Facultade de Química
Universidade de Vigo

"Computational approaches in drug discovery: expectations and reality."

Stefano Moro

Molecular Modeling Section (MMS)

Department of Pharmaceutical and Pharmacological Sciences

University of Padova

Here we are:

Virtual
Screening (VS)

Target
Structure Determination

Ligand-based
Virtual Screening (LBVS)

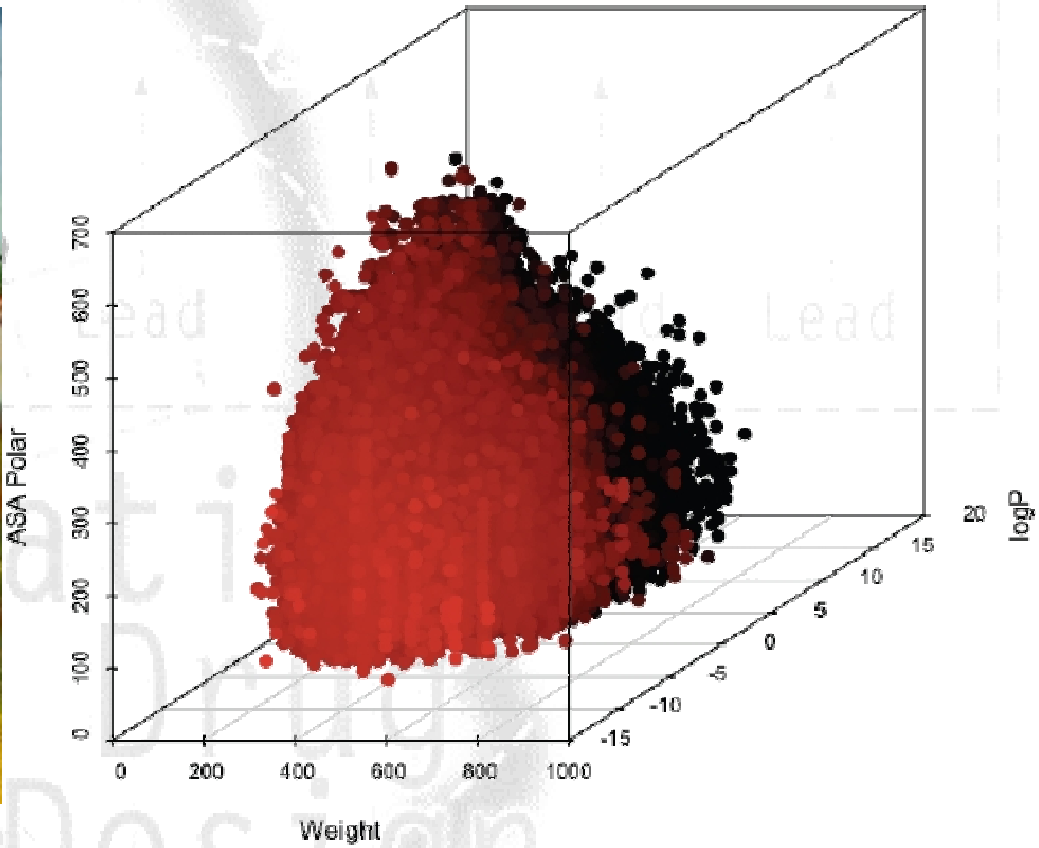
Similarity search
Pharmacophore matching

Structure-based
Virtual Screening (SBVS)

HT Molecular Docking

Back to the ratio xxx.000:1

Discovery a new drug is like to…



... molecular haystacks:

A division of the American Chemical Society

Chemical Abstract Service

(www.cas.org)

Date 06/06/2012 12:24:40 EST

Count

67,756,577 commercially available chemicals

… molecular haystacks:

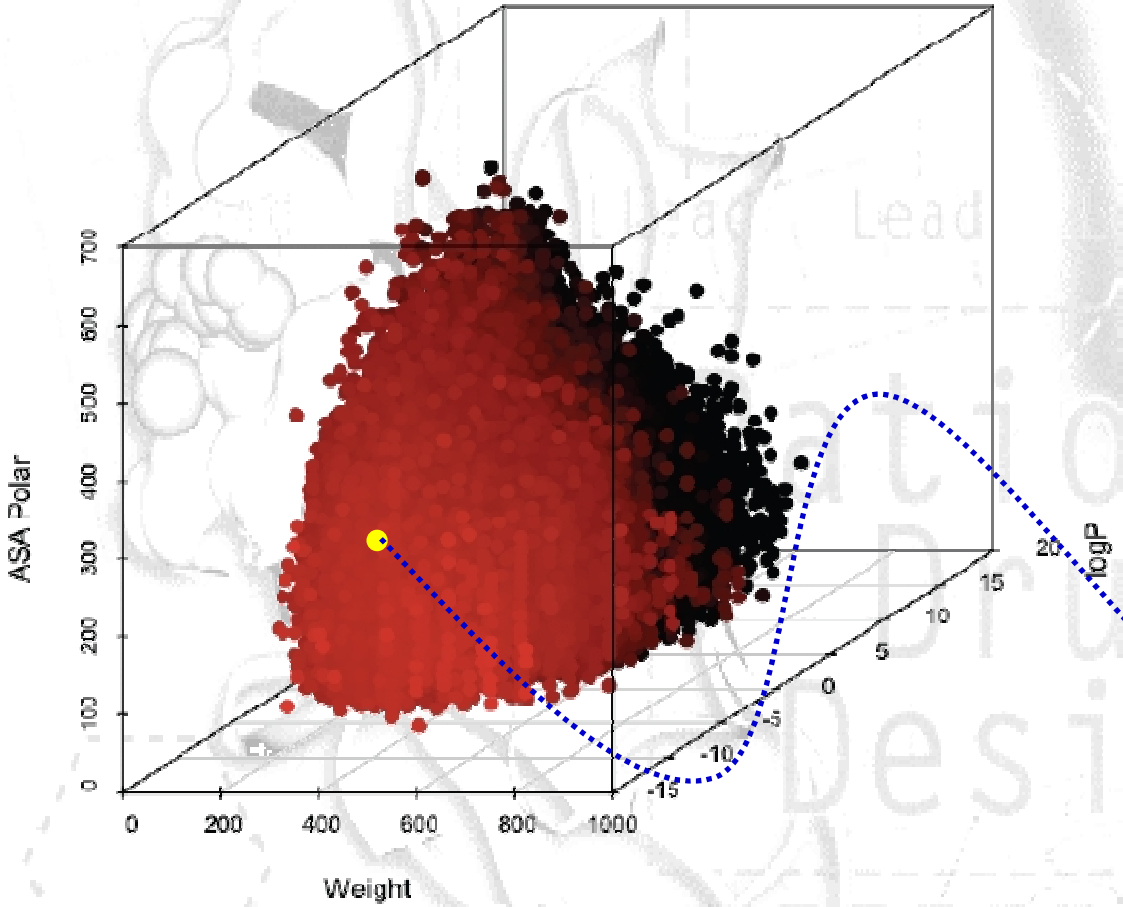
PubChem http://pubchem.ncbi.nlm.nih.gov/

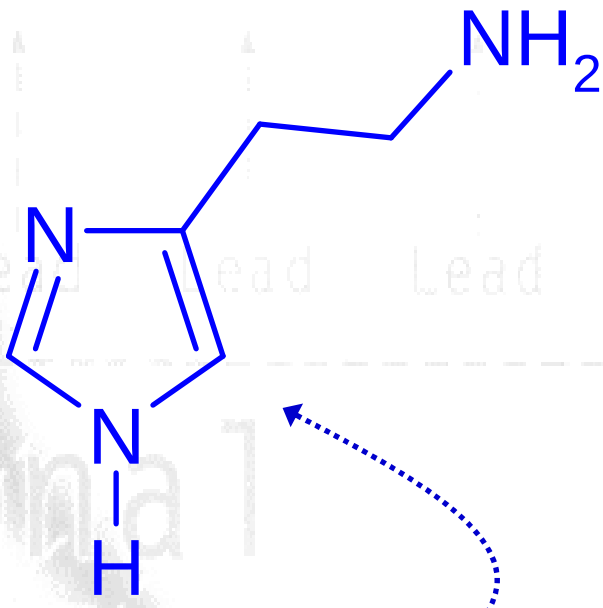
eMolecules http://www.emolecules.com/

ChemSpider Building community for chemists http://www.chemspider.com/

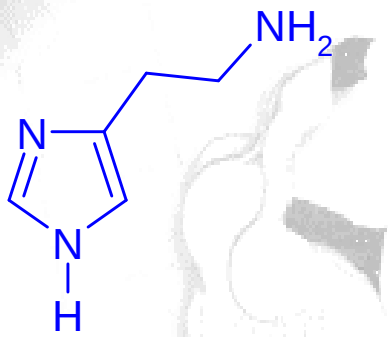
MM.INC http://mms.dsfarm.unipd.it/MMsINC.html

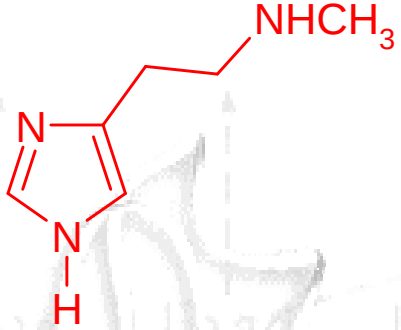
If you have only ONE active compound, could you suggest a possible strategy to fish out novel active analogs?

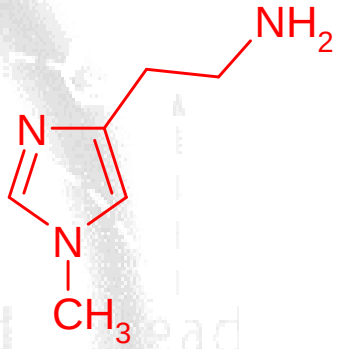


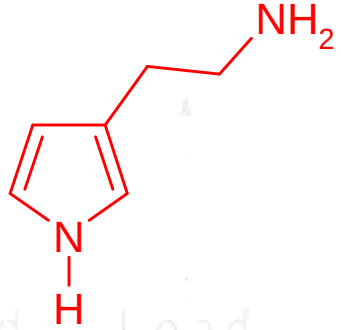


Probably using your chemical nose…

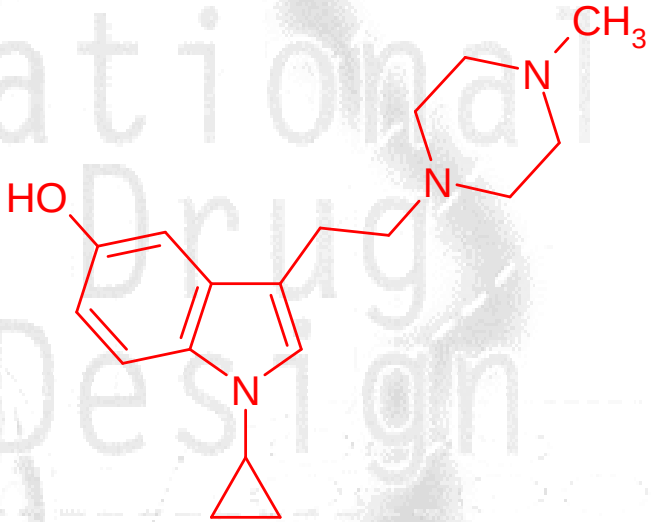




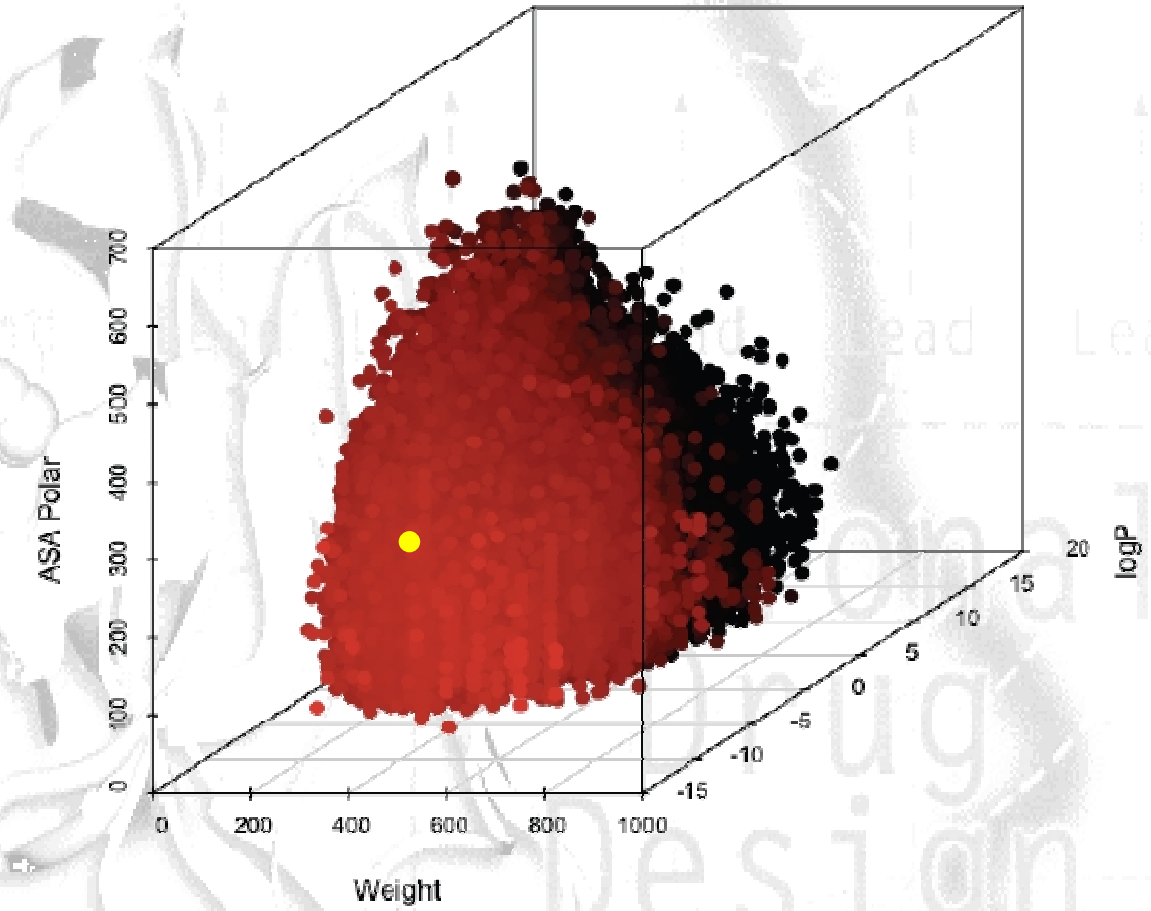




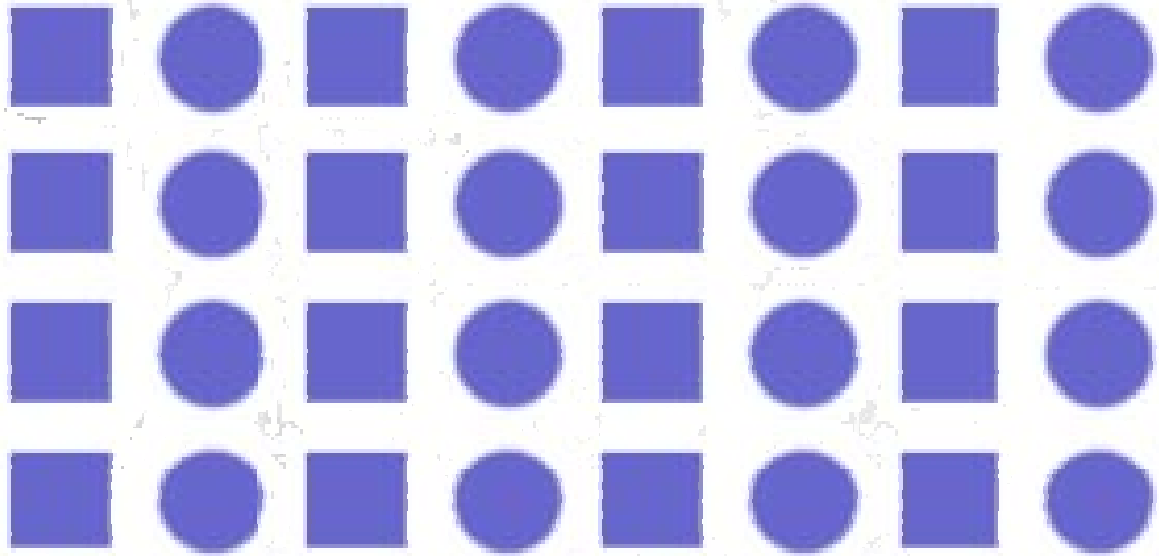
… reaching also more complex hypothesis:



or more smartly?

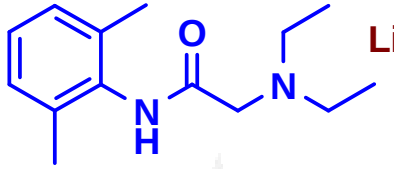


A simple exercise: could you tell me what you are looking?

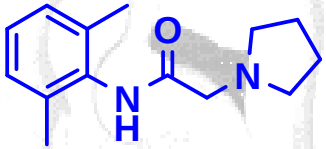


Similarity law: similar objects are typically group together.

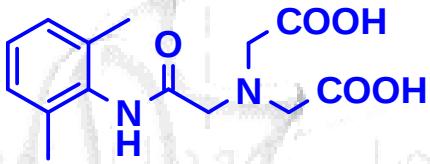
an, sometimes, this is true also in medchem:



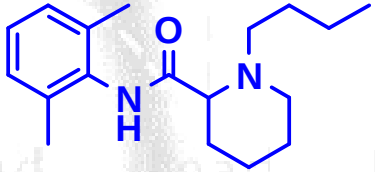
Lidocaine



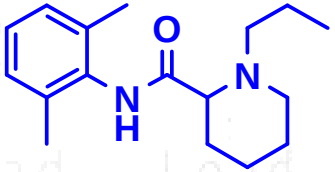
Pyrrocaine



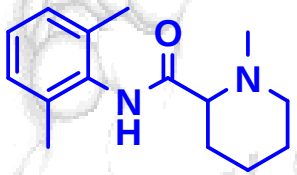
Lidofenin



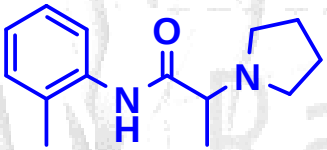
Bupivacaine



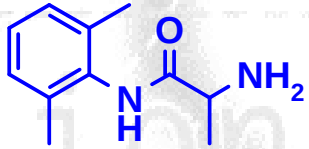
Ropivacaine



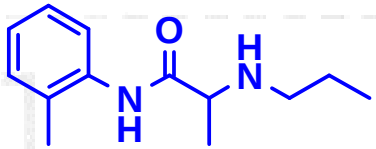
Mepivacaine



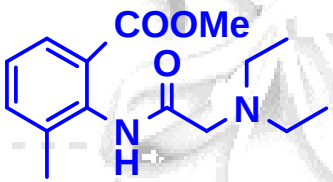
Aptocaine



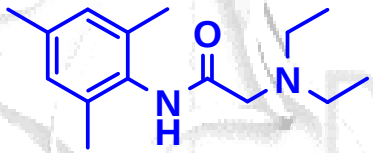
Tocainide



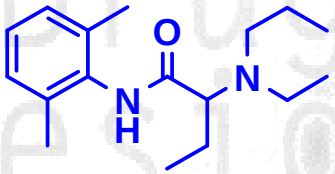
Prilocaine



Tolycaine

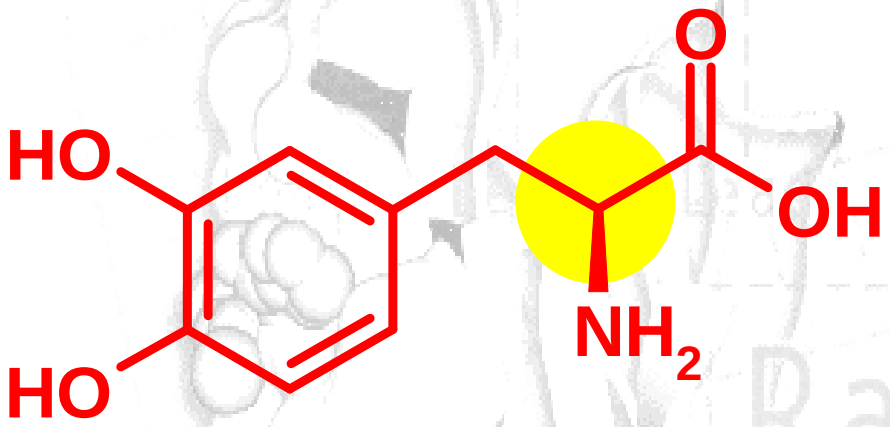


Trimecaine

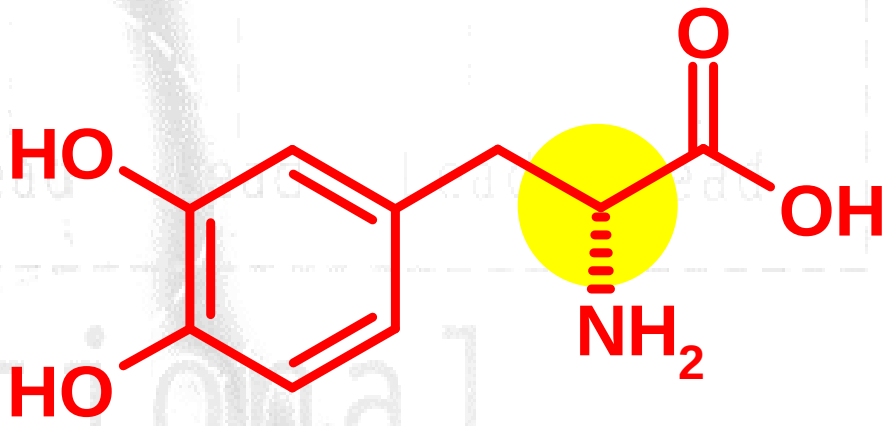


Etidocaine

Even if we MUST remember the "similarity paradox":



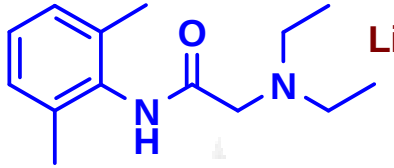
L-DOPA



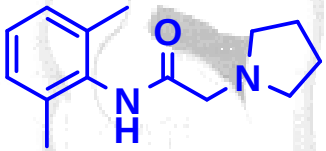
D-DOPA

If we want to virtualize *chemical similarity* searching, we have to describe exactly what similarity is!

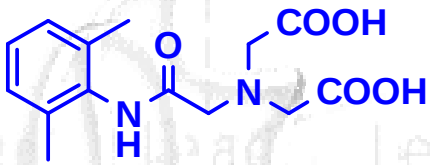
Are you able to scale *chemical similarity* of these analogs respect lidocaine?



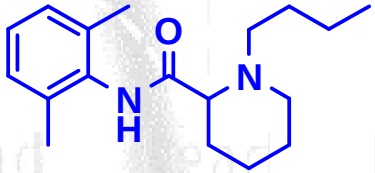
Lidocaine



Pyrrocaine

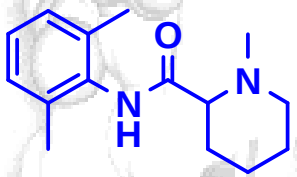


Lidofenin

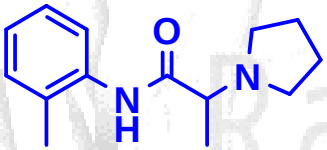


Bupivacaine

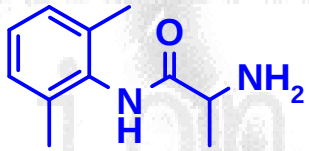
Ropivacaine



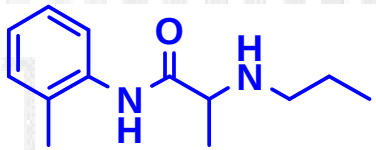
Mepivacaine



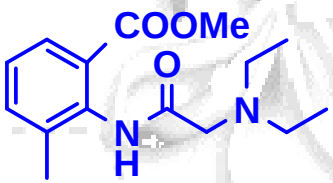
Aptocaine



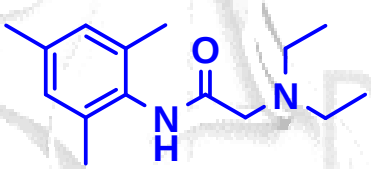
Tocainide



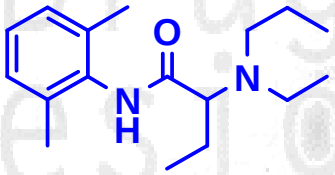
Prilocaine



Tolycaine

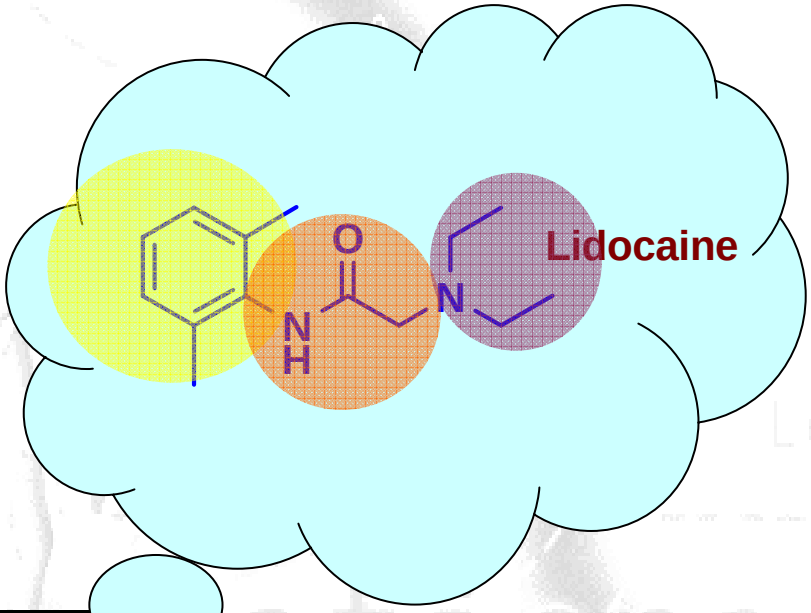


Trimecaine

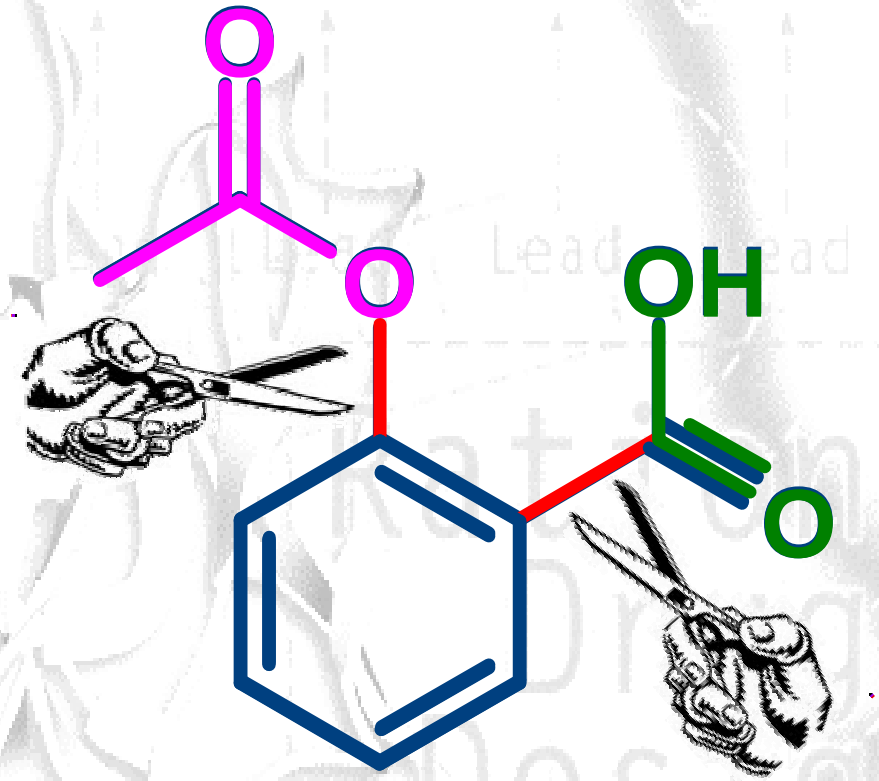


Etidocaine

Tell me what you are thinking:



Exactly "*molecular fragmentation*" could be a possible solution!

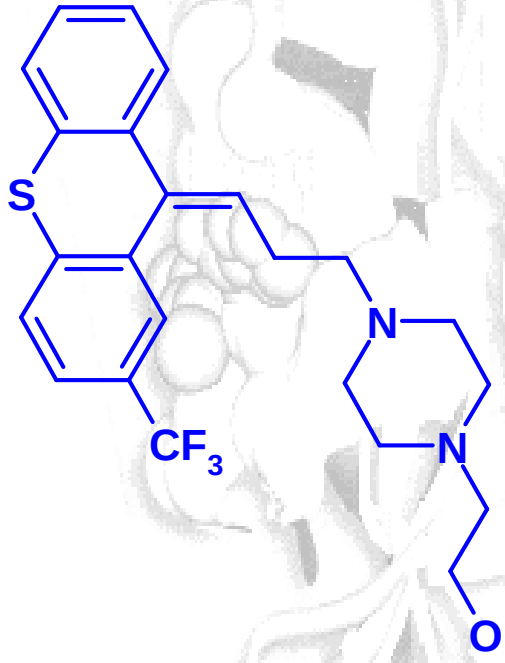


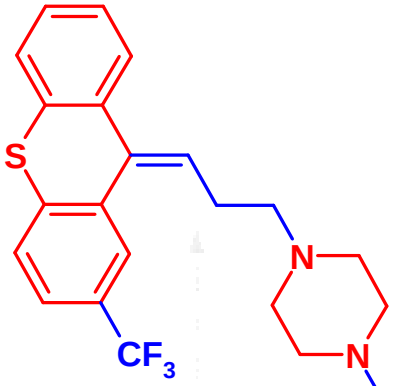
We need some rules:

1:

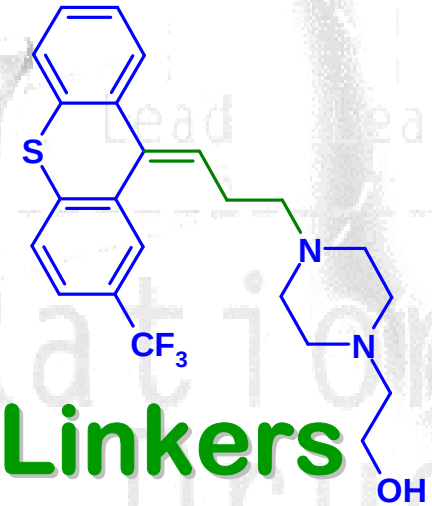
2:

3:

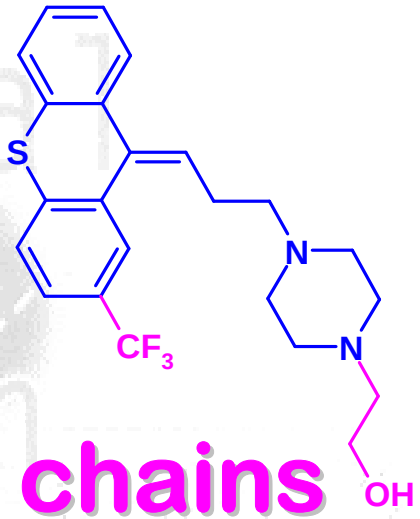




Rings



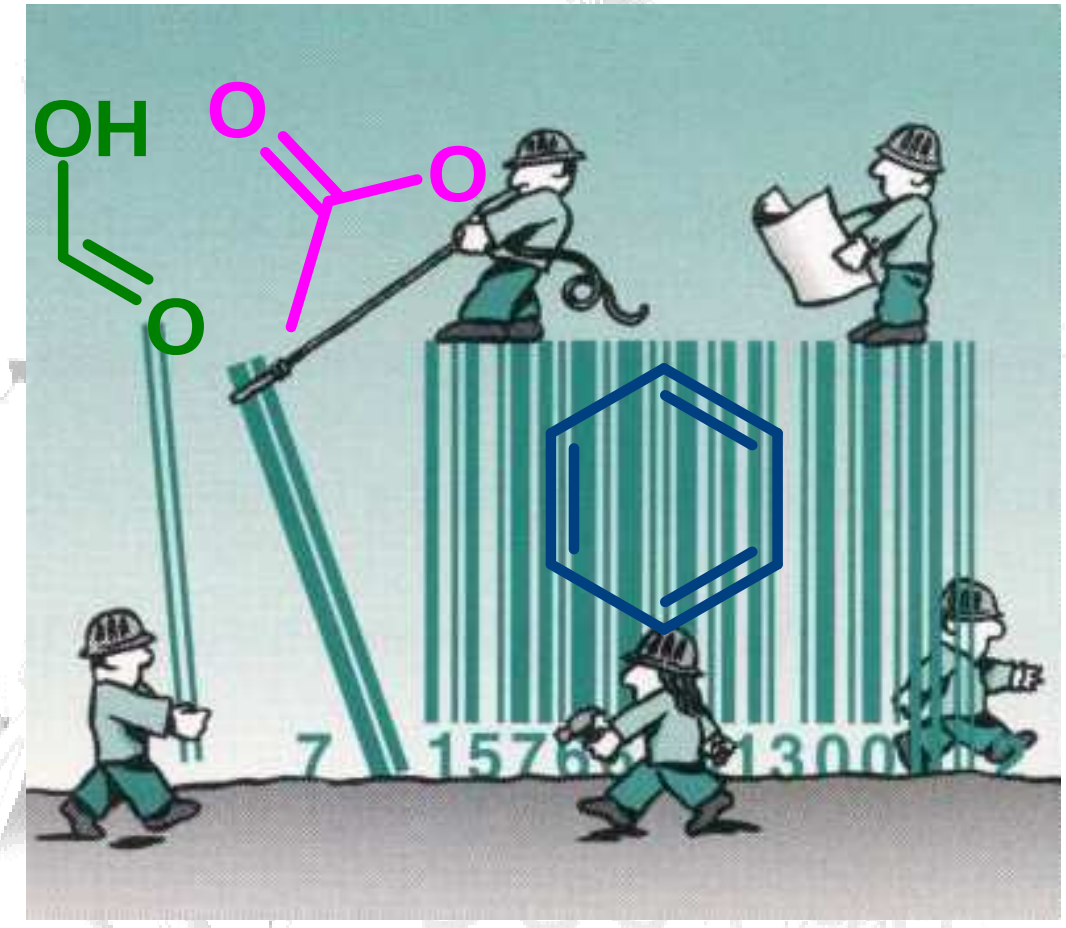
Linkers



Side chains

G.W. Bemis, M. A. Murcko, J. Med. Chem. 1996, 39, 2887 –2893.

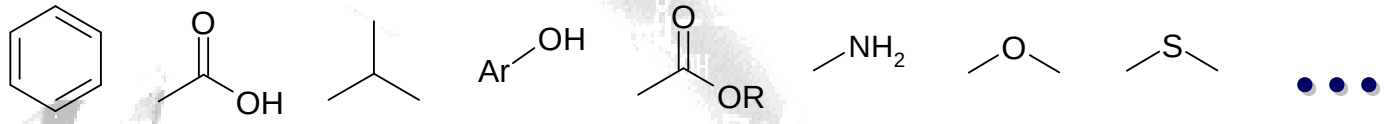
The sequence of fragments is like a molecular bar code…



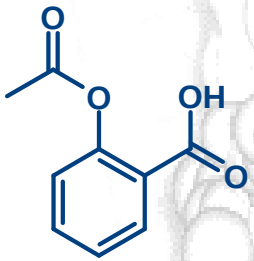
How we can build it up…

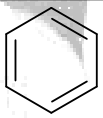
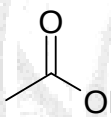
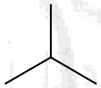
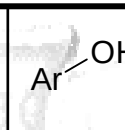
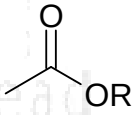
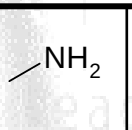
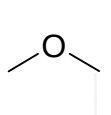
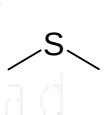
A useful concept: *structural keys*

1. Define all possible chemical fragments (*structural keys*):

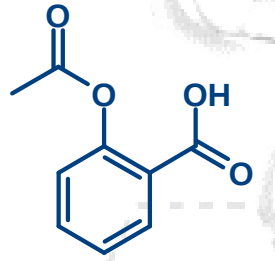


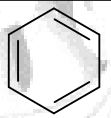
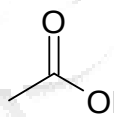
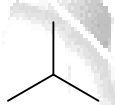
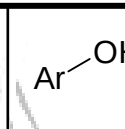
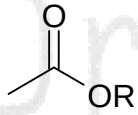
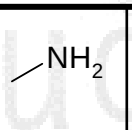
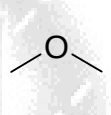
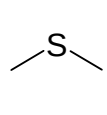
2. Assigne un *bit* for the each corrispondence in our molecule:

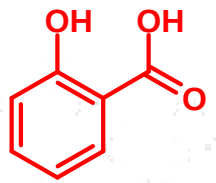


							
1	1	0	0	1	0	0	0

3. Reiterate this procedure for each molecule of the database :



							
1	1	0	0	1	0	0	0
1	1	0	1	0	0	0	0

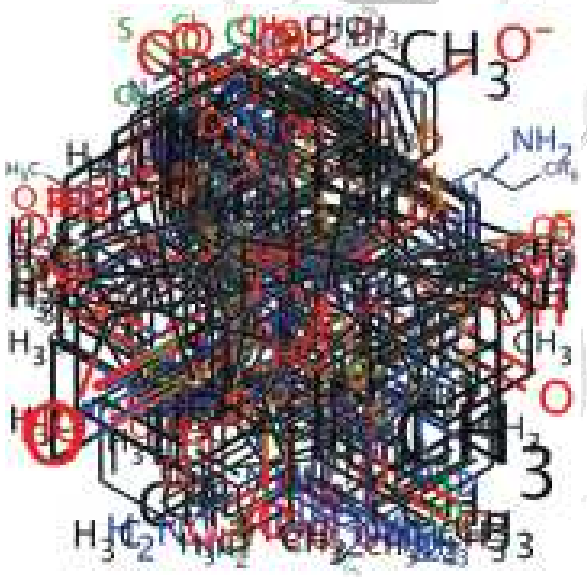


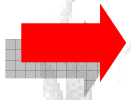
Any standard?

MACCS (Molecular ACCess System) a collection of 166 functional groups representing the most accessible medchem chemical space.

MACCS keys, MDL Information Systems Inc., San Leandro, CA.

… so, now it is very easy to transform my molecular database in a collection of "*structural keys*"!

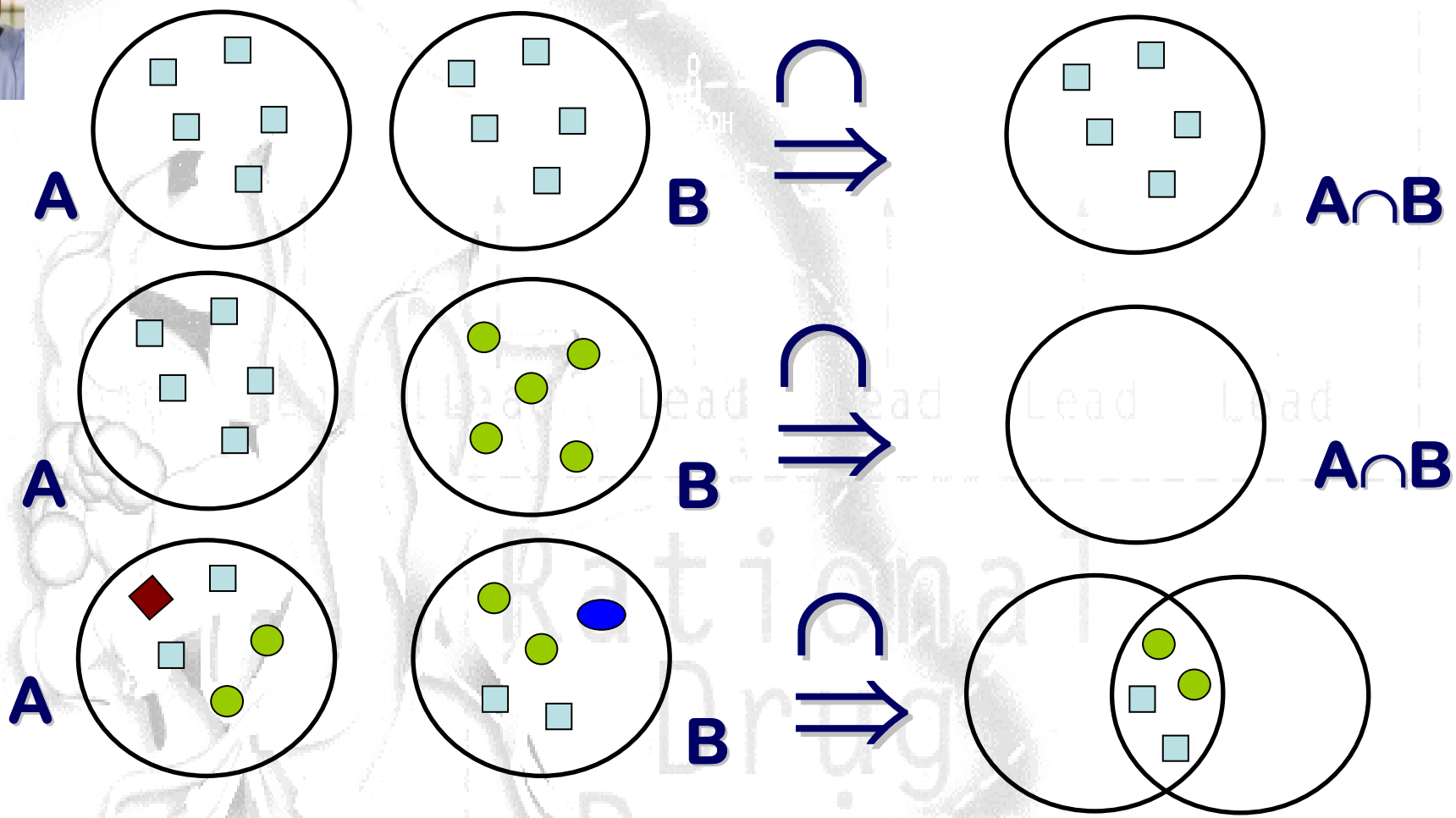




00000001000011010000001010100000000001100000100001000010000010000
01000101100100100101100110100111001111010000001100000001100010000
01000101000111010100001100001010000100110000101000000001001000000
00011011100111011111101000001000100001101101100000000100110100000
01000101001101000100000000100000000010010000000100100001000101000
01000111000111010001000010111011001101100100100011010011000010000
01011101001101010101111110000100000111111000100001000010001010000
01000101001111010100001000100000000100100000101001000010001010000
00010001000101000101001000000000000010100000100001000001000000000
01000101000100110000000000000000000010100000010000000000000000000
01000101000101000000000000001010000100100000000001000000000000000
01010101011111001111101000000000001101010001110010000110010100000
01000101000110000100000110000000000100010000001100000000001100000
00000001000000000100001000000000000001010100000000100000100100000
01000101000101000000001000000000000010000000000000100001000011000
00010001000011000100101000000101001010111000100001000010001010000
01000111000101000100001000010011100100100000100011000000000101000
01010101000101000101001000000000000010010000010010100100100010000

… well, but why this transformation cuold help us in measuring chemical similarity?

Do you remember the Eulero-Venn diagrams?

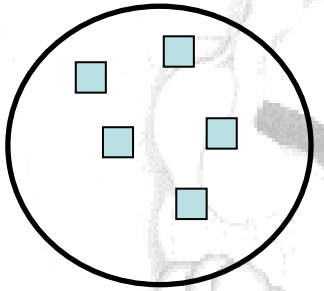


… and the intersection set A∩B?

Now, I can introduce you the Jaccard e Tanimoto index!

$$\text{Jaccard o Tanimoto Index} = \frac{A \cap B}{A_{solo} + B_{solo} + A \cap B}$$

Tanimoto, T.T. (1957) IBM Internal Report 17th Nov see also ***Jaccard, P. (1901) Bulletin del la Société Vaudoisedes Sciences Naturelles 37, 241-272*** .

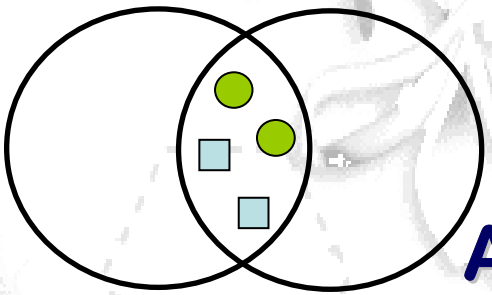


A∩B

= 1

A∩B

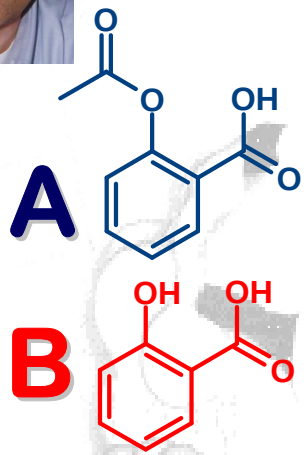
= 0

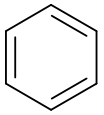
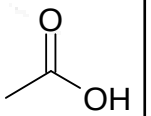
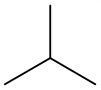
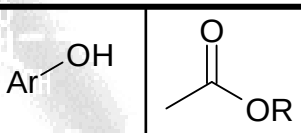
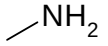
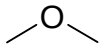
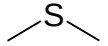


A∩B

$$= \frac{4}{1+1+4} = 0.66$$

Eureka… now it's really easy!



							
1	1	0	0	1	0	0	0
1	1	0	1	0	0	0	0

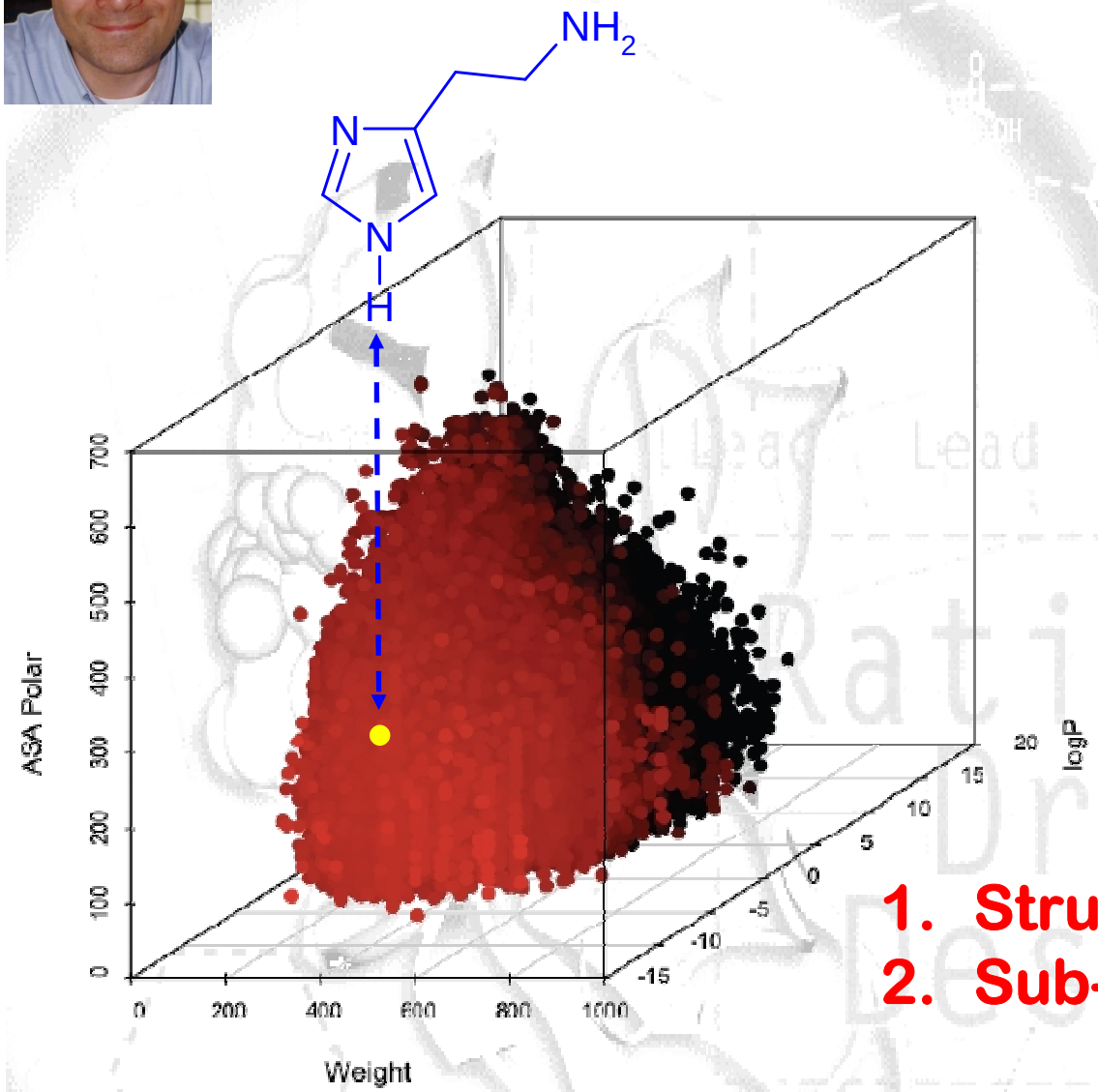
Fragments only in A = 1

Fragments only in B = 1

Fragments in A∩B = 2

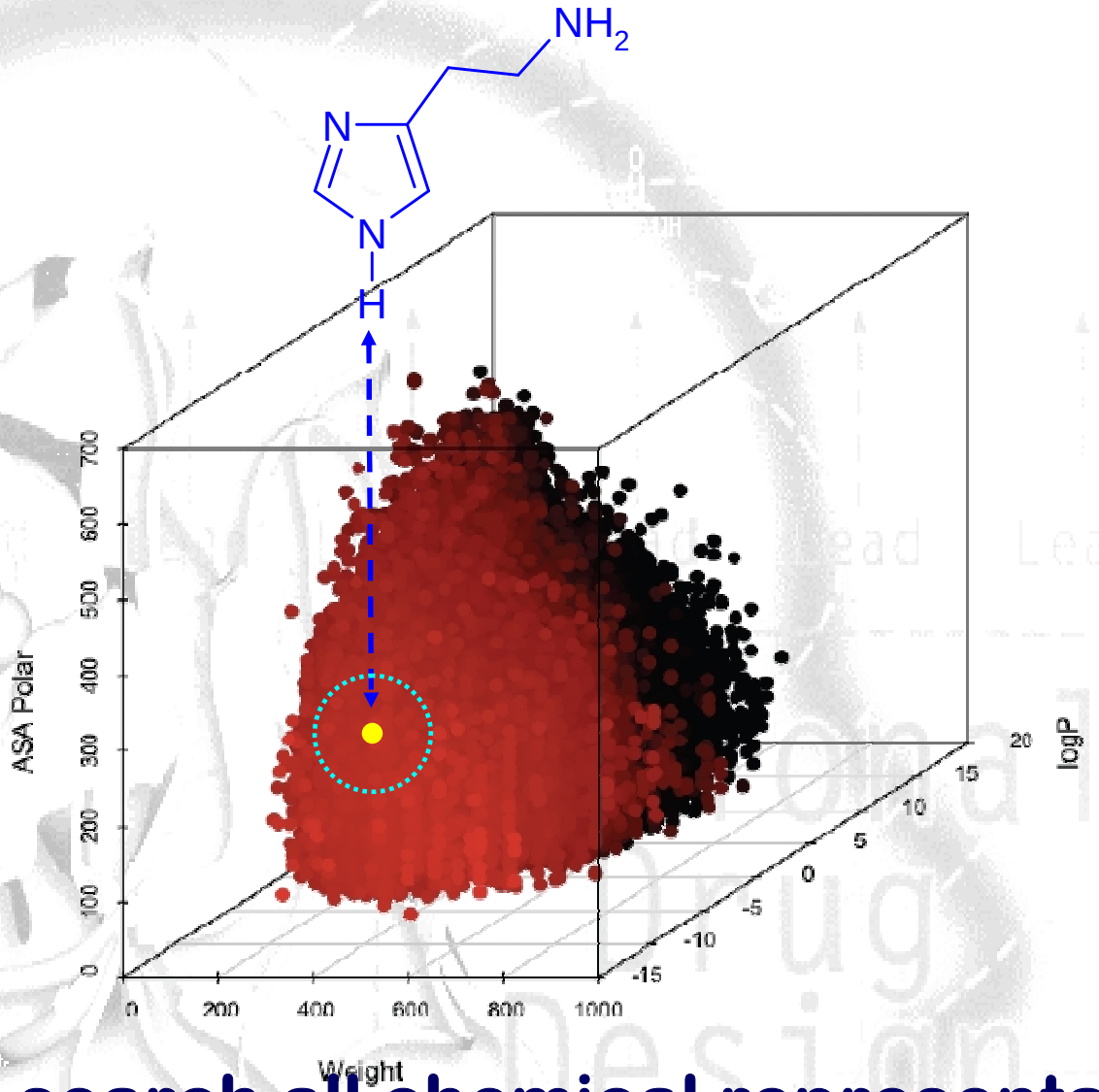
Tanimoto similarity index $= \frac{2}{1+1+2} = 0.5$

Now we can smartly explore our molecular haystack!



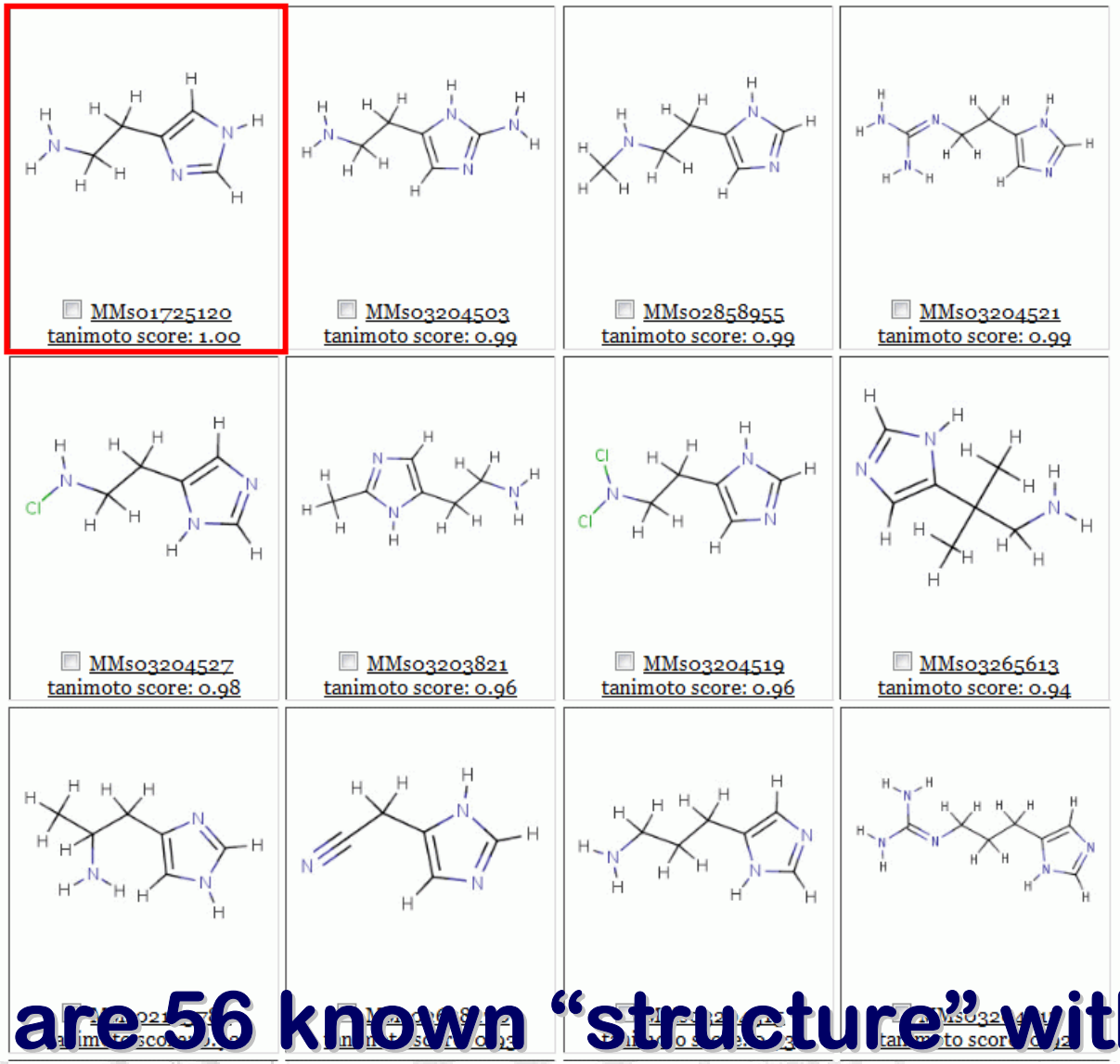
1. Structure similarity search;
2. Sub-structure search.

Here the first example:



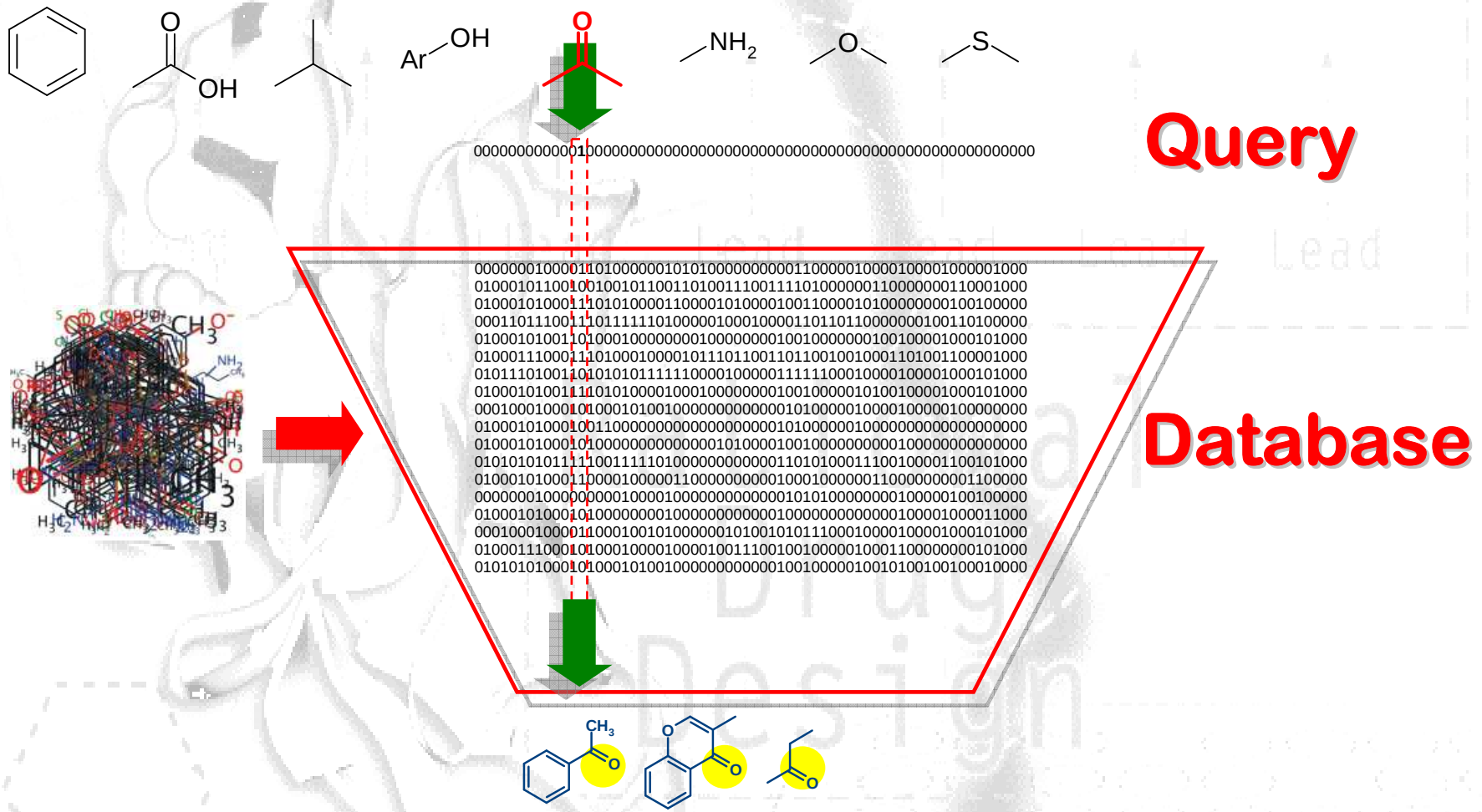
Please, search all chemical representations with Tanimoto index > 0.85?

and look a possible result:

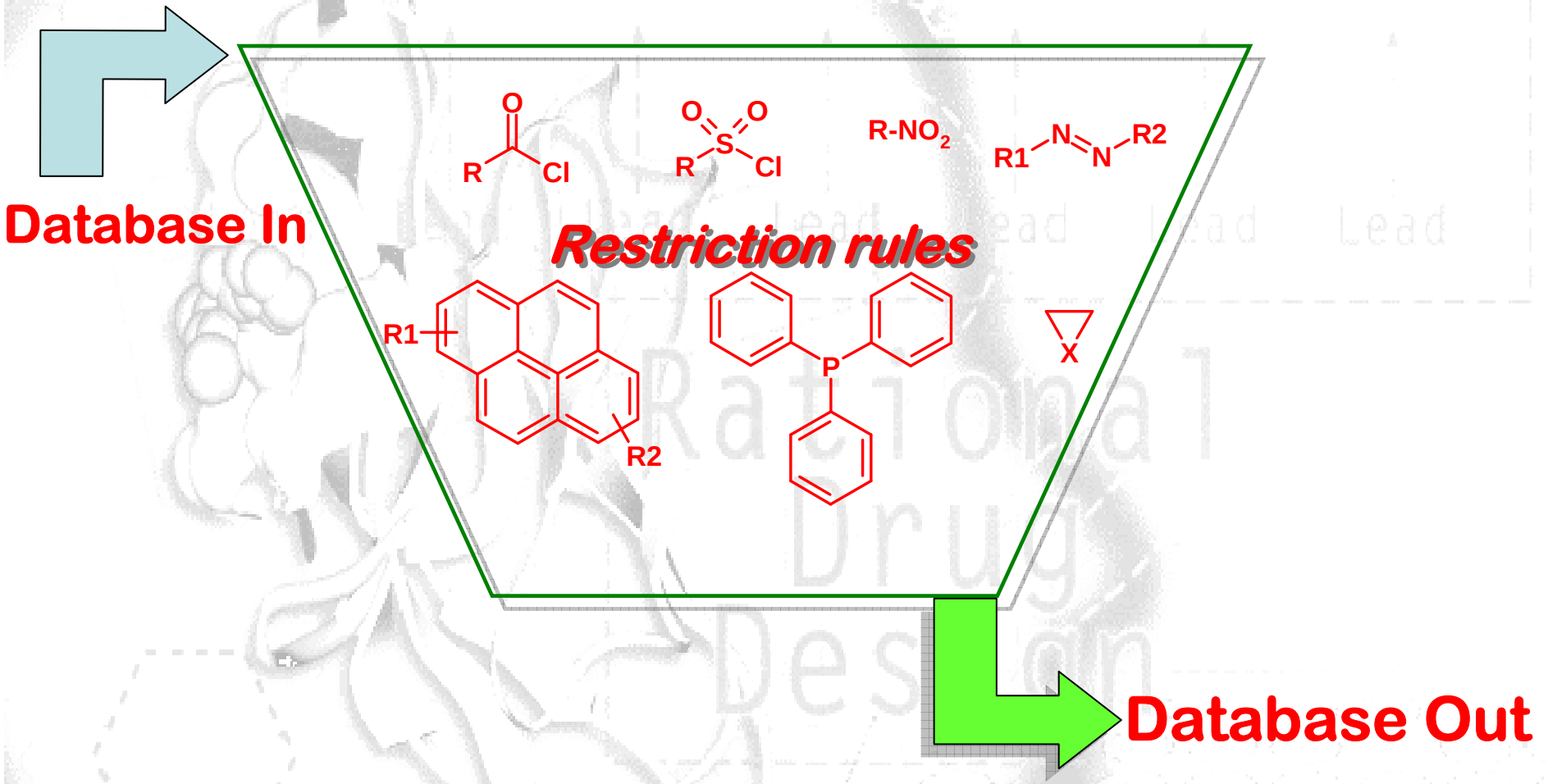


There are 56 known "structure" with a Tanimoto similarity index> 0.85!

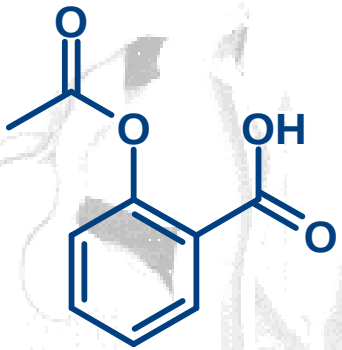
... but we can also run a sub-structure search!

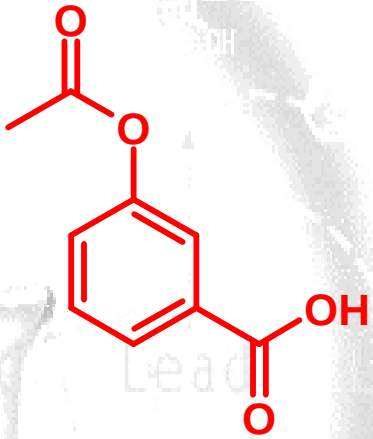


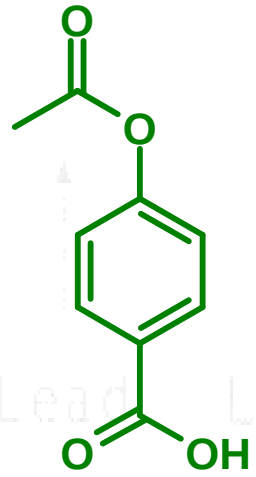
.... or we can easily remove compound with undesirable fragments (chemical or biological reactivity, toxicity etc.)!

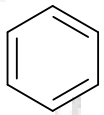
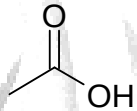
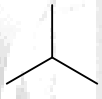
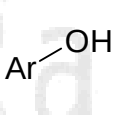
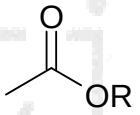
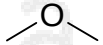
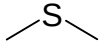


… as for any tool, there is a limitation!

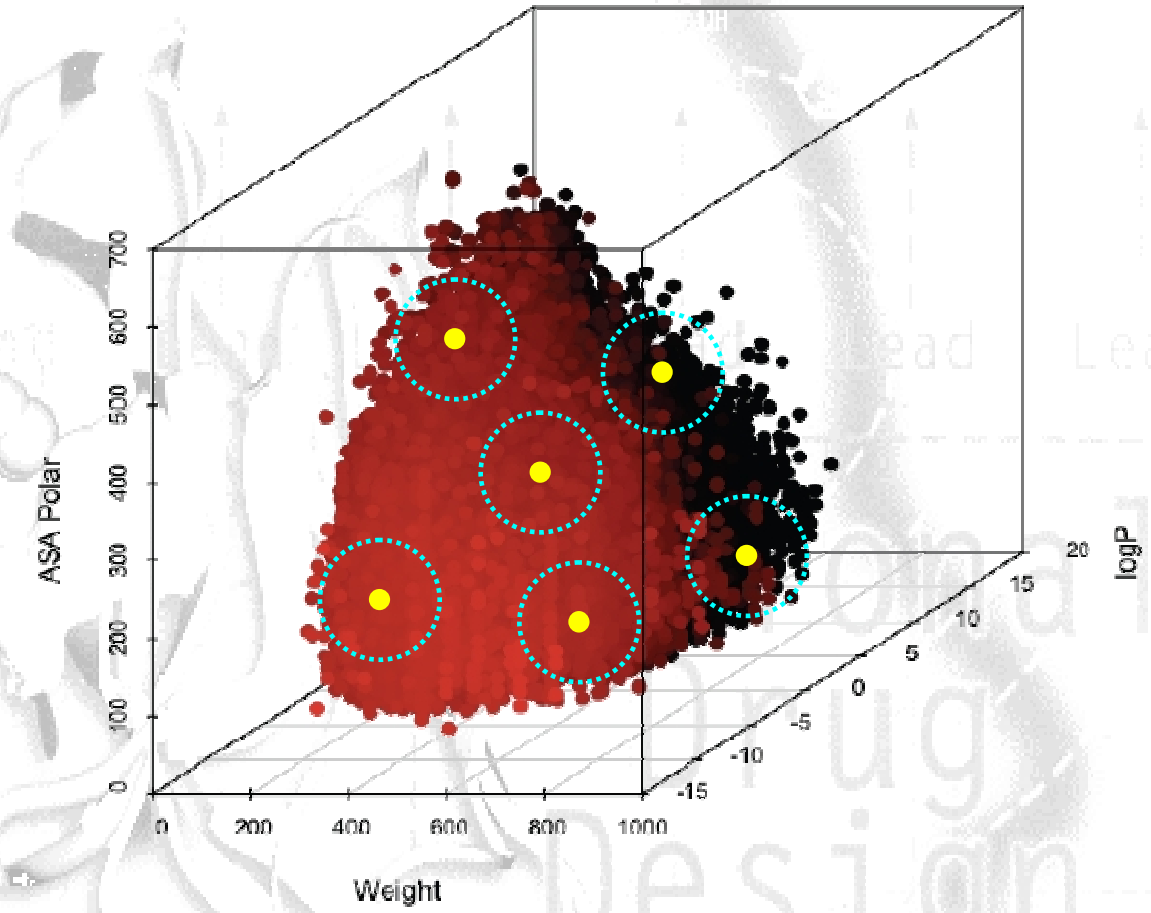






							
1	1	0	0	1	0	0	0
1	1	0	0	1	0	0	0
1	1	0	0	1	0	0	0

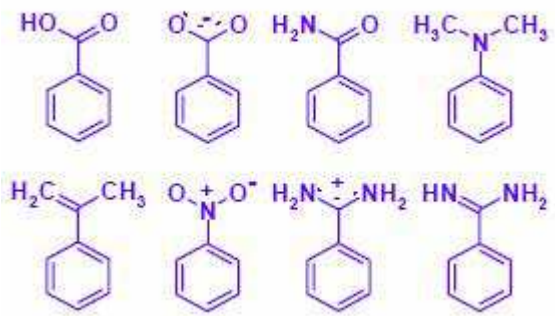
And don't forget:



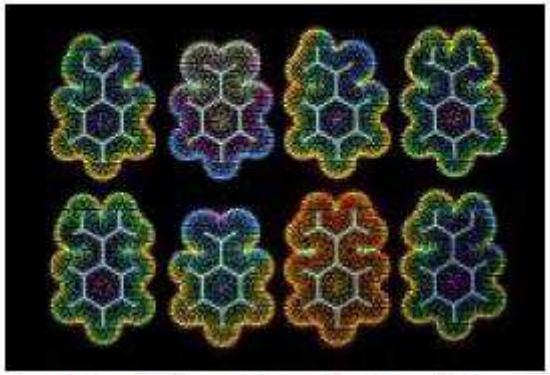
1- similarity = diversity

Structurally similar compounds can have very different 3D chemical properties!

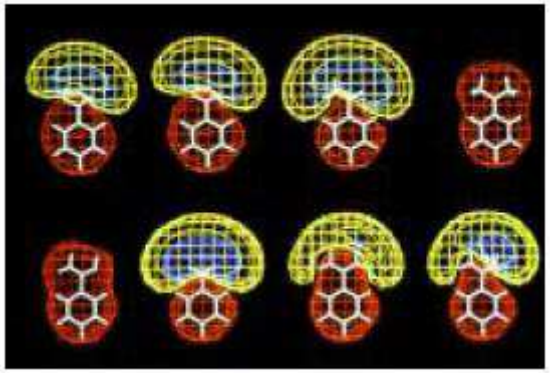
Similarity and Diversity



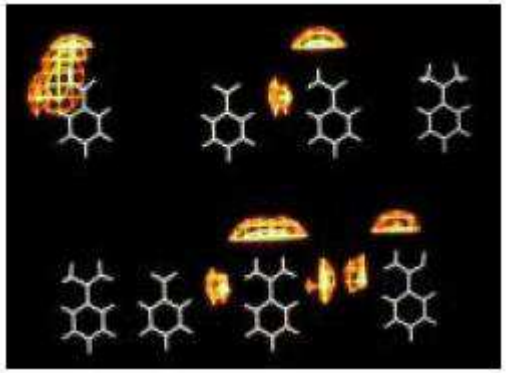
Volumes and Surface Potentials



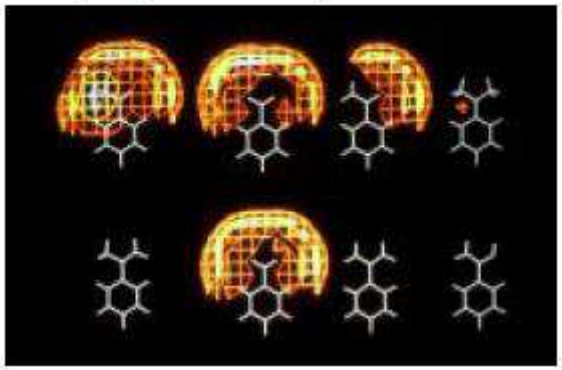
Hydrophobic and Polar Regions



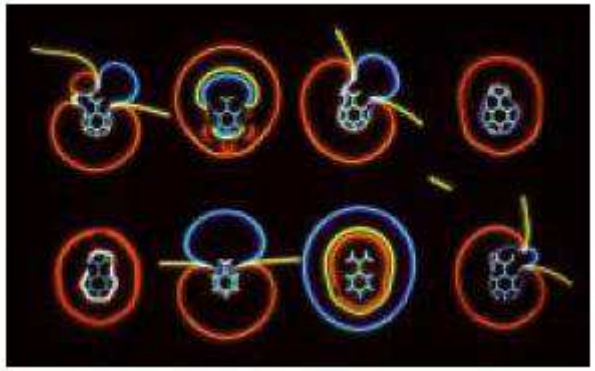
Hydrogen Bond Donor Potentials



Hydrogen Bond Acceptor Potentials



Molecular Electrostatic Potentials (MEP)



Kubinyi, H., Chemical Similarity and Biological activity

Here we are again:

Virtual
Screening (VS)

Target
Structure Determination

Ligand-based
Virtual Screening (LBVS)

Similarity search
Pharmacophore matching

Structure-based
Virtual Screening (SBVS)

HT Molecular Docking

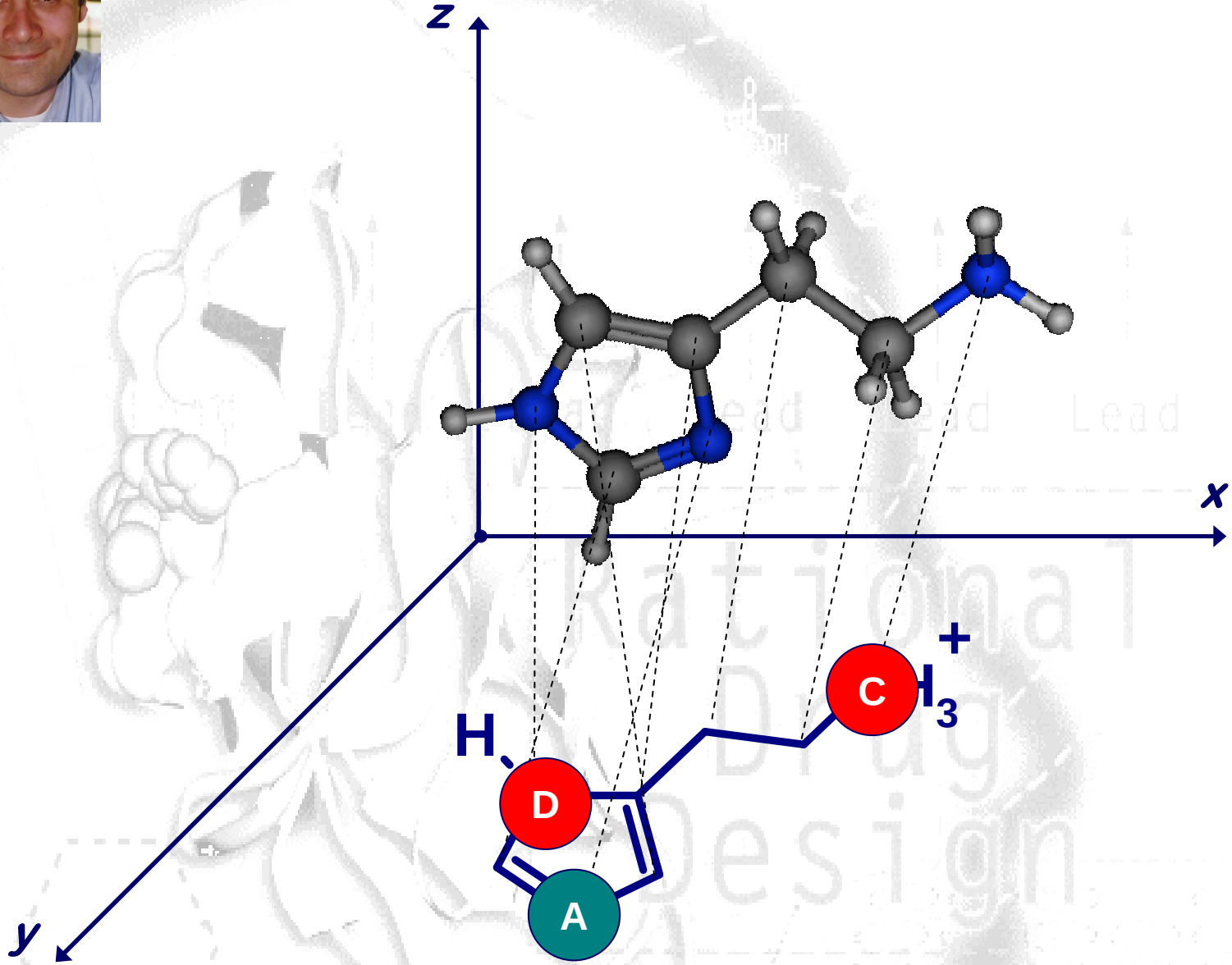
Pharmacophore definition:

A "*pharmacophore*" is a three-dimensional substructure of a molecule that carries ("phoros") the essential features responsible for a drug's ("pharmacon") biological activity. Alternatively described as an ensemble of interactive functional groups with a defined geometry. Basically, one tries to talk the protein language by finding the "structural and chemical complementaries" (pharmacophore hypothesis) to target receptors.

.... a quick refresh: what is the goal of every SAR study?

The generation of pharmacophoric hypothesis (models)!!!!!

the shadow of the reality:



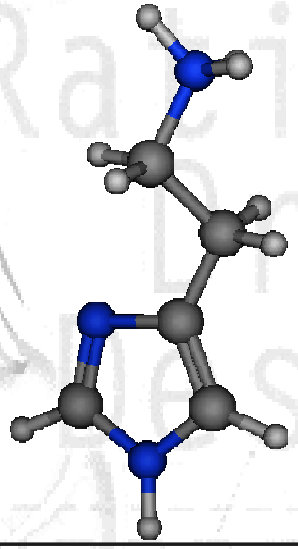
Pharmacophore definition:
1. *conformational selection*

Ligand-Target 3D info

- X-ray info
- Docking

Ligand 3D info

- X-ray info
- Conformational Analysis
- Rigid active analogs

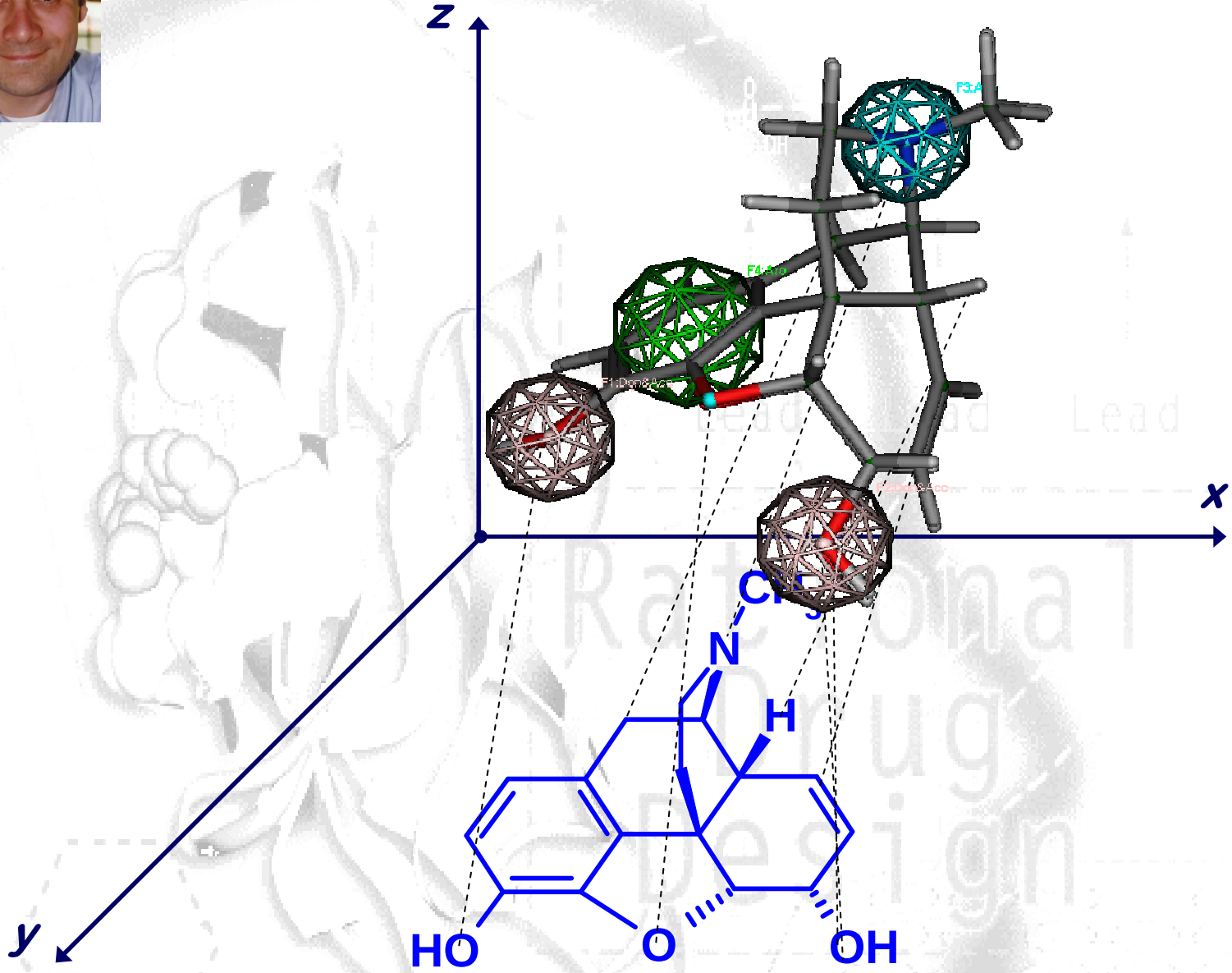


Two very interesting concepts:

Stability as a measure of the geometrical deformability of an object;

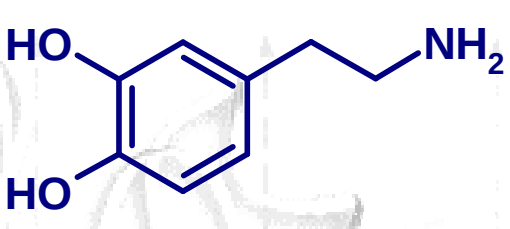
Rigidity as a measure of the reduction degree of the geometrical deformability of an object.

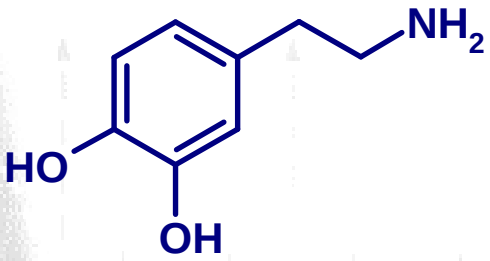
The magic behavior of cyclization:

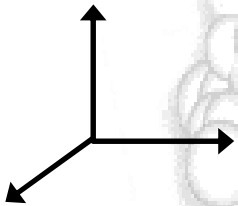


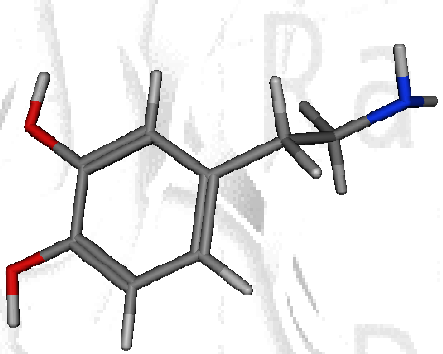
A little memo: “Rigid Active analogs”

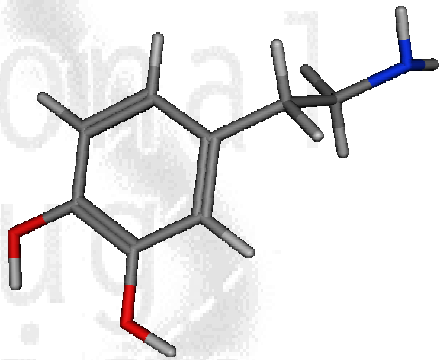
Dopamine and its rigid analogs





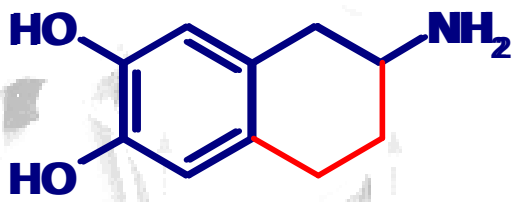


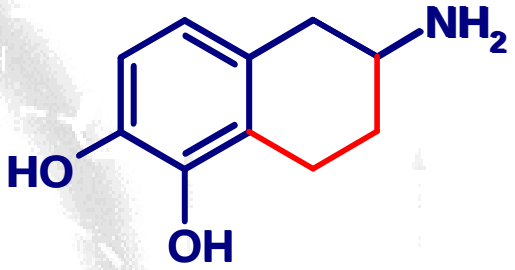


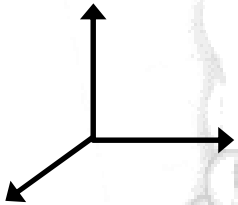


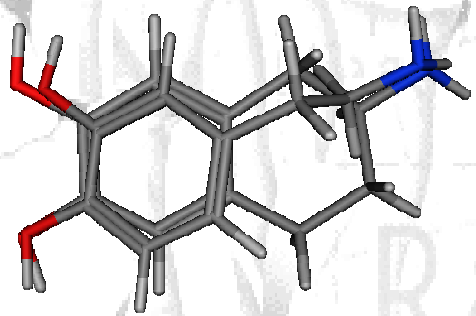
A little memo: “Rigid Active analogs”

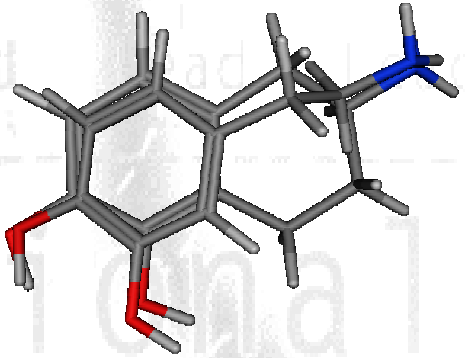
Dopamine and its rigid analogs



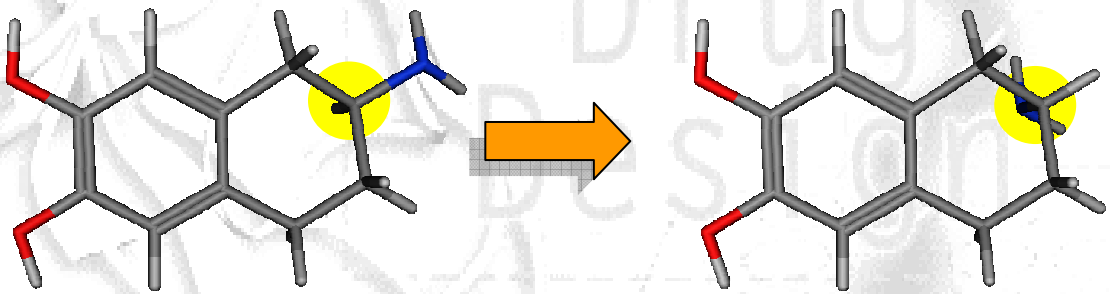




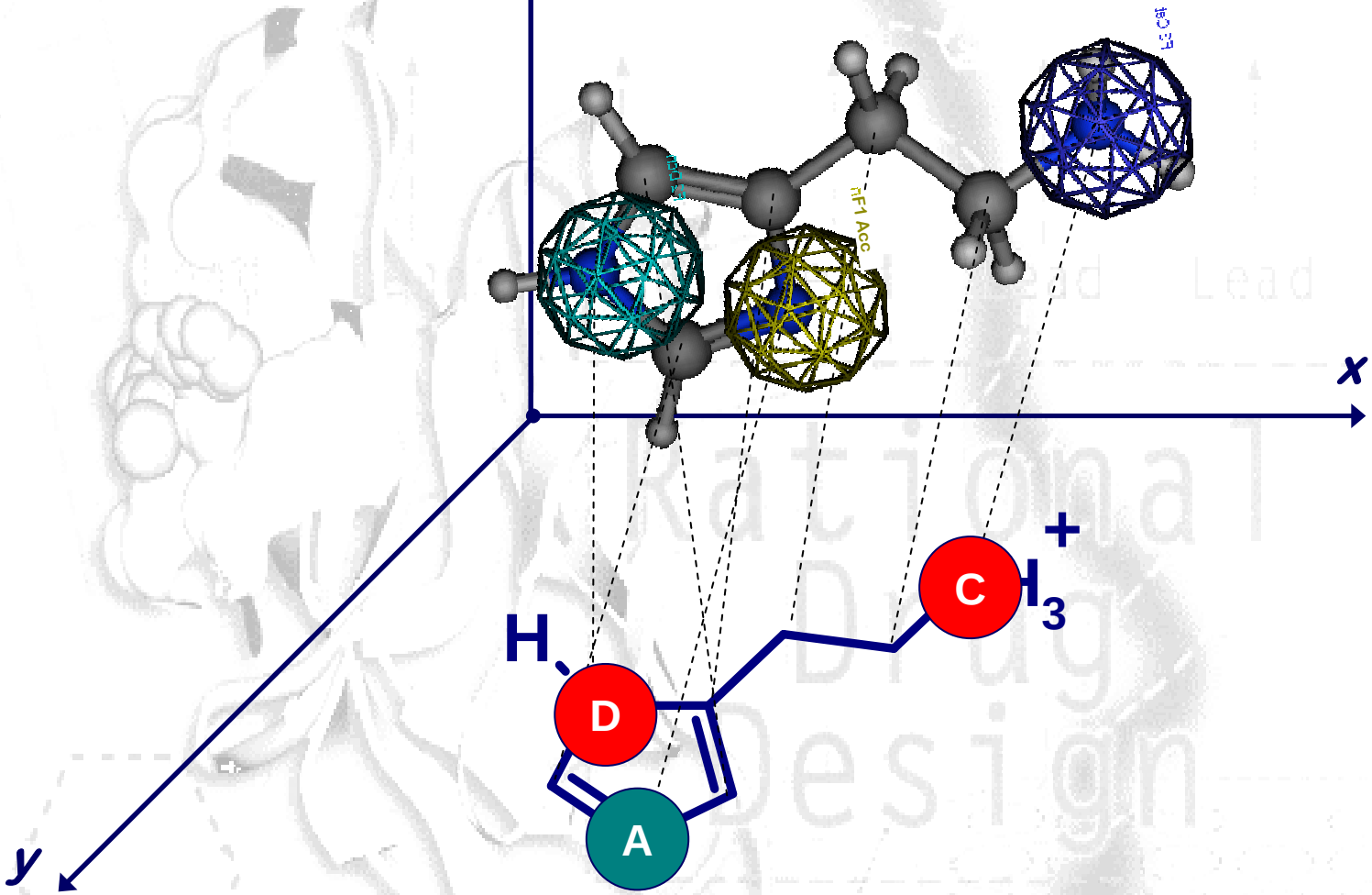




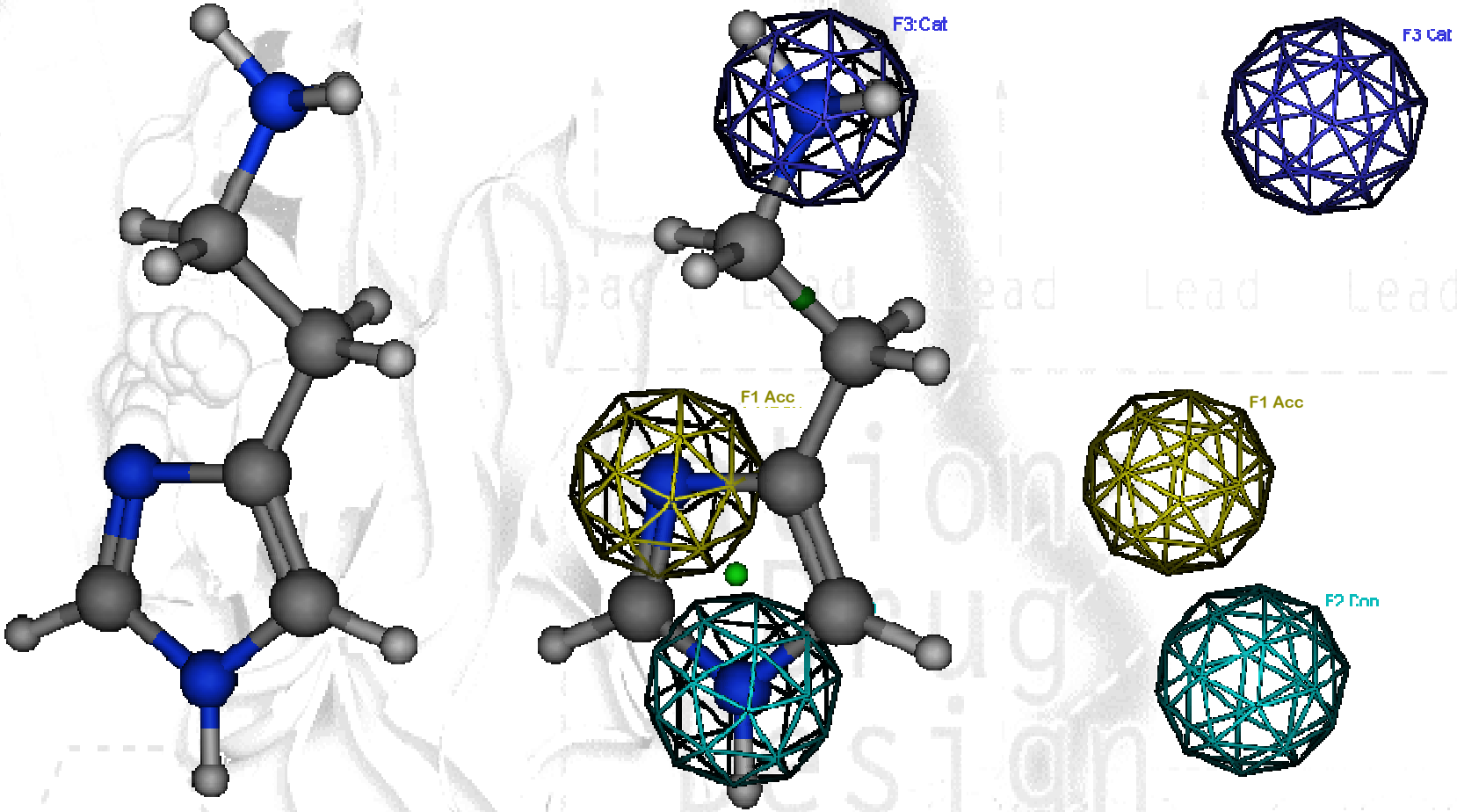
With a little 3D chemical trouble:



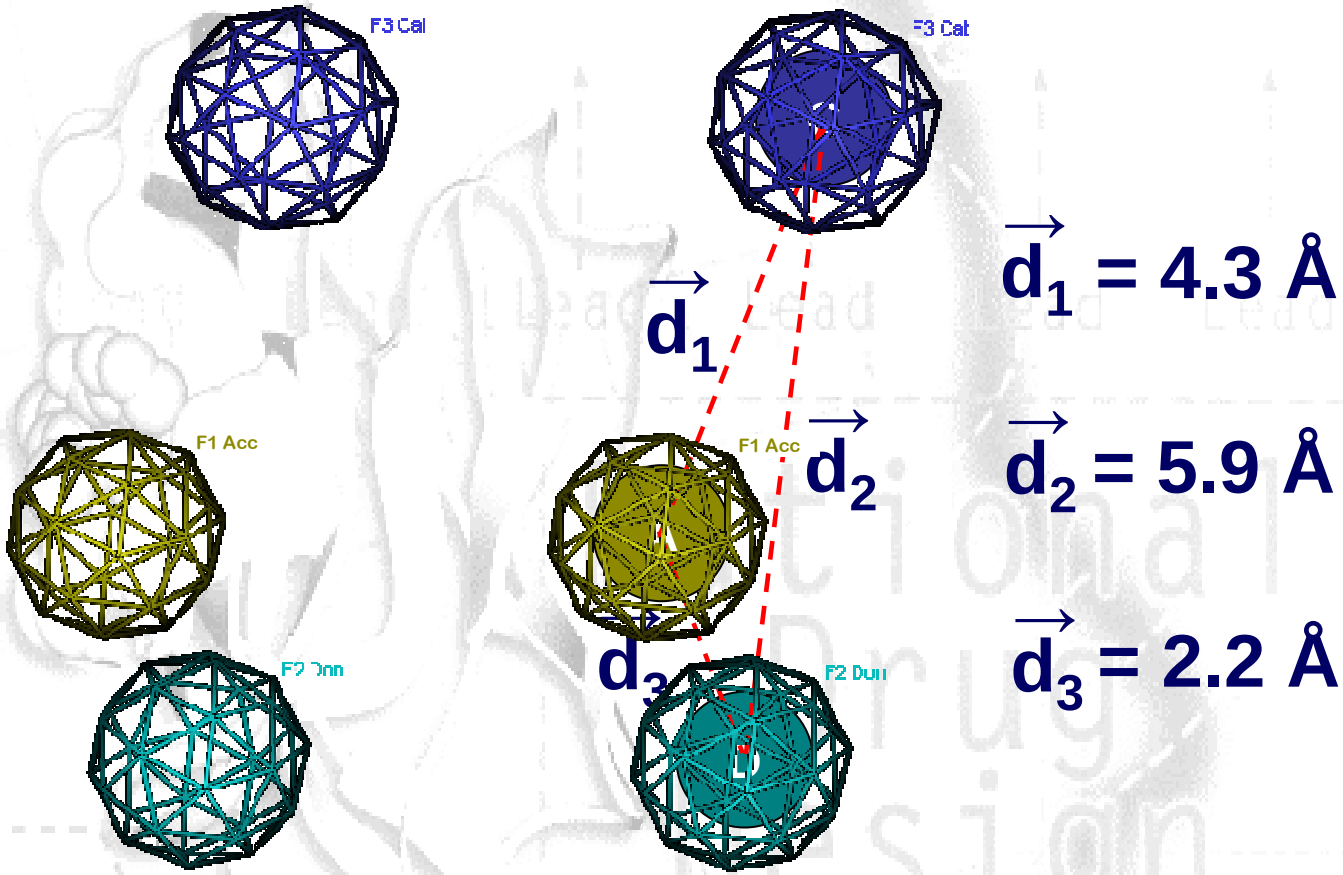
Pharmacophore definition:
2. *Pharmacophoric keys selections*



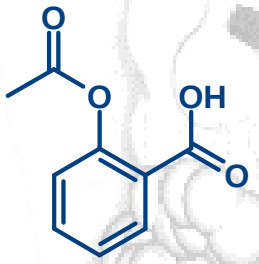
Pharmacophore definition:
2. *Pharmacophoric keys selections*

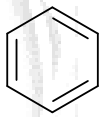
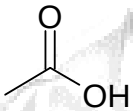
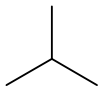
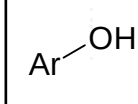
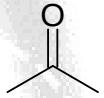
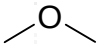
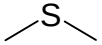


Pharmacophore definition:
3. *Interaction triangle*



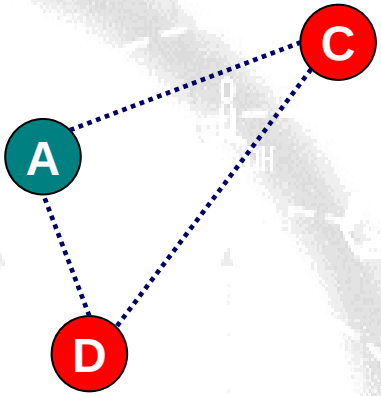
Pharmacophore definition:
4. *From structural key to pharmacophoric key*



							
1	1	0	0	1	0	0	0
Ar	Ac	H	D	A	A	A	H

Ar = aromatic
Ac = acid
H = hydrophobic
D = H-bonding donor
A = H-bonding acceptor
C = cation
An = anion

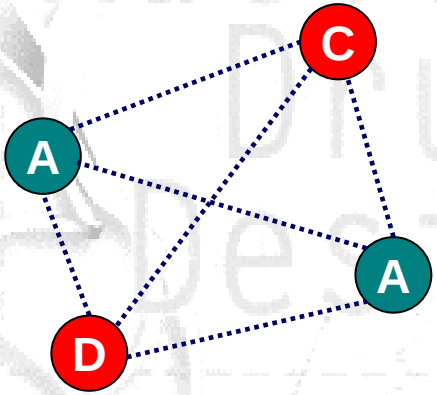
The triangle saga :



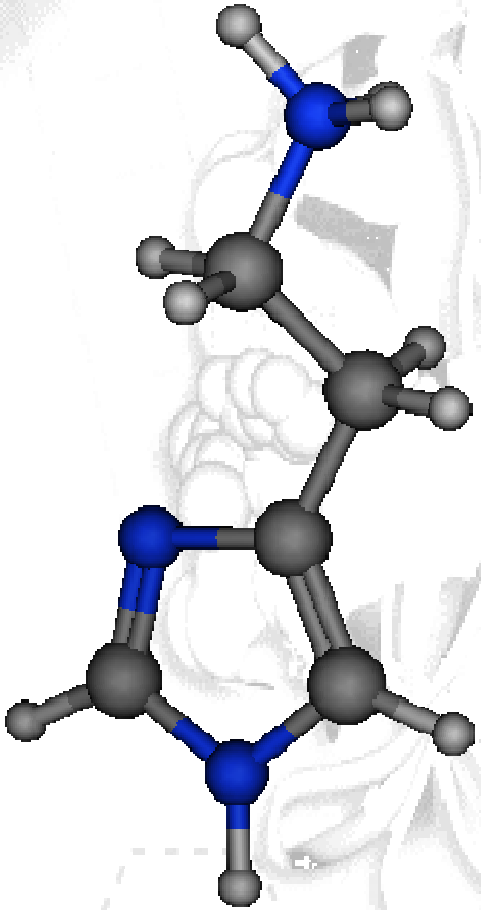
Any pharmacophoric triangle can be described as a three characters string: **ACD**

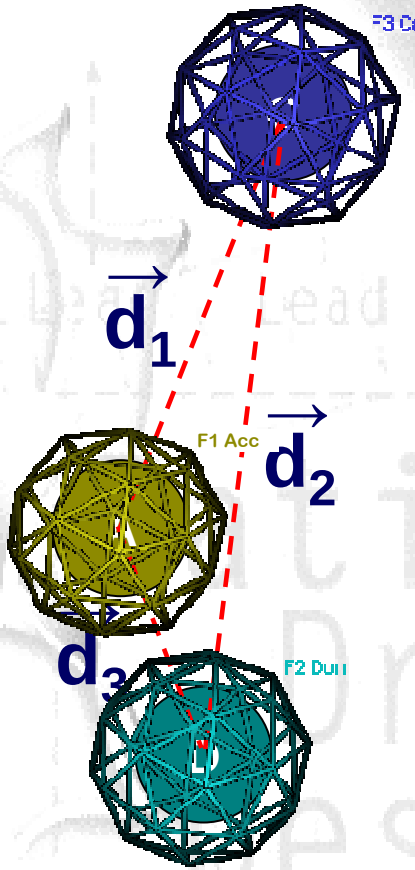
For any triangle we cal calculated numerical descriptors: **perimeter, sides length.**

Any polygons can be subdivided into a sum of triangles:



Here is another interesting 3D→1D chemical representation transformation!

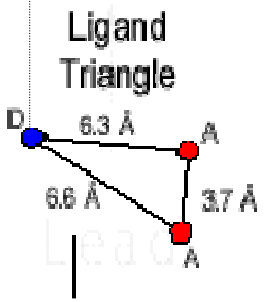


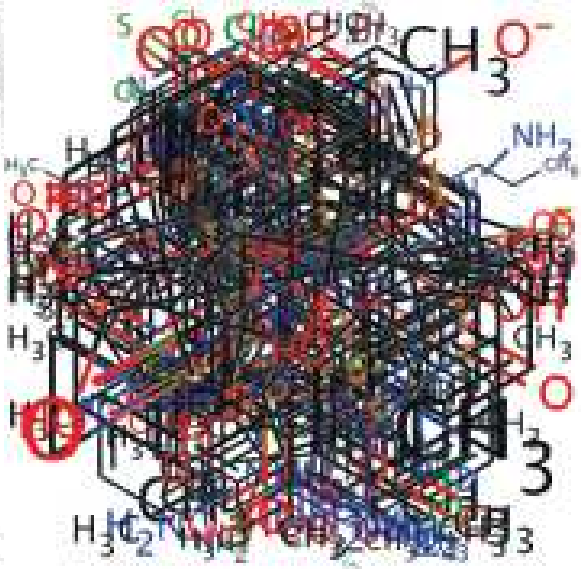


ACD 462 12

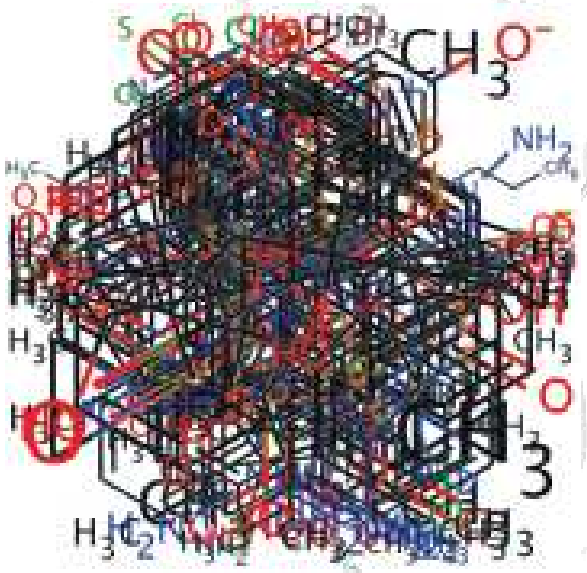
3D 2D 1D

Pharmacophore definition:
5. *Geometric hashing scheme*





What we need for a good fishing?



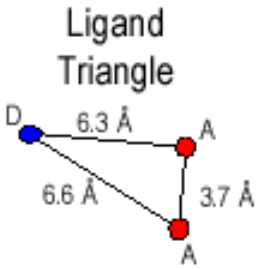
① generation of a collection of good conformers for each chemicals;

② generation of all plausible pharmacophoric hypothesis;

③ generation of all possible pharmacophoric triangles;

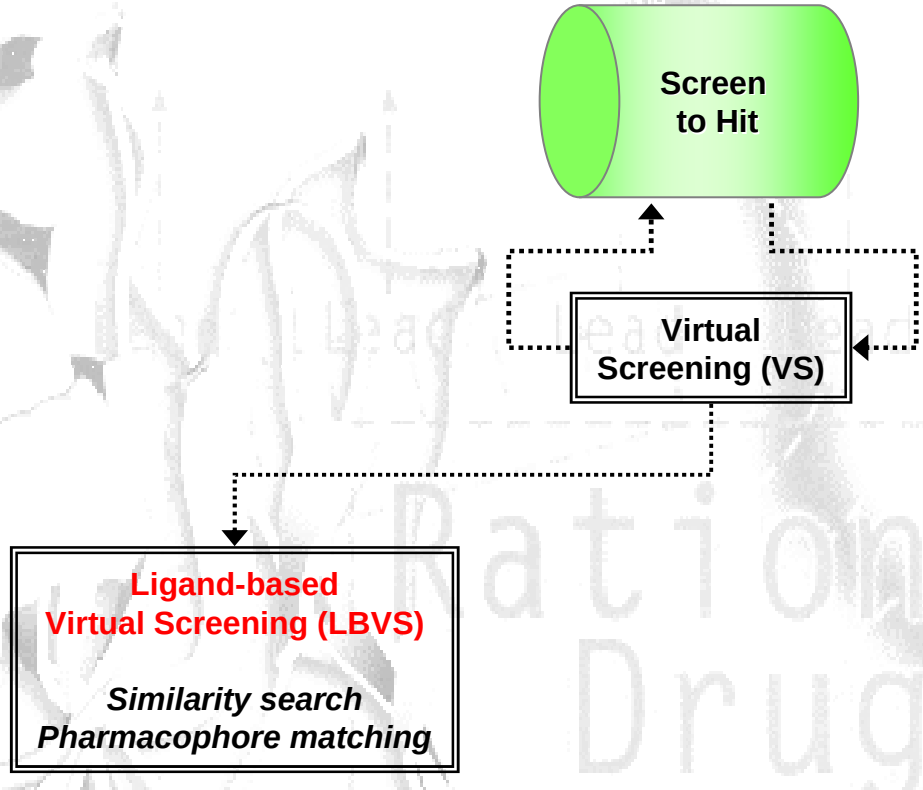
④ generation of all possible pharmacophoric strings.

Pharmacophore definition:
5. *Geometric hashing scheme*

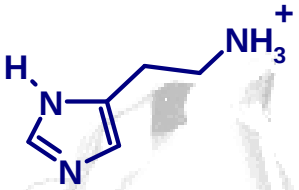


D = HBond Donor
A = HBond Acceptor
H = Hydrophobic Point

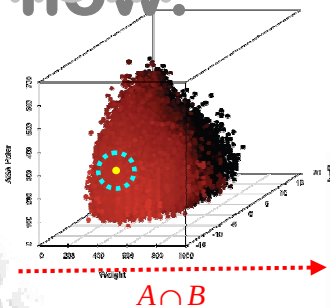
Here is a possible work-flow:



Here is a possible work-flow:



0000111110010101010



$$\frac{A \cap B}{A_{solo} + B_{solo} + A \cap B}$$