

Computational Methods in Drug Discovery

Computer-aided methods are used in the pharmacochemistry research to identify ligands of orphan targets (targets for which no ligand is known)

And

to identify new series of ligands on non-orphan targets.

Computational Methods in Drug Discovery

With the increase in the number of available targets thanks to the deciphering of the genome

computer-based methods will become a tool essential for a rapid and efficient identification of ligands prone to be “lead” molecules.

These methods are not to be substituted for the experimental ones but instead to limit in a rational way the chemical “space” so as to find the ligand sought-after.

Computational Methods in Drug Discovery

The rational computer-based methods and chance methods such as HTS and combinatorial chemistry are totally unlike.

HTS is limited
by the screening throughput and
by the limitations of the diversity in the chemical libraries.

The probability that a given library statistically produced by combinatorial chemistry covers the chemical “space” is rather weak
(This chemical space of drug-like molecules has been estimated to be in excess of 10^{60} molecules)

The computer-based methods permit to reduce the size of the ensemble of molecules to be really tested and to propose new tracks for the de novo design.

Computational Methods in Drug Discovery

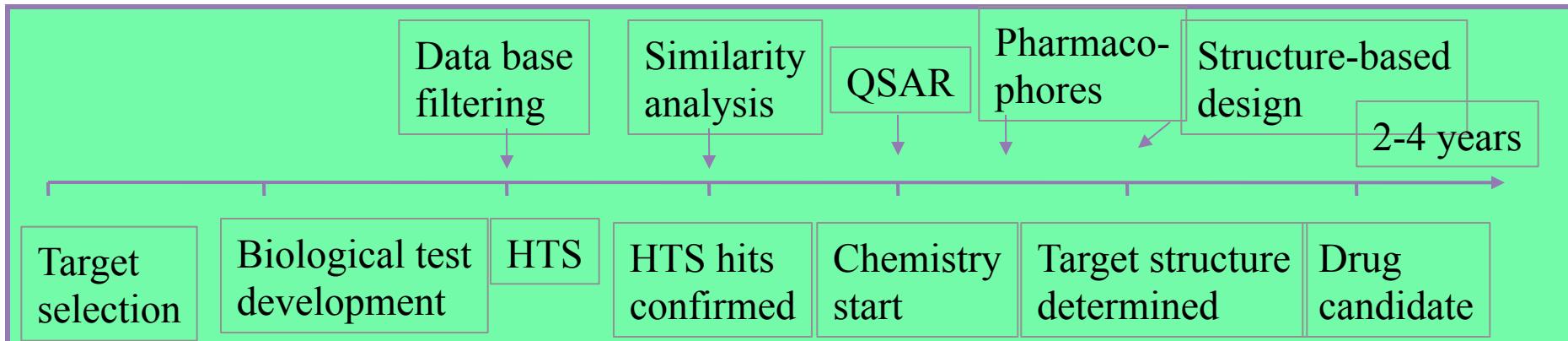
Computer-based methods play a part
in almost all steps of the process of ligand design to speed it up.

Computer-aided drug discovery (CADD)/design methods

have played a major role in the development of therapeutically important small molecules for several decades.

They may help in significantly **decreasing the number of compounds that should be screened experimentally** with high-throughput screening (HTS).

Computational Methods in Drug Discovery



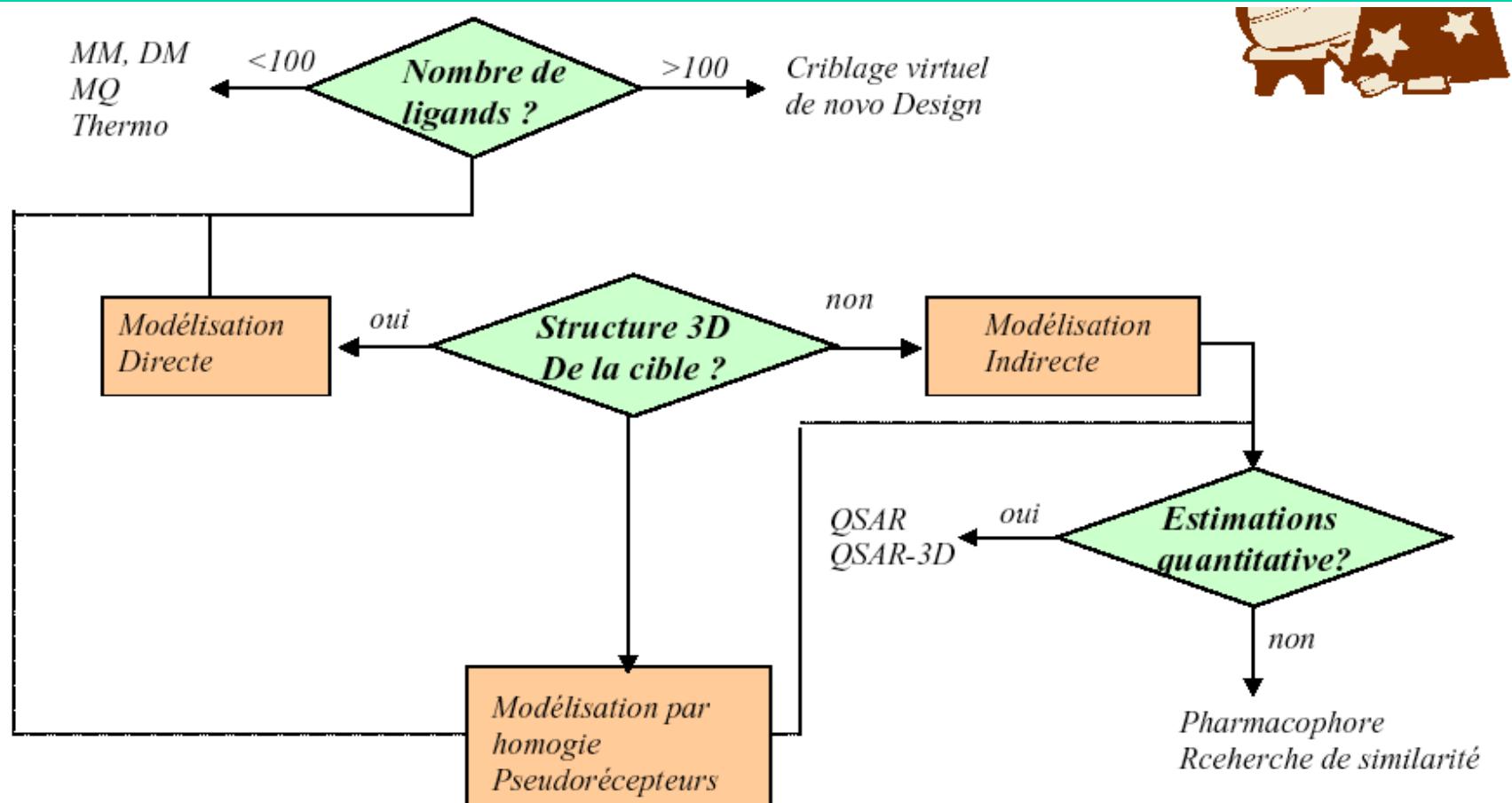
Computer tools can contribute:

- After the identification of the target and the setup of an activity test by **glancing through the databases** of existing molecules likely to meet the need.
- After the HTS you can use **similarity analysis** by screening data bases of available ligands comparable to the positive molecules obtained by HTS.
- After the phase of synthesis chemistry has started **QSAR** tools are frequently used as well as **pharmacophore** screening.
- When the **structure of the target** has been determined as well as possibly structures of complexes of the target with some synthesised ligands optimisation and design of de novo ligands can start.

Computational Methods in Drug Discovery

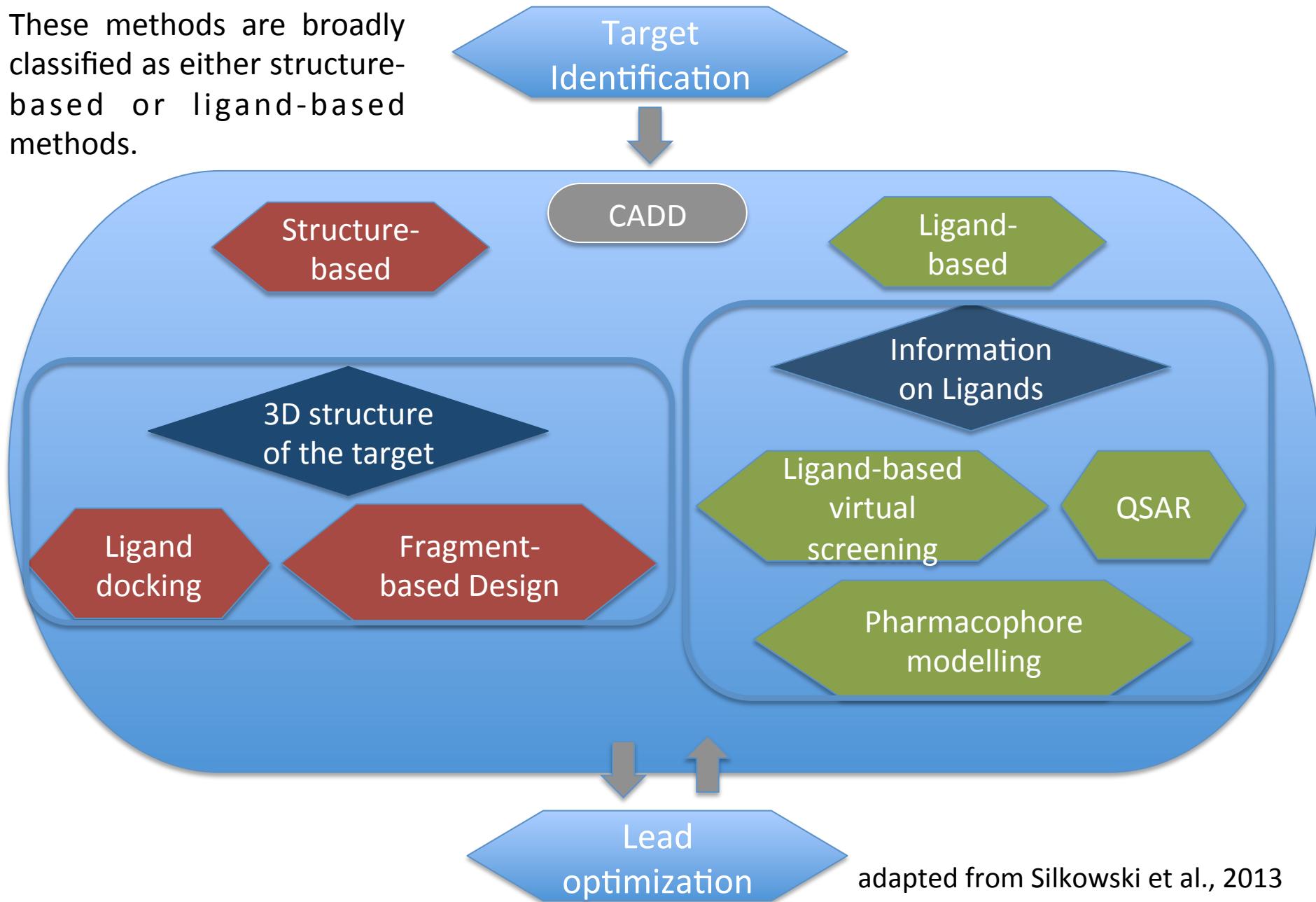
Two different families of computer-based methods can be distinguished :

- Methods which do not use the structure of the target (ligand-based)
- Methods which use the structure of the target (target-structure-based)



Computational Methods in Drug Discovery

These methods are broadly classified as either structure-based or ligand-based methods.



adapted from Silkowski et al., 2013

Computational Methods in Drug Discovery

CADD can be classified into two general categories:

Structure-based

Structure-based CADD relies on the knowledge of the target protein structure

to select compounds based on their binding energies.

Ligand-based

Ligand-based CADD exploits the knowledge of known active and inactive molecules through

- chemical similarity searches
- or
- construction of predictive, quantitative structure-activity relation (QSAR) models.

Computational Methods in Drug Discovery

Structure-based

Ligand-based

generally preferred

where high-resolution structural data of the target protein are available, i.e., for soluble proteins that can readily be crystallized.

generally preferred

when no or little structural information is available, as often for membrane protein targets.

Ligand-based computer-aided drug discovery (LB-CADD) approach involves:

- Choose ligands known to interact with the target of interest
- Use a set of reference structures collected from this ensemble of ligands
- Analyze their 2D or 3D structures.

Overall goal:

- represent these compounds with their most important physicochemical properties for their desired interactions
- Discard extraneous information not relevant to the interactions

Two fundamental approaches of LB-CADD are

Selection of novel compounds based on chemical similarity to known active compounds using some similarity measure

Construction of a QSAR model that predicts biologic activity from chemical structure

Difference between the two approaches:

QSAR weights the features of the chemical structure of the compounds according to their influence on the biological activity of interest.

Ligand-based

Construction of a QSAR
model that predicts
biologic activity from
chemical structure

Among the computational chemistry methods:
Quantitative Structure Activity Relationships (QSARs).

QSARs:

**equations that help to predict biological “activity” from
chemical structure of ligands.**

Ligand-based

Construction of a QSAR
model that predicts
biologic activity from
chemical structure

What is QSAR ?

QSAR studies try to quantitatively link the variations of the biological activity of molecules (ligands) with changes in their molecular descriptors.

The molecular descriptors are characteristics featuring electronic, spectroscopic or other properties (hydrophobicity).

The biological activity and the physicochemical properties are connected by some mathematical function F :

Biological activity = F (Physicochemical Properties)

The aim is to determine these relationships and to apply them on new chemical entities.

Ligand-based

Construction of a QSAR model that predicts biologic activity from chemical structure

Where does the QSAR principle come from ?

How a molecule behaves? ← from its structure.

Small molecules containing from one to four carbon atoms are gases at room temperature:

methane CH_4 , ethane C_2H_6 , propane C_3H_8 , butane C_4H_{10}

Adding more carbons: the substance becomes:
hexane, C_6H_{14} : a liquid and octadecane, $\text{C}_{18}\text{H}_{38}$: a solid.

Adding one oxygen atom to methane (CH_4): Methanol (CH_3OH) is a liquid

This behaviour is not limited to predicting whether the molecule is a solid, liquid or gas, but predicting more various and complex properties.

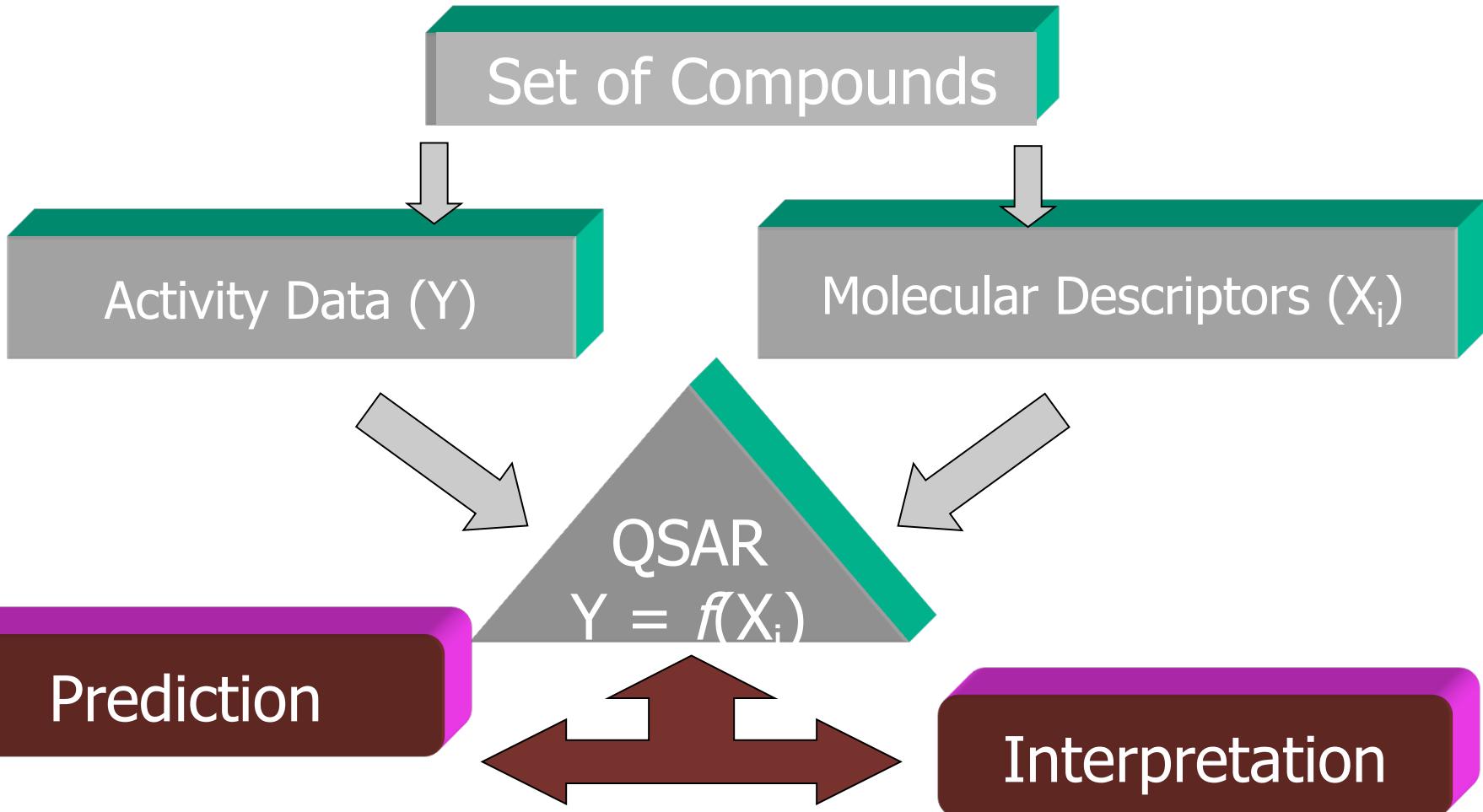
Construction of a QSAR
model that predicts
biologic activity from
chemical structure

What does one need to do QSAR ?

- A number of molecules whose activity is known is required.
 - Physicochemical descriptors for these molecules
- and a mathematical method appropriate to obtain a model is also needed.

Ligand-based

Construction of a QSAR model that predicts biologic activity from chemical structure



Construction of a QSAR model that predicts biologic activity from chemical structure

Molecular descriptors can be

- structural
- physicochemical

Properties such as:

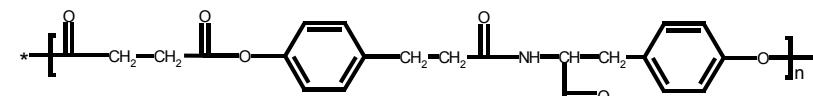
- ✧ molecular weight
- ✧ geometry
- ✧ volume, surface areas, electronegativities, polarizabilities
- ✧ functional group composition
- ✧ solvation properties

Ligand-based

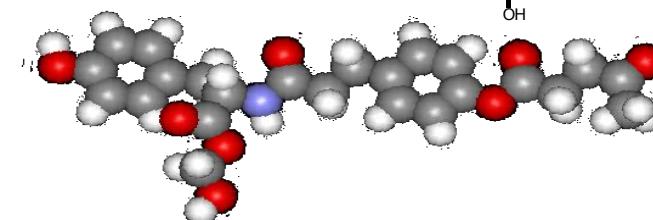
Construction of a QSAR model that predicts biologic activity from chemical structure

Types of Molecular Descriptors

2-D structural formula

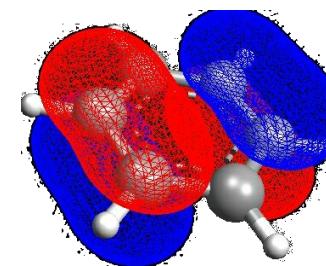


Geometrical

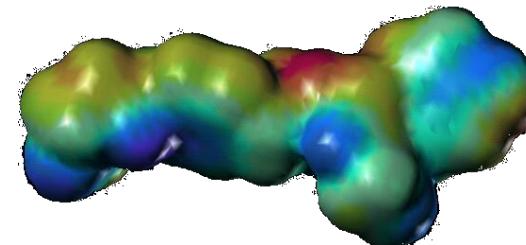


3-D shape and structure

Quantum Chemical



Electrostatic



Thermodynamic

Construction of a QSAR model that predicts biologic activity from chemical structure

Molecular descriptors are classified according to their “dimensionality”

1D scalar physicochemical properties such as molecular weight

2D molecular constitution-derived descriptors

3D molecular conformation-derived descriptors

4D extension of 3D-QSAR that treats each molecule as an ensemble of different conformations

Construction of a QSAR model that predicts biologic activity from chemical structure

Prediction of Psychochemical Properties.

The simplest ones, such as molecular weight and number of hydrogen bond donors, are relatively simple to compute.

More complex descriptors such as solubility are more difficult to compute.

Prediction of physicochemical properties is a critical step in developing effective molecular descriptors.

Construction of a QSAR model that predicts biologic activity from chemical structure

- The mathematical methods used are

Least square fit, principal component analysis, partial least squares, genetic algorithms or neuronal networks

Simple Linear Regression

$$\text{Activity} = a_0 + a_1 (\text{Mol Vol}_i)$$

Multiple Linear Regression

$$\text{Activity} = a_0 + a_1 (\text{Mol Vol}_i) + a_2 (\log P) + a_3 (\mu_i) + \dots$$

Partial Least-Squares (PLS) Regression

$$\text{Activity} = a_0 + a_1 (\text{PC1}) + a_2 (\text{PC2}) + a_3 (\text{PC3}) + \dots$$

Neuronal Networks

Construction of a QSAR
model that predicts
biologic activity from
chemical structure

Comparative field molecular analysis (CoMFA) (Cramer et al., 1988) :

3D-QSAR technique

- ✓ aligns molecules and their molecular interaction fields to relate them to biological activity.

Construction of a QSAR model that predicts biologic activity from chemical structure

Methodology

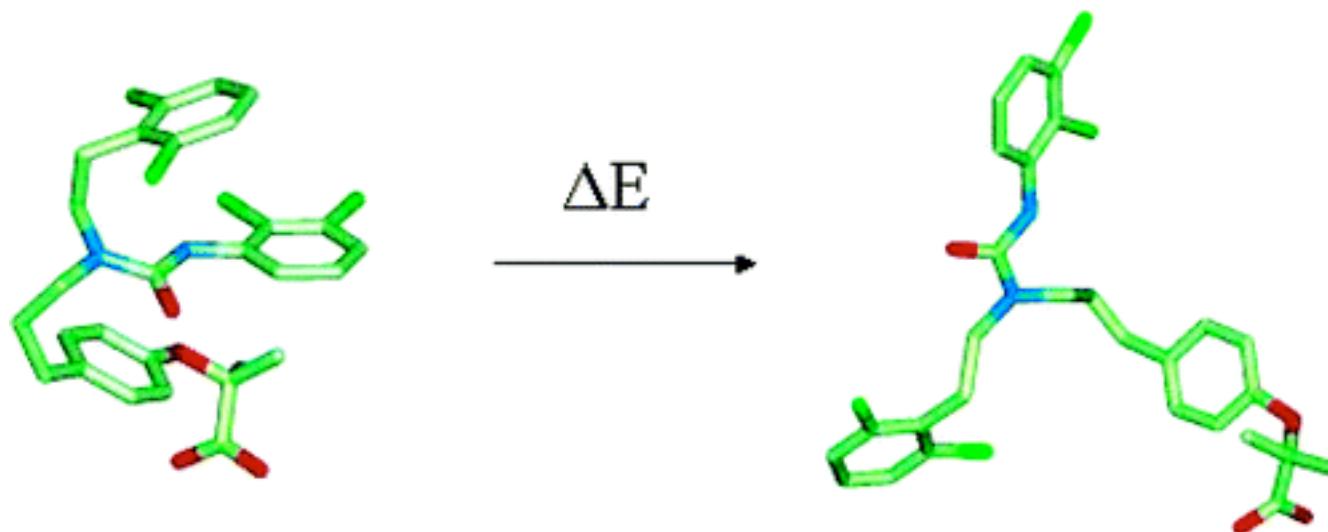
- 1. Choice of a set of representative molecules (ligands) == Training set**
- 2. Definition of the bioactive conformation of each molecule (3D information)**
- 3. Alignment of molecules**
- 4. Definition of a 3D grid including the totality of the molecules (larger than the volume of the molecules)**
- 5. Calculating for each molecule in every point of the grid the interaction energy with a probe atom → steric (van der Waals) and electrostatic fields**
- 6. Importing all the energy values in a table**
- 7. Correlation between the variations of the biological activity and the variations of the fields Biological Activity = f (fields)**

Ligand-based

Construction of a QSAR model that predicts biologic activity from chemical structure

CoMFA: Comparative Molecular Field Analysis

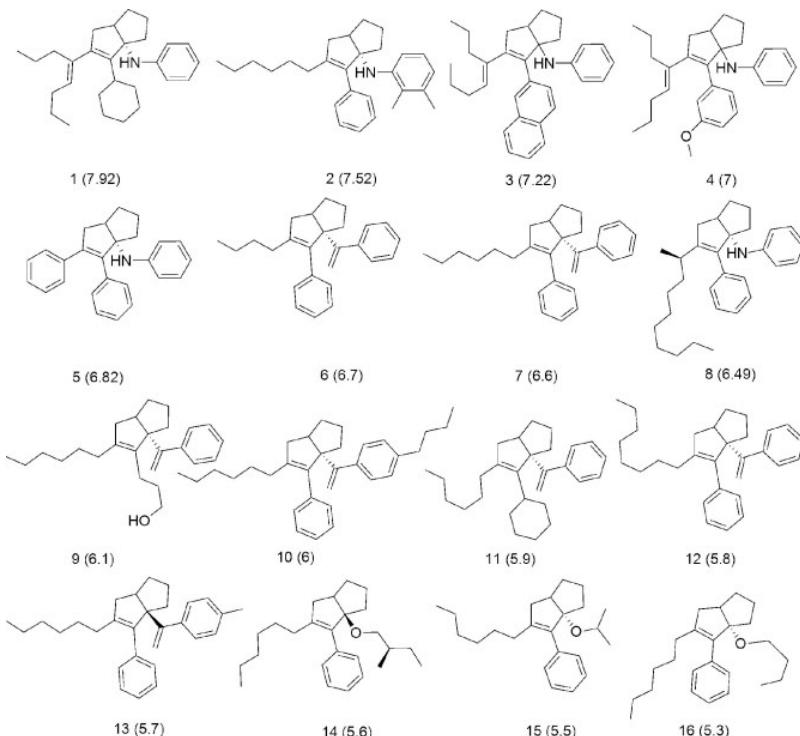
2. Definition of the bioactive conformation of each molecule (3D information)



Ligand-based

Construction of a QSAR model that predicts biologic activity from chemical structure

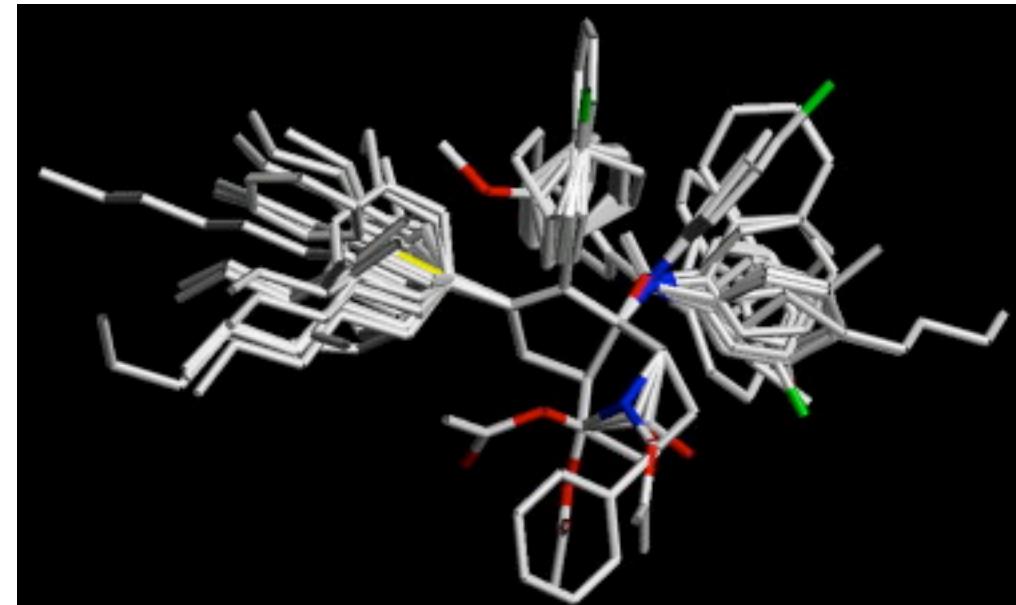
CoMFA: Comparative Molecular Field Analysis



Liver receptor homolog-1 (LRH-1)
(nuclear receptors) agonists

3. Alignment of molecules

The data are separated into a training set for which a CoMFA model is derived and a test set that will prove the predictivity of the model ill be proved



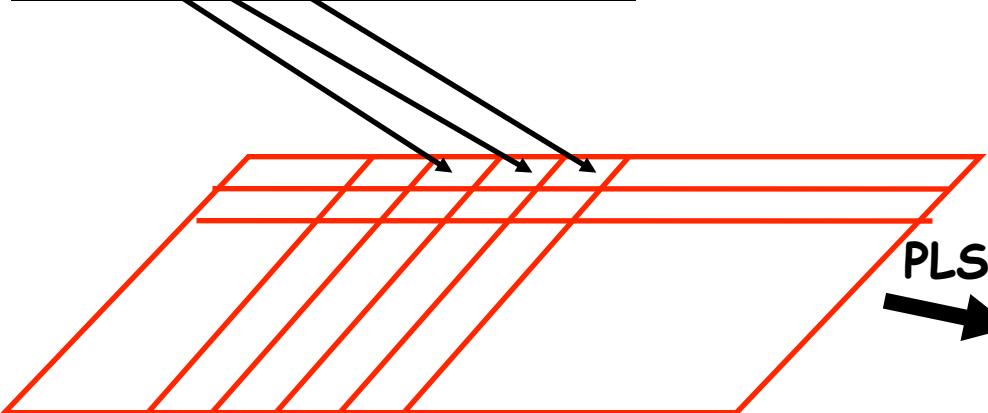
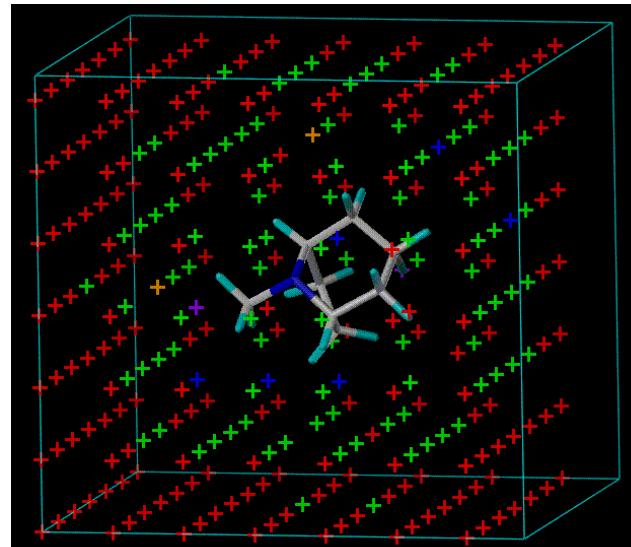
Ligand-based

Construction of a QSAR model that predicts biologic activity from chemical structure

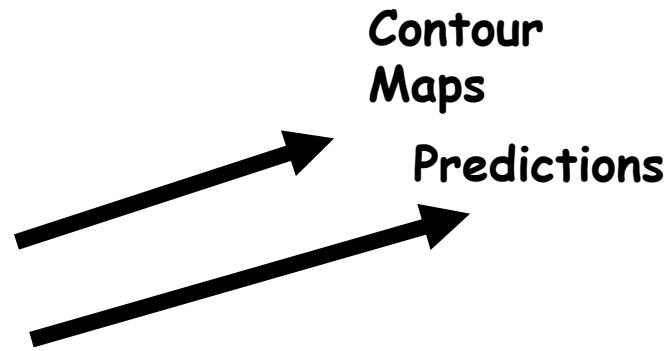
4. Definition of a 3D grid including the totality of the molecules

5. Calculating for each molecule in every point of the grid the interaction energy with a probe atom steric (van der Waals) and electrostatic fields

6. Importing all the energy values in a table



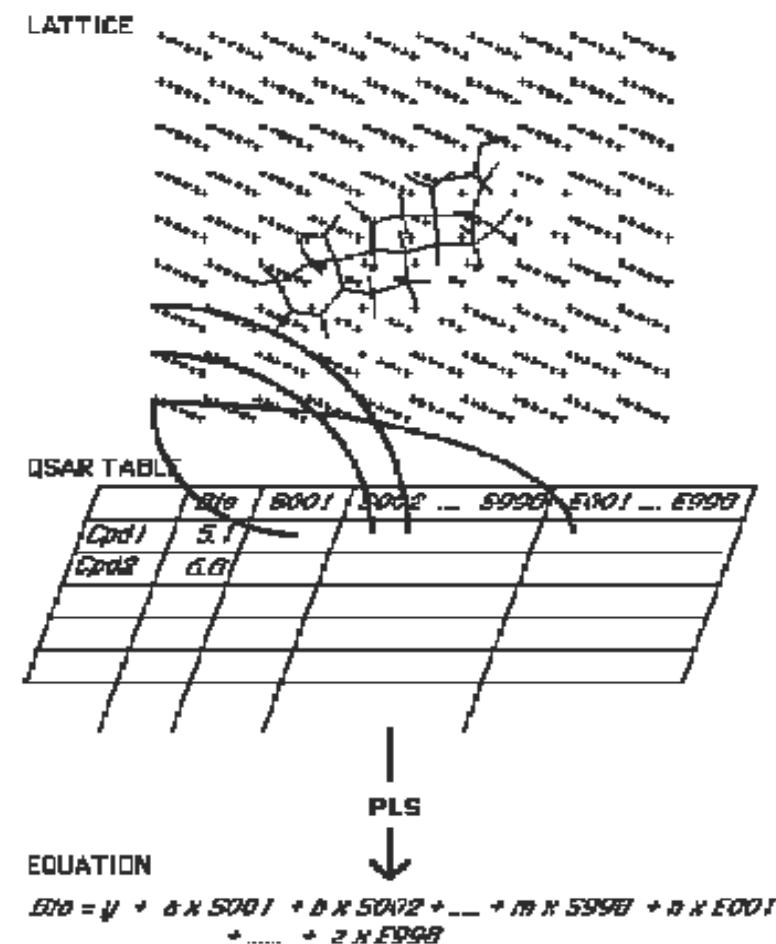
PLS
QSAR
equation



Construction of a QSAR model that predicts biologic activity from chemical structure

7. Correlation between the variations of the biological activity and the variations of the fields

Biological Activity = f (fields)



Ligand-based

Construction of a QSAR
model that predicts
biologic activity from
chemical structure

Success of QSAR depends:

on the quality of the initial set of active/inactive compounds

And

on the choice of descriptors

And

the ability to generate the appropriate mathematical relationship.

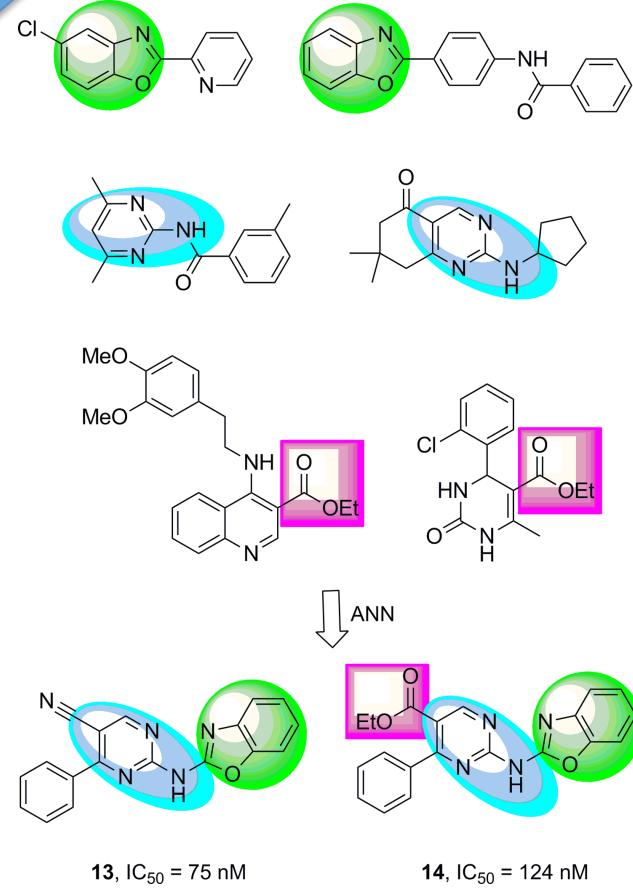
Ligand-based

Construction of a QSAR model that predicts biologic activity from chemical structure

QSAR models identified novel positive and negative allosteric modulators of mGlu5 implicated in neurologic disorders using data set from HTS (Mueller et al. (2012)).

The descriptors included scalar, 2D, and 3D descriptor categories.

QSAR-based virtual screening of mGlu5 negative allosteric modulators yields lead compounds that contain substructure combinations taken across several known actives used for model generation.



Ligand-based

Selection of compounds
based on chemical
similarity to known
active compounds using
some similarity measure

Virtual screening is a **toolbox of methods** to select appropriate candidates.

Input: **chemical structures and calculated properties** of the compounds

Virtual screening: applied to **virtual libraries of almost any size**.

Underlying assumption: **similar structures have similar biological activity**.

Selection of compounds
based on chemical
similarity to known
active compounds using
some similarity measure

Ligand Structure Searches

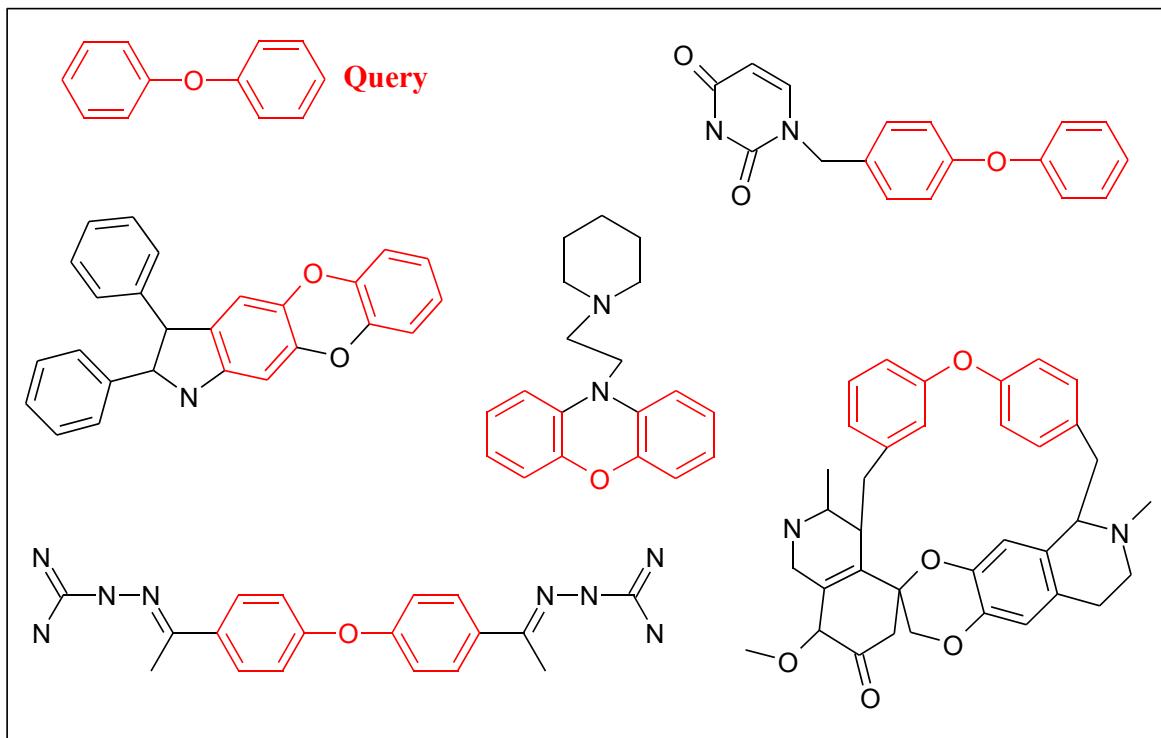
Strategies

- 2D Substructure searches
- 3D Substructure searches
- Pharmacophore matching

using chemotype information from first generation hits

Selection of compounds
based on chemical
similarity to known
active compounds using
some similarity measure

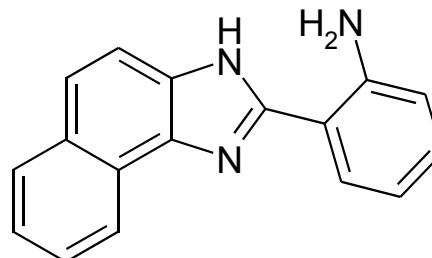
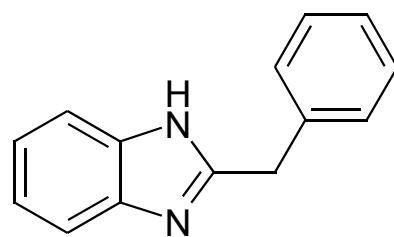
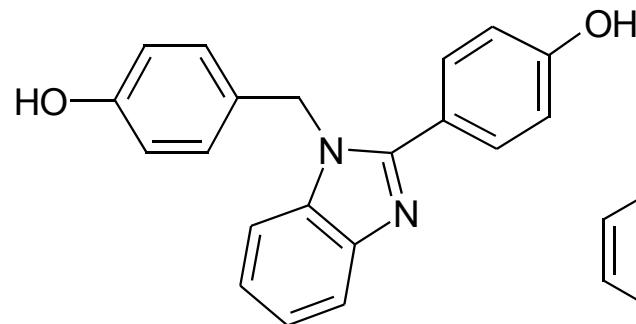
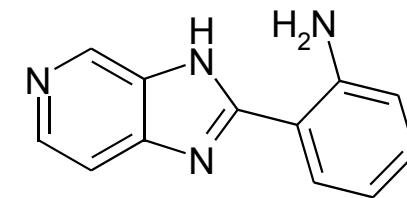
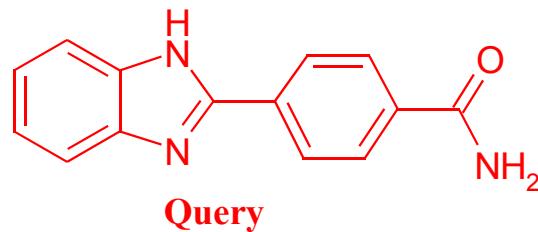
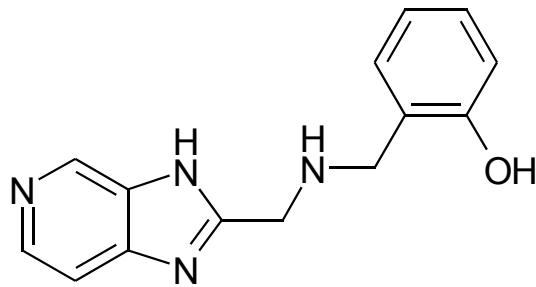
2D Substructure Searching



Ligand-based

Selection of compounds
based on chemical
similarity to known
active compounds using
some similarity measure

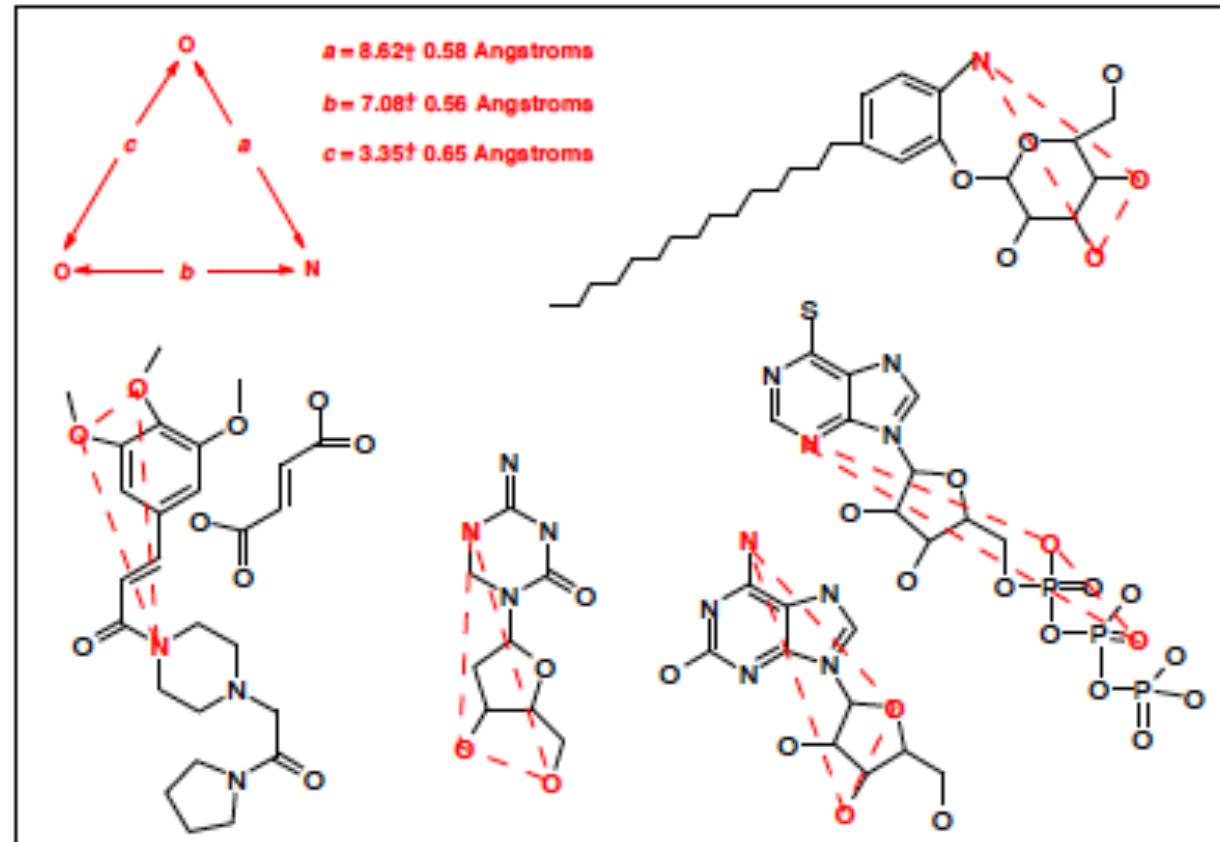
2D Similarity Searching



Selection of compounds based on chemical similarity to known active compounds using some similarity measure

3D Substructure Searching

Search database of molecules for ones with similar 3D shape and chemistry



Ligand-based

Selection of compounds based on chemical similarity to known active compounds using some similarity measure

Pharmacophore Searching

A pharmacophore:

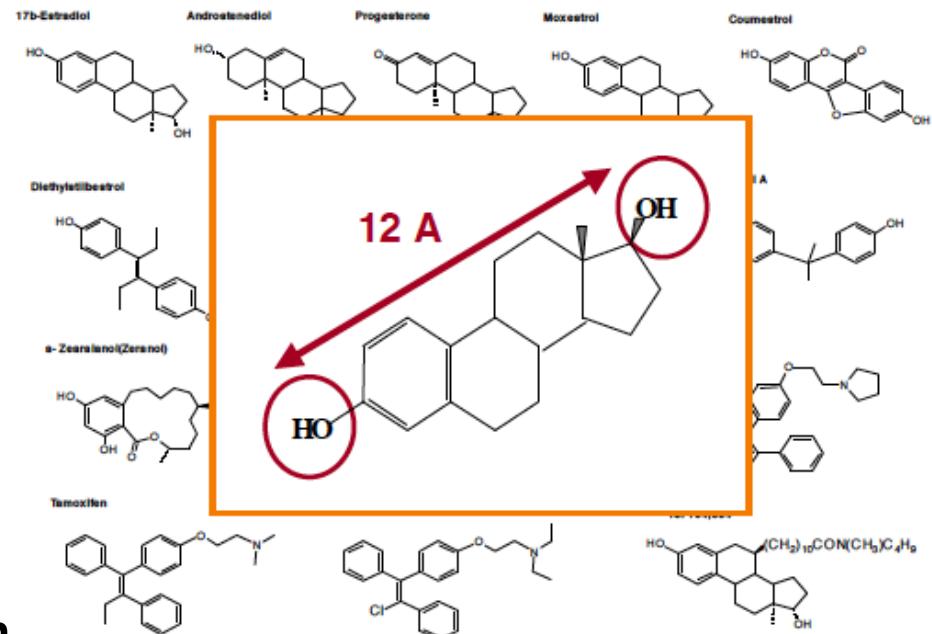
contains information about functional groups that interact with the target as well as information regarding the type of noncovalent interactions and interatomic distances between these functional groups/interactions.

This arrangement can be derived in a ligand-based approach or in a structure-based manner

Selection of compounds based on chemical similarity to known active compounds using some similarity measure

1. Selection of representative molecules
2. Choose a reference molecule
3. Definition of the bioactive conformation of the chosen molecules
4. Superposition of all molecules on the reference
5. Pharmacophore Feature Extraction
6. Design of new molecules

Pharmacophore Searching

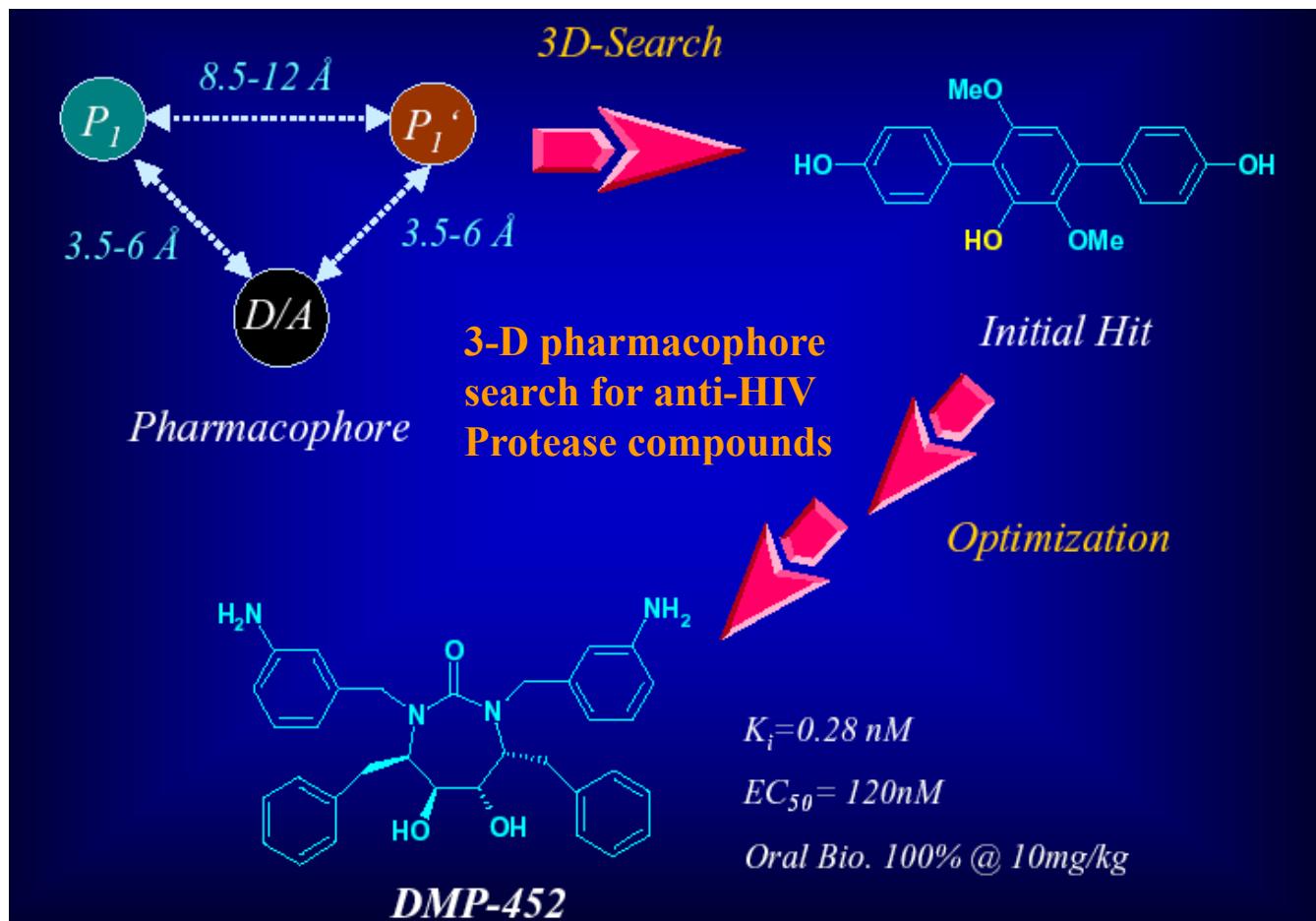


Ligands for Estrogen Receptor Slide courtesy of Bill Welsh

Ligand-based

Selection of compounds
based on chemical
similarity to known
active compounds using
some similarity measure

Pharmacophore Searching

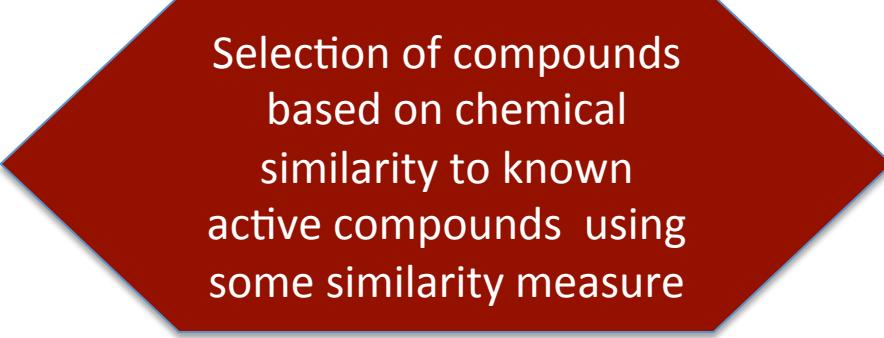


Selection of compounds
based on chemical
similarity to known
active compounds using
some similarity measure

Virtual screening can be used to eliminate or to select molecules in a given data base which corresponds to a negative or positive selection.

One can introduce characteristics like

- stability, solubility, lipophilicity and metabolismable to influence absorption, distribution, metabolism and excretion (ADME) of the drugs.



Selection of compounds
based on chemical
similarity to known
active compounds using
some similarity measure

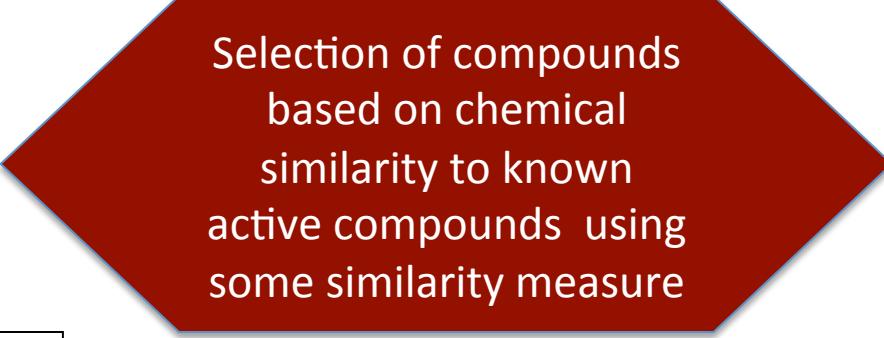
Positive selection

The Rule of Five from Lipinski is an empirical method to evaluate the bioavailability of a compound.

Four elements permit to postulate that a molecule will be absorbed are :

- the molecule cannot possess more than 5 hydrogen bond donors
- the molecular weight cannot exceed 500 g/mol
- the measure of lipophilicity of the molecule cannot exceed a certain threshold ($\log P > 5$)
- the molecule cannot possess more than 10 hydrogen bond acceptors

These properties can be quickly computed and can be applied to filter a large chemical library.



Selection of compounds
based on chemical
similarity to known
active compounds using
some similarity measure

Positive selection

The Rule of Five from Lipinski is an empirical method to evaluate the bioavailability of a compound.

Four elements permit to postulate that a molecule will be absorbed are :

- the molecule cannot possess more than 5 hydrogen bond donors
- the molecular weight cannot exceed 500 g/mol
- the measure of lipophilicity of the molecule cannot exceed a certain threshold ($\log P > 5$)
- the molecule cannot possess more than 10 hydrogen bond acceptors

These properties can be quickly computed and can be applied to filter a large chemical library.

Selection of compounds
based on chemical
similarity to known
active compounds using
some similarity measure

Negative selection

Likewise Lipinski's rule can be used to keep molecules able to lead to bioavailable compounds, screening can also be used to exclude compounds known to cause stability and/or toxicity problems.

REOS ("Rapid Elimination Of Swill") permits to guide the search towards ligands not only active but also prone to become good leads in terms of pharmacochemistry.

Selection of compounds based on chemical similarity to known active compounds using some similarity measure

Criteria used to
eliminate
undesirable
molecules are :

The presence of
some functional
groups known to be
toxic
(chlorosulfonyl,
pyrene, nitro
group, ...)

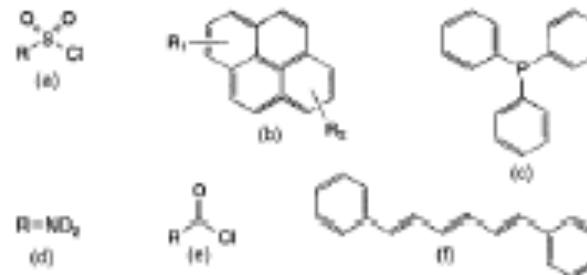


Figure 4: Exemples de groupes fonctionnels indésirables pouvant être éliminés par le filtre REOS. (a) chlorosulfonyl; (b) pyrène; (c) triphénylphosphine; (d) groupe nitro; (e) chlorure d'acide; (f) 1,6-diphényl 1,3,5-hexène. (Adapté de : W.P. Walters *et al.*, 1998, *Drug Discov. Today*, **3**, p. 171).

Property	Minimum	Maximum
Molecular weight	200	500
LogP	-5	5
Hydrogen-bond donors	0	5
Hydrogen-bond acceptors	0	10
Formal charge	-2	2
Number of rotatable bonds	0	8
Number of heavy atoms	15	50

Long aliphatic chain	O-O	1,2 dicarbonyl
Sulfonyl halide	Nitro group	Aldehyde
Primary alkyl halide	Epoxide	Sulphonate ester
Perhalo compound	-N=C=O	-N=N ⁺ =N-
	Isocyanate	Azide

The Concept of „Privileged Structures“

„these structures appear to contain common features which facilitate binding to various ... receptor surfaces, perhaps through binding elements different from those employed for binding of the natural ligands

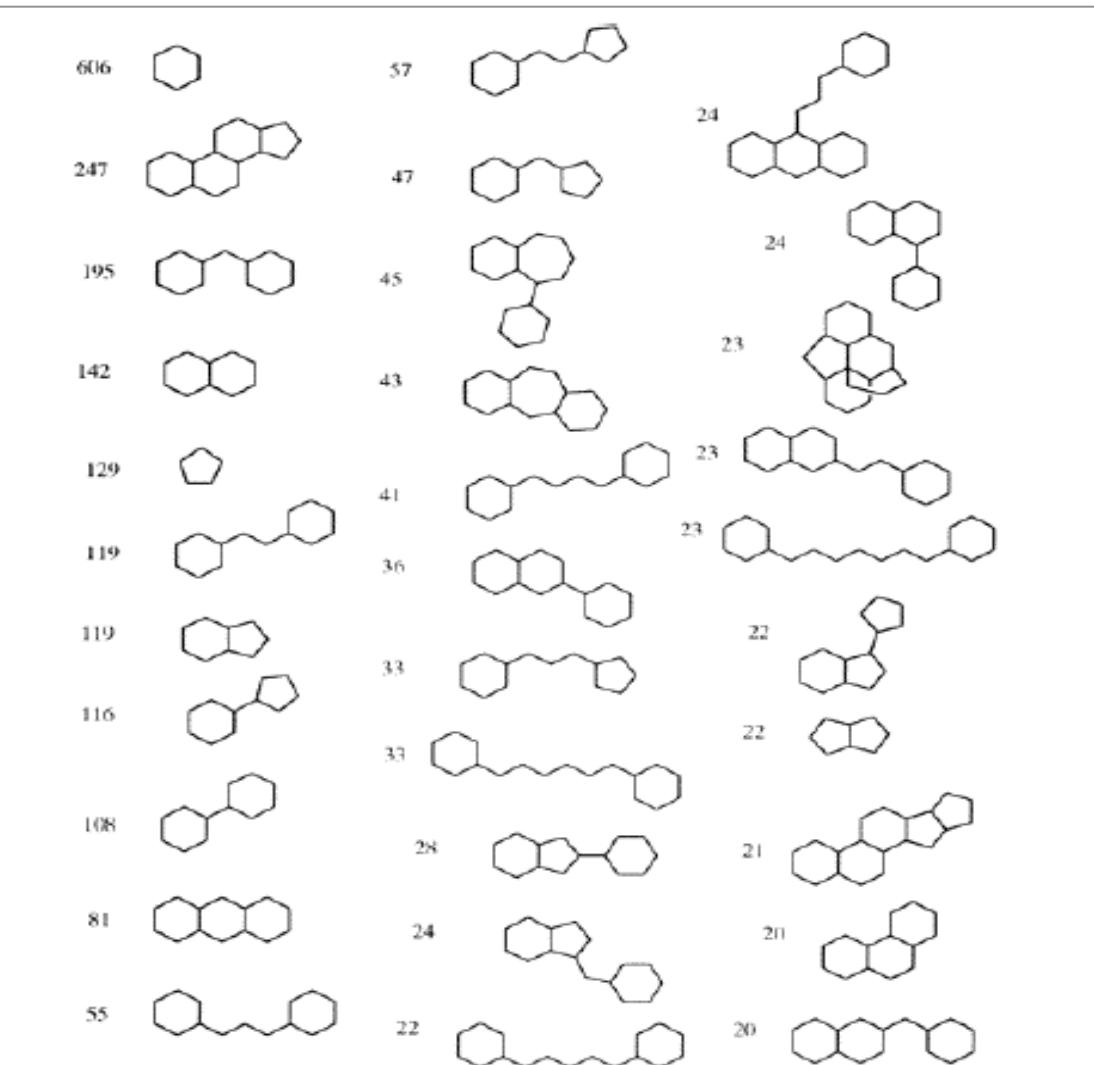
... what is clear is that certain „privileged structures“ are capable of providing useful ligands for more than one receptor and that judicious modification of such structures could be a viable alternative in the search for new receptor agonists and antagonists.“

B. E. Evans et al., J. Med. Chem. 31, 2235-2246 (1988)

Intense interest has been placed on the “privileged structures” for drug design

A “privileged structure” is a single molecular framework able to provide ligands for diverse receptors

Computational Methods in Drug Discovery



Indole, piperazine,
biphenyls, ...

Figure 5: Les 32 plateformes les plus couramment trouvées dans 2548 médicaments. Le nombre à coté des plateformes représente le nombre d'occurrence.
(Source : G.W. Bernis & M.A. Murcko, 1996, *J. Med. Chem.*, 39, p. 2889).

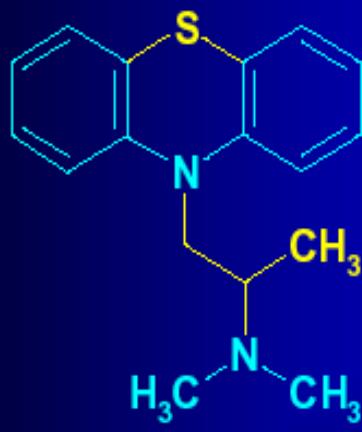
Ligand-Based Screening: Critical Issues

Starting pharmacophore(s)

multiple binding mode, overlapping binding sites, pharmacophore number
agonists vs. antagonists, substrates vs. inhibitors

What is similarity ?

Biological vs. chemical similarity



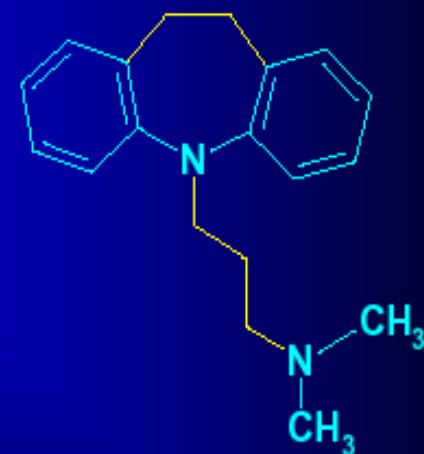
Promethazine

H1 antagonist
antiallergic



Chlorpromazine

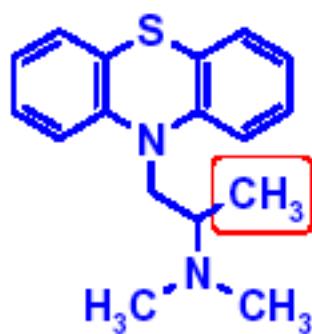
D-antagonist
meuroleptic



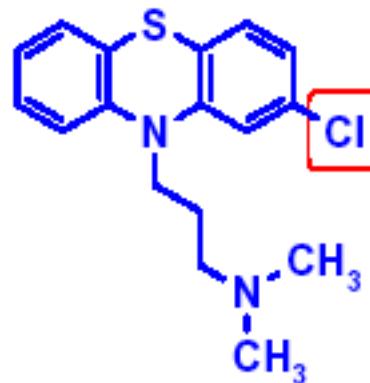
Imipramine

NA/5-HT uptake inhibitor
antidepressant

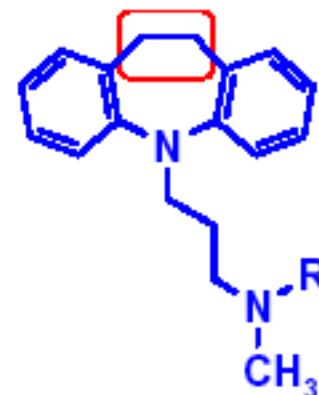
Different Modes of Action of Chemically Similar Molecules



promethazine
 $(\text{H}_1$ antagonist)



chlorpromazine
 $(\text{dopamine}$ antagonist)



a, $\text{R} = \text{CH}_3$, imipramine
b, $\text{R} = \text{H}$, desipramine
(uptake blocker)