

# Ligand-Based Drug Design



SAPIENZA  
UNIVERSITÀ DI ROMA



# Summary

## 1. Overview on LBDD

A. Definition and Methods

B. Pharmacophore Approach

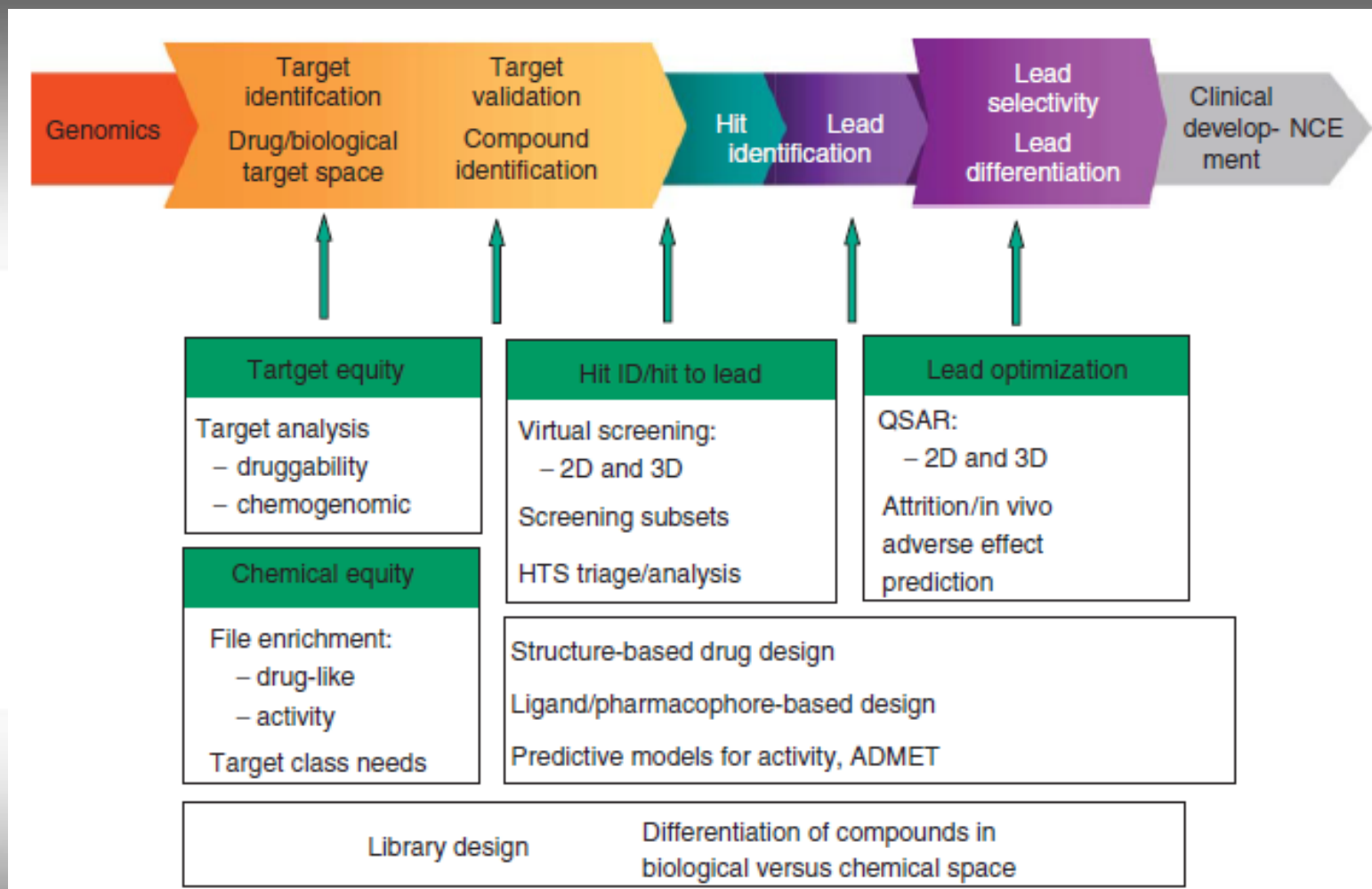
C. QSAR

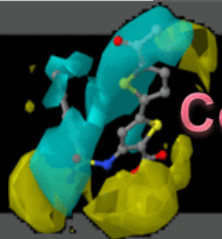
D. 3-D QSAR

## 2. Application (and demonstration) of 3-D QSAR to medicinal chemistry

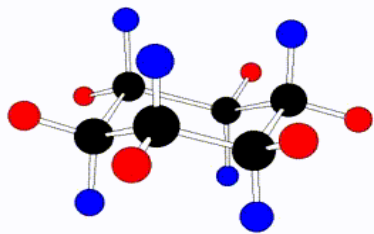


# Drug Discovery Process



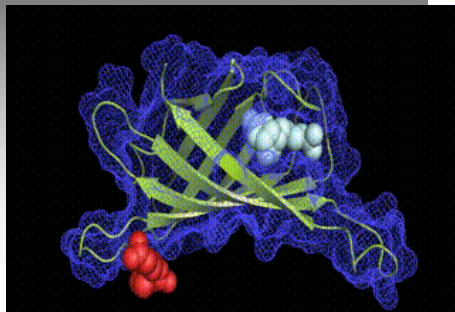


chair-chair interconversion

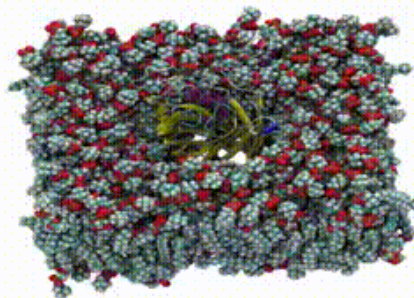


chair 1: blue H's axial

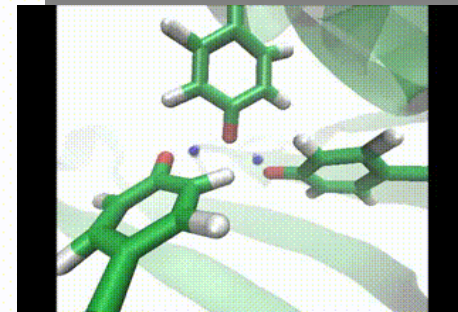
Conformational Search



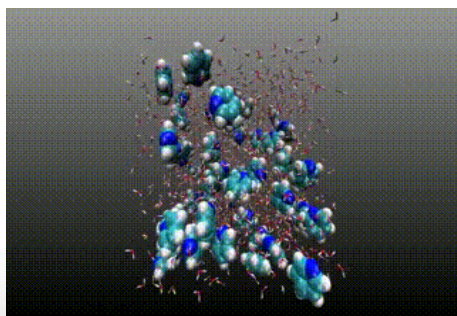
Molecular Docking



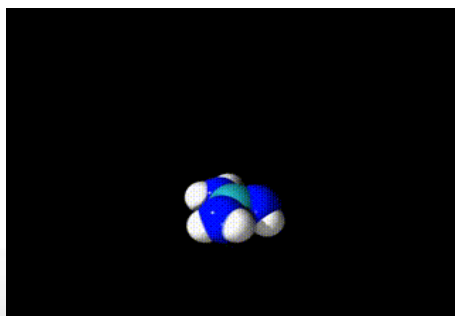
Simulate Annealing



Ab Initio QM



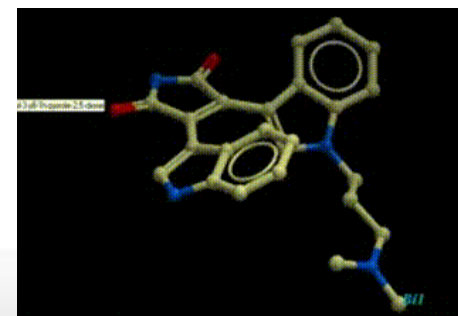
Molecular Dynamics



Graphical Visualization

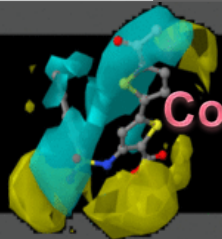


3-D QSAR

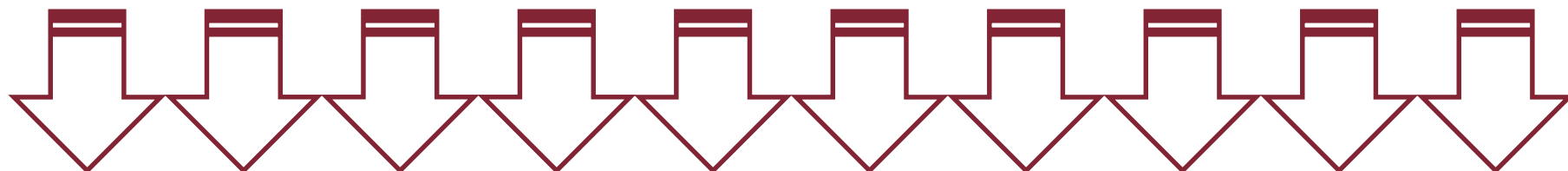
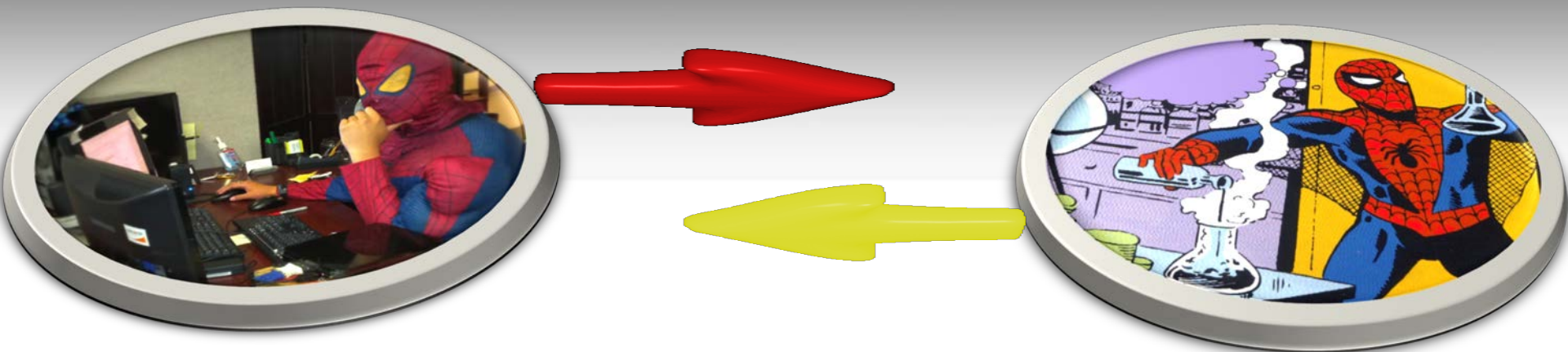


Pharmacophore

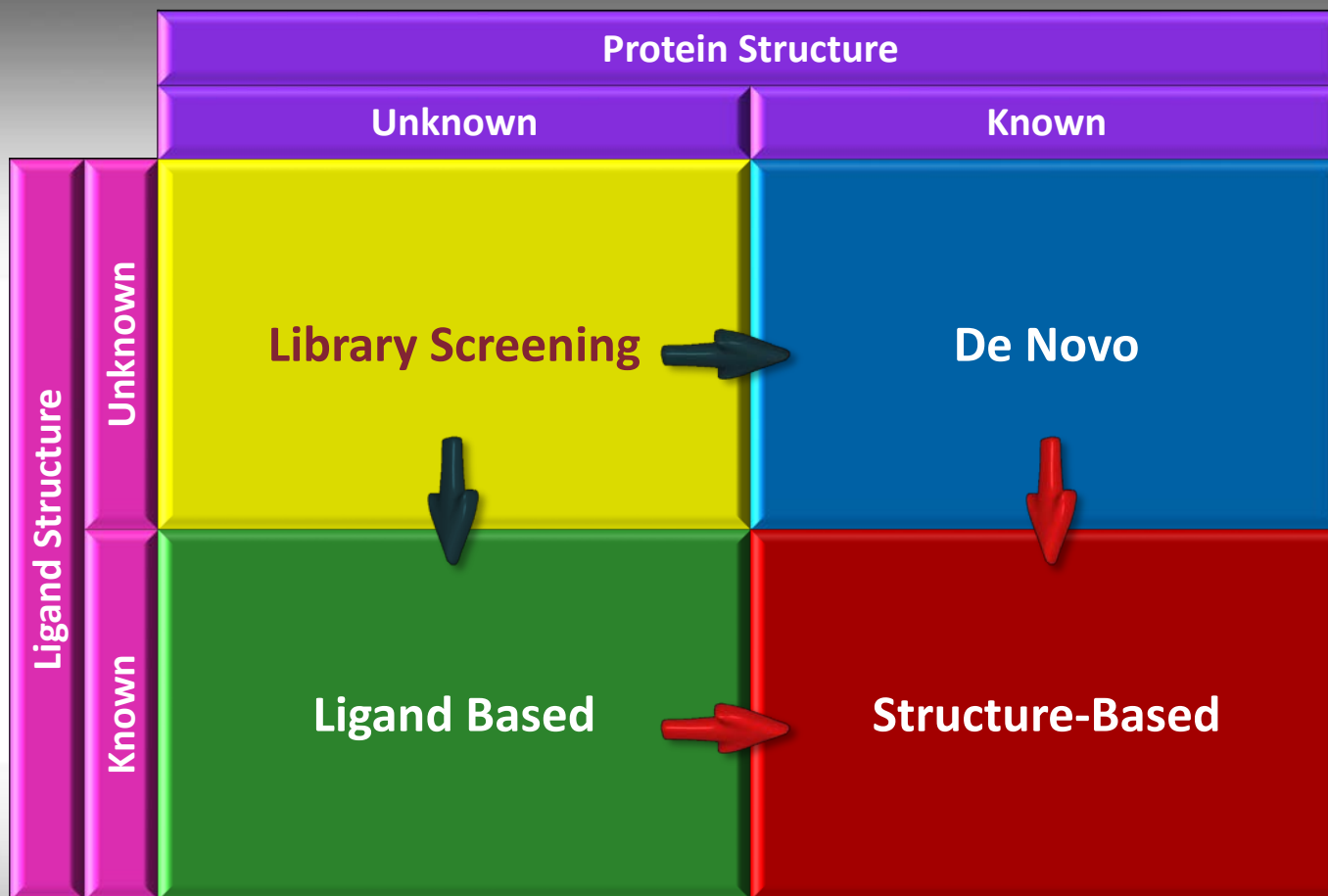
QSAR, COMBINE, Scoring Functions, Homology Modeling,.....

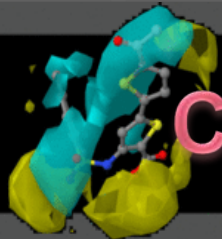


## Drug Design = Computational & Synthesis Tandem









# CADD methods in Drug Design

**Ligand-Based**

**Structure-Based**

**QSAR**

**Pharmacophore**

**3-D QSAR**

**Scoring  
Function  
Docking**

**COMBINE**





# LBDD

In the absence of three-dimensional (3-D) structures of potential drug targets, ligand-based drug design is one of the popular approaches for drug discovery and lead optimization. **3-D structure activity relationships (3-D QSAR) and pharmacophore modeling are the most important and widely used tools in ligand-based drug design** that can provide crucial insights into the nature of the interactions between drug target and ligand molecule and provide predictive models suitable for lead compound optimization.



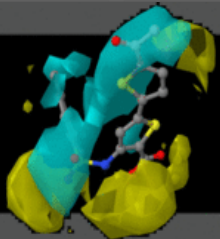


# Pharmacophore Definition



“A pharmacophore is the **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.”

C.-G. Wermuth et al., *Pure Appl. Chem.* 1998, 70: 1129-1143

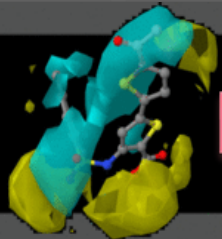


# Definition by IUPAC

A **pharmacophore** is the ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target structure and to trigger (or to block) its biological response.

A **pharmacophore** does not represent a real molecule or a real association of functional groups, but a purely abstract concept that accounts for the common molecular interaction capacities of a group of compounds towards their target structure.

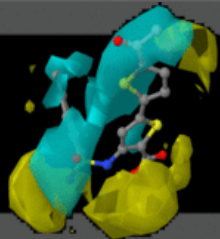
The **pharmacophore** can be considered as the largest common denominator shared by a set of active molecules. This definition discards a misuse often found in the [medicinal chemistry](#) literature which consists of naming as **pharmacophores** simple chemical functionalities such as guanidines, sulfonamides or dihydroimidazoles (formerly imidazolines), or typical structural skeletons such as flavones, phenothiazines, prostaglandins or steroids.



Interaction sites used to define a pharmacophore,

- H-bonding,
- hydrophobic
- Electrostatic

Defined by atoms, ring centers and virtual points.



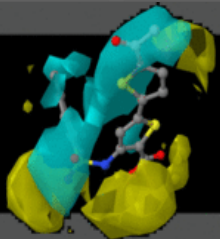
# The First Paper on Pharmacophoric Approach



## The Conformational Parameter in Drug Design: The Active Analog Approach

GARLAND R. MARSHALL, G. DAVID BARRY, HEINZ E. BOSSHARD,  
RICHARD A. DAMMKOEHLER, and DEBORAH A. DUNN

Departments of Physiology and Biophysics, of Pharmacology and of Computer  
Science and the Computer Systems Laboratory, Washington University,  
St. Louis, MO 63110



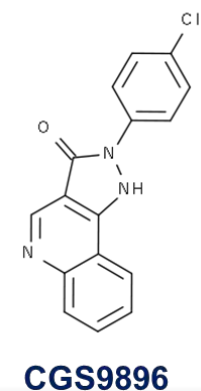
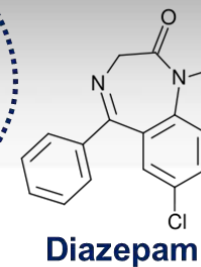
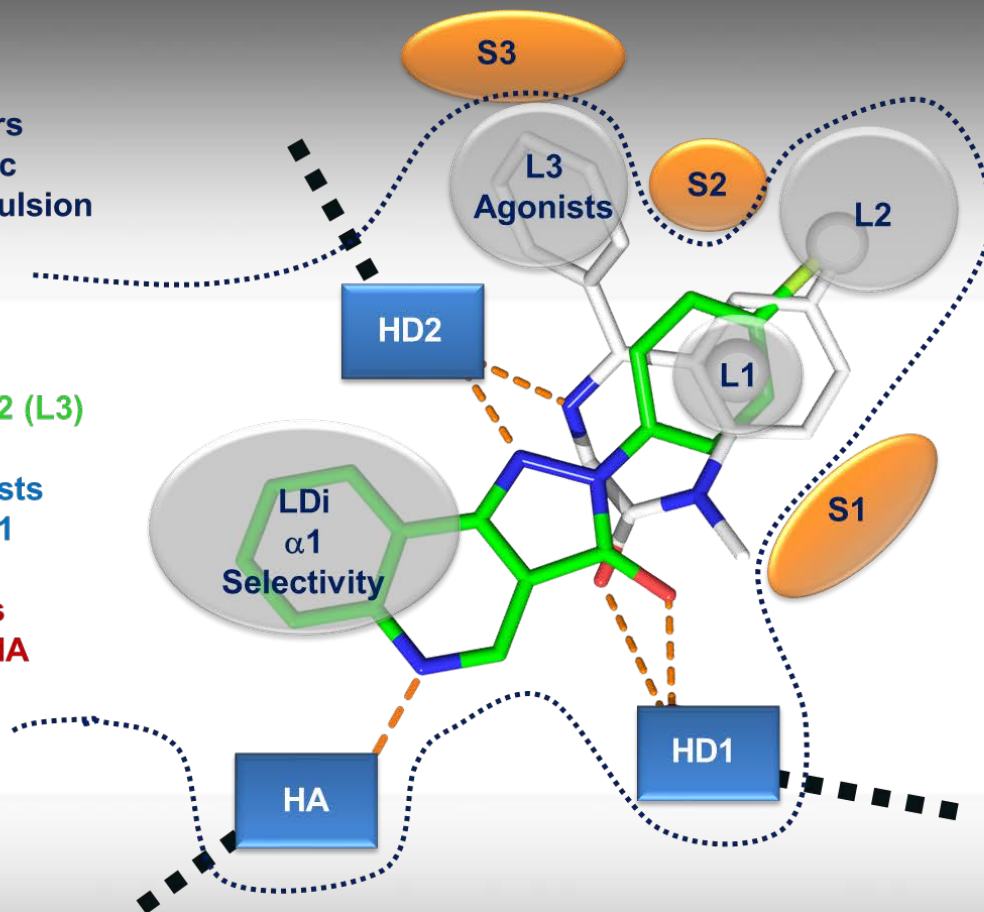
# BDZ Cook Model

HD: H Donator  
HA: H Acceptors  
L: Lipophyolic  
S: Steric Repulsion

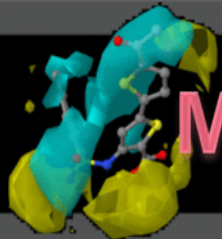
**Agonists**  
HD1, HD2, HA, L2 (L3)

**Inverse Agonists**  
HD1, HA e L1

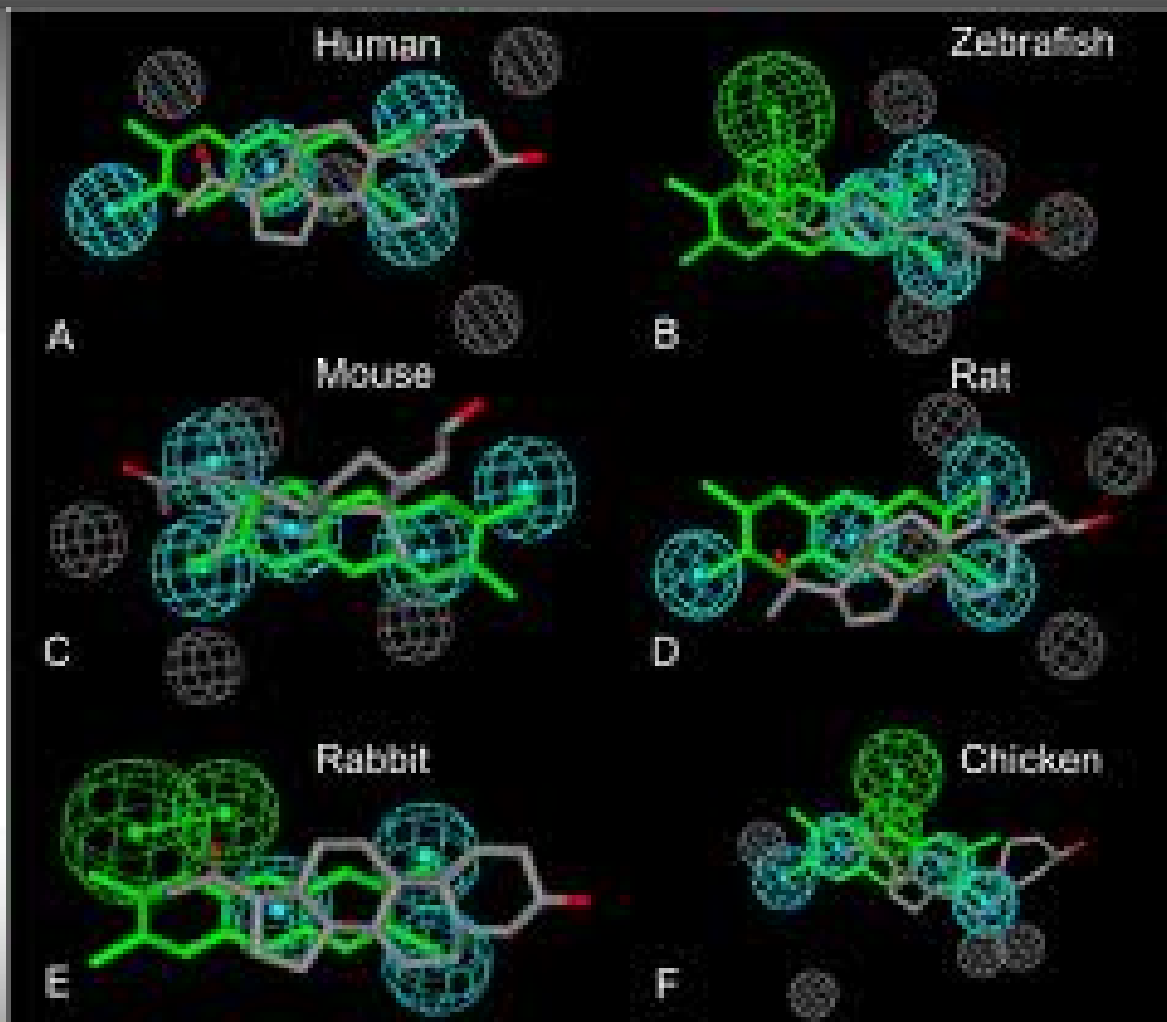
**Antagonists**  
HD1, HD2 e HA







# Modern Pharmacophore Model





# The Beginning of QSAR

## QSAR date back to the 19th century

In 1863, A.F.A. Cros at the University of Strasbourg observed that toxicity of alcohols to mammals increased as the water solubility of the alcohols decreased

In the 1890's, Hans Horst Meyer of the University of Marburg and Charles Ernest Overton of the University of Zurich, working independently, noted that the toxicity of organic compounds depended on their lipophilicity.

**Louis Hammett (1938) Electronic Parameters (Sigma-Rho)**

**Robert W. Taft (1952) Steric Parameters (Es)**

**Corvin Hansch (1964) Hydrophobic Parameter (LogP -  $\pi$ )**



# The Hansch Equation

ACS Publications  
 SciFinder subscribers: Click to sign in | Not a SciFinder subscriber? Learn more at www.cas.org

**ACCOUNTS**  
 of chemical research

Publications A-Z Home | Aut

Search

Acc. Chem

Browse the Journal | Articles ASAP | Current Issue | Submission & Review | Subscriptions

Research Article

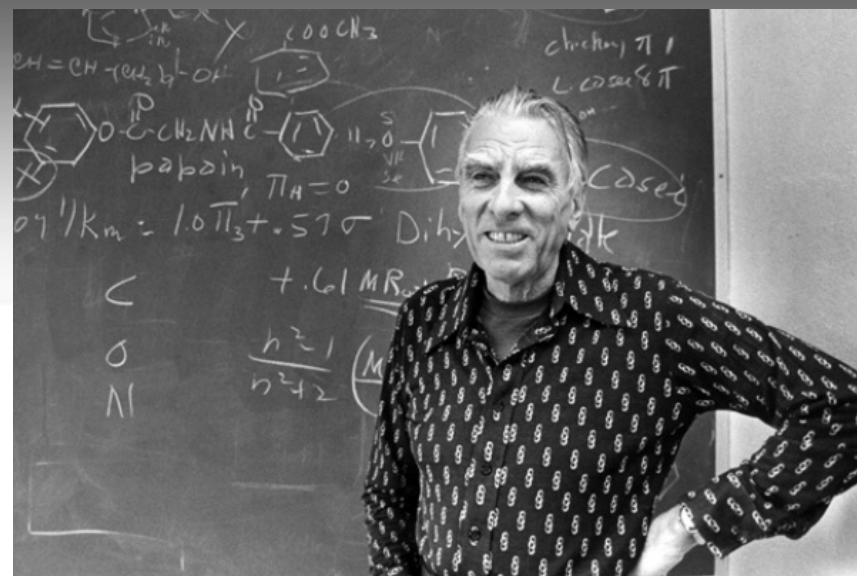
Quantitative approach to biochemical structure-activity relationships

Corwin **Hansch**

Acc. Chem. Res., 1969, 2 (8), pp 232-239  
 DOI: 10.1021/ar50020a002  
 Publication Date: August 1969

First Page Citing Articles

Hi-Res PDF [1026 KB]



## The Hansch Equation

$$\delta_X \log k_i = \delta_X \Delta G_{\text{hydrophobic}} + \delta_X \Delta G_{\text{electronic}} + \delta_X \Delta G_{\text{steric}}$$



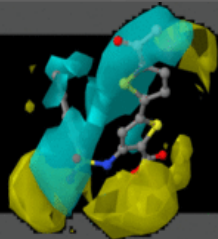
The first QSAR equations were based on the observation that partition coefficients, as expressed by  $\log P$  values, are to some extent, correlated to certain biological endpoints.

$$\log (1/C) = k_1 \log P + k_2 \sigma + k_3$$

*Conc. of compound required  
to produce a standard response  
in a given  $t$*

*Logarithm of the  
molecule's partition coefficient  
(1-octanol/water)*

*Hammet Parameter  
(molecule's electronic  
characteristics)*



# '60: QSAR Golden Age

**rcmd**  
www.rcmd.it

April 20, 1964

CORRELATION OF BIOLOGICAL ACTIVITY AND CHEMICAL STRUCTURE

1617

[CONTRIBUTION FROM THE DEPARTMENT OF CHEMISTRY, POMONA COLLEGE, CLAREMONT, CALIFORNIA]

$\rho$ - $\sigma$ - $\pi$  Analysis. **A Method for the Correlation of Biological Activity and Chemical Structure**

BY CORWIN HANSCH AND TOSHIO FUJITA<sup>1</sup>

RECEIVED AUGUST 19, 1963

Using the substituent constant,  $\sigma$ , and a substituent constant,  $\pi$ , defined as  $\pi = \log P_X - \log P_H$  ( $P_H$  is the partition coefficient of a parent compound and  $P_X$  that of a derivative), regression analyses have been made of the effect of substituents on the biological activity of benzoic acids on mosquito larvae, phenols on gram-positive and gram-negative bacteria, phenyl ethyl phosphate insecticides on houseflies, thyroxine derivatives on rodents, diethylaminoethyl benzoates on guinea pigs, and carcinogenic compounds on mice.

## *Journal of Medicinal Chemistry*

© Copyright 1964 by the American Chemical Society

VOLUME 7, NUMBER 4

JULY 6, 1964

---

**A Mathematical Contribution to Structure-Activity Studies**

SPENCER M. FREE, JR., AND JAMES W. WILSON





# Free-Wilson

## Journal of Medicinal Chemistry

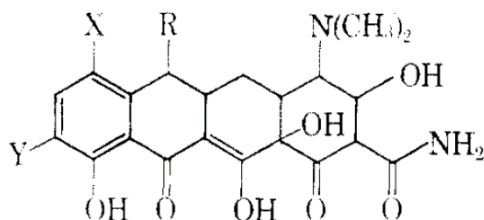
© Copyright 1964 by the American Chemical Society

VOLUME 7, NUMBER 4

JULY 6, 1964

### A Mathematical Contribution to Structure-Activity Studies

SPENCER M. FREE, JR., AND JAMES W. WILSON



where R was H or CH<sub>3</sub>; X was Br, Cl, or NO<sub>2</sub>; and Y was NO<sub>2</sub>, NH<sub>2</sub>, or CH<sub>3</sub>CONH.

#### BIOLOGICAL ACTIVITY OF TEN TETRACYCLINES

Compound	Compound identification							Biological activity	
	-R-	-X-			-Y-				
	H	CH <sub>3</sub>	NO <sub>2</sub>	Cl	Br	NO <sub>2</sub>	NH <sub>2</sub>	CH <sub>3</sub> CONH	
III	1		1			1			60
IV	1			1		1			21
V	1				1	1			15
VI	1			1			1		525
VII	1				1		1		320
VIII	1		1				1		275
IX		1	1				1		160
X		1	1					1	15
XI		1			1		1		140
XII		1			1			1	75

#### CONTRIBUTION OF STRUCTURAL CHANGES<sup>a</sup>

Side chain positions					
R		X		Y	
<i>a</i> [H]	75	<i>b</i> [Cl]	84	<i>c</i> [NH <sub>2</sub> ]	123
<i>a</i> [CH <sub>3</sub> ]	-112	<i>b</i> [Br]	-16	<i>c</i> [CH <sub>3</sub> CONH—]	18
		<i>b</i> [NO <sub>2</sub> ]	-26	<i>c</i> [NO <sub>2</sub> ]	-218



---

## Full Papers

---

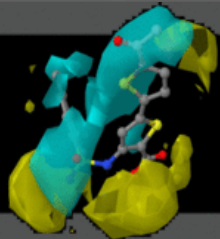
### **A Combined Hansch/Free-Wilson Approach as Predictive Tool in QSAR Studies on Propafenone-Type Modulators of Multidrug Resistance<sup>[1]</sup>**

Claudia Tmej<sup>a)</sup>, Peter Chiba<sup>b)</sup>, Mario Huber<sup>a)</sup>, Elisabeth Richter<sup>b)</sup>, Manuela Hitzler<sup>b)</sup>, Klaus-Jürgen Schaper<sup>c)</sup>, and Gerhard Ecker<sup>a)\*</sup>

*Arch. Pharm. Pharm. Med. Chem.*

© WILEY-VCH Verlag GmbH, D-69451 Weinheim, 1998

0365-6233/98/0708/0233 \$17.50 +.50/0



# QSAR Building Procedure

**Identification of Active Ligands**



**Identification of Suitable Descriptors (molecular fingerprint)**



**Establish Mathematical Expression Relating Descriptors to Activity**



**Construction and Validation of the QSAR model**



# QSAR Limitations

- **Congeneric Series**
- **Same Binding Mode**
- **Lack of 3-D Structural Information**
- **Statistical Limitation**



**(i) Multivariable linear regression analysis (MLR)**

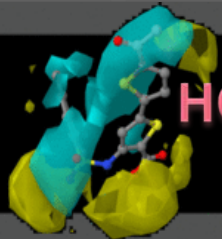
**(ii) Principal component analysis (PCA)**

**(iii) Partial least square analysis (PLS)**



## least square applications





## Squared Correlation Coefficient

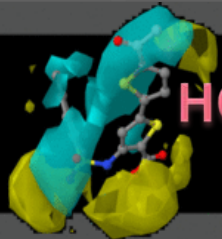
$R^2$  or  $r^2$

$$r^2 = 1 - \frac{\sum_{i=1}^N (Y_{\text{exp},i} - Y_{\text{calc},i})^2}{\sum_{i=1}^N (Y_{\text{exp},i} - \bar{Y})^2}$$

## Cross-Validated $R^2$

$Q^2$  or  $q^2$

$$q^2 = 1 - \frac{\sum_{i=1}^N (Y_{\text{exp},i} - Y_{\text{pred},i})^2}{\sum_{i=1}^N (Y_{\text{exp},i} - \bar{Y})^2}$$



## Fitting

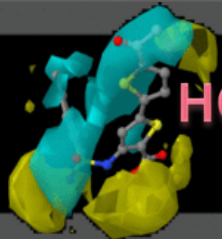
Squared Correlation  
Coefficient

$R^2$  or  $r^2$

$$r^2 = 1 - \frac{\sum_{i=1}^N (Y_{\text{exp},i} - Y_{\text{calc},i})^2}{\sum_{i=1}^N (Y_{\text{exp},i} - \bar{Y})^2}$$

$$0 \leq r^2 \leq 1$$

$$r^2 = \frac{ESS}{TSS} \equiv 1 - \frac{RSS}{TSS} \equiv 1 - \frac{\sum_{i=1}^N (Y_{\text{exp},i} - Y_{\text{calc},i})^2}{\sum_{i=1}^N (Y_{\text{exp},i} - \bar{Y})^2}$$



## CV (Cross-Validation)

Cross-validated  $R^2$

$Q^2$  or  $q^2$

$$q^2 = 1 - \frac{\sum_{i=1}^N (Y_{\text{exp},i} - Y_{\text{pred},i})^2}{\sum_{i=1}^N (Y_{\text{exp},i} - \bar{Y})^2}$$
$$SDEP = \sqrt{\frac{\sum_{i=1}^N (Y_{\text{exp},i} - Y_{\text{pred},i})^2}{N}}$$

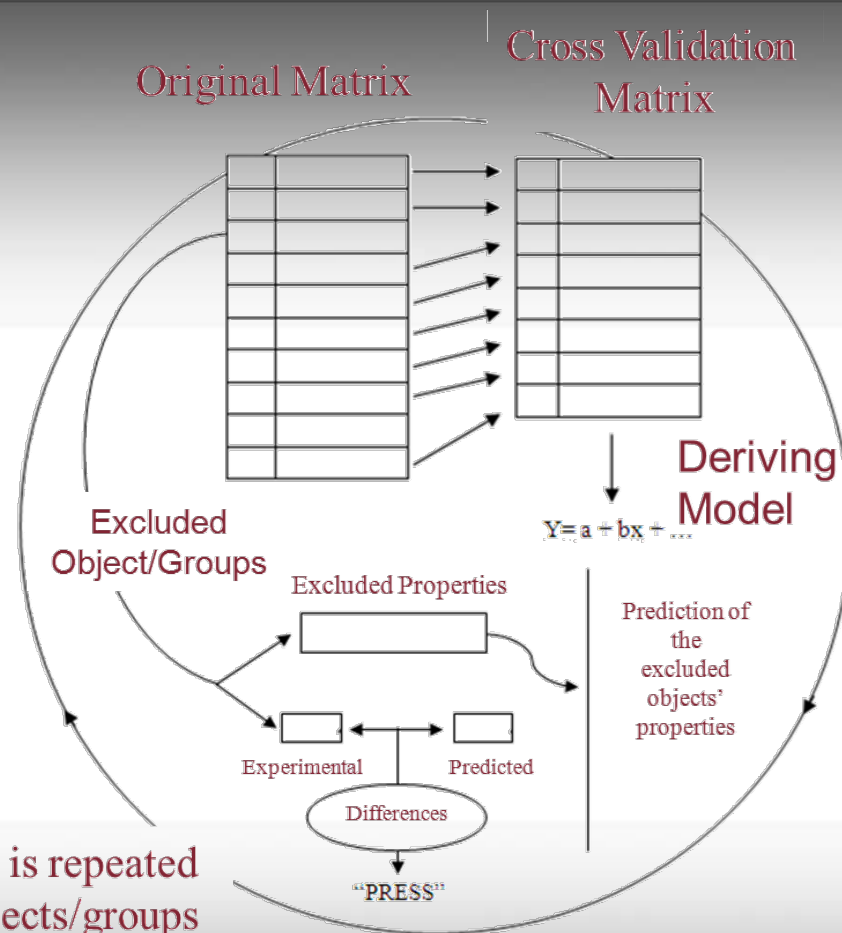
$$-\infty \leq q^2 \leq 1$$

The predictive ability of a model is estimated using a reduced set of structural data

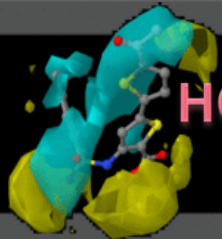


# Common Cross-Validation

- **K-fold cross-validation**
- **Repeated random sub-sampling validation**
- **Leave-one-out cross-validation**

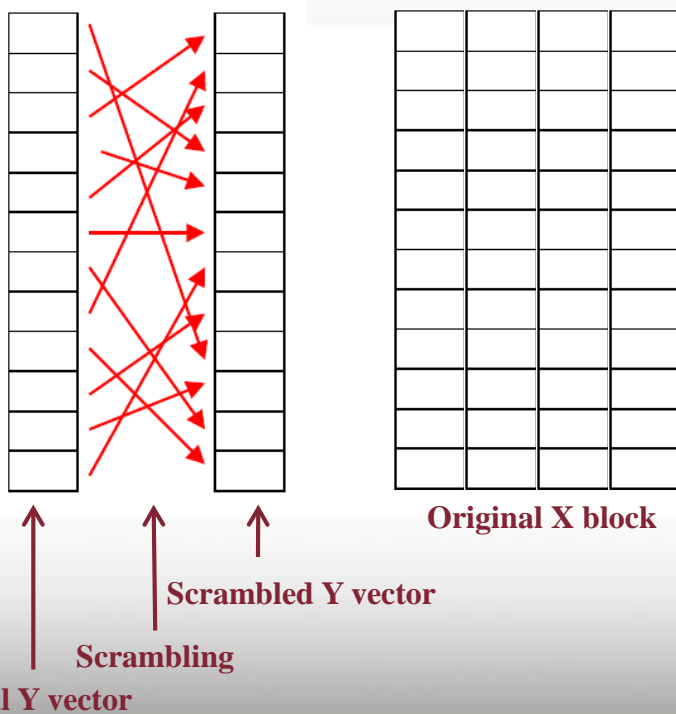


The process is repeated until all objects/groups are predicted



## Y-Scrambling

A statistical test of prediction tools, in which models are fitted for randomly reordered property/activity values and compared with the model obtained for the actual property/activity values.



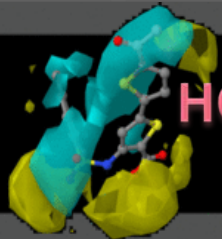
A new model is obtained for such permuted data,  $R^2$  and  $Q^2$  are then recalculated.

This step is repeated for a sufficient number of times (iterations):

**a good number being 50 to 100.**

Values obtained in the above fashion are compared with the true values obtained for the model that was fitted on the real data.





## External Test-Set

# SDEP (Standard Deviation Error of Prediction)

$$SDEP = \sqrt{\frac{\sum_{i=1}^N (Y_{\text{exp},i} - Y_{\text{pred},i})^2}{N}}$$

N° of predicted compounds