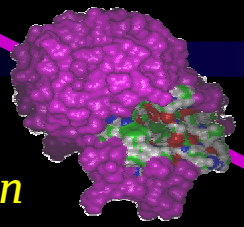
Tools for rational drug design

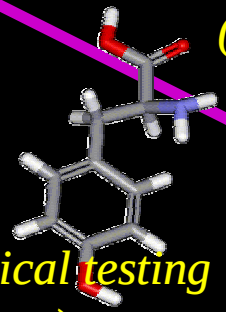
E. Mikros
University of Athens

Drug Discovery & Development

Identify disease



Isolate protein involved in disease (2-5 years)



Preclinical testing (1-3 years)

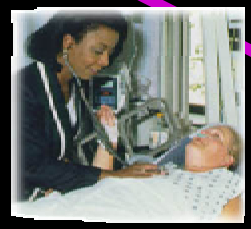
Find a drug effective against disease protein (2-5 years)

Scale-up

File IND

Human clinical trials (2-10 years)

Formulation



File NDA

FDA approval (2-3 years)

- Discovery of new targets
  - **target discovery,**
  - **target identification and validation**
  - **proteomics**,
- Discovery of new therapeutic molecules
  - **lead discovery,**
  - **lead identification and optimization**

Technology is impacting this process

Identify disease

GENOMICS, PROTEOMICS & BIOPHARM.

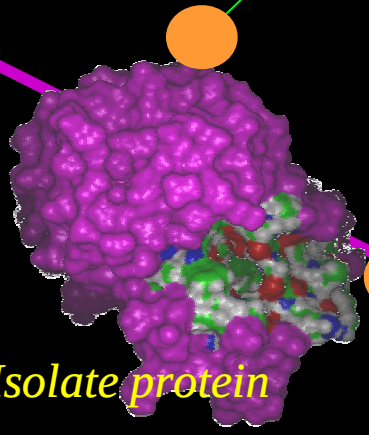
Potentially producing many more targets and "personalized" targets

HIGH THROUGHPUT SCREENING

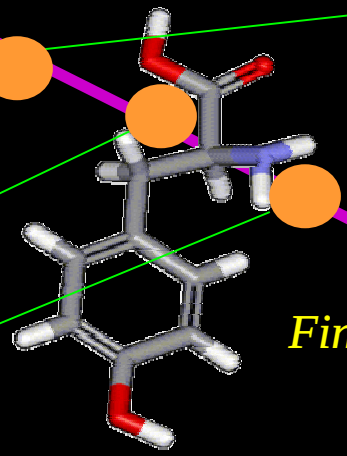
Screening up to 100,000 compounds a day for activity against a target protein

VIRTUAL SCREENING

Using a computer to predict activity



Isolate protein



Find drug

COMBINATORIAL CHEMISTRY

Rapidly producing vast numbers of compounds

MOLECULAR MODELING

Computer graphics & models help improve activity

Preclinical testing

IN VITRO & IN SILICO ADME MODELS

Tissue and computer models begin to replace animal testing

In vitro, in vivo ……..in the market

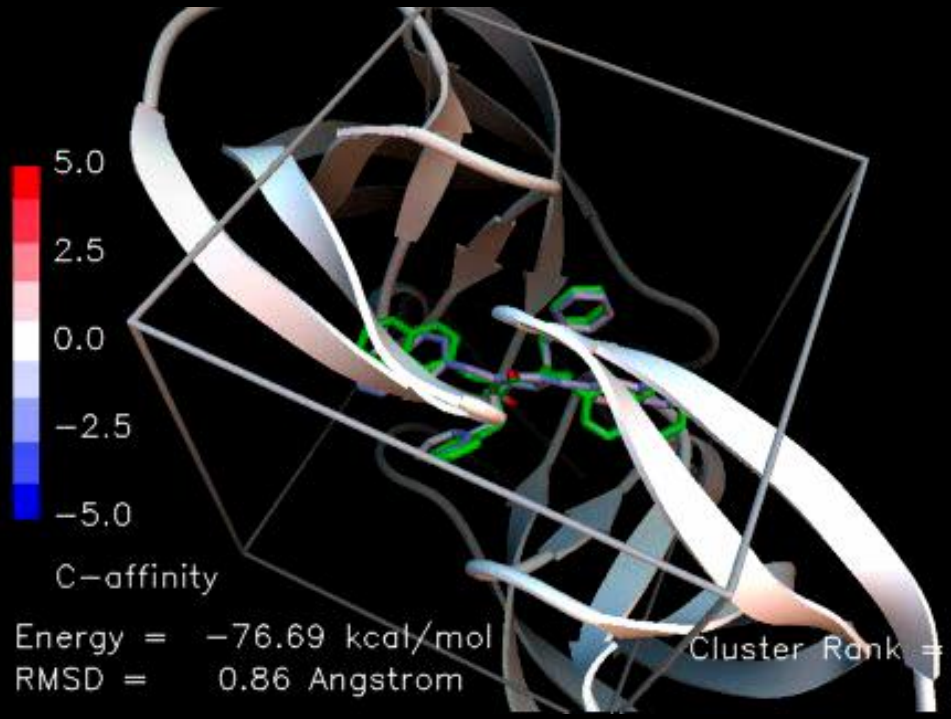
In silico ?

- Gene expression analysis
- Prediction of gene function
- Protein structure
- Virtual screening
- ADME prediction

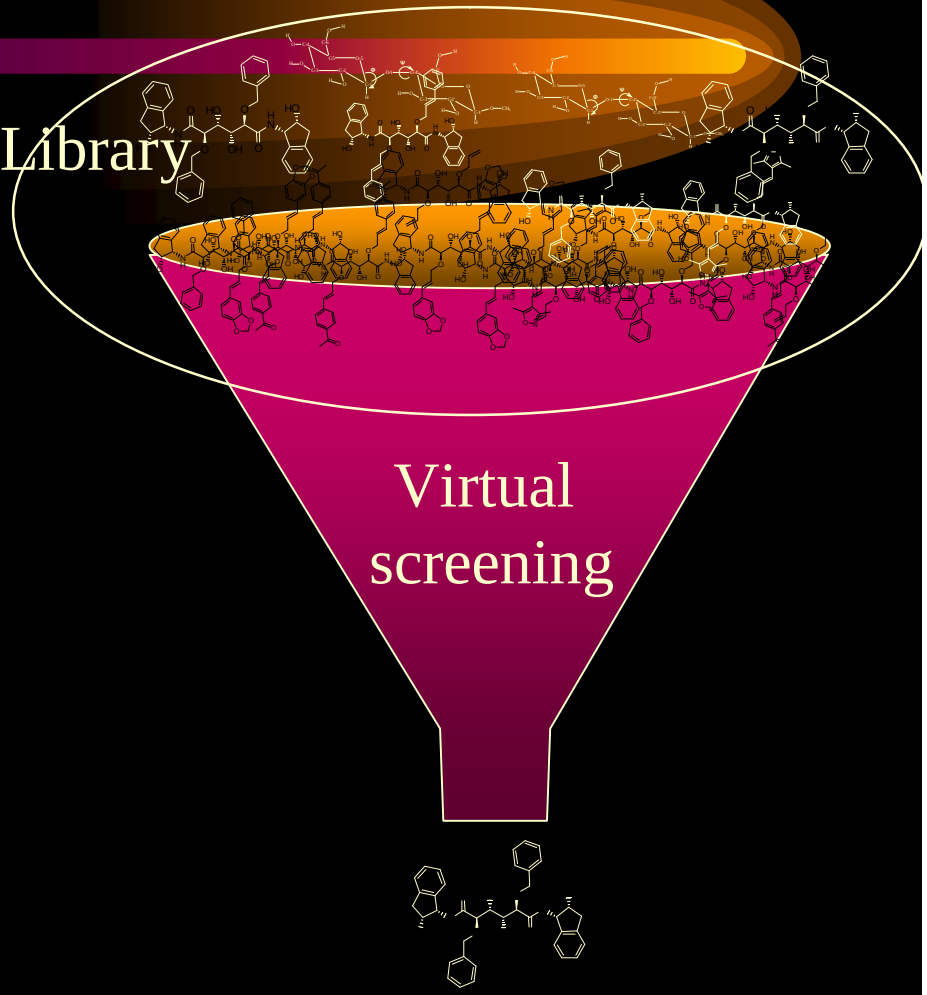
"….it is impossible to make every molecule that could be made ….."

Structure-Based Drug Design

Receptor Structure

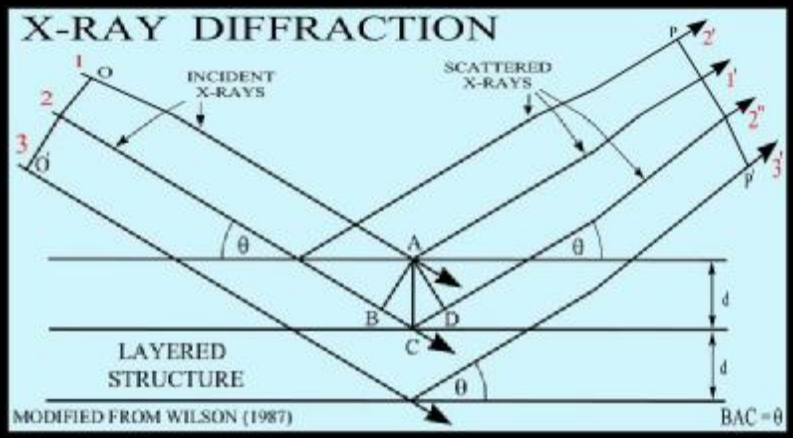


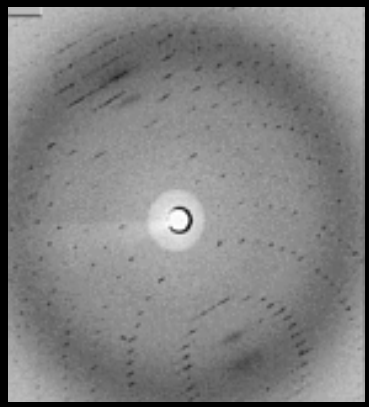
Chemical Library

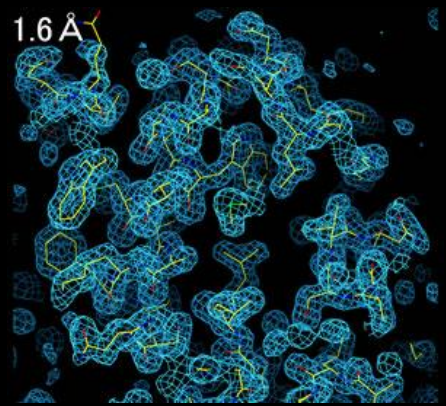


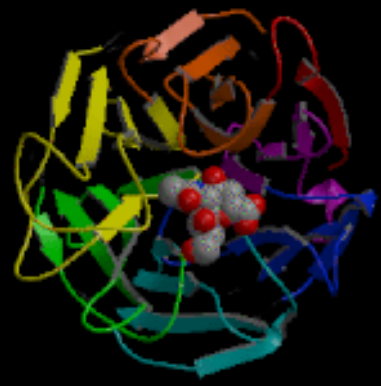
Lead compound

Protein Structure



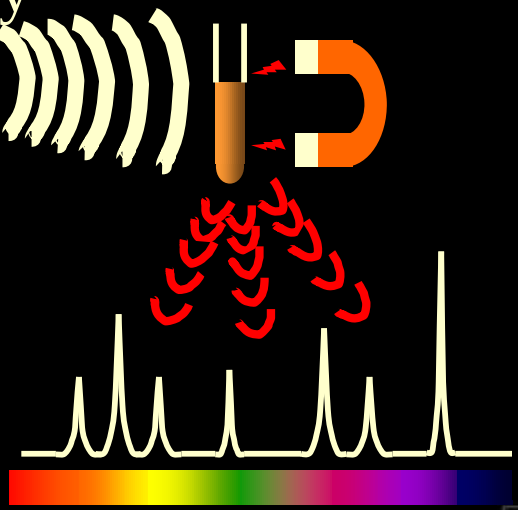


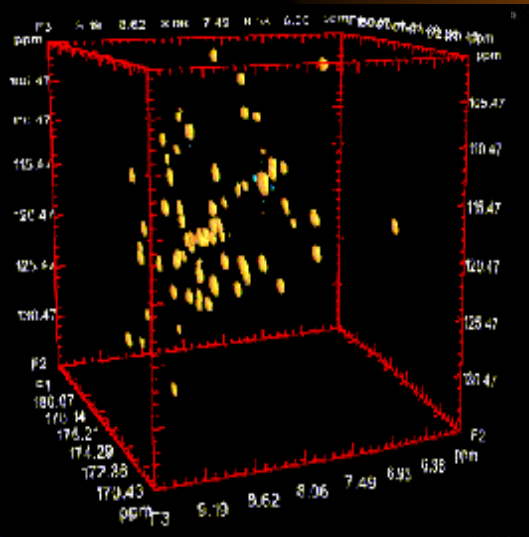


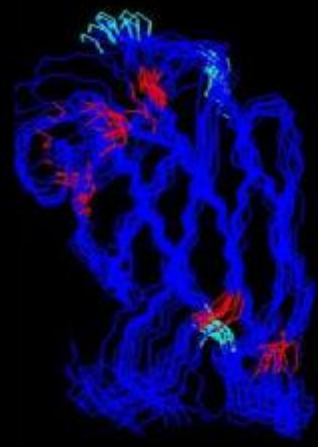


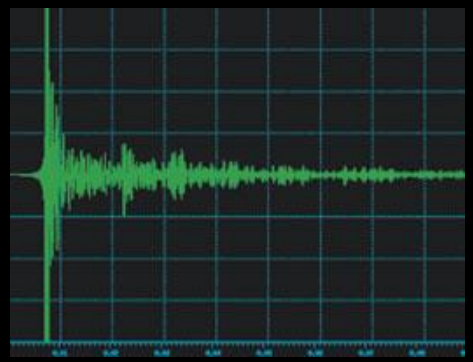
Protein Structure

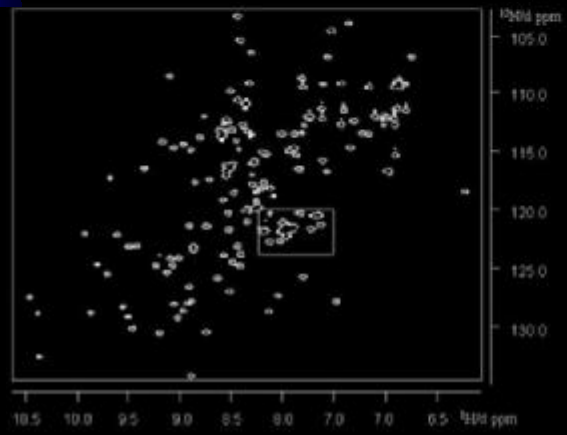
NMR spectroscopy









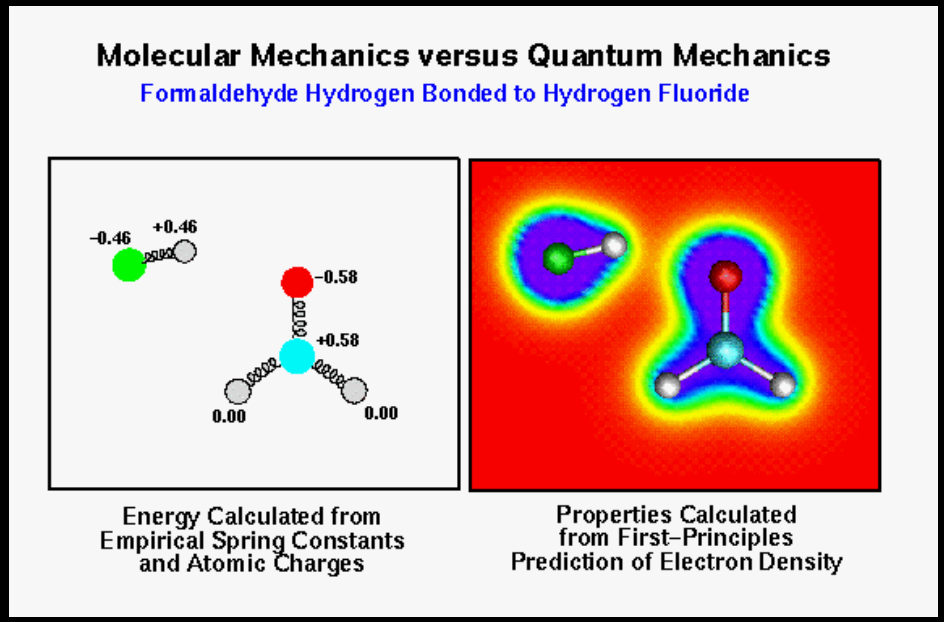


Virtual screening

Prediction of Receptor Ligand Interactions

Simulation of the Structure

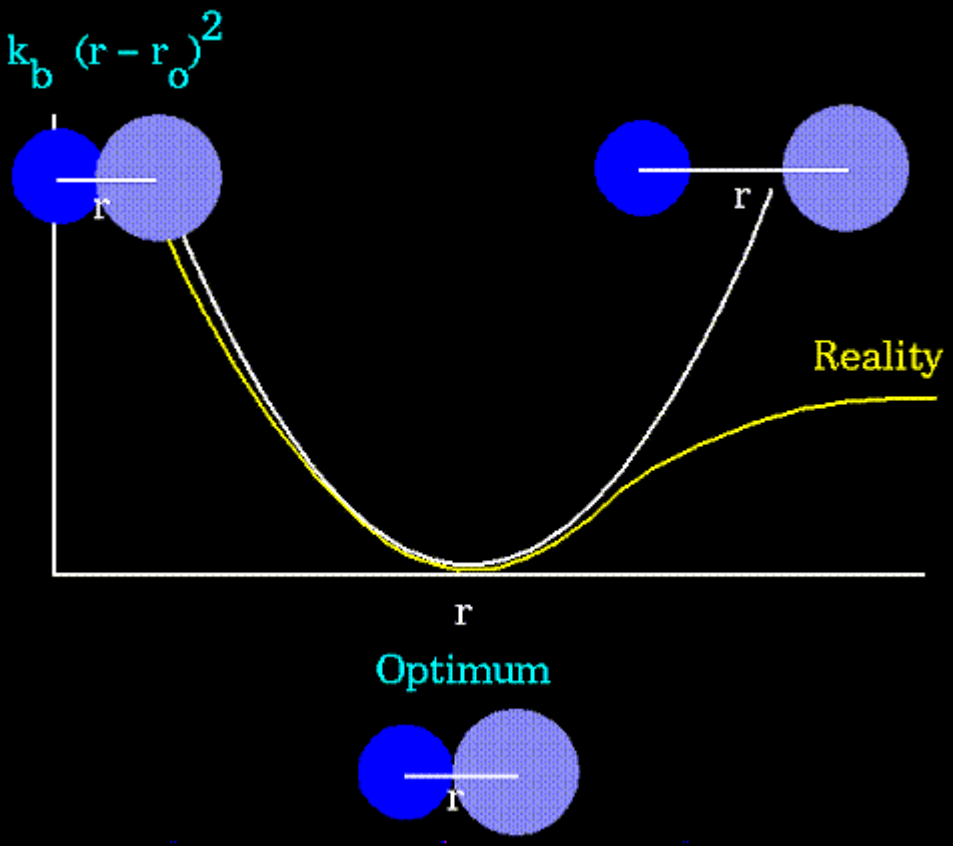
Calculation of the energy



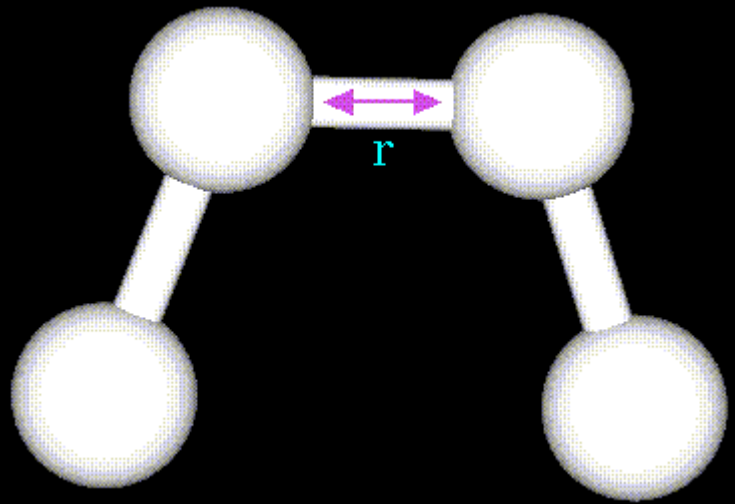
- Quantum Mechanics
  - *ab initio*
  - Semi-empirical

- Molecular Mechanics

Molecular Mechanics



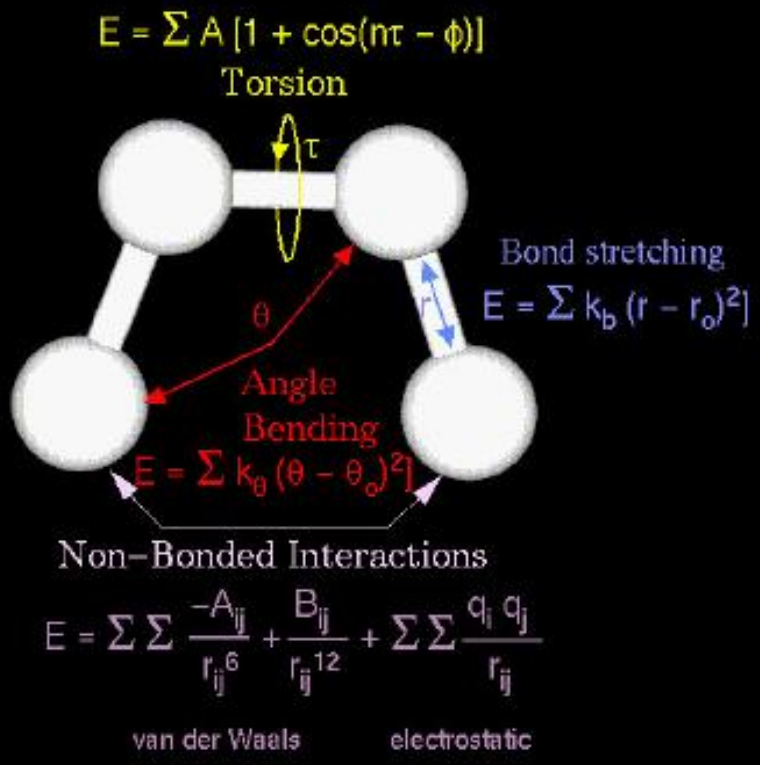
$$E = \sum_{\text{bonds}} k_b (r - r_o)^2$$



Energy is a function of the coordinates. ⟷ **Coordinates are function of the energy.**

Potential Energy Function

FORCE FIELD



$$U = \sum_{all\ bonds} \frac{1}{2} K_b (b - b_0)^2$$

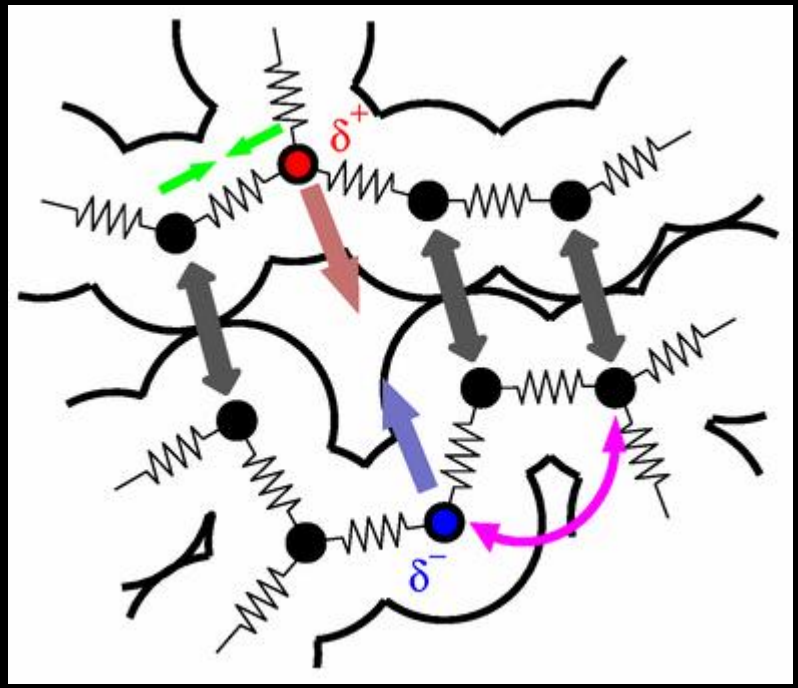
$$+ \sum_{all\ angles} \frac{1}{2} K_q (q - q_0)^2$$

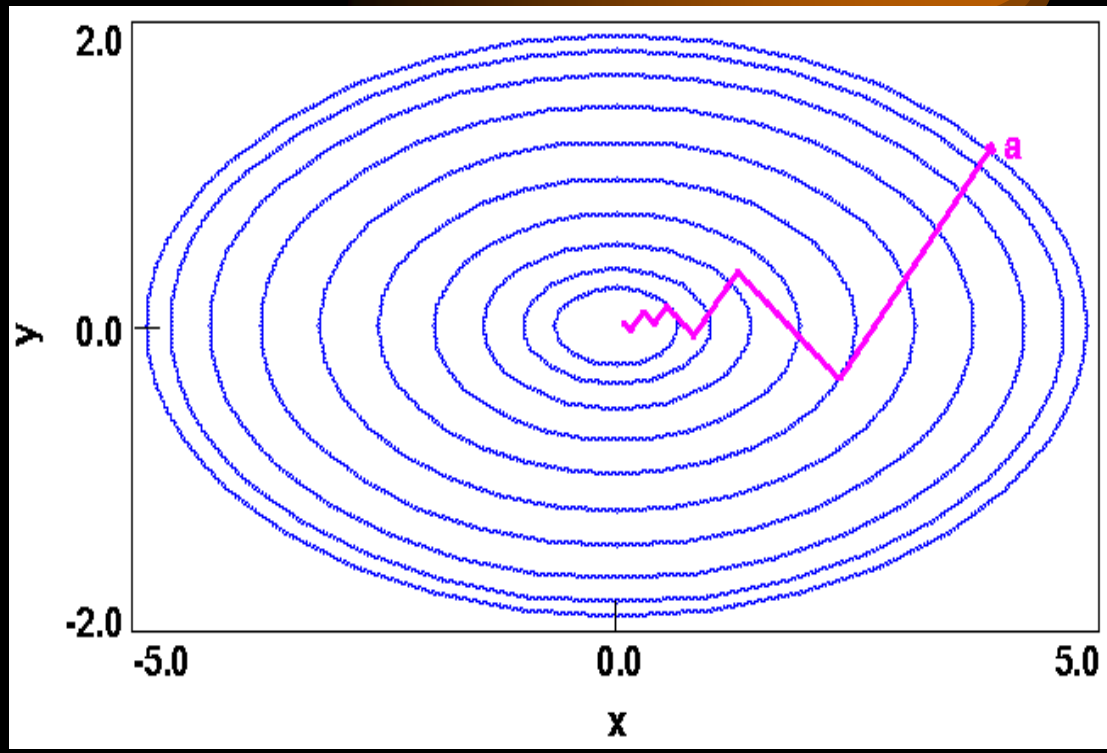
$$+ \sum_{all\ torsions} K_f [1 - \cos(nf)]$$

$$+ \sum_{i,j\ nonbonded} e_{ij} \left[\left(\frac{R_{ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{ij}}{r_{ij}} \right)^{6} \right]$$

$$+ \sum_{i,j\ nonbonded} \frac{q_i q_j}{4pe_0 e r_{ij}}$$

Energy Minimization

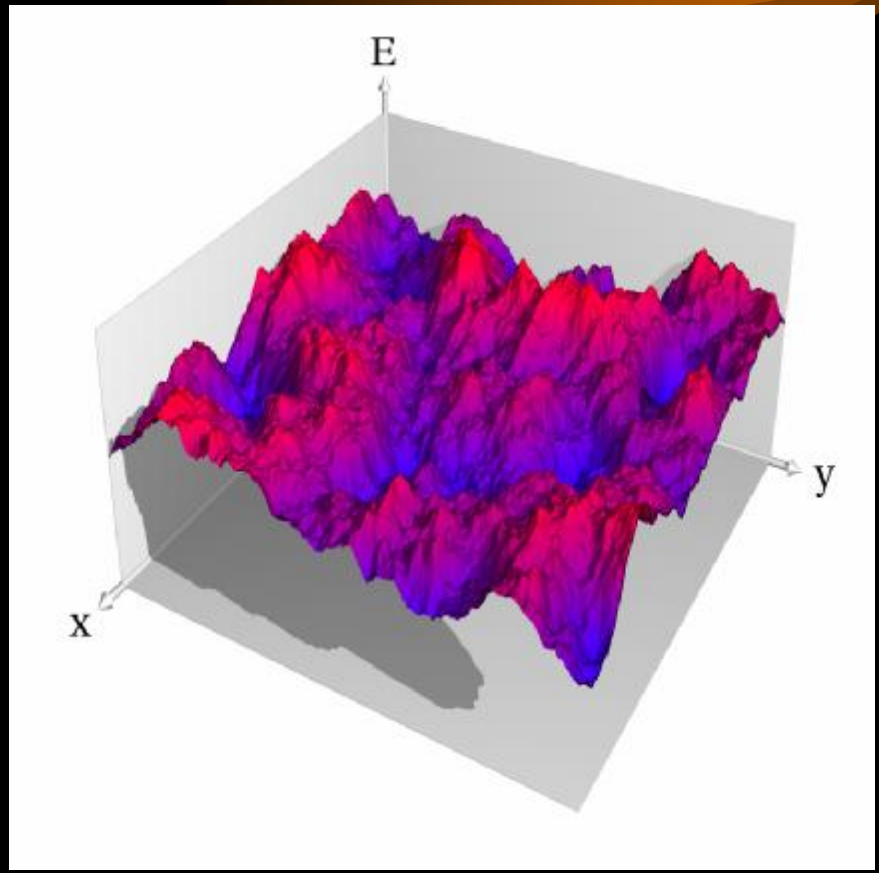




"evolution" can be performed by systematic variation of the atom positions towards the lower energy directions. This procedure is called "structure optimization" or "energy minimization"

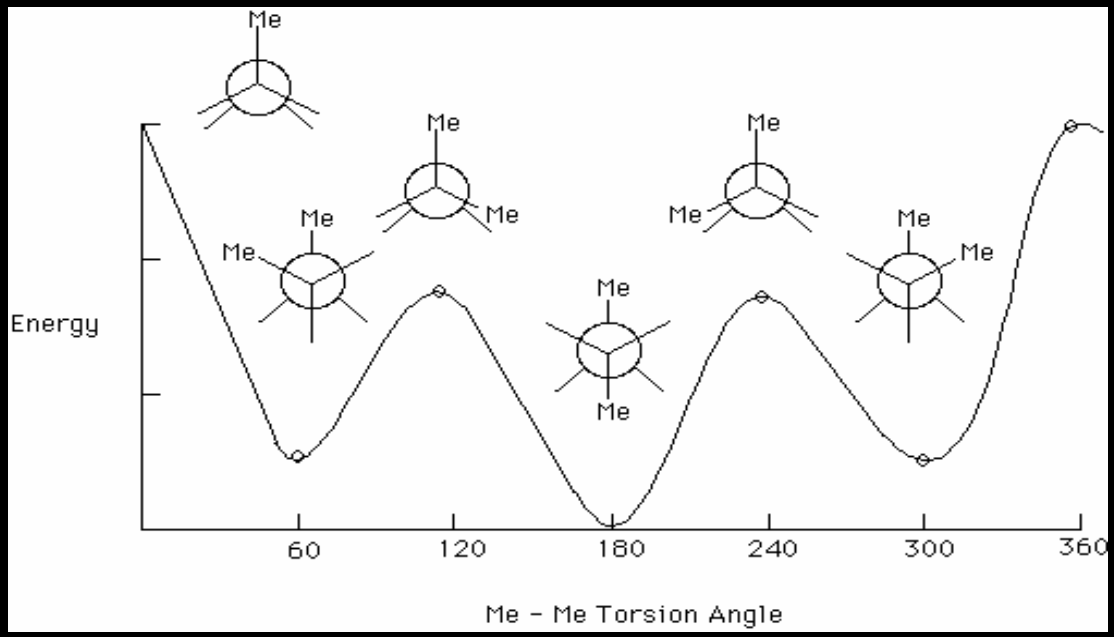
Conformational Search

- Systematic Search
- Molecular Dynamics
- Simulated Annealing
- Monte Carlo



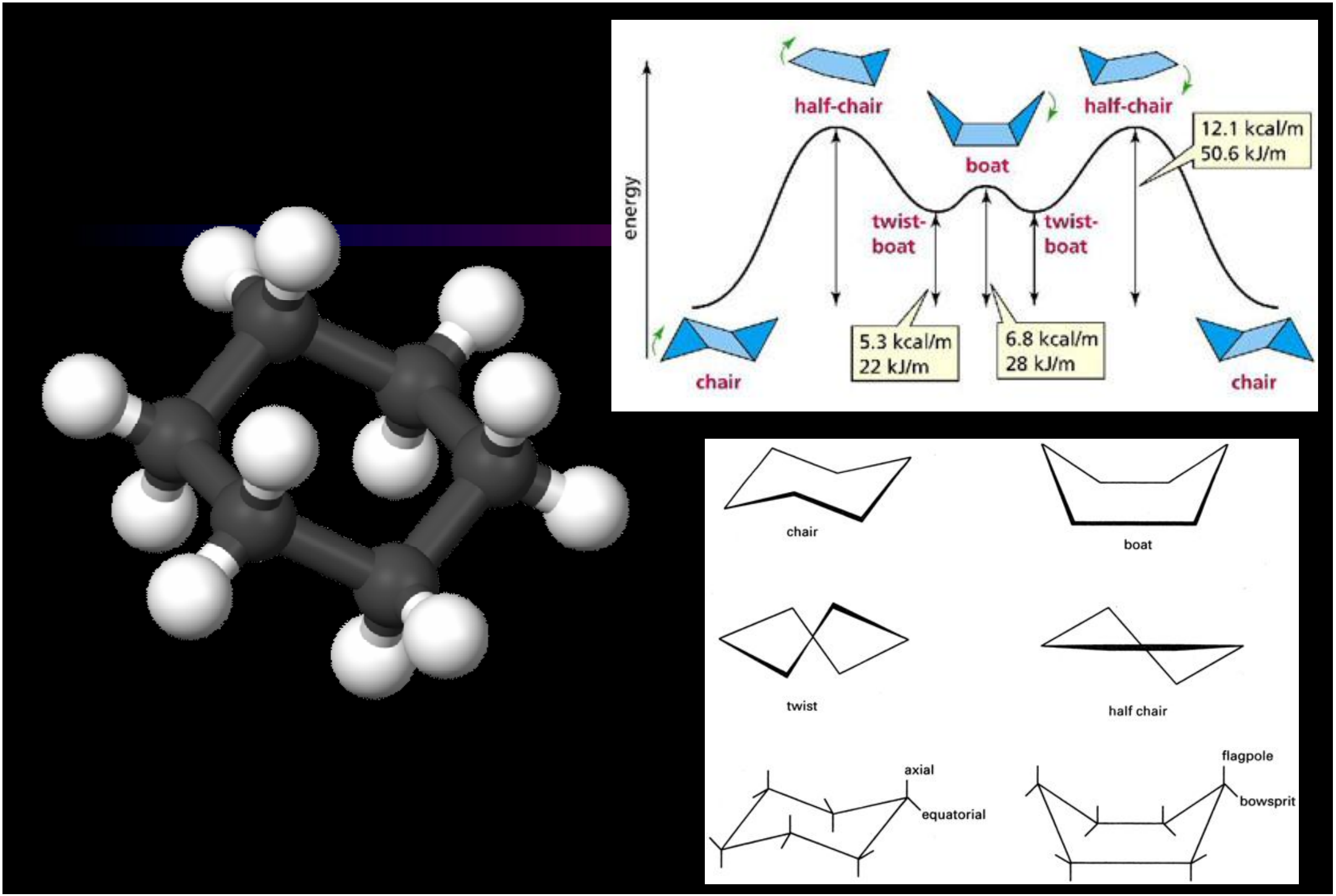
exploring the energy landscape

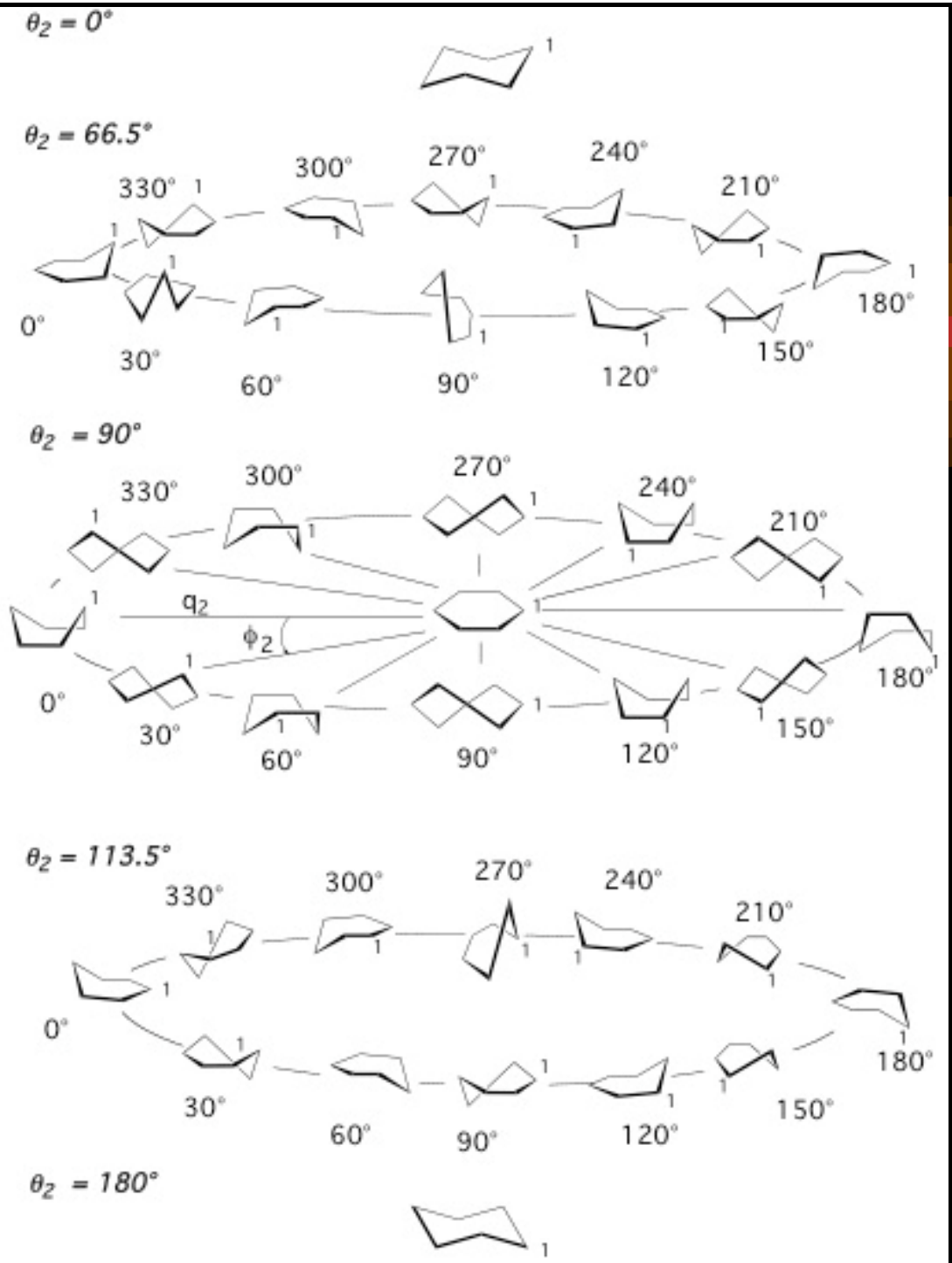
Conformations

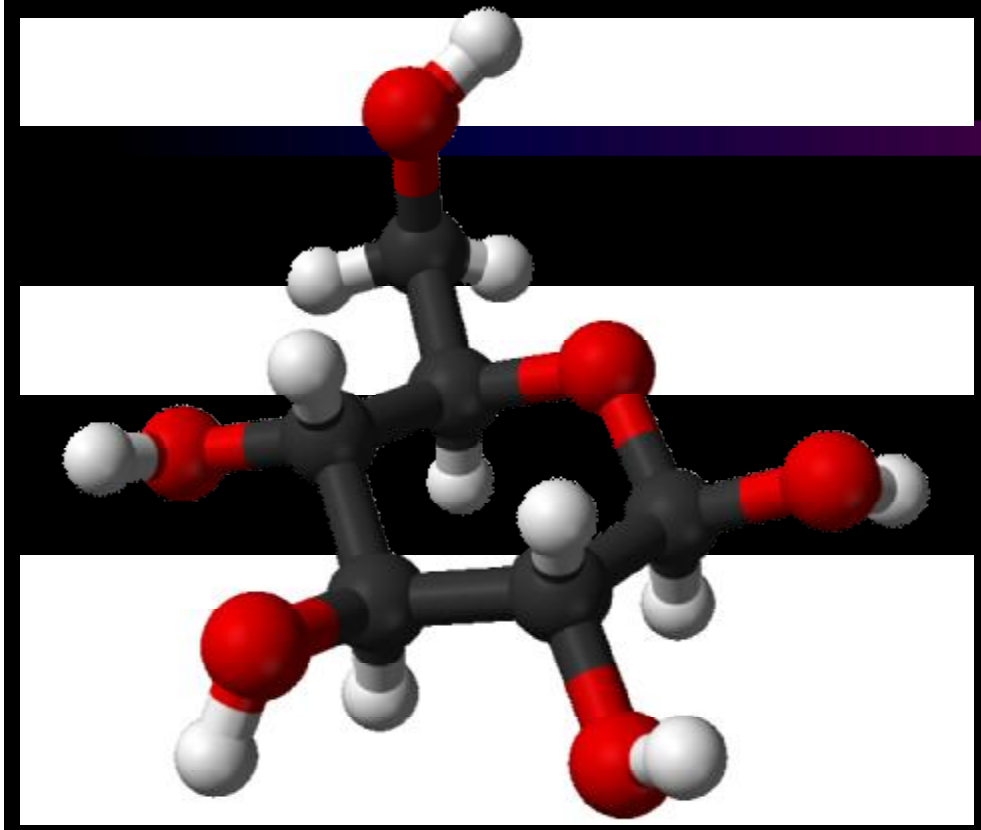


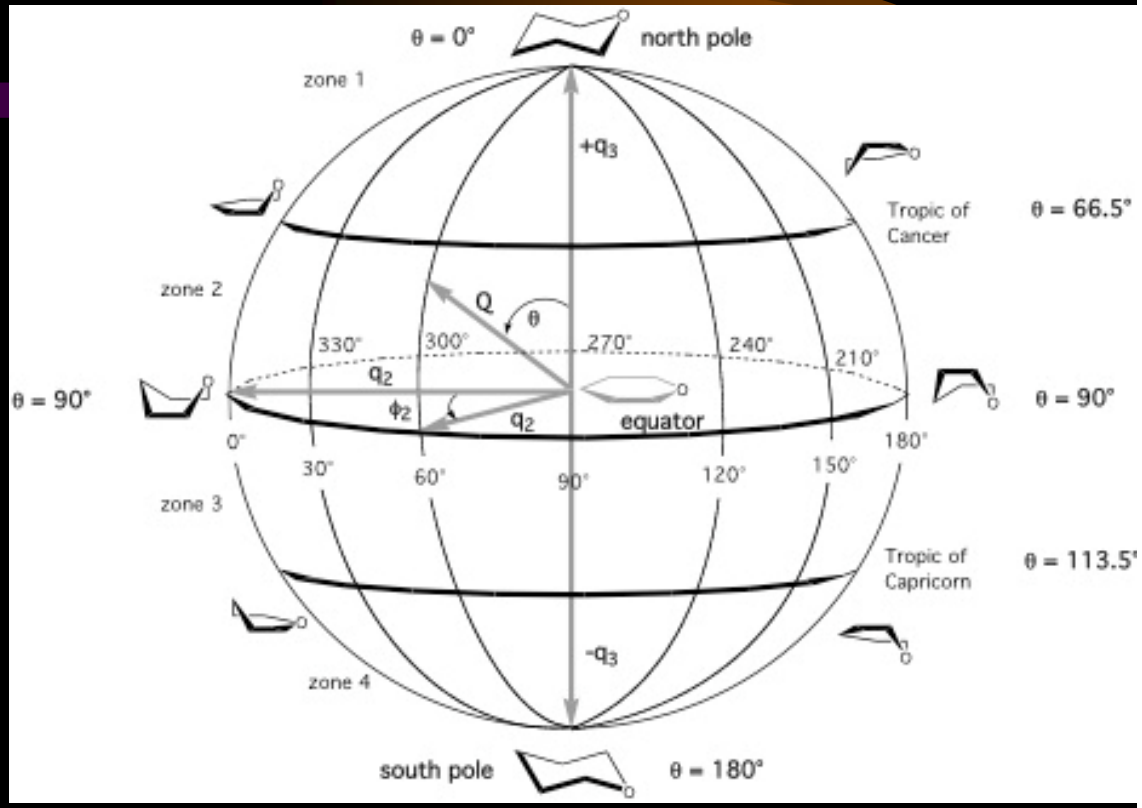
- Equilibrium between conformers.
- **Conformational Space**
  - ALL possible conformations under given constraints.

$$P = \exp\left(-\frac{E(B) - E(A)}{k_B T}\right)$$


half-chair
boat
half-chair
12.1 kcal/m
50.6 kJ/m
energy
twist-
boat
twist-
boat
5.3 kcal/m
22 kJ/m
6.8 kcal/m
28 kJ/m
chair
chair
chair
boat
twist
half chair
axial
equatorial
flagpole
bowsprit

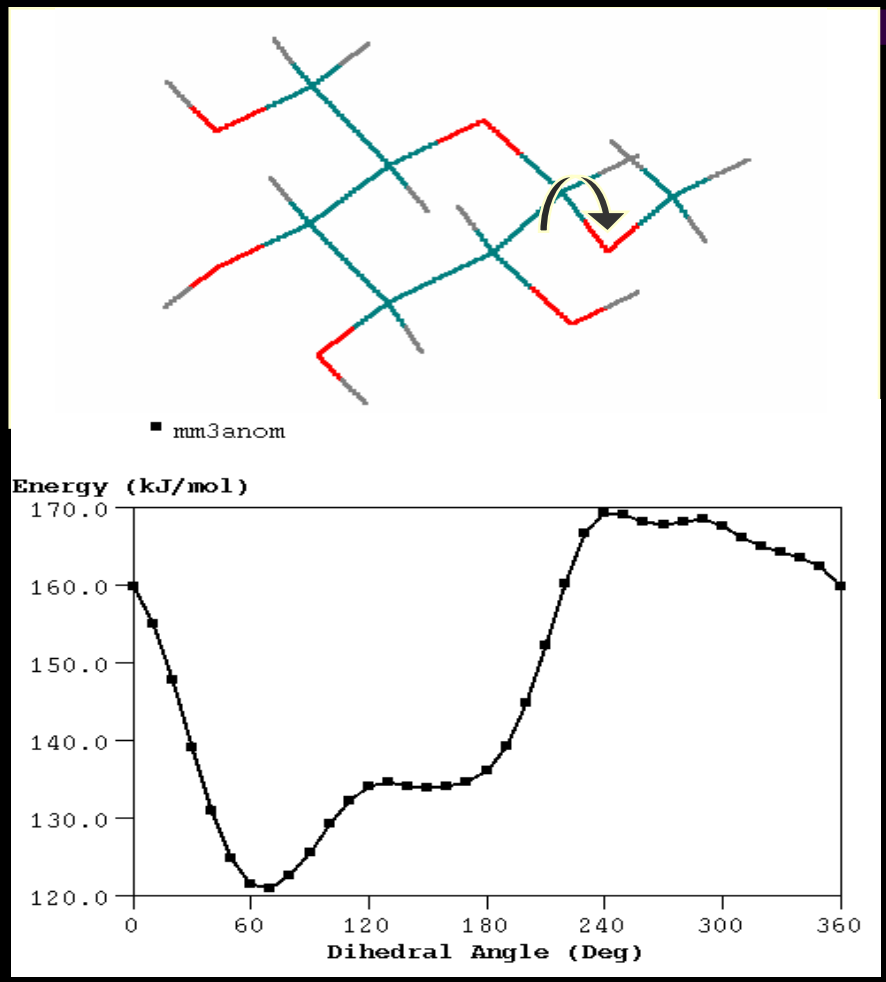

θ2 = 0°
θ2 = 66.5°
330°
300°
270°
240°
210°
180°
150°
120°
90°
60°
30°
0°
θ2 = 90°
q2
ϕ2
θ2 = 113.5°
θ2 = 180°



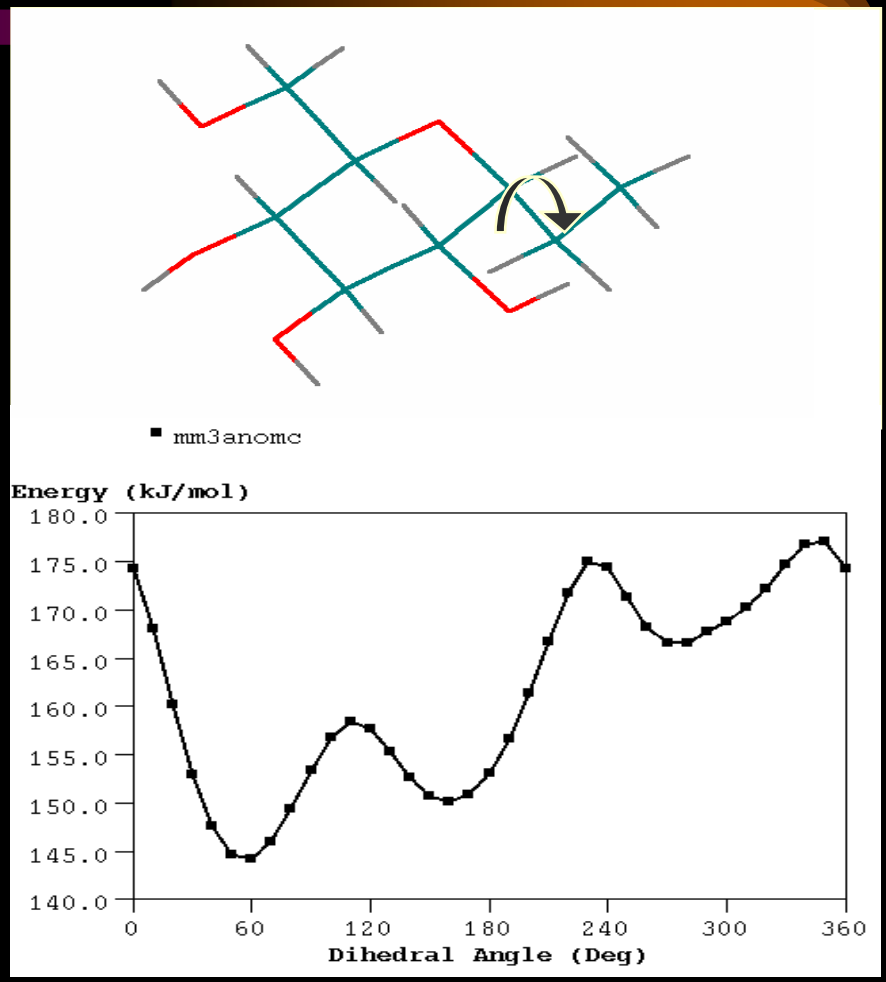

θ = 0°
north pole
zone 1
+q3
Tropic of Cancer
θ = 66.5°
zone 2
Q
θ
330°
300°
270°
240°
210°
θ = 90°
q2
ϕ2
q2
equator
θ = 90°
0°
30°
60°
90°
120°
150°
180°
zone 3
Tropic of Capricorn
θ = 113.5°
-q3
zone 4
south pole
θ = 180°

Systematic Search

O-Glucosides



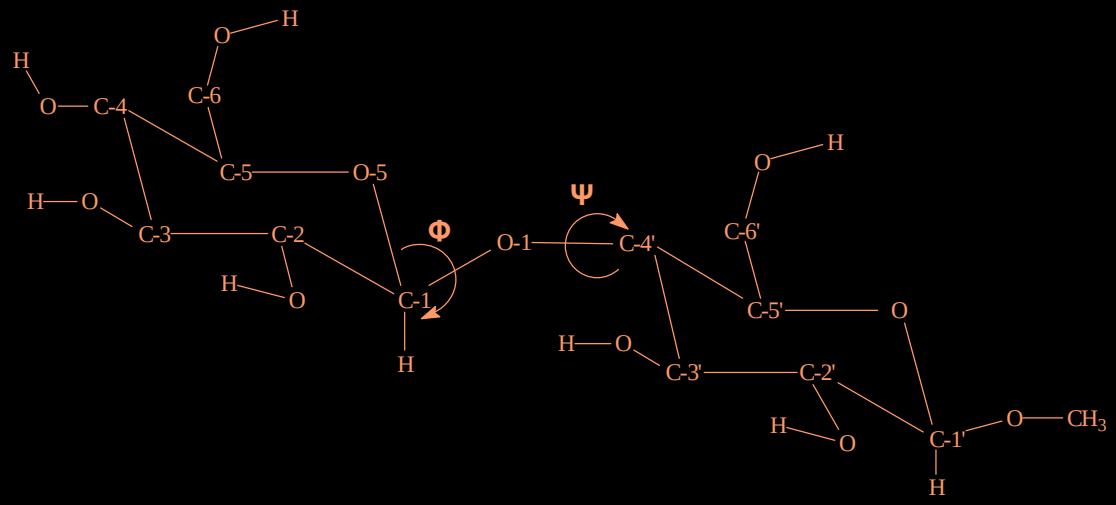
C-Glucosides



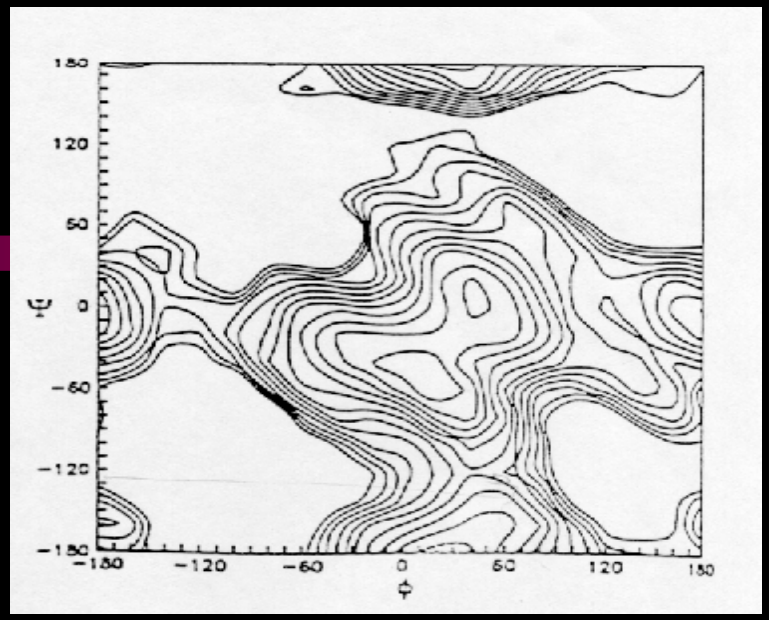
One rotatable bond Energy profile

Relaxed map

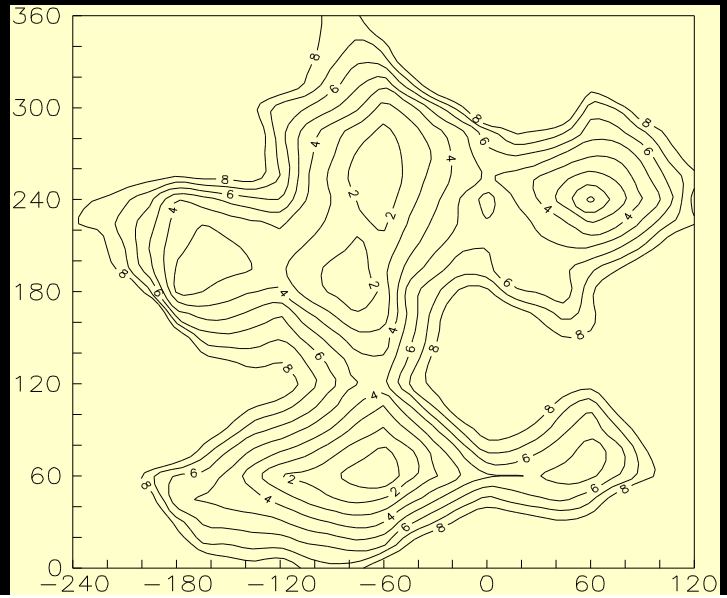
11.664 structures



O-Cellobiose

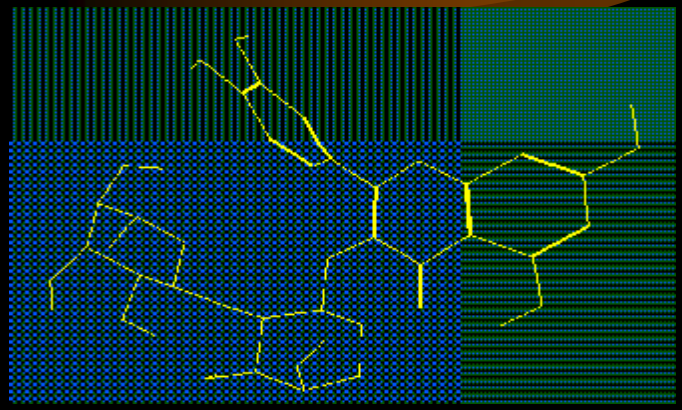


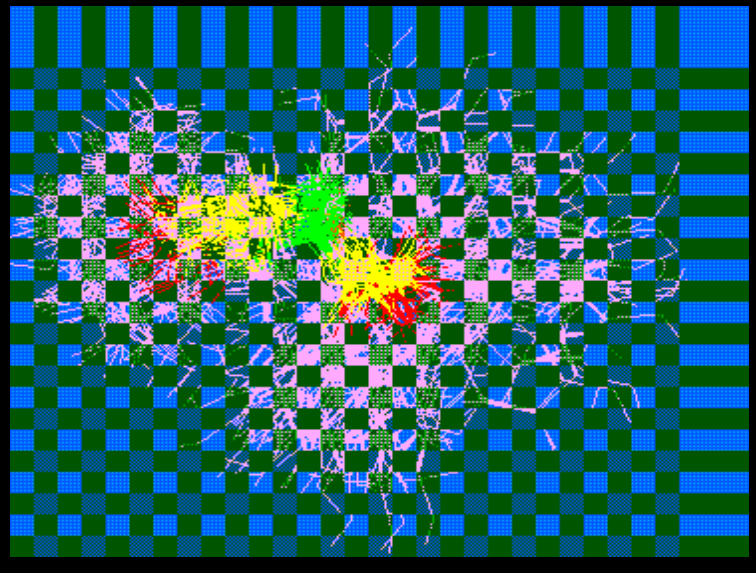
C-Cellobiose

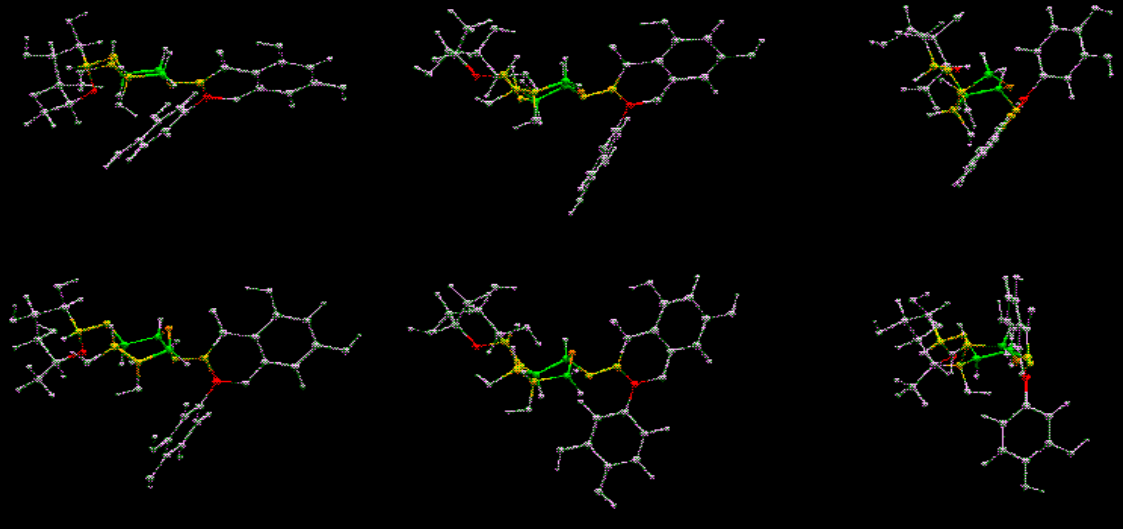


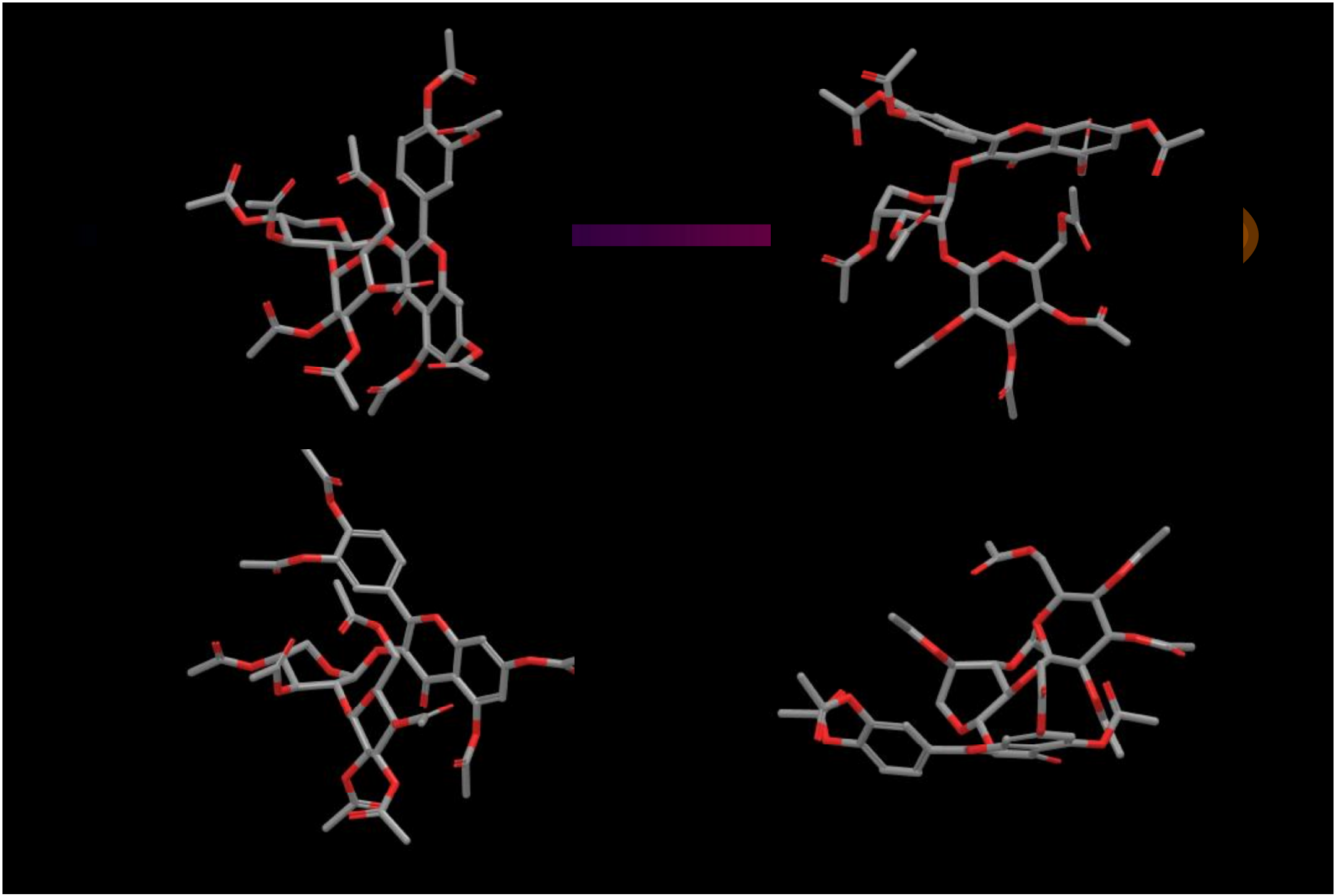
Quercetin-3-(α-L-Rha-2-1-α-L-Ara)

Monte Carlo 10.000 steps
15 Degrees of Freedom
3500 Structures up to 50 Kcal/mol





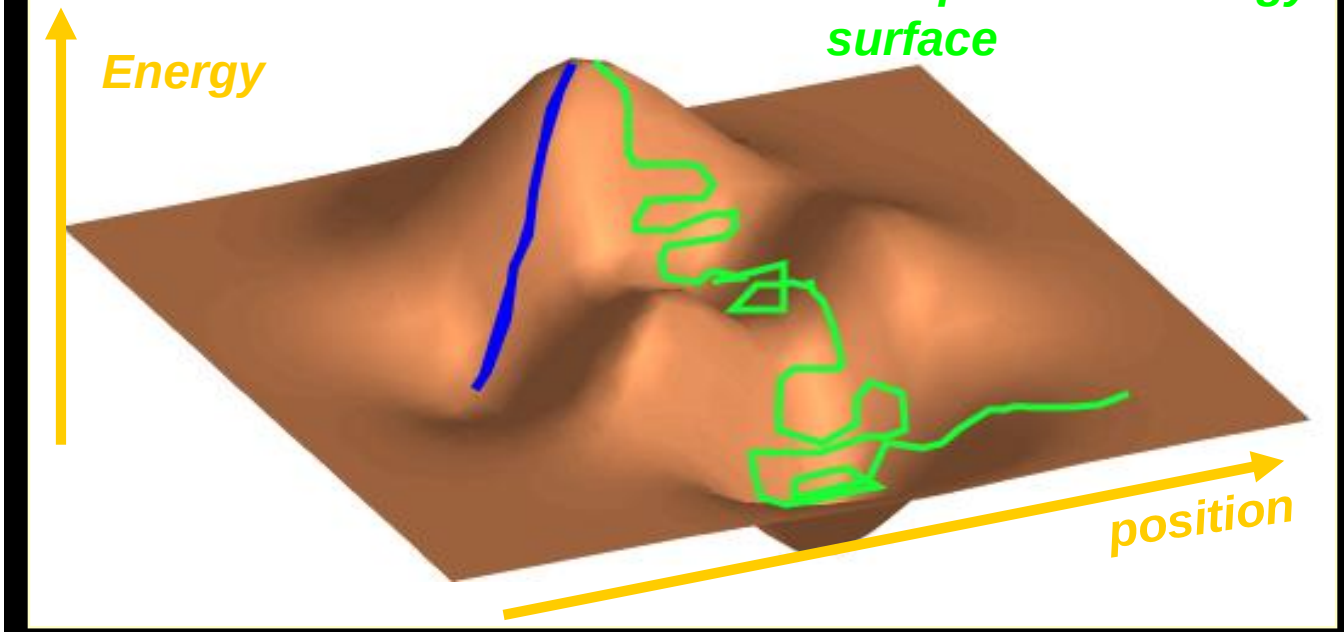


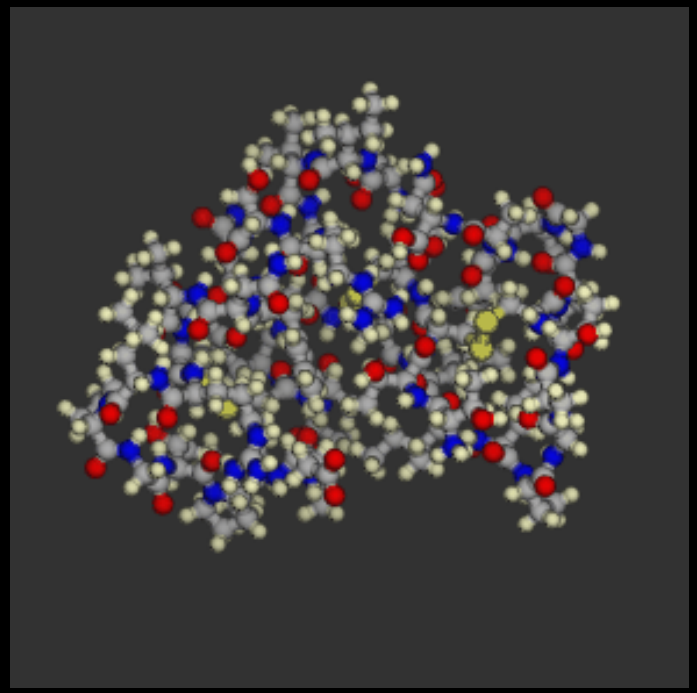


Molecular Dynamics

$$\frac{1}{2}mv^2 = \frac{3}{2}kT$$

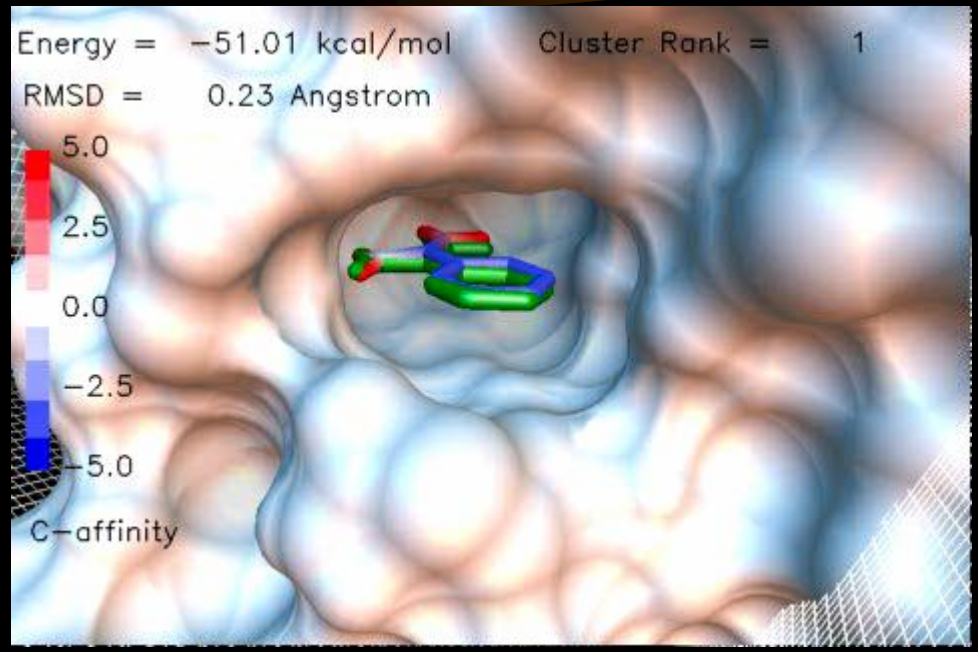
Molecular dynamics uses thermal energy to explore the energy surface



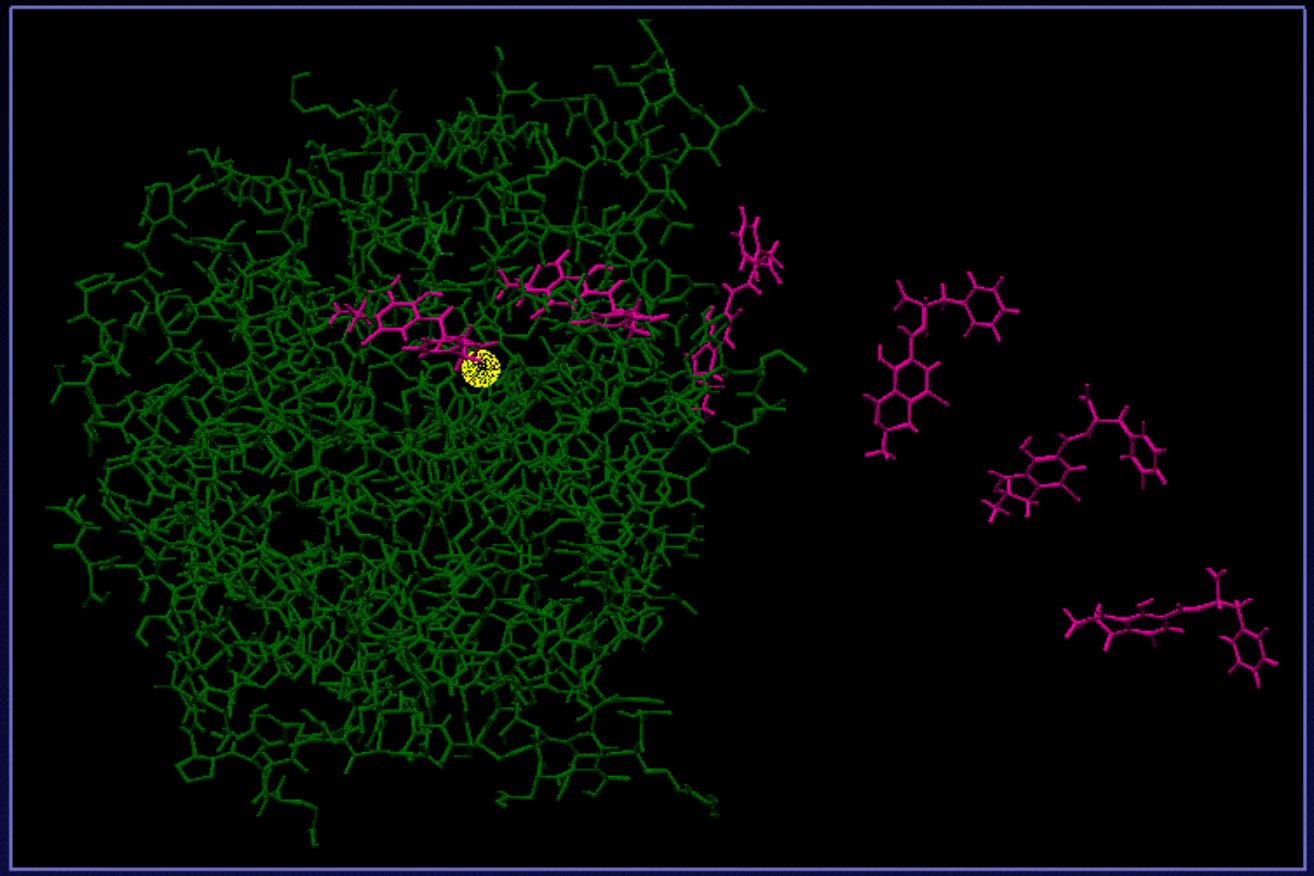


Receptor-Ligand Interactions

- Docking Calculations
  - Molecular Dynamics
  - Monte Carlo
  - Genetic Algorithms
  - Combinatorial Docking
  - Low Mode Search



Structure–based design (known 3D structure of the protein)



$$E_{total} = \sum_{bonds} K_r (r - r_{eq})^2 + \sum_{angles} K_\theta (\theta - \theta_{eq})^2 +$$

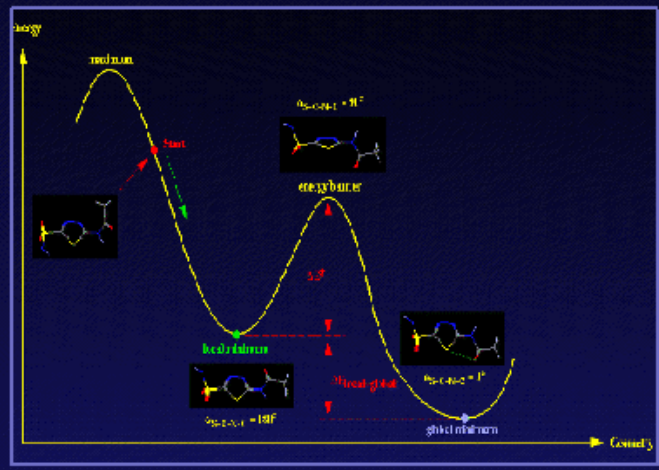
$$\sum_{dihedrals} \frac{V_n}{2} [1 - \cos(n\phi - \gamma)] + \sum_{nb\ pairs} \frac{q_i \cdot q_j}{4\pi\varepsilon_0 D(r) r_{ij}} + \sum_{nb\ pairs} \frac{A}{r_{ij}^{12}} - \frac{B}{r_{ij}^{6}}$$

$$+ \sum_{H\text{-}bonds} \left(\frac{C}{r_{H\text{-}Acc}^{12}} - \frac{D}{r_{H\text{-}Acc}^{10}} \right) \cdot \cos^2(\theta_{Don\text{-}H\text{-}Acc}) \cdot \cos^n(\omega_{H\text{-}Lig\text{-}LP})$$

$$+ \sum_{metal\text{-}ligand\ pairs} \frac{q_i^{CT} \cdot q_j^{CT}}{4\pi\varepsilon_0 D(r) r_{ij}} + \sum_{metal\text{-}ligand\ pairs} \left(\frac{E}{r_{M\text{-}Lig}^{12}} - \frac{F}{r_{M\text{-}Lig}^{10}} \right)$$

$$+ (E_{MC} - E_{LFS}) \cdot \prod_{indep.\ angles} \cos^2(\psi_{Lig\text{-}M\text{-}Lig'} - \psi_{eq}) \cdot \frac{1}{n} \sum_{1^{st}\ shell\ ligands} \cos^n(\omega_{M\text{-}Lig\text{-}LP})$$

Structure optimization



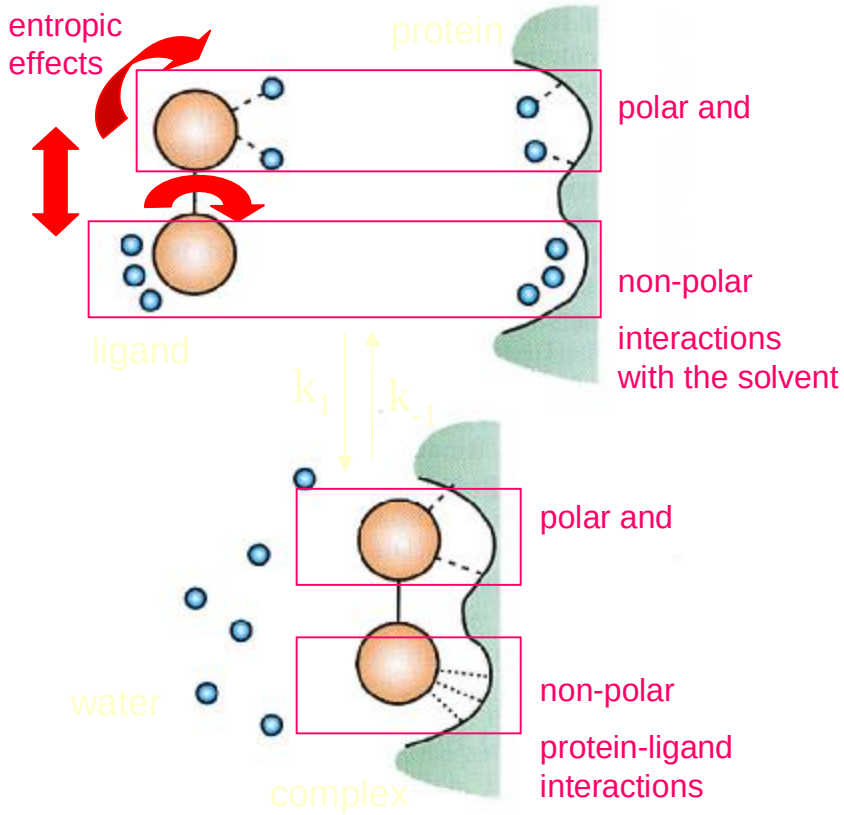
Fitting a small molecule to a macromolecular binding site

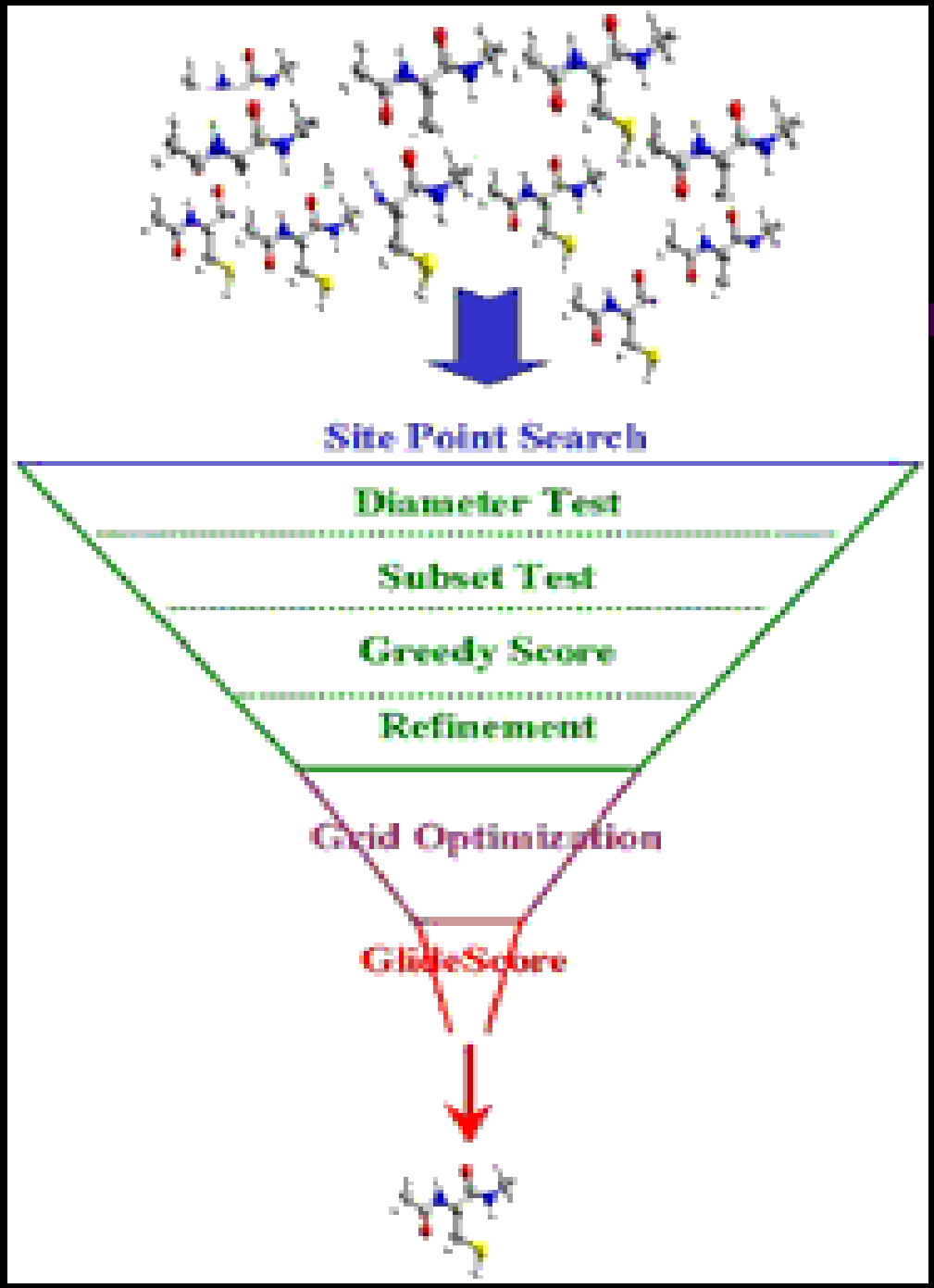
$$E_{binding} = E_{rec-lig} - \Delta G_{solv,lig} - \Delta E_{int,lig} - T\Delta S_{lig} - E_{ind.fit}$$

Free Energy

$$K_{bind} = \frac{k_1}{k_{-1}} = \frac{[C]}{[P][L]}$$

$$\Delta G_{bind} = -RT \ln K_{bind}$$





Virtual Screening

Glide Software

- Application of a series of filters that rapidly funnel down the possible ligand positions and orientations to a manageable number for detailed examination.

Halgren TA t al, *J Med Chem*. 2004, 47, 1750

Συνάρτηση βαθμολόγησης GLIDE

- **Ημιεμπειρική συνάρτηση αξιολόγησης GLIDESCORE**
- **Αθροισμα όρων που περιγράφουν αλληλεπιδράσεις μεταξύ πρωτεΐνης, μικρομορίου και διαλύτη**
- **Εκτίμηση συγγένειας πρόσδεσης $\Delta G_{\text{πρόσδεσης}}$**

$$\begin{aligned}\Delta G_{\text{bind}} = {} & C_{\text{lipo-lipo}} \sum f(r_{\text{lr}}) + \\ & C_{\text{hbond-neut-neut}} \sum g(\Delta r)\, h(\Delta\alpha) + \\ & C_{\text{hbond-neut-charged}} \sum g(\Delta r)\, h(\Delta\alpha) + \\ & C_{\text{hbond-charged-charged}} \sum g(\Delta r)\, h(\Delta\alpha) + \\ & C_{\text{max-metal-ion}} \sum f(r_{\text{lm}}) + C_{\text{rotb}} H_{\text{rotb}} + \\ & C_{\text{polar-phob}} V_{\text{polar-phob}} + C_{\text{coul}} E_{\text{coul}} + \\ & C_{\text{vdW}} E_{\text{vdW}} + \text{solvation terms}\end{aligned}$$

The ZINC Database

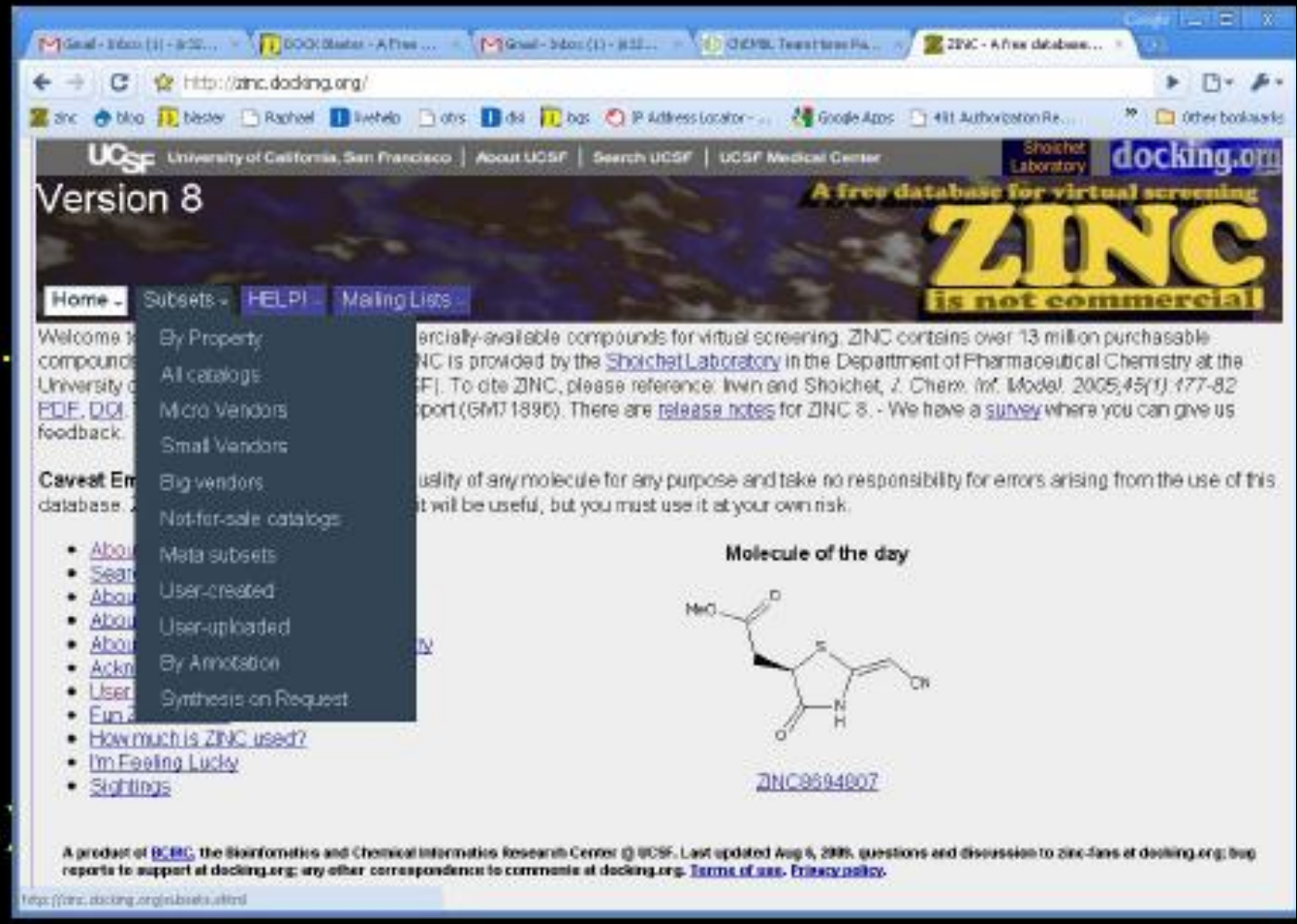
http://zinc.docking.org

21 million compounds
commercially available
structures calculated
multiple conformations
properties (charge, solv, etc..
links to suppliers

Free to the community

Multiple subsets
8.8 M drug-like (Lipinski)
3.4 M lead-like (Oprea…)
450 K fragment-like (Astex, …)

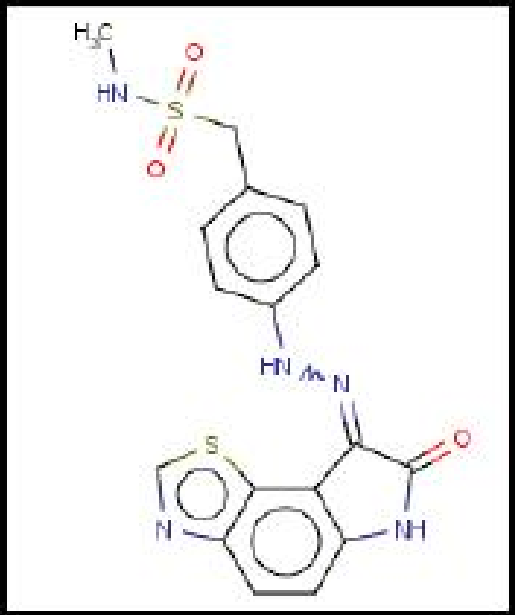
Availlable in popular formats
SMILES, SDF, mol2, flexibase

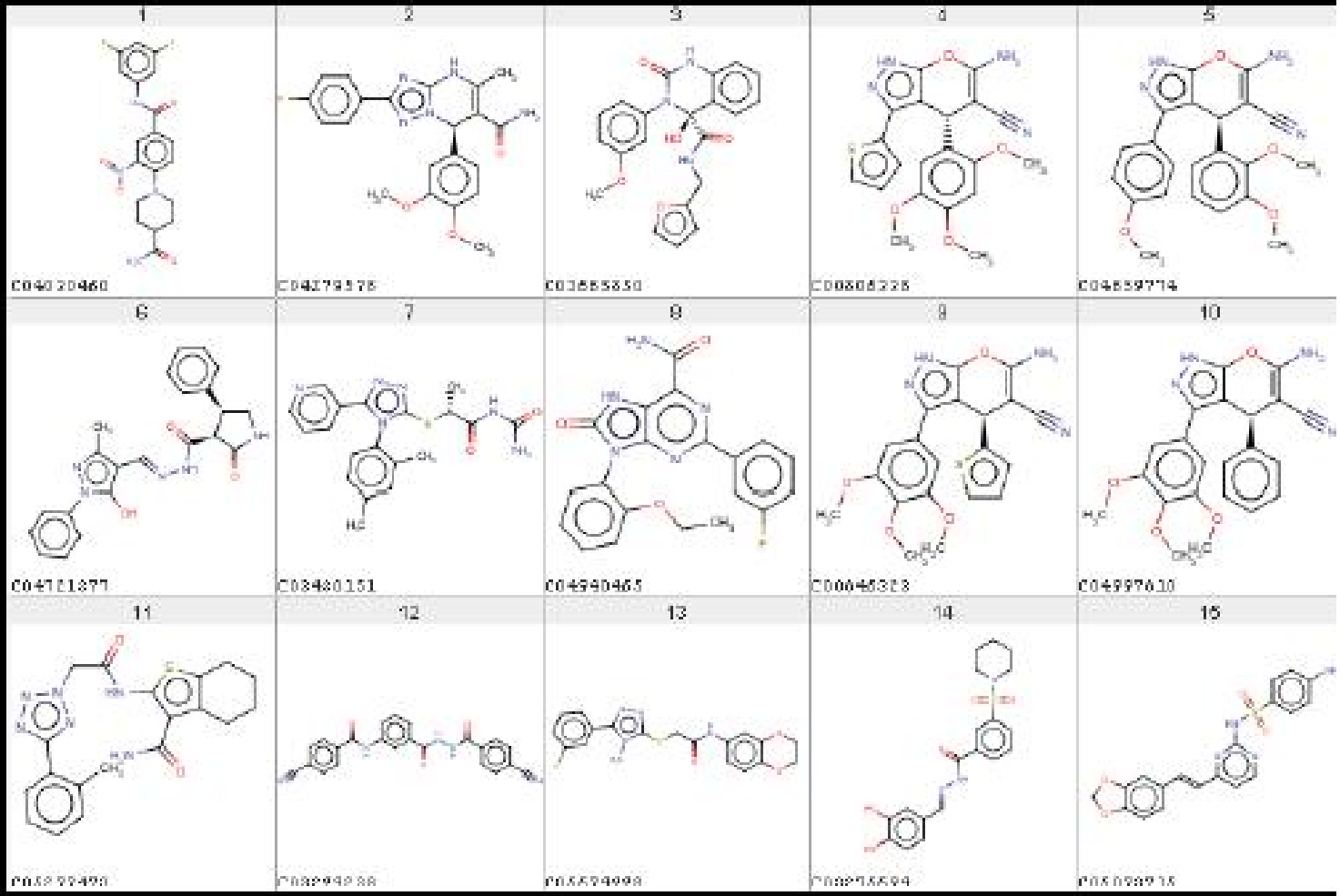


Updated continuously (10,000 new today)
Over 2 million new compounds per year
Over 1 million depletions per year

What is a docking decoy?

Similar physical properties, but chemically distinct, thus unlikely to bind.



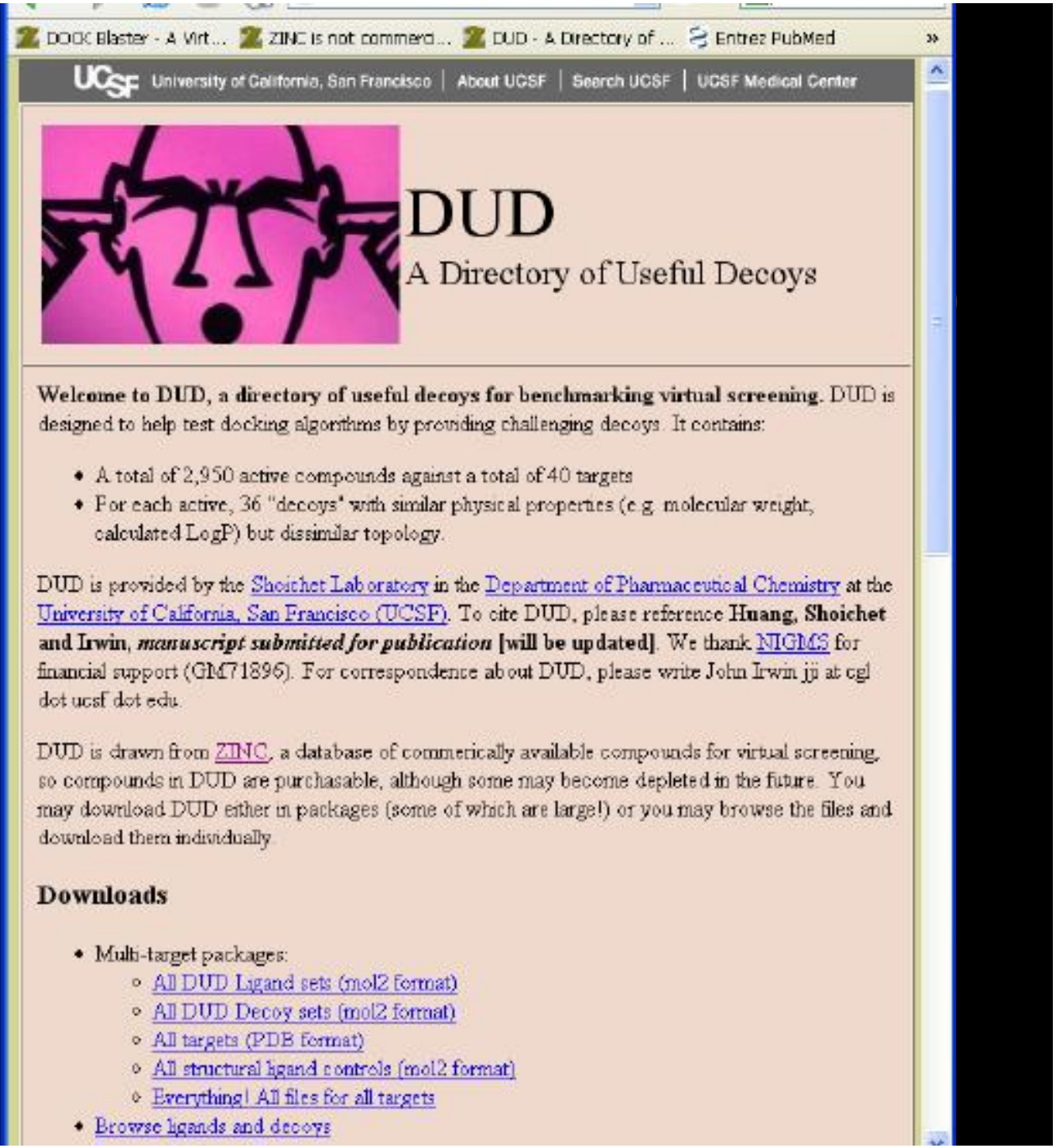


DUD is free

40 targets
2,950 ligands
95,358 decoys

mol2 format
All docking files

dud.docking.org



VIRTUAL SCREENING

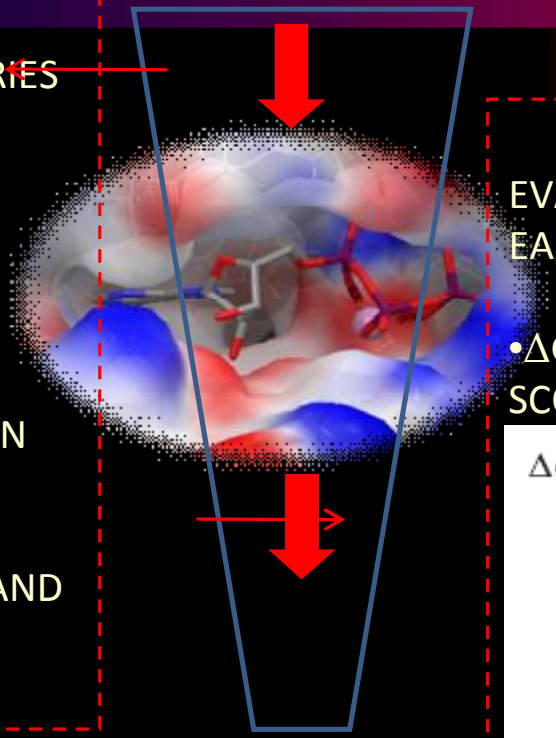
LIBRARY COMPOUNDS

3D STRUCTURES

STEP 1: DOCKING

EVALUATION OF THE BINDING GEOMETRIES (RELATIVE ORIENTATION) OF EACH COMPOUND TO THE TARGET MACROMOLECULE

- STOCHASTIC MONTE CARLO SAMPLING
- RIGID REPRESENTATION OF THE PROTEIN BASED ON GRID CALCULATIONS
- FLEXIBLE REPRESENTATION OF THE LIGAND

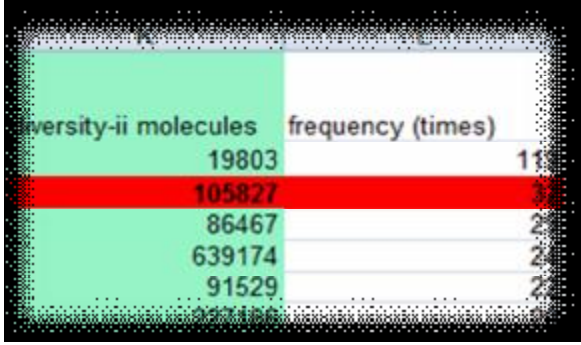


STEP 2: SCORING

EVALUATION OF THE BINDING AFFINITY OF EACH COMPOUND

- ΔG APPROXIMATED BY SEMIEMPIRICAL SCORING FUNCTION:

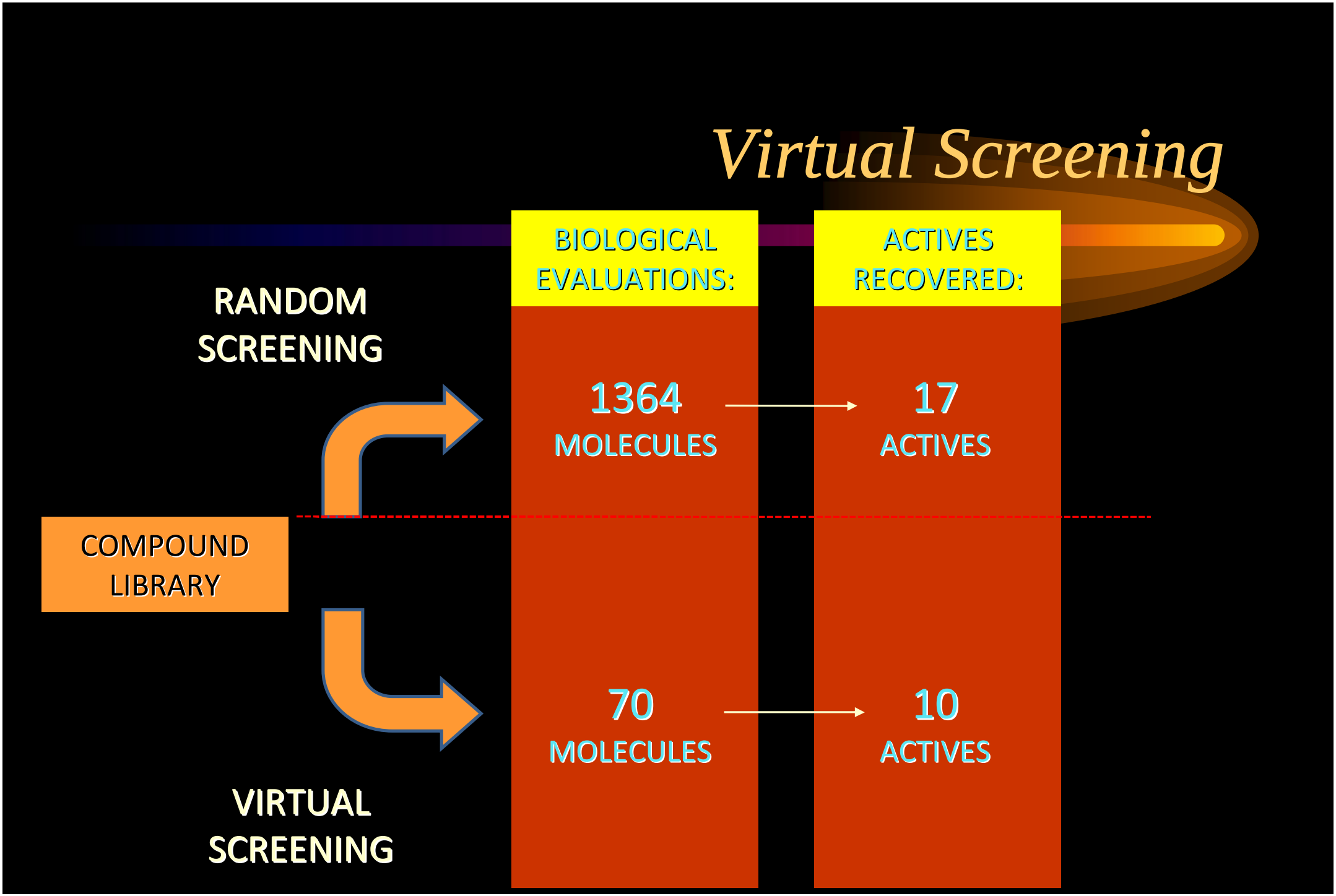
$$\Delta G_{\text{bind}} = C_{\text{lipo-lipo}} \sum f(r_{\text{lr}}) + C_{\text{hbond-neut-neut}} \sum g(\Delta r)\, h(\Delta\alpha) + C_{\text{hbond-neut-charged}} \sum g(\Delta r)\, h(\Delta\alpha) + C_{\text{hbond-charged-charged}} \sum g(\Delta r)\, h(\Delta\alpha) + C_{\text{max-metal-ion}} \sum f(r_{\text{lm}}) + C_{\text{rotb}} H_{\text{rotb}} + C_{\text{polar-phob}} V_{\text{polar-phob}} + C_{\text{coul}} E_{\text{coul}} + C_{\text{vdW}} E_{\text{vdW}} + \text{solvation terms}$$

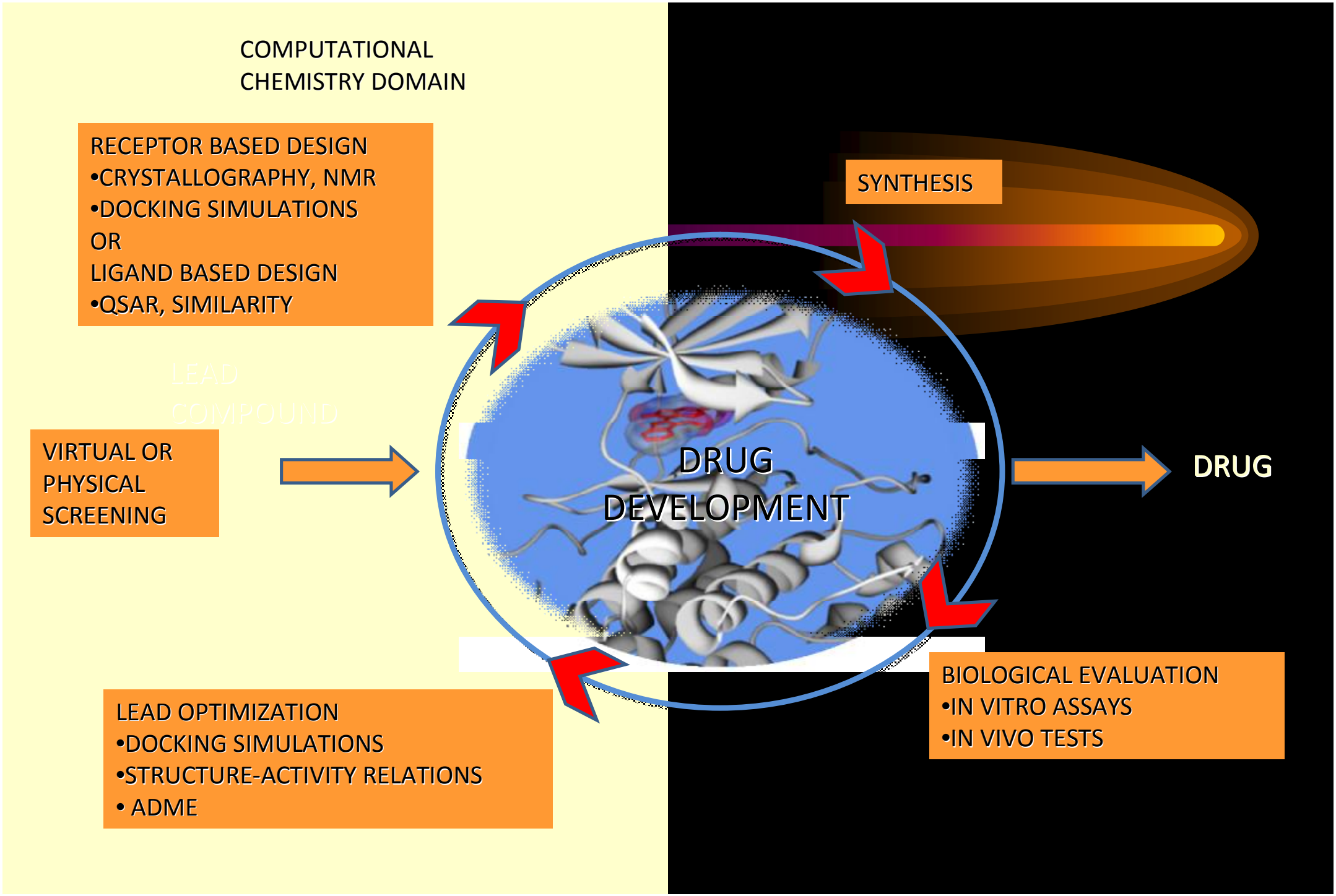


versity-ii molecules	frequency (times)
19803	11
105827	3
86467	2
639174	2
91529	2

RANKING OF THE COMPOUNDS

GLIDE v.3
SCHRÖDINGER Inc.


Virtual Screening
BIOLOGICAL EVALUATIONS:
ACTIVES RECOVERED:
RANDOM SCREENING
1364 MOLECULES
17 ACTIVES
COMPOUND LIBRARY
70 MOLECULES
10 ACTIVES
VIRTUAL SCREENING


COMPUTATIONAL
CHEMISTRY DOMAIN
RECEPTOR BASED DESIGN
•CRYSTALLOGRAPHY, NMR
•DOCKING SIMULATIONS
OR
LIGAND BASED DESIGN
•QSAR, SIMILARITY
SYNTHESIS
LEAD
COMPOUND
VIRTUAL OR
PHYSICAL
SCREENING
DRUG
DEVELOPMENT
DRUG
BIOLOGICAL EVALUATION
•IN VITRO ASSAYS
•IN VIVO TESTS
LEAD OPTIMIZATION
•DOCKING SIMULATIONS
•STRUCTURE-ACTIVITY RELATIONS
• ADME

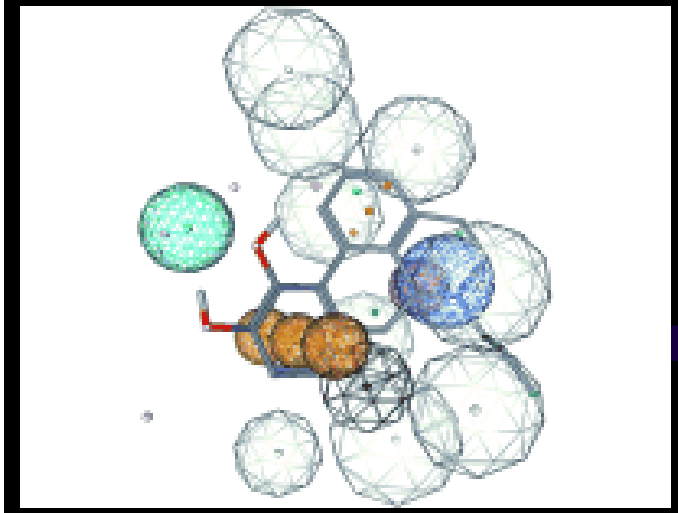
Drug likeness

- Enriching screening libraries with drug-like compounds
- “fail fast, fail cheap” strategy
- Manual classification is time-consuming and bias
- Computational approaches speeds up the screening, reduce the size and improves the quality of combinatorial libraries
- *Assumption: typical drugs have something in common that other compounds lack*

Cheminformatics

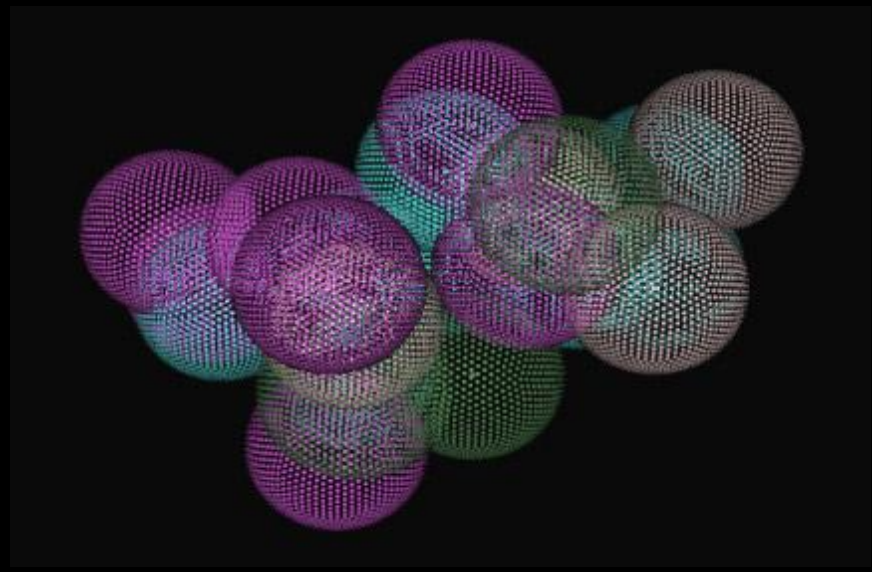
- Lipinski Rule of Five
- *Poor absorption and permeation are more likely to occur when there are*
  - *more than 5 hydrogen-bond donors,*
  - *more than 10 hydrogen-bond acceptors,*
  - *the molecular mass is greater than 500,*
  - *the log P value is greater than 5.*

Lipinski *et al*., *Adv. Drug Deliv. Rev*. 23, 3-25 (1997)



Ligand Based Drug Design

- Physicochemical Properties
  - H-bond donor-acceptor, ClogP, logD, charge distribution, pK
- Pharmacophore concept



Binary Fingerprints

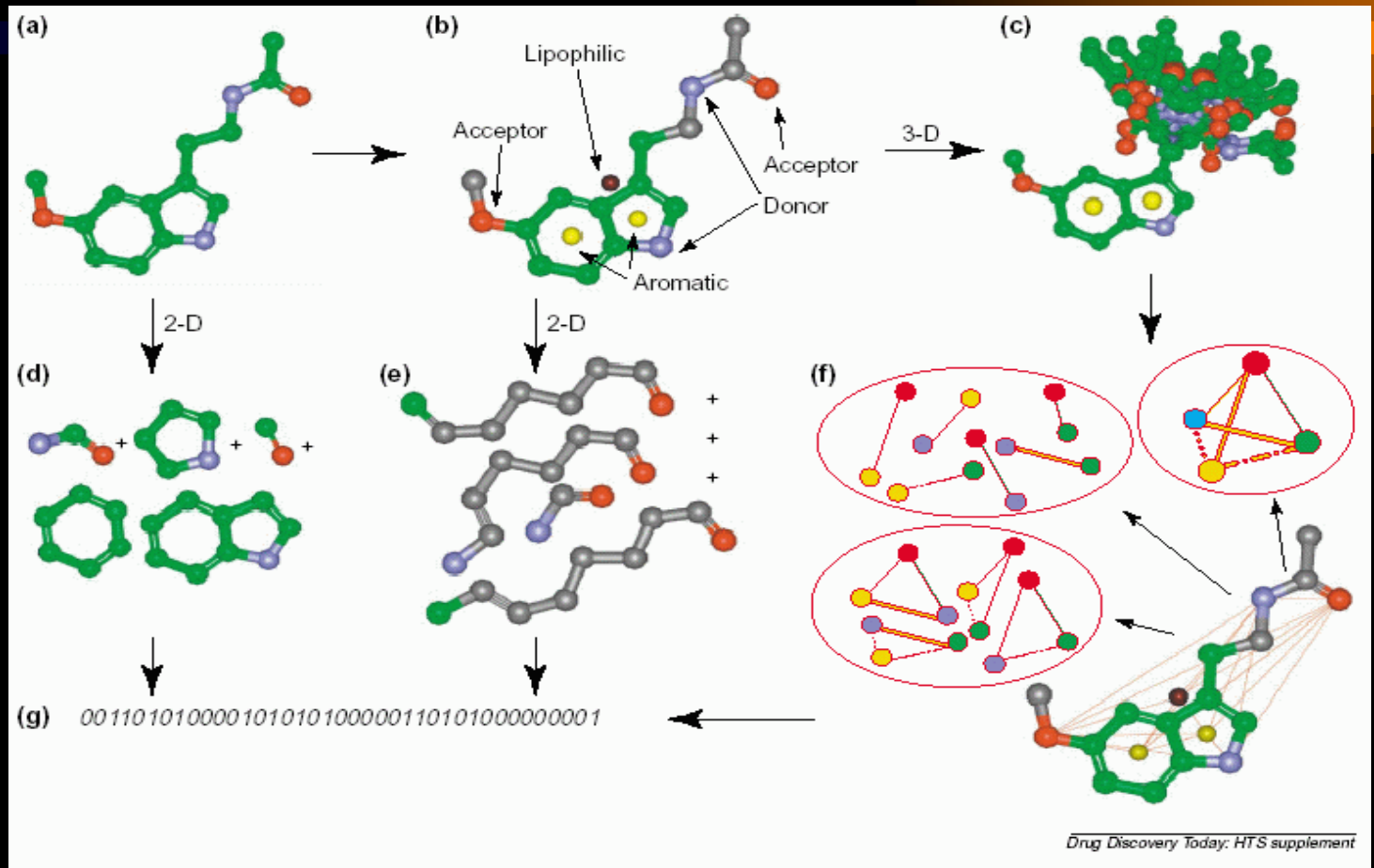
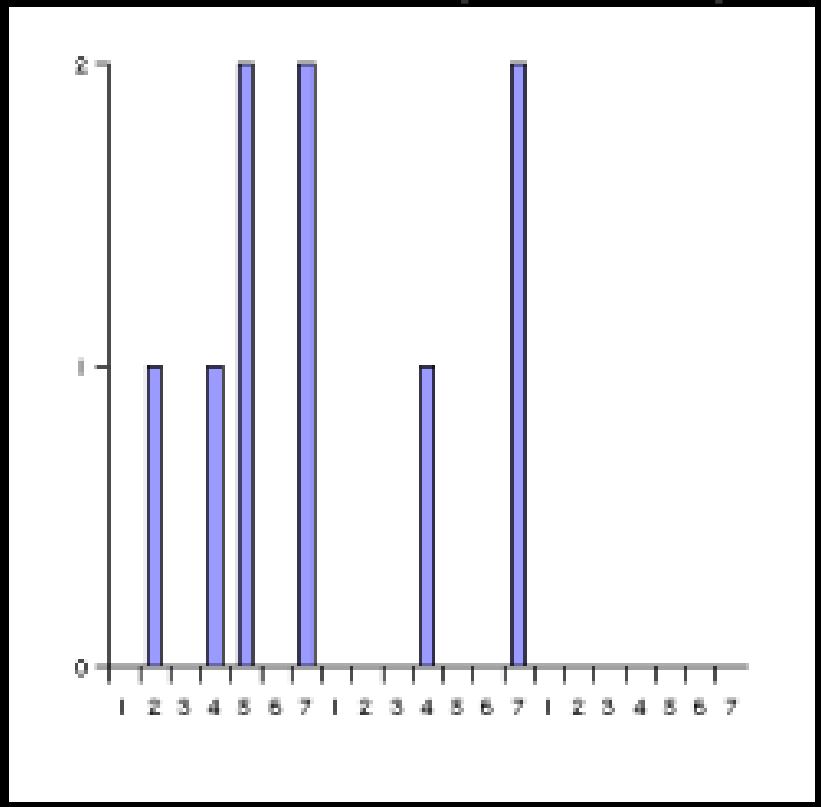


Figure 2. Schematic illustration of primary methods used in molecular fingerprint creation. **(a)** Create 2-D and 3-D model of molecule; **(b)** deconstruct the molecule into pharmacophoric elements; **(c)** generate conformational models; **(d)** deconstruct the molecule into topological/substructural elements; **(e)** determine distance between pharmacophoric groups using bond counts; **(f)** determine 2-, 3- or 4-center distance combinations of pharmacophoric groups for each conformer; and **(g)** determine the presence or absence of each descriptor element and combine to create a binary fingerprint.

Chemaxon Pharmacophore fingerprints

The pharmacophore feature pair histograms of captopril, and the pharmacophore fingerprint of captopril.



0	1	0	1	2	0	2	0	0	0	1	0	0	2	0	0	0	0	0	0	0

Similarity and Dissimilarity

Similarity and Dissimilarity

2D similarity based on groups + connectivity

e.g., Daylight fingerprints or MDL keys

2D similarity = Tanimoto index

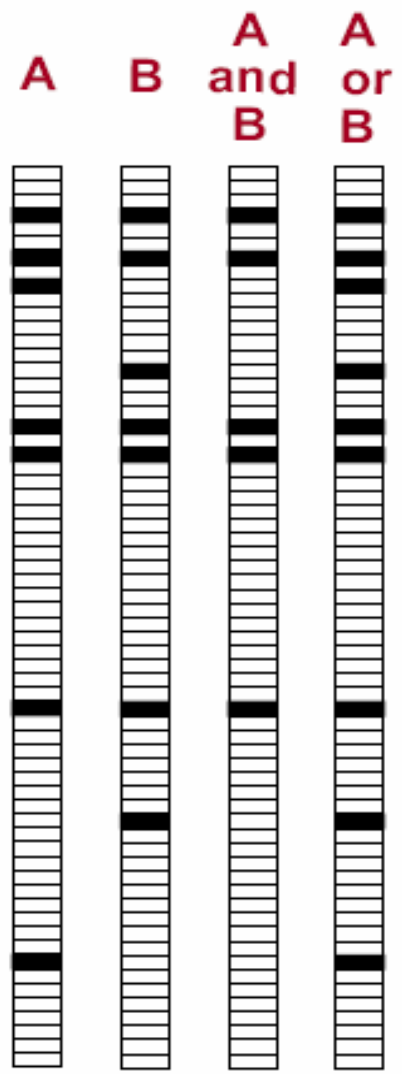
$$\frac{N_{AB}}{N_A + N_B - N_{AB}} = \frac{\text{\# bits set in A and B}}{\text{\# bits set in A or B}} =$$

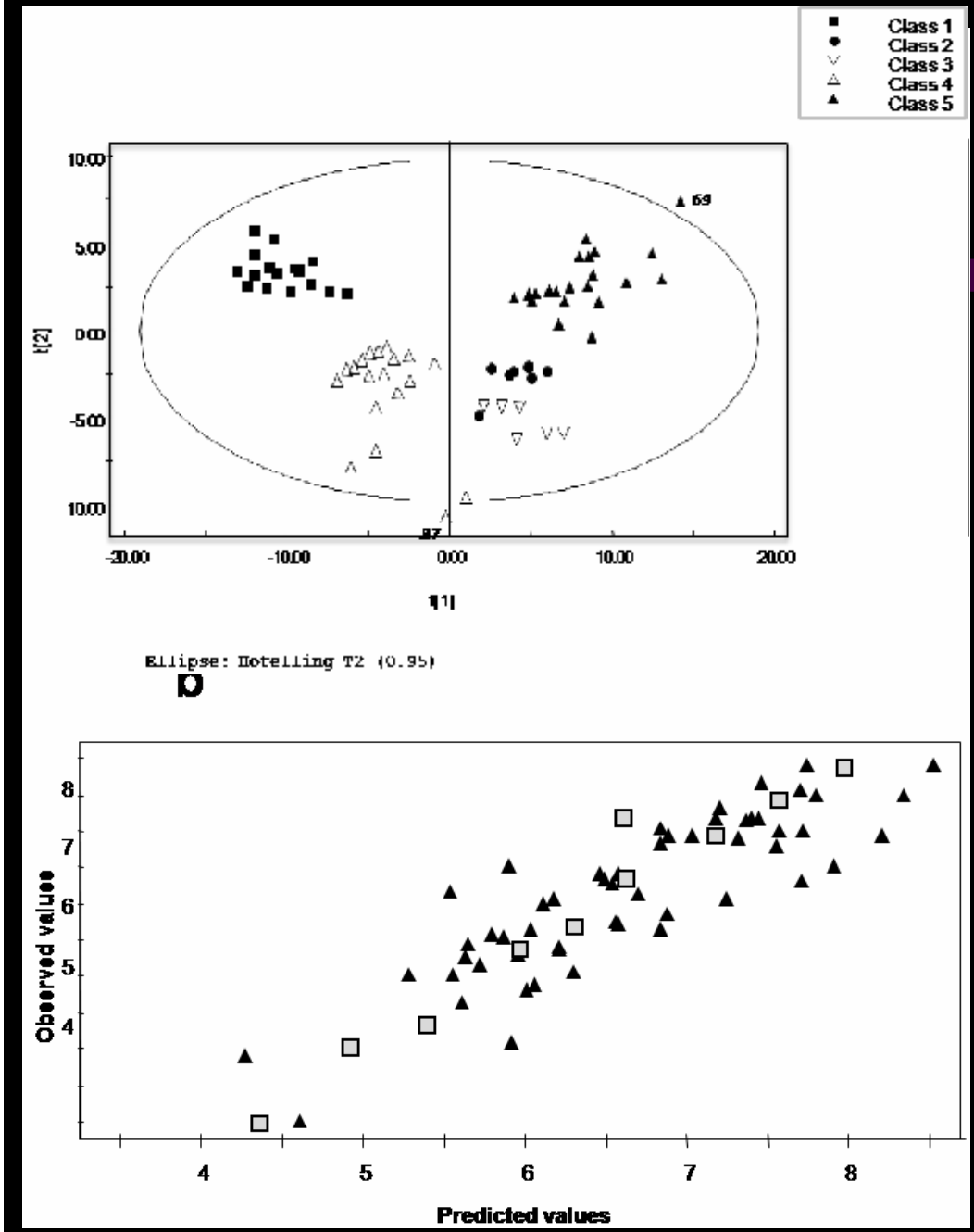
$$\frac{\text{\# keys common in A and B}}{(\text{\# keys in A})\ (\text{\# keys in B}) - (\text{\# keys common in A and B})}$$

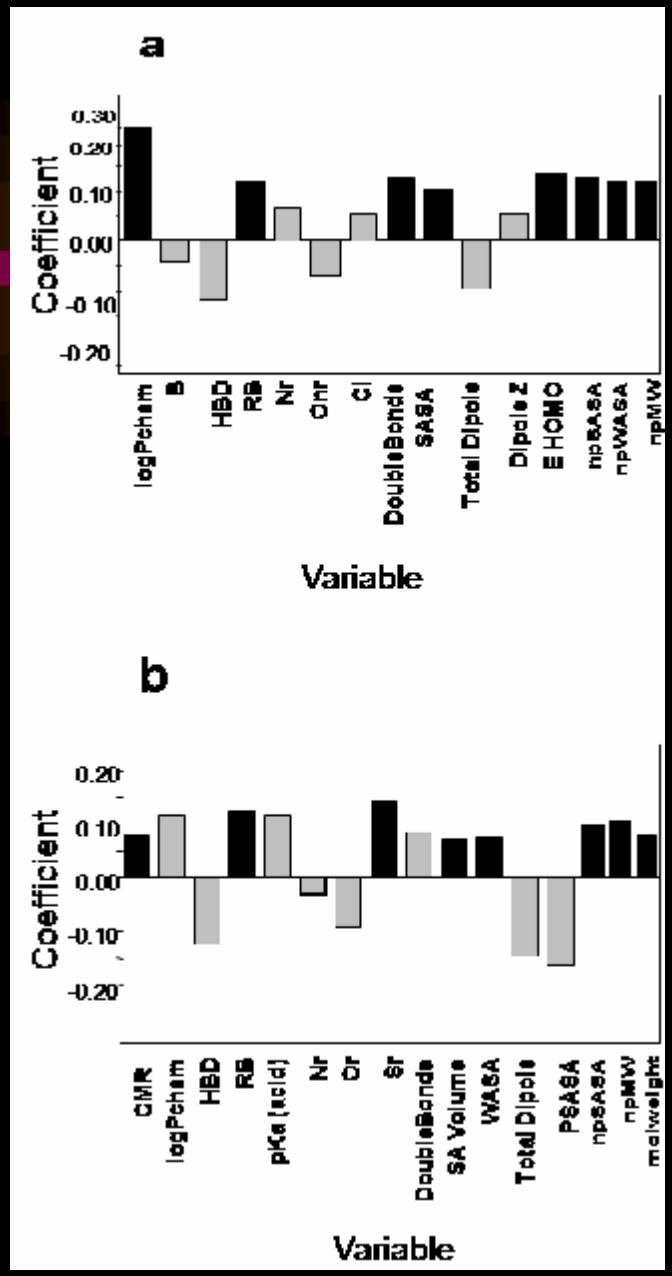
0 <= Tanimoto Index (i, j) <= 1

e.g. (example): T(A,B) = 5 / 9 = 0.555

2D dissimilarity = 1 - Tanimoto Index

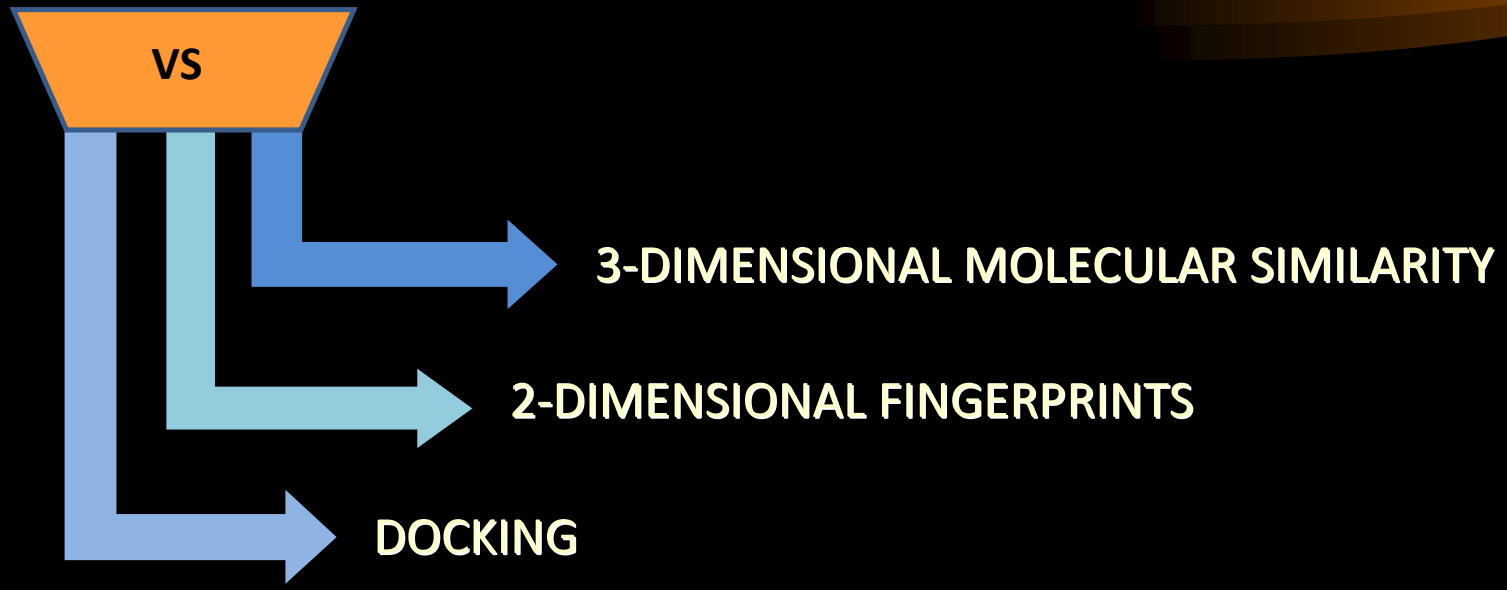



Class 1
Class 2
Class 3
Class 4
Class 5
t[2]
t[1]
69
87
Ellipse: Hotelling T2 (0.95)
b
Observed values
Predicted values


a
Coefficient
logPchem
B
HBD
RB
Nr
Or
Cl
DoubleBonds
SASA
Total Dipole
Dipole Z
E HOMO
npSASA
npWASA
npMW
Variable
b
Coefficient
CMR
logPchem
HBD
RB
pKa (acid)
Nr
Or
Sr
DoubleBonds
SA Volume
WASA
Total Dipole
PSASA
npSASA
npMW
molweight
Variable

RESULTS: VIRTUAL SCREENING

NCI Repository: 260.000 compounds

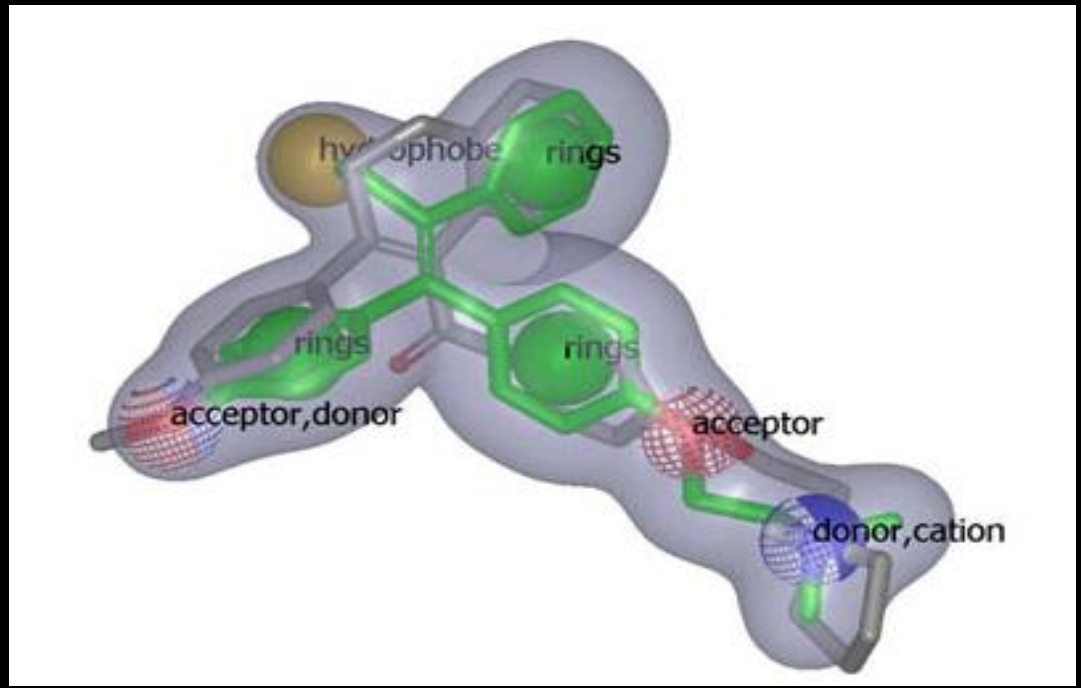


RESULTS: VIRTUAL SCREENING

NCI Repository: 260.000 compounds

VS

3-DIMENSIONAL MOLECULAR SIMILARITY

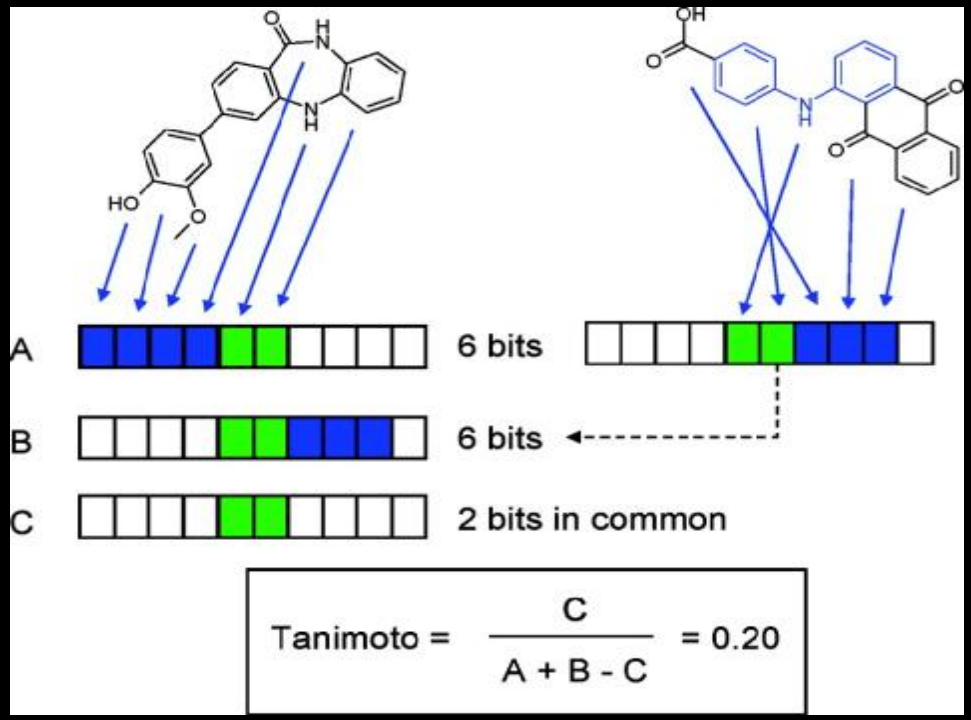


RESULTS: VIRTUAL SCREENING

NCI Repository: 260.000 compounds

VS

2-DIMENSIONAL FINGERPRINTS

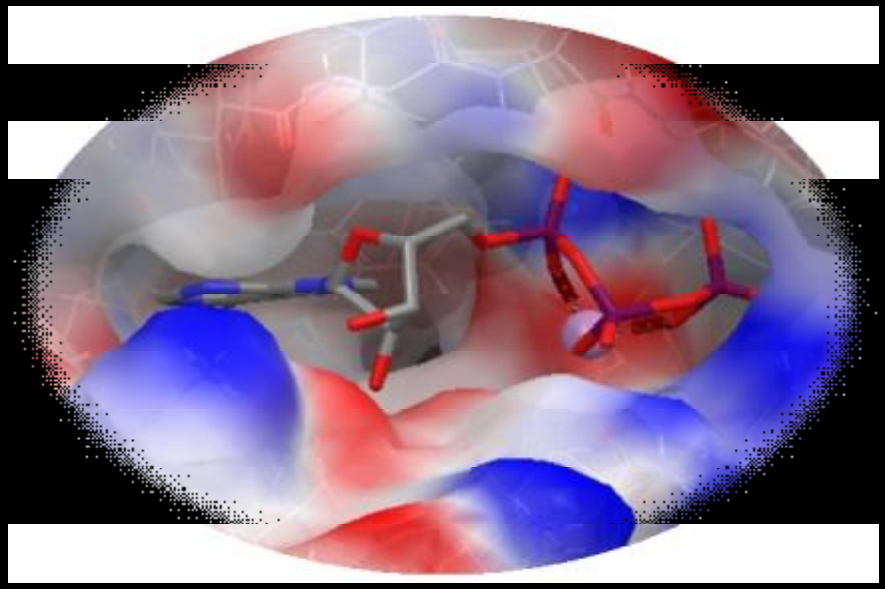


RESULTS: VIRTUAL SCREENING

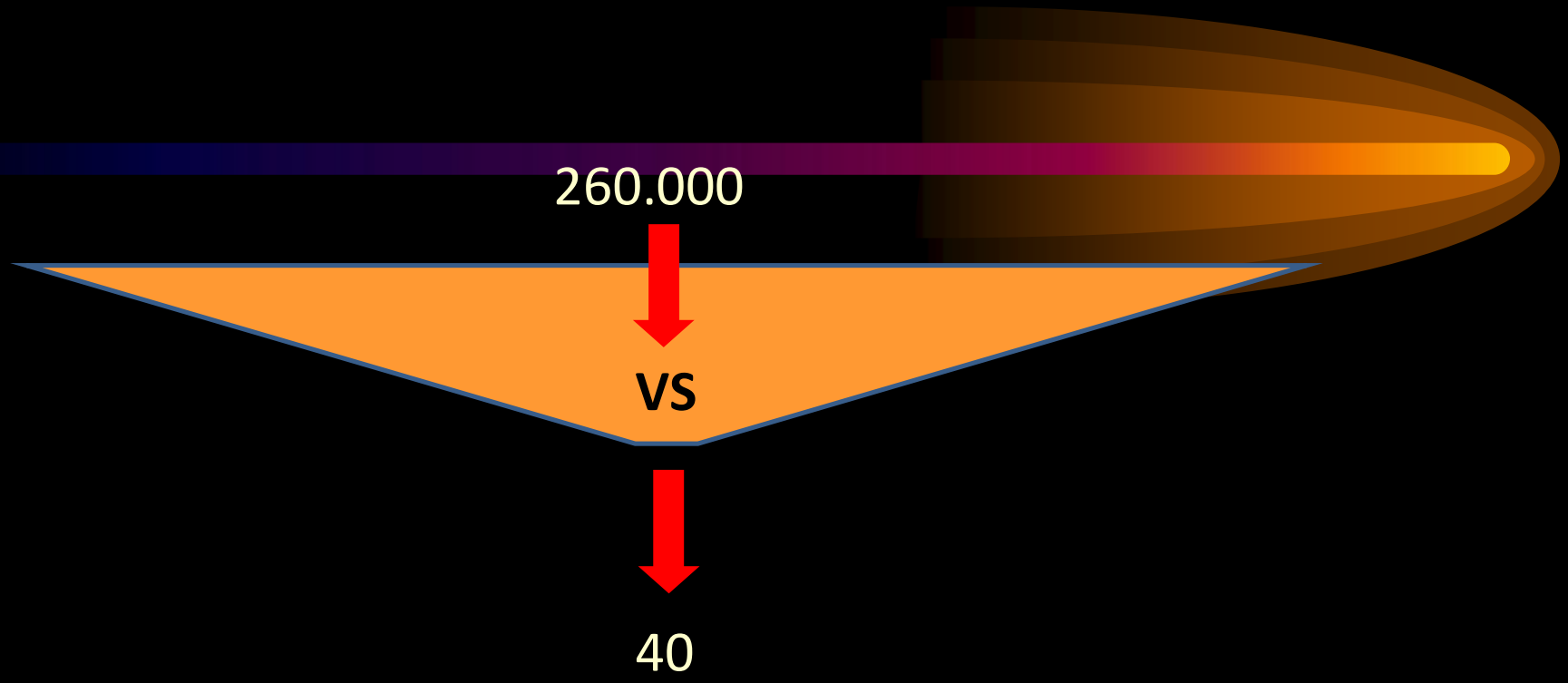
NCI Repository: 260.000 compounds

VS

DOCKING



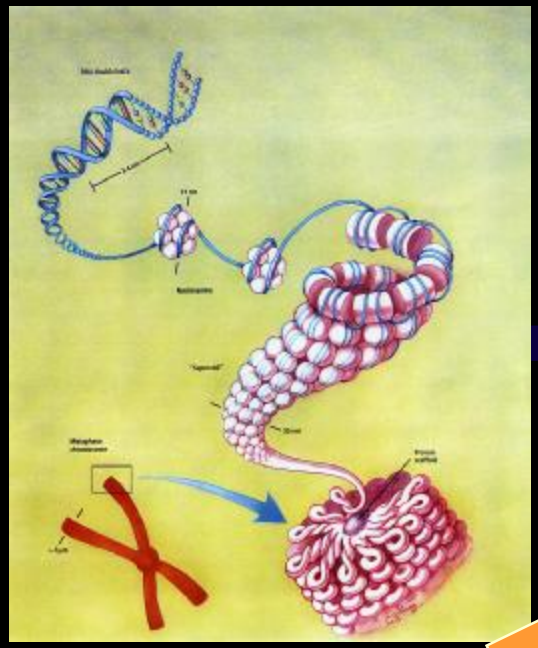
RESULTS: VIRTUAL SCREENING



Consensus of 3 different screening methodologies

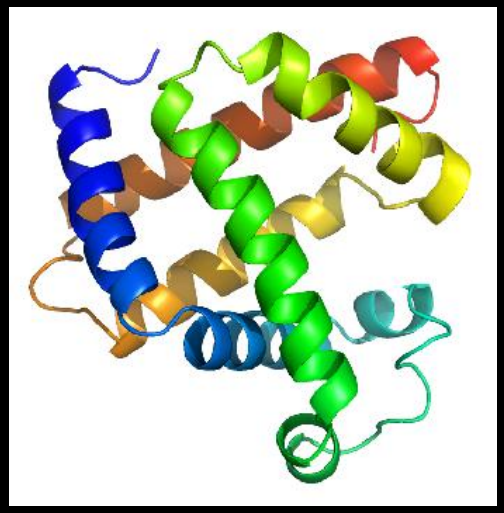
Selected compounds for
in vitro screening:

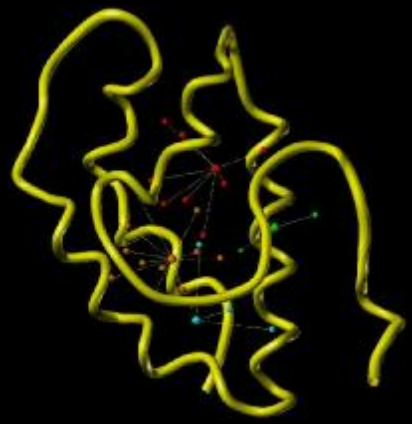
0.00015% of total collection



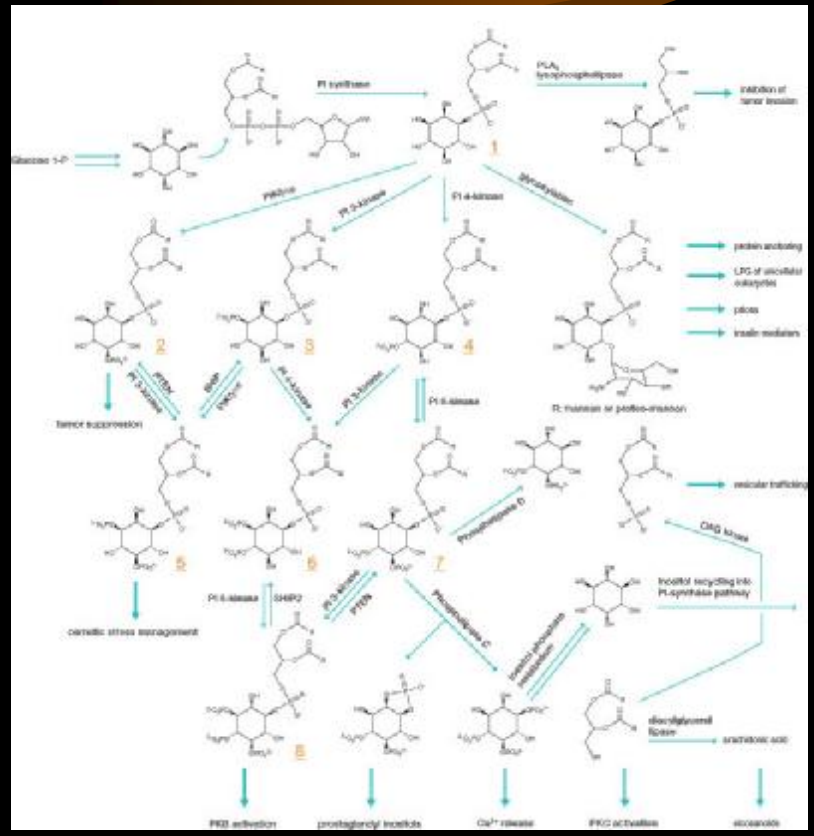
genome

Metabonomics





proteome

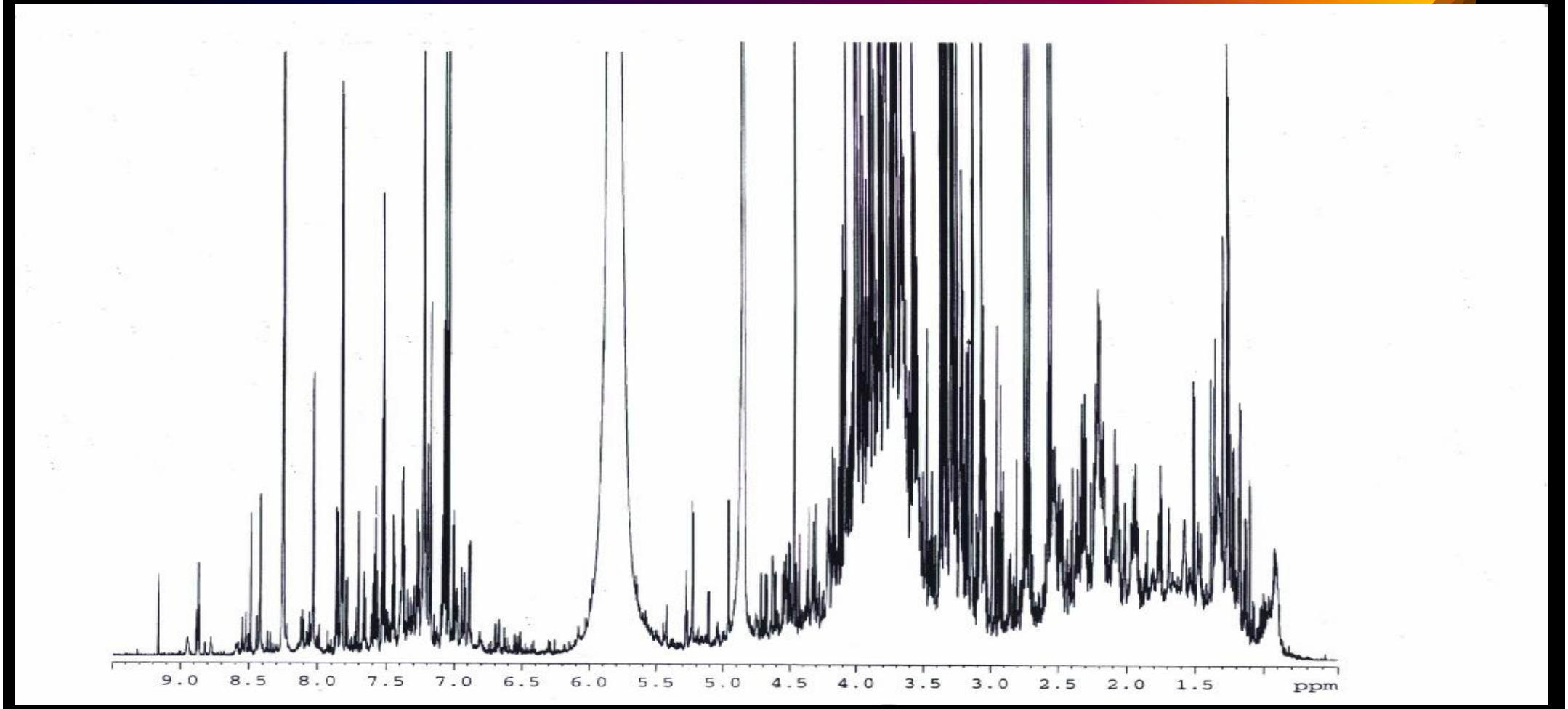


metabolome

Metabonomics

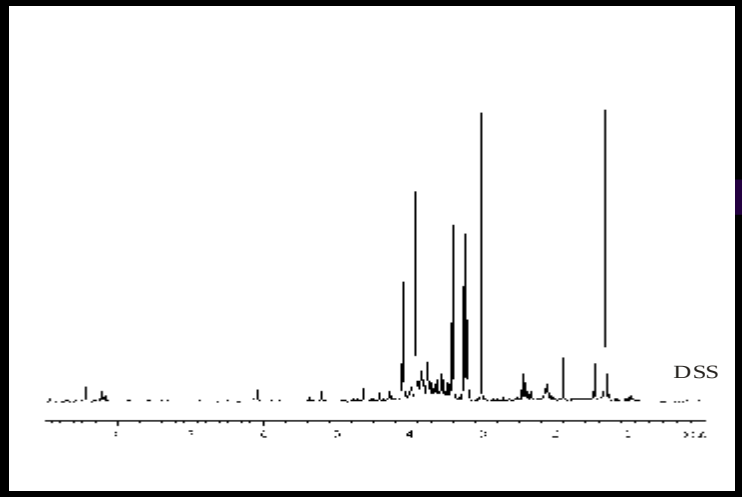
Quantitative measurement of multivariate metabolic responses of multicellular systems to pathophysiological stimuli or genetic modification

Metabonomics



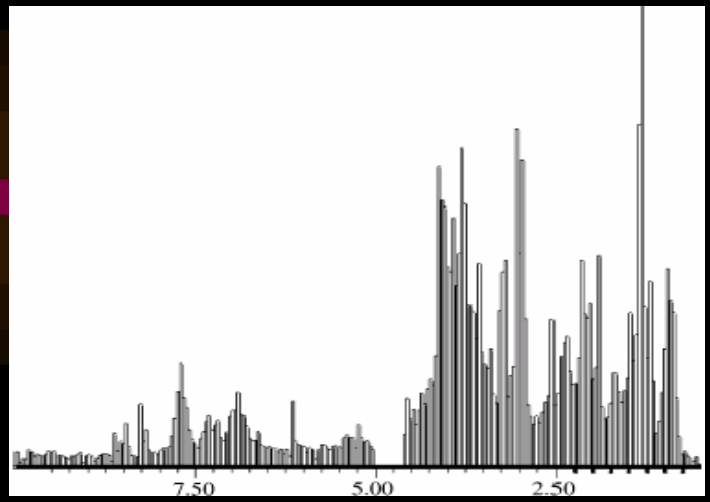
^{1}H NMR Spectrum of Untreated Human Urine

A procedure for Metabonomics



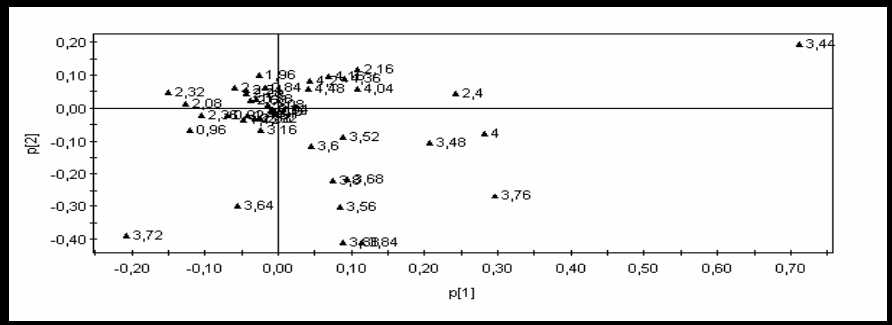
Tissue or biofluid sample

Measure the metabolite profile

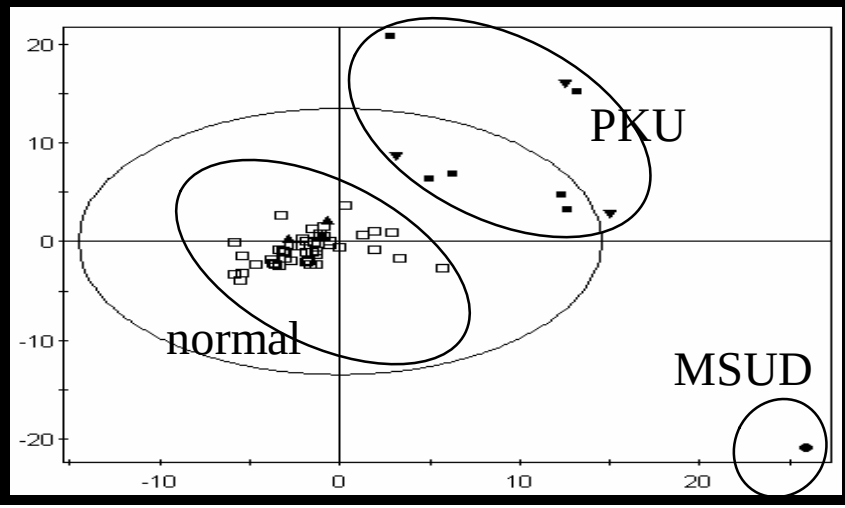


Explore metabolite profile to gain mechanistic insight

Treat metabolite profile as statistical ‘object’ for classification purposes

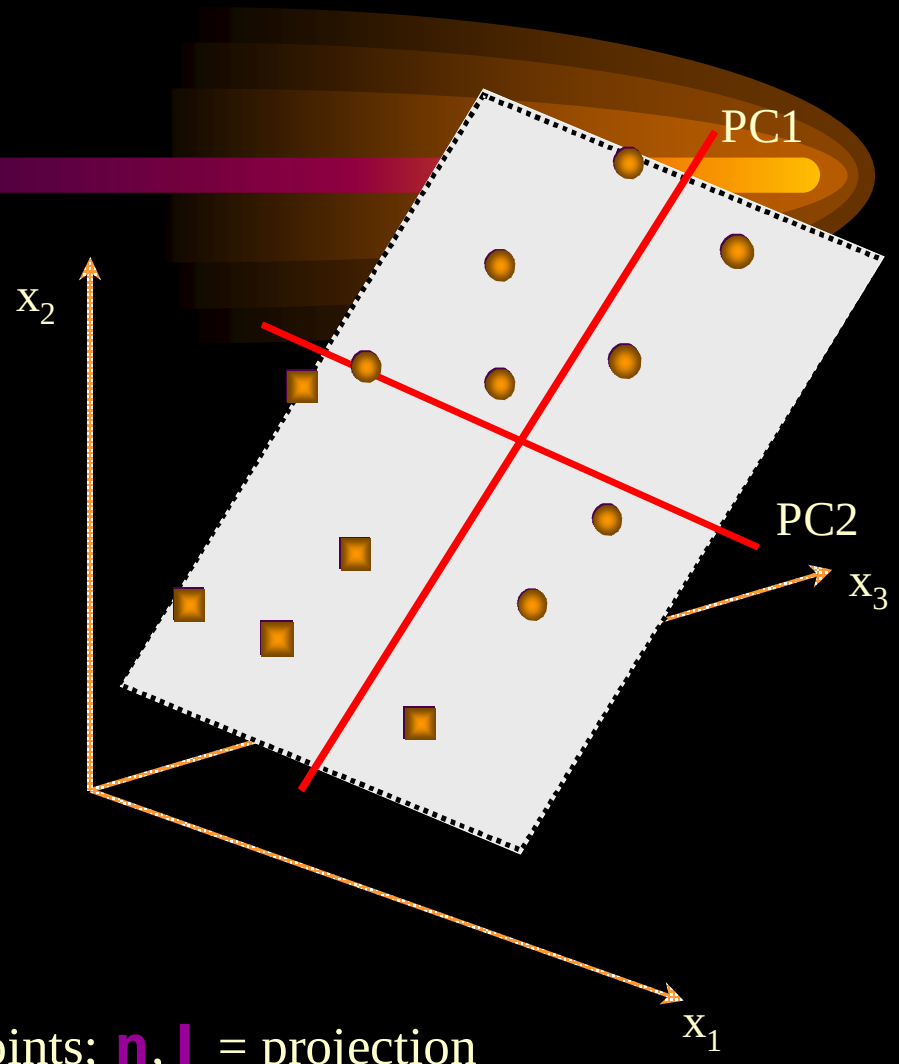


PCA



Application of NMR spectroscopy combined with principal component analysis in detecting inborn errors of metabolism using blood spots. A metabonomic approach
M.A. Constantinou, E. Papakonstantinou, M. Spraul, K. Shulpis, M.A. Koupparis, E. Mikros
Analytica Chimica Acta, 511, 303-312, 2004

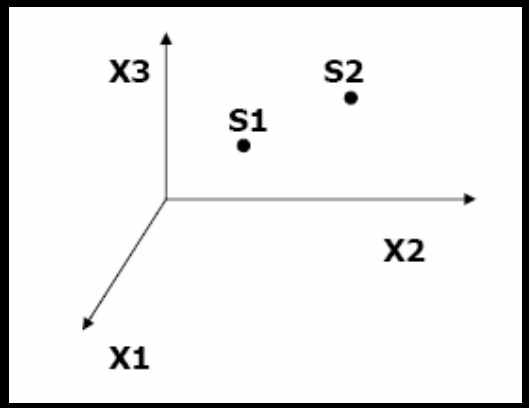
PCA theory – step by step



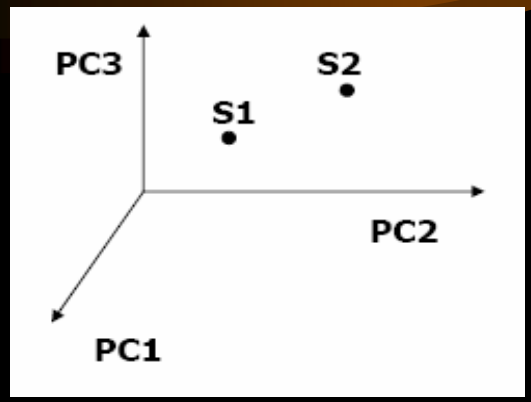
- Two PCs make a plane (window) in the K-dimensional variable space. The points are projected down onto the plane which is lifted out and viewed as a two dimensional plot.
- This is the scores plot
  - similarities or differences between samples can now be seen.

- A corresponding loading plot describes the variables relationships
  - allows interpretation of the scores plot by showing which variables are responsible for similarities and differences between samples.

n, **l** =data points; **n**, **l** = projection



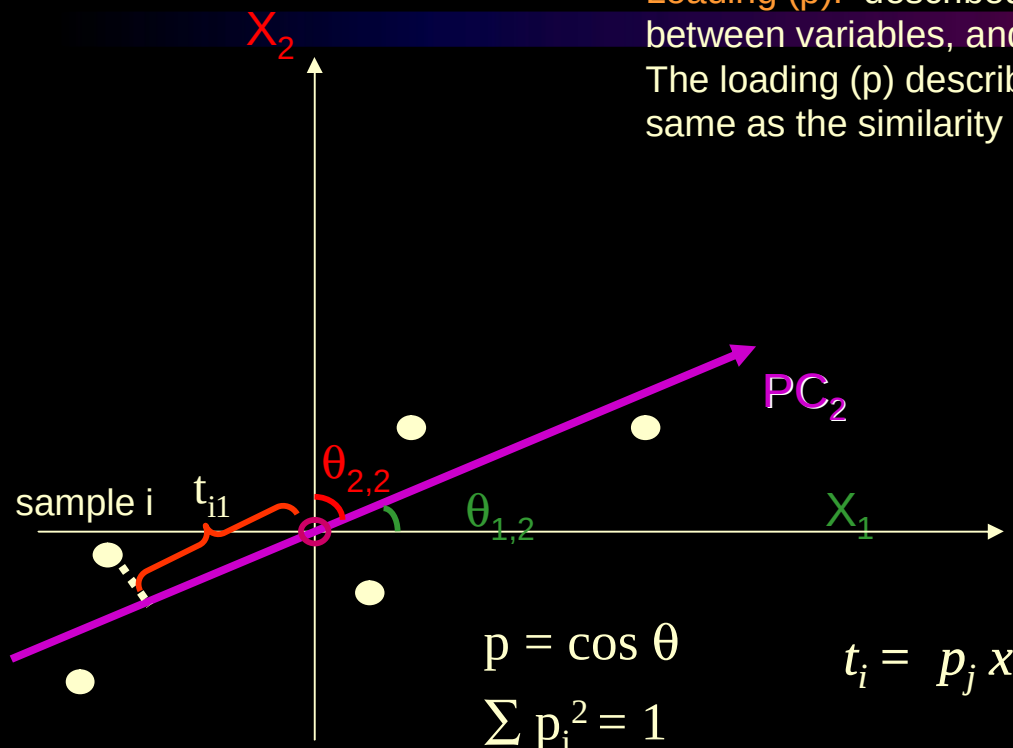
	X1	X2	X3
S1	a11	a12	a13
S2	a21	a22	a23



	PC1	PC2	PC3
S1	t11	t12	t13
S2	t21	t22	t23

The Loadings Plots

Loading (p): described the variation in the variable direction i.e. similarity/ dissimilarity between variables, and also explains the variation in scores.
The loading (p) describes the original variables importance for respective PC. This is the same as the similarity in direction between the original variable and the PC.

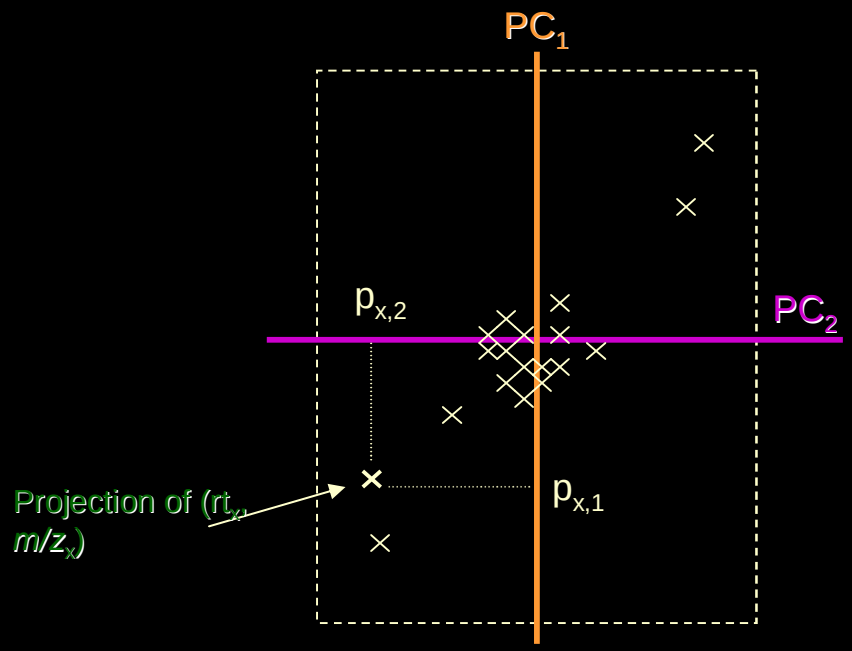


$p = \cos \theta$

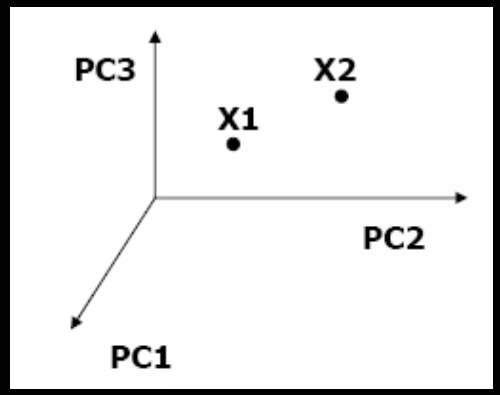
$\sum p_i^2 = 1$

$t_i = p_j x_{i,j}$

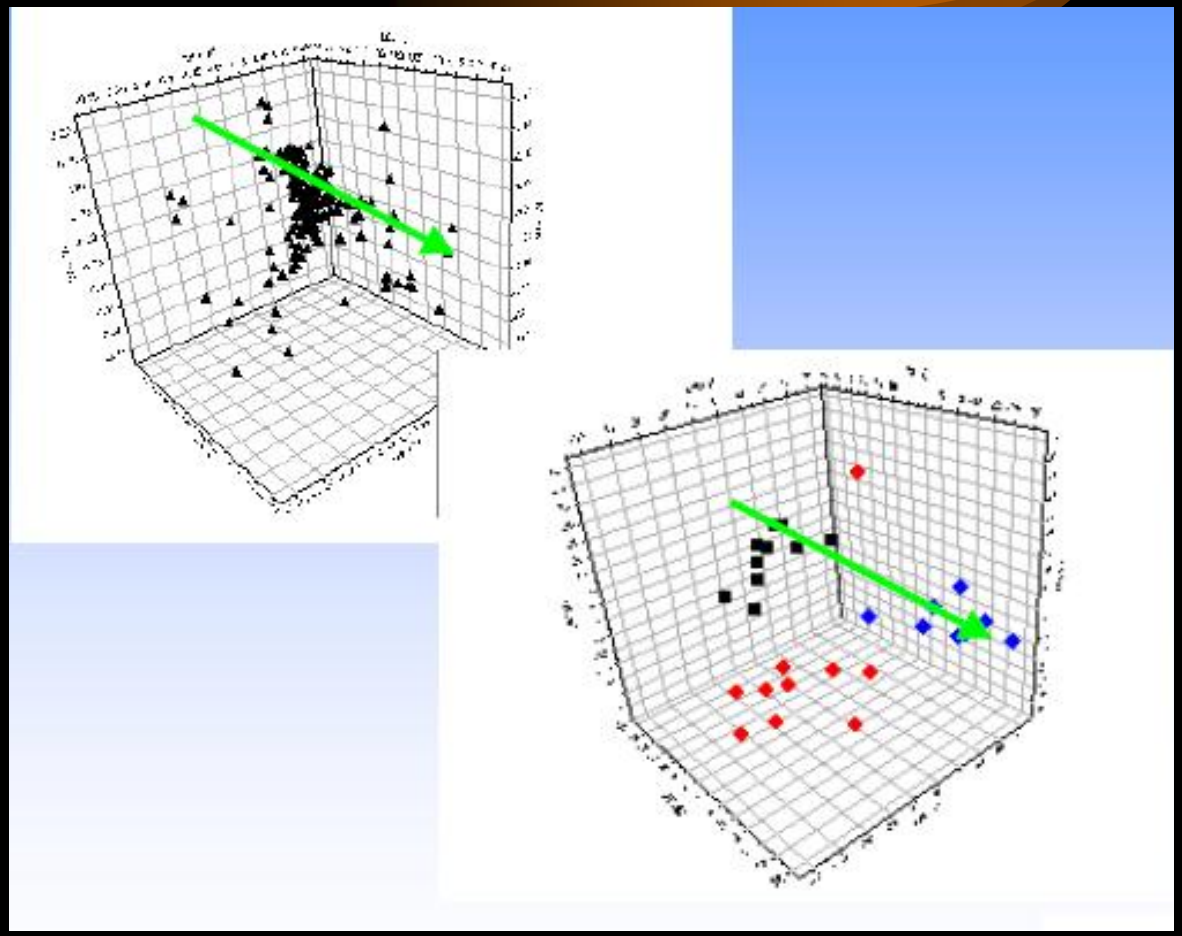
The loading (p) is described as the cosine of the angle between the original variable and the PC.



With $p_{x,1} = \cos(\theta_{x,1})$ and $p_{x,2} = \cos(\theta_{x,2})$
and $\theta_{x,1}$: angle between axe (rt_x, m/z_x) and PC1
and $\theta_{x,2}$: angle between axe (rt_x, m/z_x) and PC2

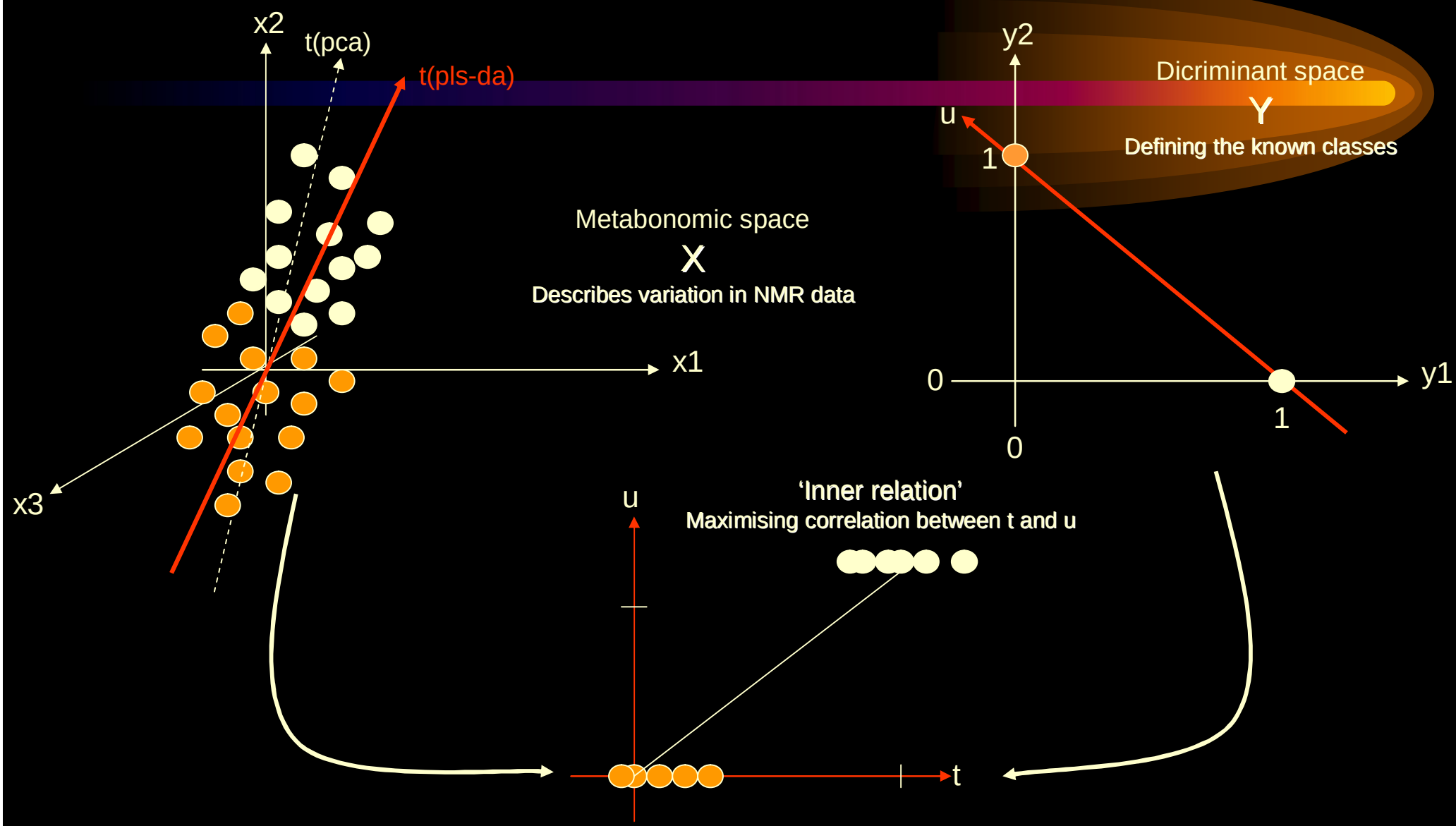


	PC1	PC2	PC3
X1	p11	p12	p13
X2	p21	p22	p23



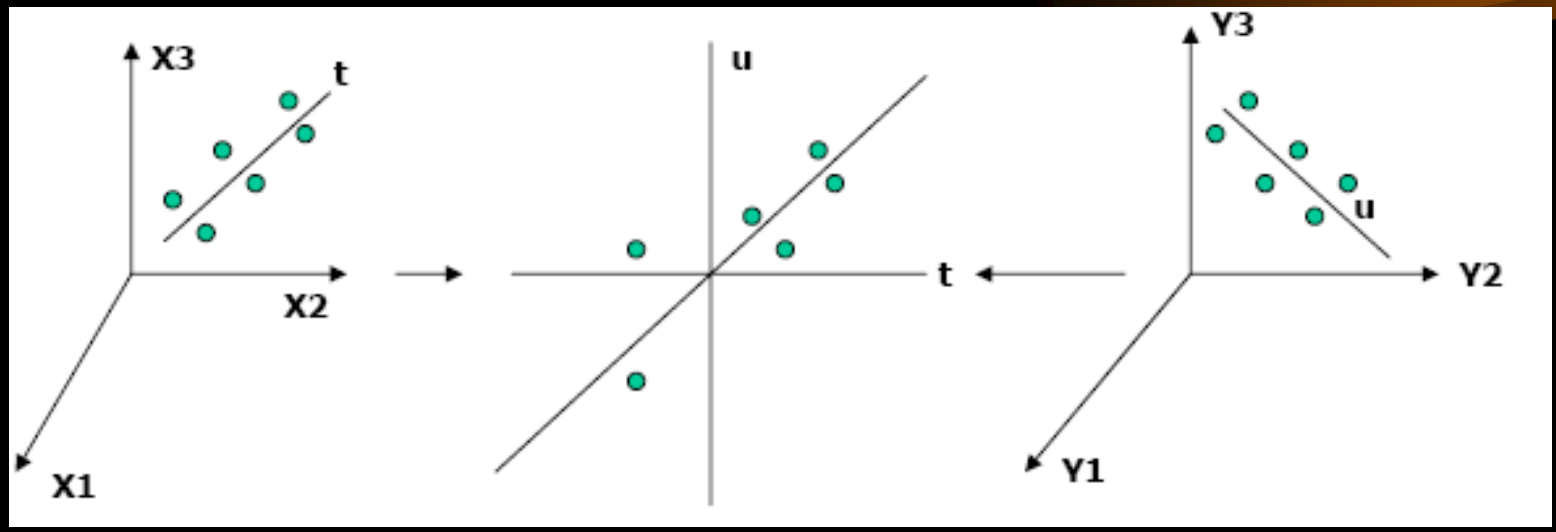
PLS-DA

Partial Least Squares or Projection to latent structure.



Partial Least Squares or Projection to latent structure.

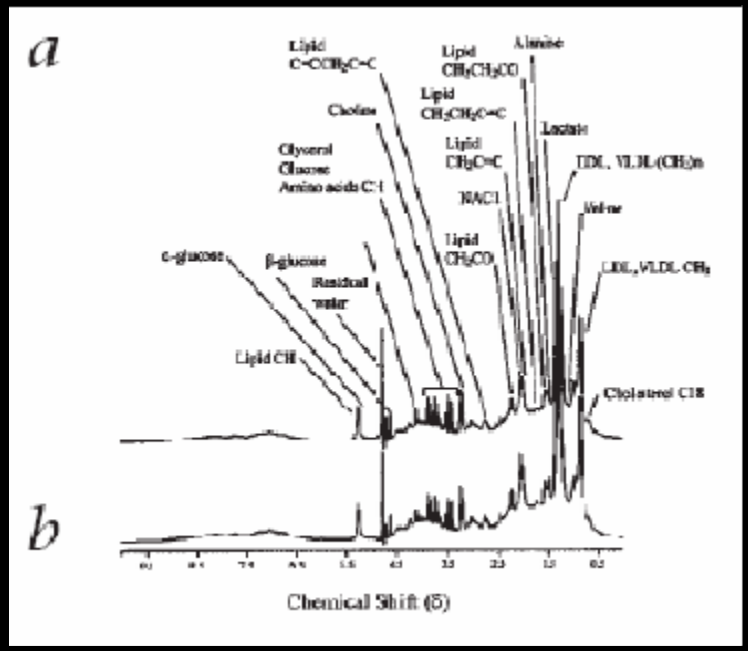
Partial least squares (PLS) is a method for constructing predictive models when the factors are many and highly collinear.

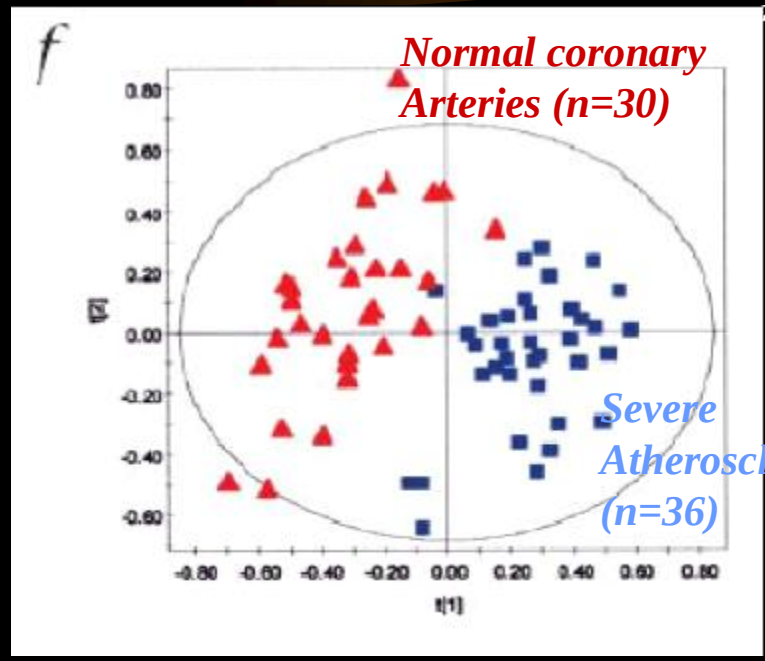


Models both the X & Y matrices simultaneously to find the latent variables in x that will predict the latent variables in Y the best. These PLS-Components are similar to principal components and will also be referred to as PCs.

Clinical Diagnosis

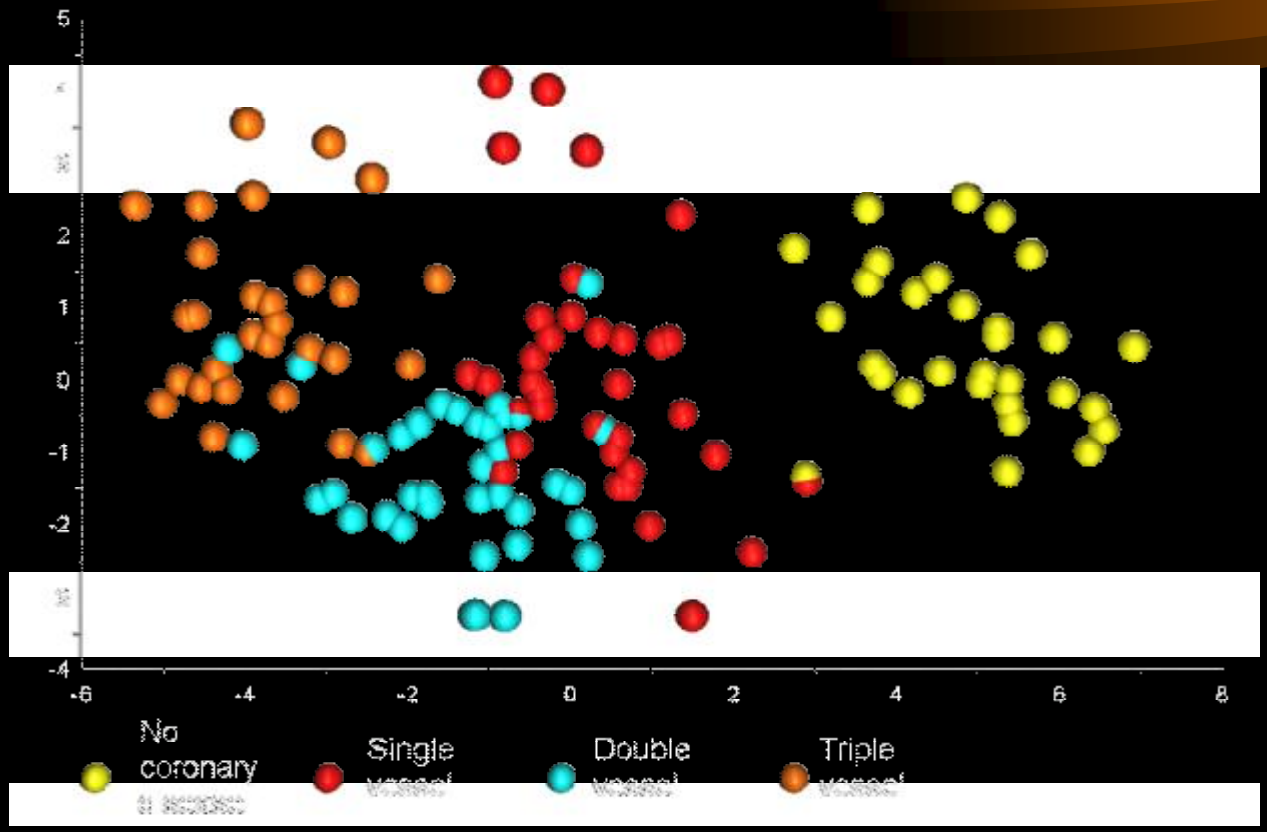
Predicting Coronary Artery Disease In Humans



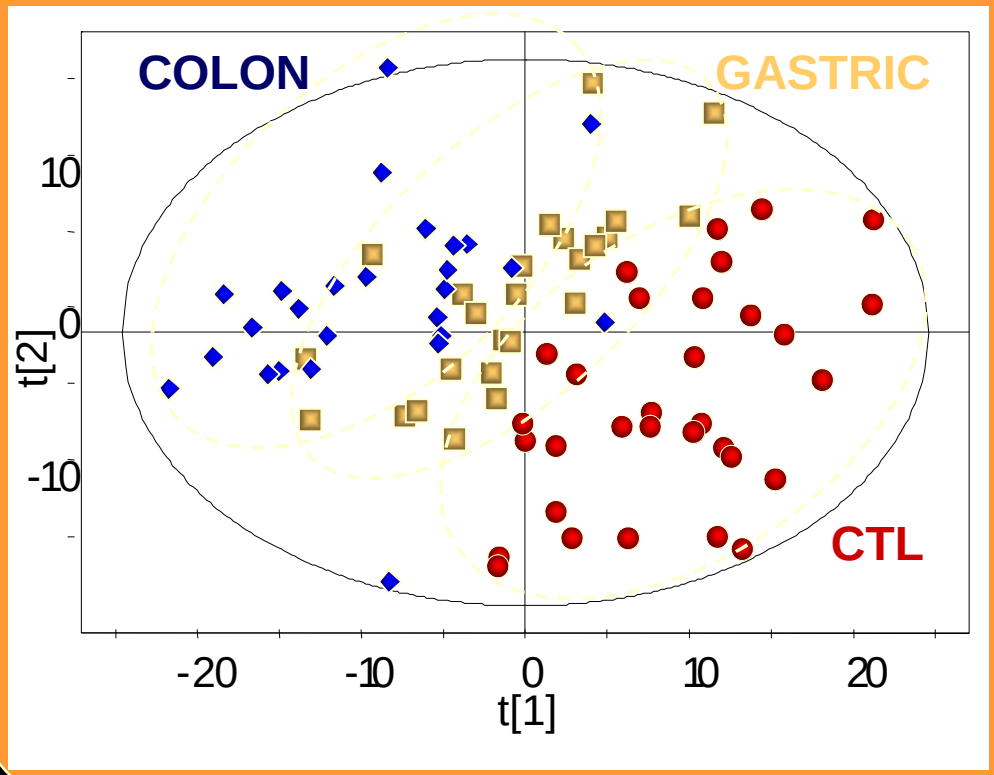


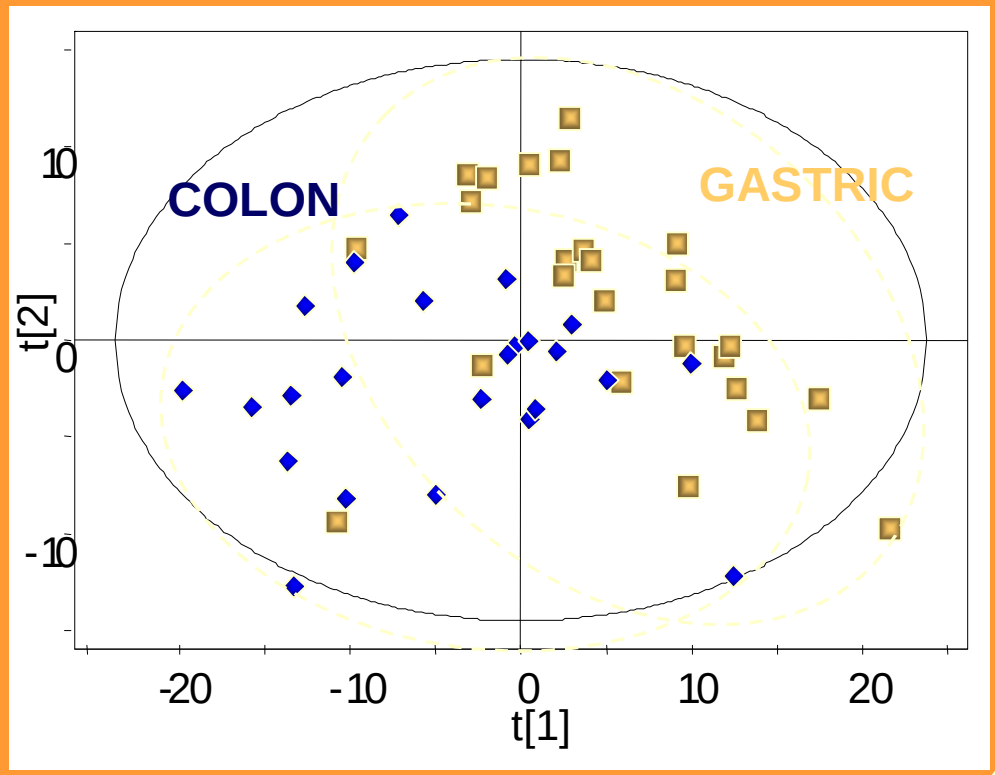
Rapid and noninvasive diagnosis of the presence and severity of coronary heart disease using 1H-NMR-based metabonomics
Joanne T. Brindle, Henrik Antti, Elaine Holmes, George Tranter, Jeremy K. Nicholson, Hugh W.L. Bethell, Sarah Clarke, Peter M. Schofield, Elaine McKilligin, David E. Mosedale & David J. Grainger Nature Medicine 8, 1439 - 1445 **(2002)**

Predicting Coronary Artery Disease In Humans



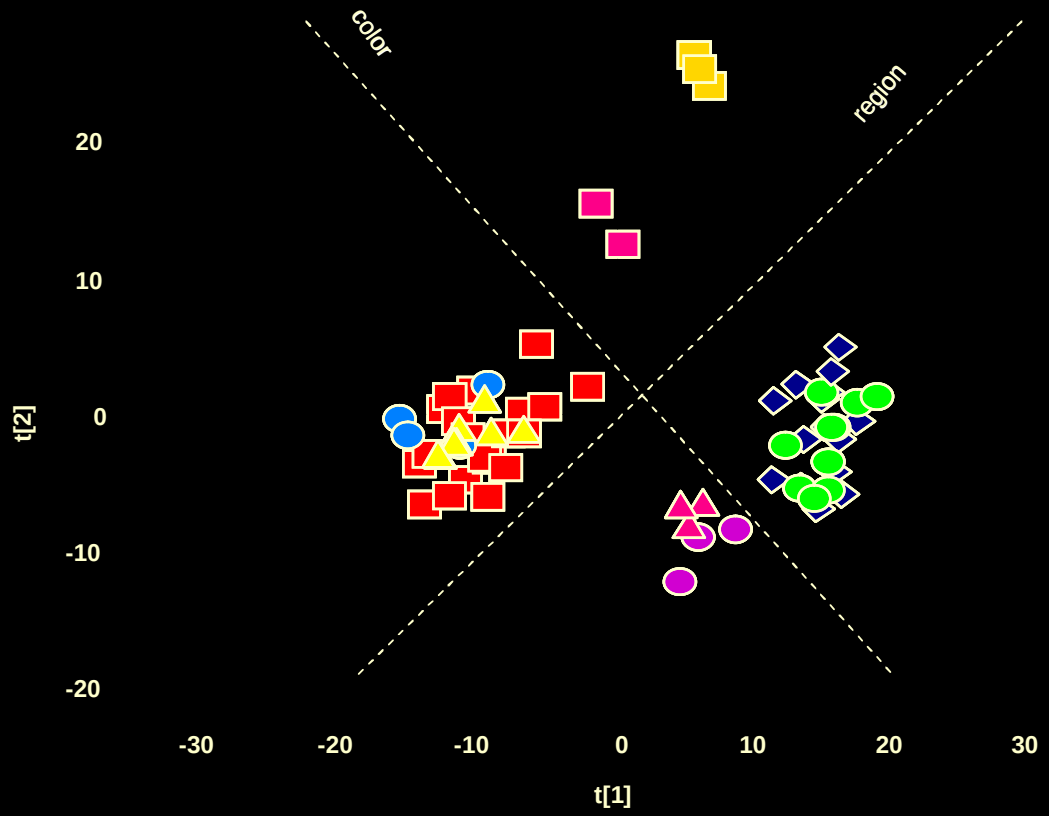
gastric and colon cancer





PCA

Greek Wines classification



1H NMR-Based Metabonomics for the Classification of Greek Wines According to Variety, Region and Vintage – Comparison with HPLC Data.
Anastasiadi, M;*J. Agr. Food. Chem.* (2009); 57; 11067-11074