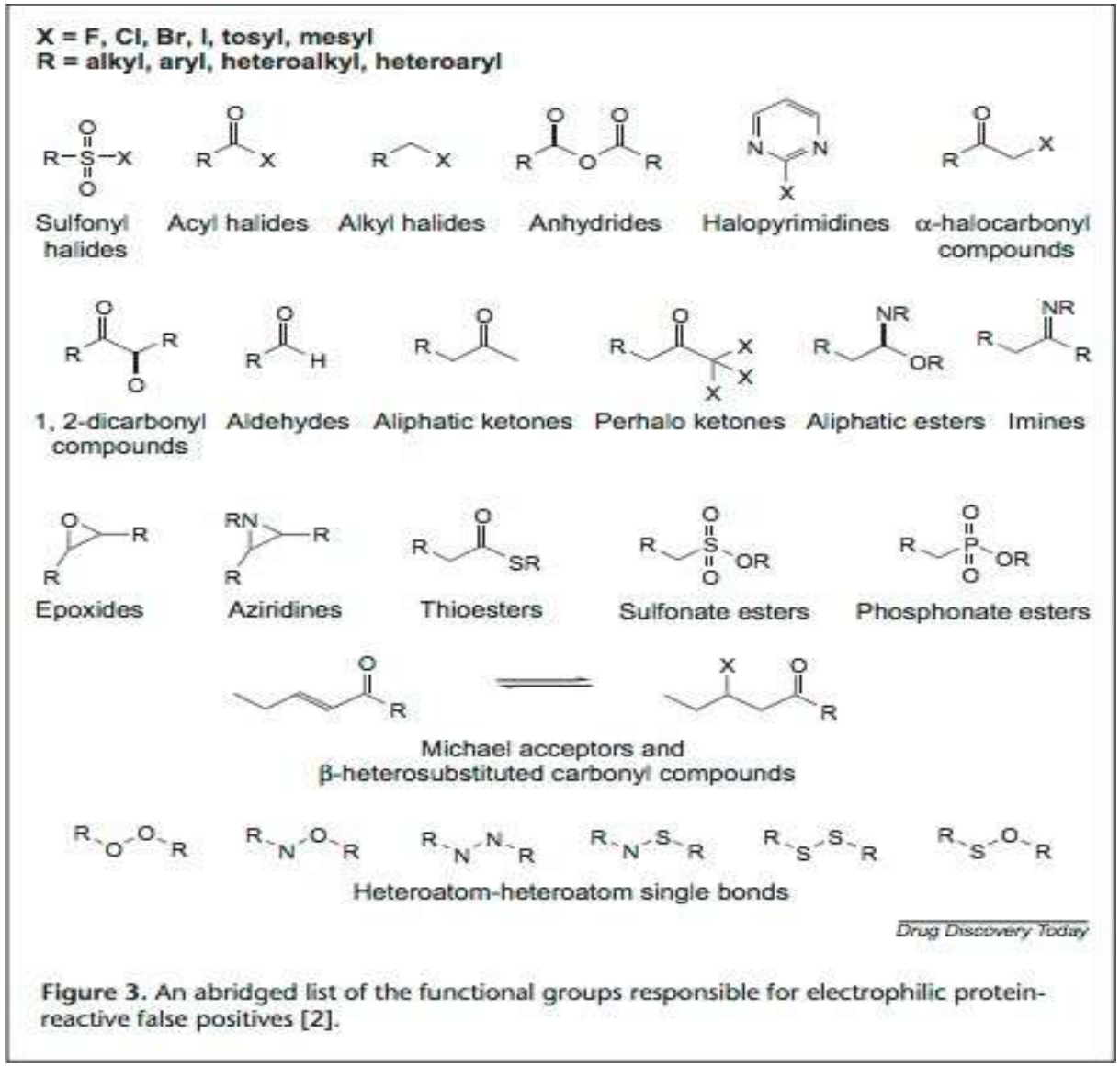


Figure 3. An abridged list of the functional groups responsible for electrophilic protein-reactive false positives [2].

Filter out: Synthesis Intermediates, Chelators

'warhead' agents - functional groups which shows high reactivity to proteins due which there is high attrition rate in drug development.

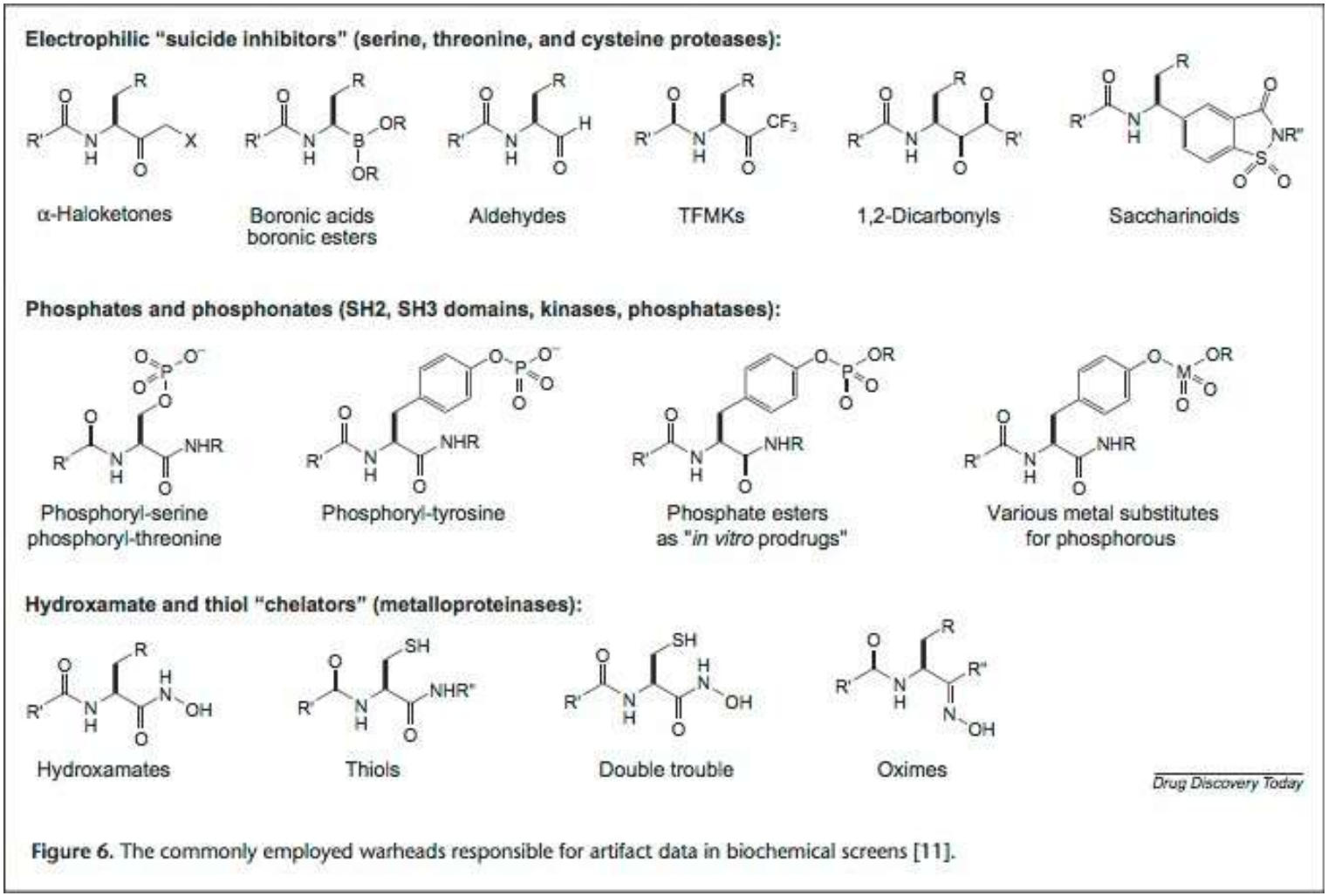
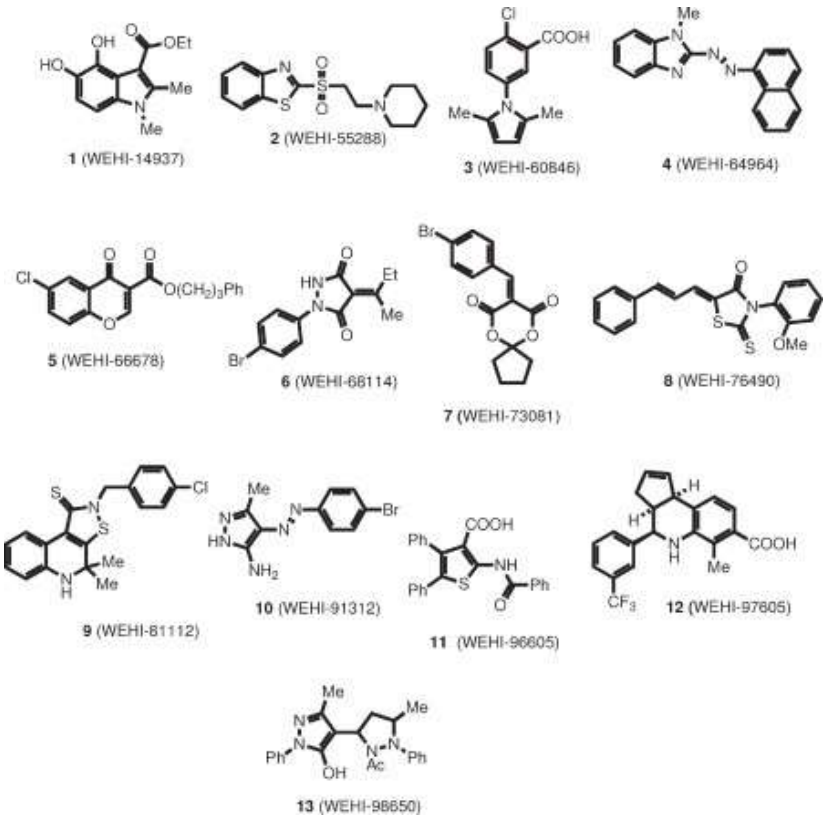


Figure 6. The commonly employed warheads responsible for artifact data in biochemical screens [11].

PAINS Filter

- PAINS = “Pan-Assay Interference Compounds”
- Problematic scaffolds – has cost their Institute time and $$



Rules-of-Thumb for Hit Selection & Lead Optimization

parameter	rules-of-thumb	comment	programs	key references
oral bioavailability ("rule of 5")	MW ≤ 500 Da ClogP ≤ 5 H-bond donors ≤ 5 #(N + O) ≤ 10	violation of these limits decreases oral bioavailability	Biobyte ClogP[85,86] or ACD LogP v4.0[12]	Lipinski (1997)[1] Wenlock (2003)[12]
oral bioavailability	Nrot ≤10 PSA ≤ 140 Å^2	violation of these limits decreases oral bioavailability	tPSA[62] (nitrogen and oxygen only)	Veber (2002)[13]
oral bioavailability ("Golden Triangle")	MW ≤ 500 variable LogD (LogD range: 0 − 5)	violation of these limits decreases oral bioavailability	experimental LogD	Johnson (2009)[35]
toxicity	ClogP ≤ 3 PSA ≥ 75 Å^2	violation of these limits increases the risk of toxicity	Biobyte ClogP v4.3[85] tPSA[62] (nitrogen and oxygen only)	Hughes (2008)[2]
toxicity	LLE ≥ 5	low ligand-lipophilicity efficiency can lead to increased promiscuity	Biobyte ClogP[85]	Leeson (2007)[19] Leach (2006)[23]
membrane permeability	PSA ≤ 120 Å^2	violation of this limit decreases membrane permeability	Quanta 3D (nitrogen and oxygen only)	Kelder (1999)[61]
membrane permeability	MW ≤ 500 variable LogD (LogD range: 0.5 − 5)	violation of these limits decreases membrane permeability	ACD PhysChem Batch[87] or AZlogD[88]	Bhal (2007)[34] Waring (2009)[36]
blood−brain barrier penetration	PSA ≤ 70 Å^2	violation of this limit decreases brain penetration	Quanta 3D (nitrogen and oxygen only)	Kelder (1999)[61]
solubility	Fsp3 ≥ 0.4	increased fraction of sp3 hybridized carbons (Fsp3) increases solubility	Pipeline Pilot 7.5	Lovering (2009)[51]
general "developability"	number of aromatic rings ≤ 3	increase in aromatic ring count decreases solubility and increases protein binding	none listed	Ritchie (2009)[52]

Muchmore, SW et al. "Cheminformatic Tools for Medicinal Chemists" J. Med. Chem. (2010) 53, 4830 – 4841

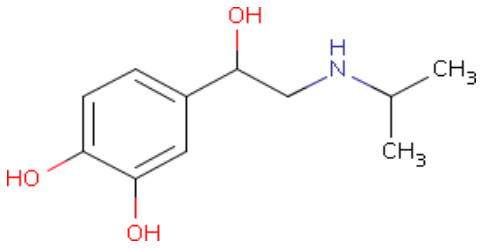
Similarity Searching

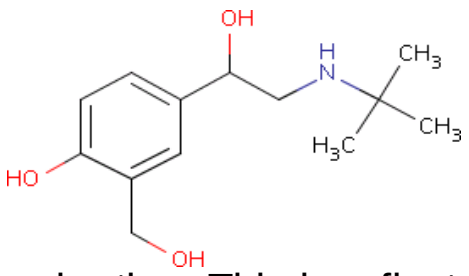
What is it ??

Chemical, pharmacological or biological properties of two compounds match.

The more the common features, the higher the similarity between two molecules.

Chemical





The two structures on top are chemically similar to each other. This is reflected in their common sub-graph, or scaffold: they share 14 atoms

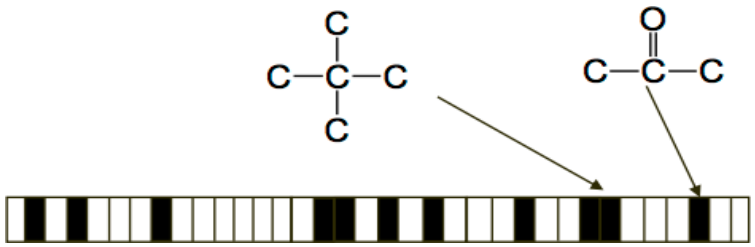
Pharmacophore

The two structures above are less similar chemically (topologically) yet have the same pharmacological activity, namely they both are Angiotensin-Converting Enzyme (ACE) inhibitors

What is required for a similarity search ?

- A Database SQL or NoSQL (Postgres, MySQL, MongoDB) or flat file of descriptors eg: ChemFP
- Chemical Cartridge to generate fingerprints(descriptors) for molecules (RDKit, openbabel)
- Similarity function to calculate similarity(Jaccard, Dice, Tversky) this can be written in c,c++ or python as a function inside SQL databases.

2D fingerprints: molecules represented as binary vectors



- Each bit in the bit string (binary vector) represents one molecular fragment. Typical length is ~1000 bits
- The bit string for a molecule records the presence (“1”) or absence (“0”) of each fragment in the molecule
- Originally developed for speeding up substructure search

− for a query substructure to be present in a database molecule each bit set to “1” in the query must also be set to “1” in the database structure

- Similarity is based on determining the number of bits that are common to two structures

Calculate molecular similarity

Sequences/vectors of bits, or numeric values that can be compared by distance functions, similarity metrics .

E= Euclidean distance

T = Tanimoto index /Jaccard

$$E(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

$$T(x, y) = \frac{B(x \,\&\, y)}{B(x) + B(y) - B(x \,\&\, y)}$$

3D based similarity

- Shape-based
  ROCS (Rapid Overlay of Chemical Structures)
  Silicos-it.com (Shape it)

- Computationally more expensive than 2D methods

- Requires consideration of conformational flexibility

  - Rigid search - based on a single conformer

  - Flexible search

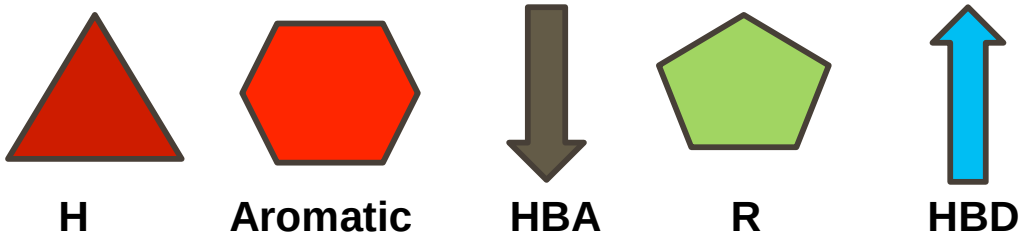
    - Conformation explored at search time

    - Ensemble of conformers generated prior to search time with each conformer of each molecule considered in turn

    - How many conformers are required

Multiple actives known: pharmacophore searching

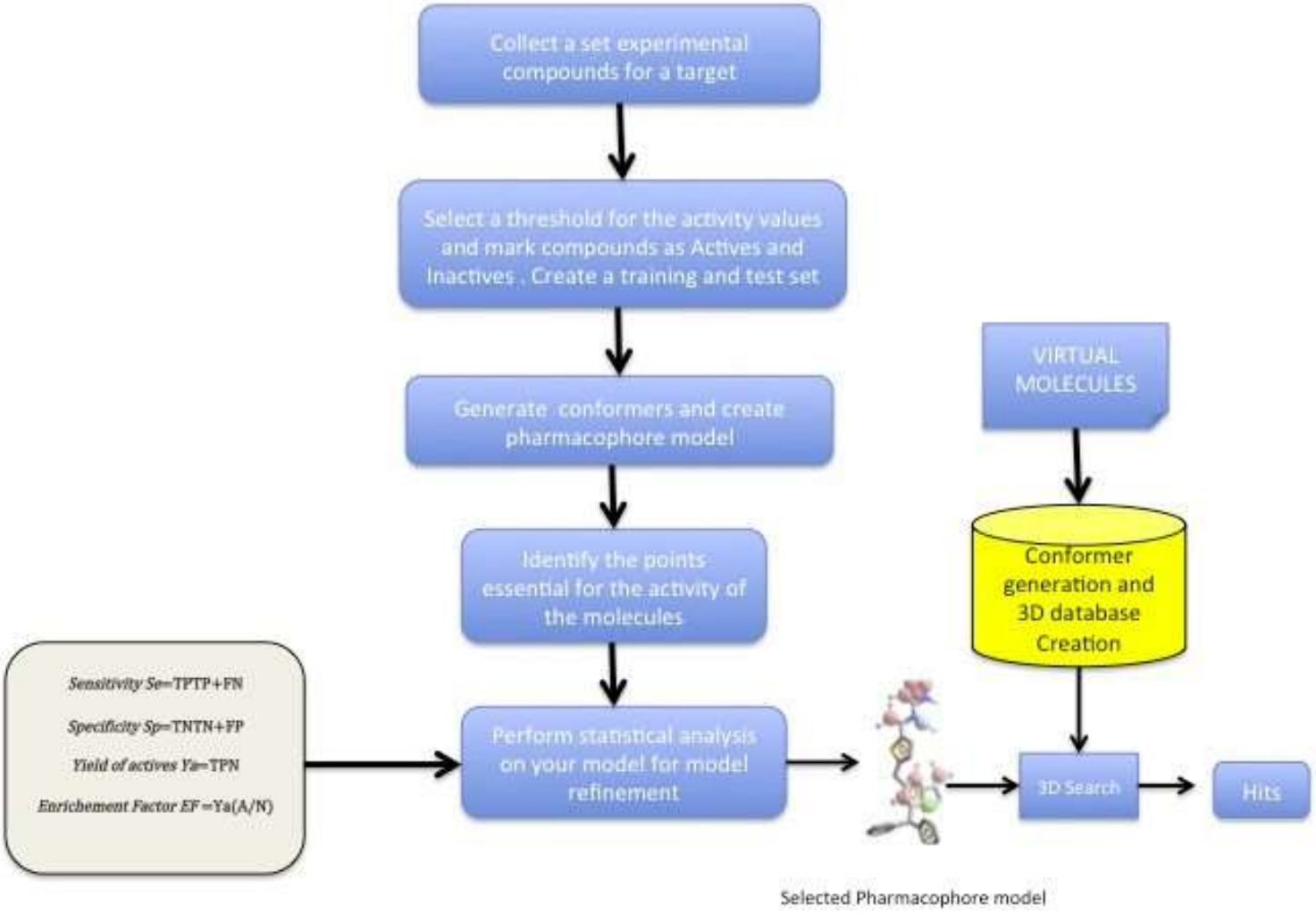
- **IUPAC Definition: "**An ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target and to trigger (or block) its biological response"



- In drug design, the term 'pharmacophore' refers to a set of features that is common to a series of active molecules
- Hydrogen-bond donors and acceptors, positively and negatively charged groups, and hydrophobic regions are typical features

We will refer to such features as 'pharmacophoric groups'

Workflow of pharmacophore modeling



Tools to perform pharmacophore searching

1) Catalyst (Accelrys)
2) Phase (Schrodinger)
3) LigandScout (Inte:Ligand)
4) PharmaGist
5) Pharmer
6) SHAFTS

Many Actives and inactives known : Machine learning methods

- **SAR Modeling**
- Use knowledge of known active and known inactive compounds to build a predictive model

- Quantitative-Structure Activity Relationships (QSARs)

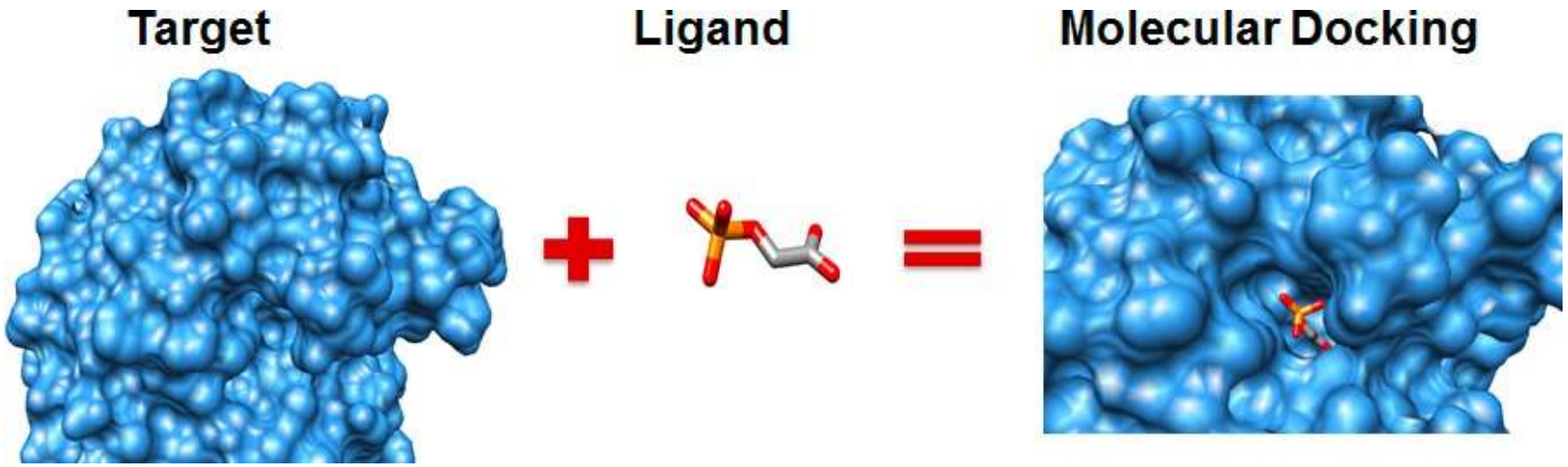
  - Long established (Hansch analysis, Free-Wilson analysis)

  - Generally restricted to small, homogeneous datasets eg lead optimization.

- Structure-Activity Relationships (SARs)

  - "Activity" data is usually treated qualitatively
  - Can be used with data consisting of diverse structural classes and multiple binding modes
  - Some resistance to noisy data (HTS data)
  - Resulting models used to prioritize compounds for lead finding (not to identify candidates or drugs)

Protein Ligand Docking



Computational method which mimics the binding of a ligand to a protein.

It predicts ..

a) the **pose** of the molecule in the binding site

b) The **binding affinity** or **score** representing the strength of binding

Pose and Binding Site

- **Binding Site (or “active site”)**

- the part of the protein where the ligand binds .
- generally a cavity on the protein surface.
- can be identified by looking at the crystal structure of the protein bund with a known inhibitor.

- **Pose (“binding mode”)**

- the geometry of the ligand in the binding site
- Geometry- location , orientation and conformation of the molecule

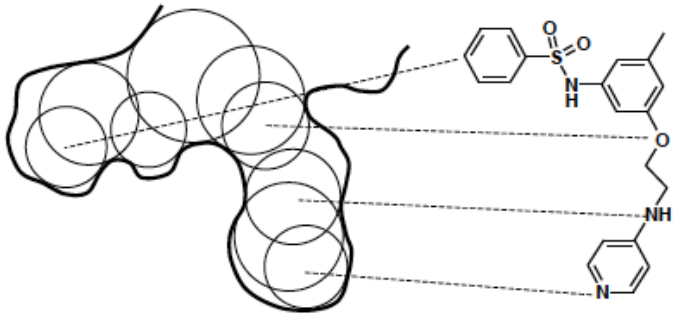
Protein Ligand Docking

- How does a ligand (small molecule) bind into the active site of a protein?

- Docking algorithms are based on two key components
  - search algorithm
    - to generate “poses” (conformation, position and orientation) of the ligand within the active site
  - scoring function
    - to identify the most likely pose for an individual ligand
    - to assign a priority order to a set of diverse ligands docked to the same protein – estimate binding affinity

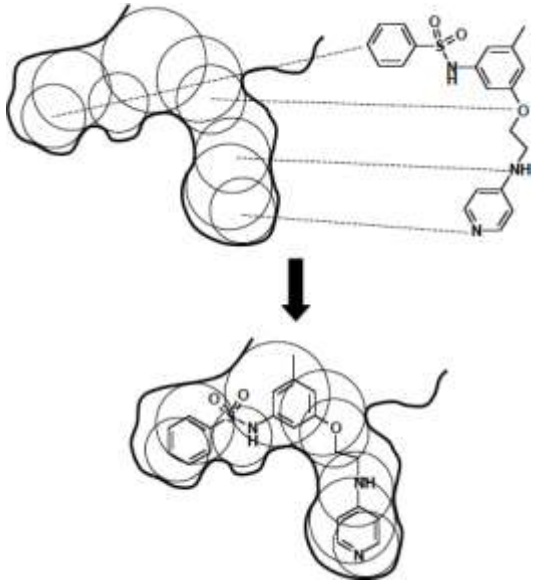
DOCK (Kuntz et al. 1982)

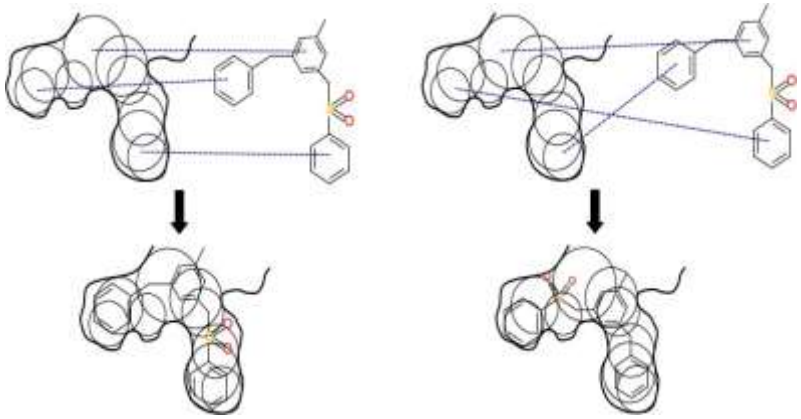
- Rigid docking based on shape
- A negative image of the cavity is constructed by filling it with spheres
- Spheres are of varying size
- Each touches the surface at two points
- The centres of the spheres become potential locations for ligand atoms.



DOCK

- Ligand atoms are matched to sphere centers so that distances between atoms equals distances between sphere centers.
- The matches are used to position the ligand within the active site.
- If there are no steric clashes the ligand is scored.
- Many different mappings (poses) are possible
- Each pose is scored based on goodness of fit
- Highest scoring pose is presented to the user





Energetics of protein-ligand binding

a) Ligand-receptor binding is driven by

- electrostatics (including hydrogen bonding interactions)
- dispersion or van der Waals forces
- hydrophobic interactions
- desolvation: surfaces buried between the protein and the ligand have to be desolvated
- Conformational changes to protein and ligand
- ligand must be properly orientated and translated to interact and form a complex
- loss of entropy of the ligand due to being fixed in one conformation.

b) Free energy of binding

$$\Delta G_{bind} = \Delta G_{solvent} + \Delta G_{conf} + \Delta G_{int} + \Delta G_{rot} + \Delta G_{t/r} + \Delta G_{vib}$$

Thank you