From 3D coordinates to molecular surface

What we are still orphans:

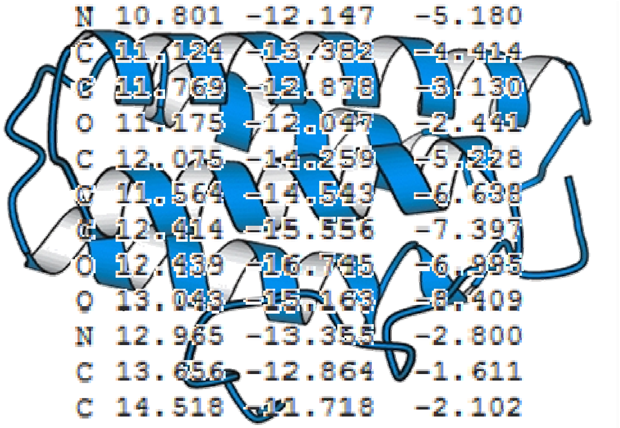
- Virtualize molecular topology (shape and volume);
- Virtualize the generation of alternative conformers;
- Virtualize the evaluation of the stability of each conformer.

We are coordinates hunters!

I would like to start from here!

NMR Spectroscopy

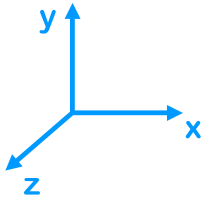
X-Ray Crystallography



```
N 10.801 -12.147  -5.180
C 11.124 -13.382  -4.414
C 11.769 -12.878  -3.130
O 11.175 -12.047  -2.441
C 12.075 -14.259  -5.228
C 11.564 -14.543  -6.638
C 12.414 -15.556  -7.397
O 12.439 -16.745  -6.995
O 13.043 -15.163  -8.409
N 12.965 -13.355  -2.800
C 13.656 -12.864  -1.611
C 14.518 -11.718  -2.102
```

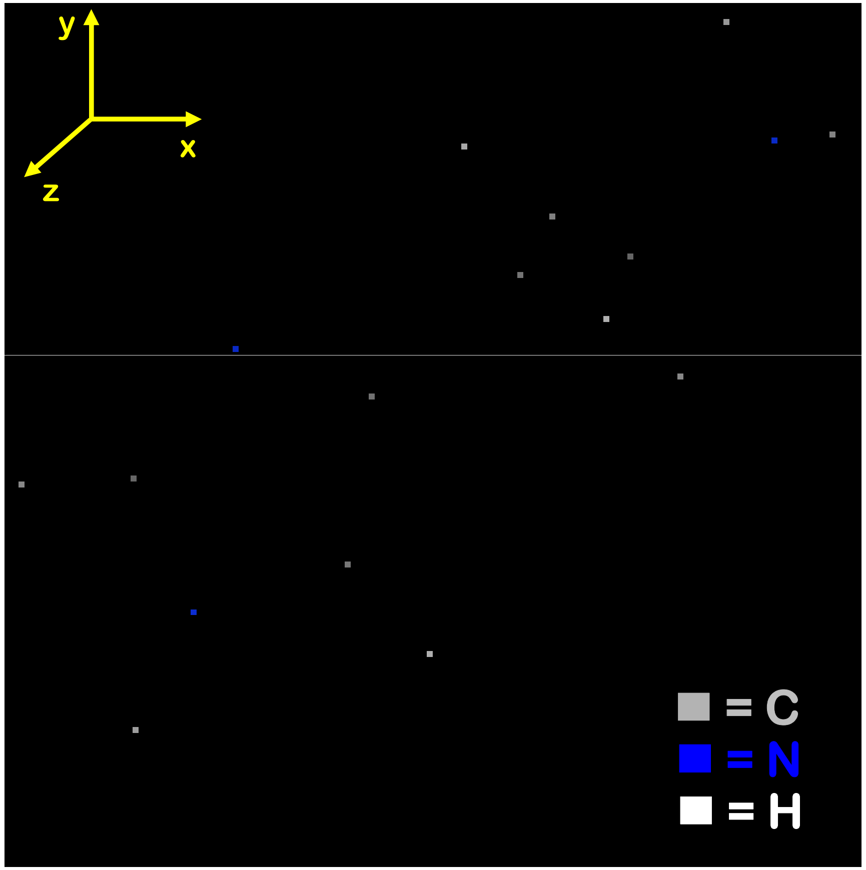
Compu Methods

3D



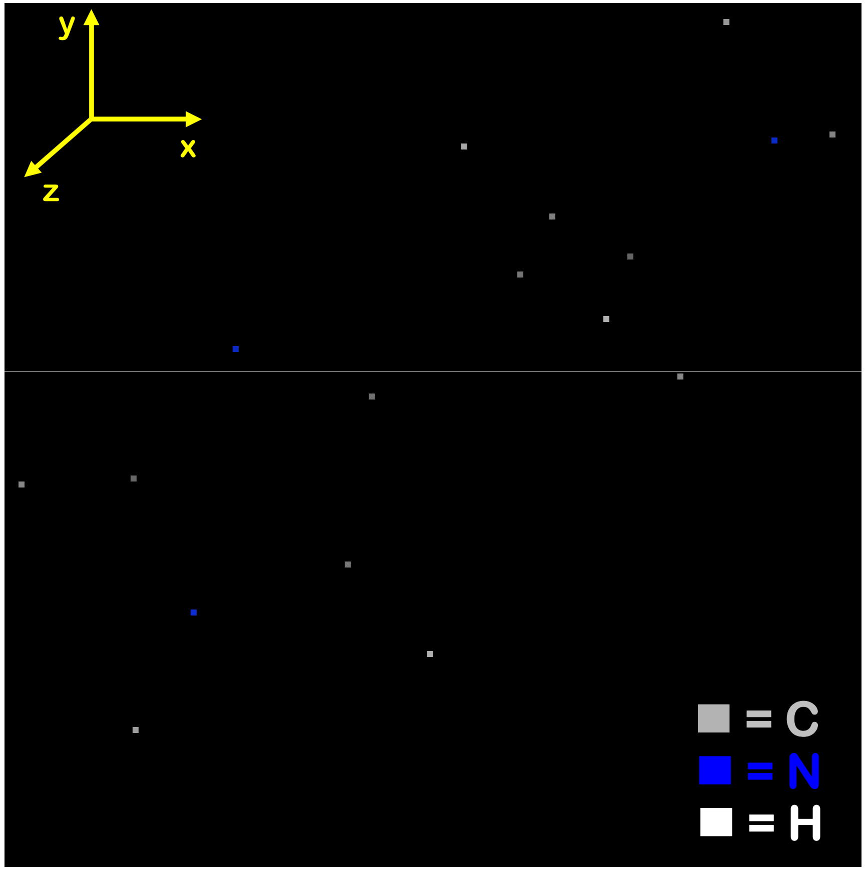
Let's go back to our Cartesian coordinate ...

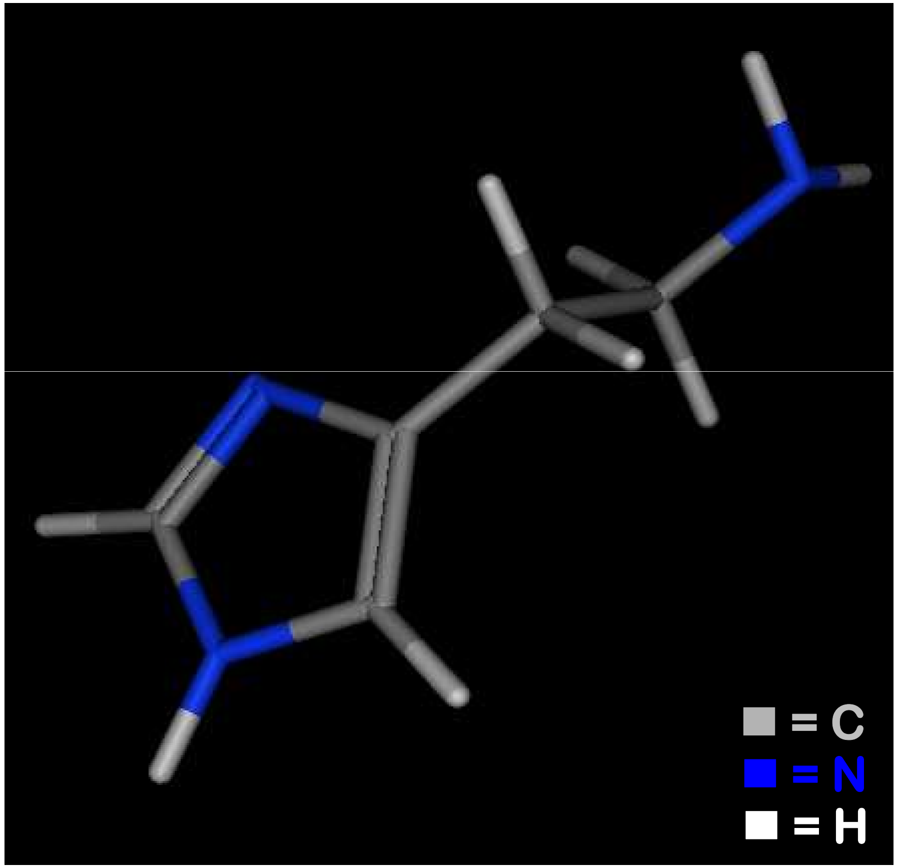
0.9760	0.5530	0.1180	C
2.1810	1.4240	0.1010	C
2.9110	1.3580	-1.2420	C
2.2460	1.6700	-2.0550	H
3.2180	0.3270	-1.4520	H
4.1000	2.2110	-1.2400	N
3.8190	3.1840	-1.1170	H
4.5460	2.1620	-2.1560	H
2.8610	1.1260	0.9100	H
1.8740	2.4560	0.3100	H
-0.1610	0.9170	-0.5670	N
-1.0220	-0.0510	-0.3510	C
-2.0310	-0.1010	-0.7380	H
-0.4920	-1.0270	0.4440	N
-0.9630	-1.8680	0.7470	H
0.7870	-0.6560	0.7530	C
1.4290	-1.2620	1.3760	H



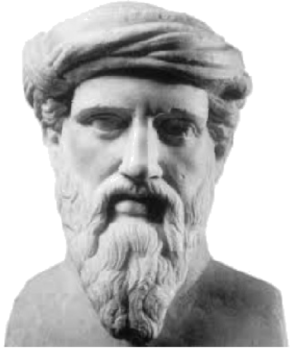
Rendering!

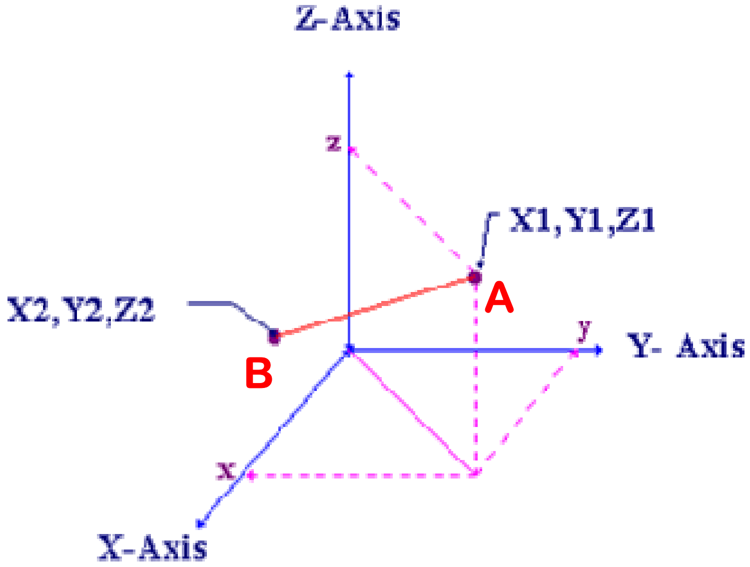
..or show me *privileged* distances!



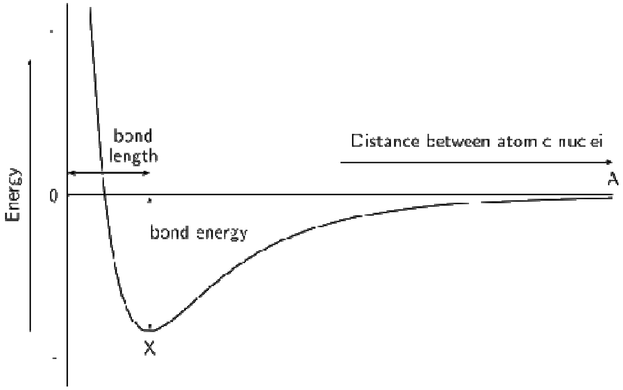


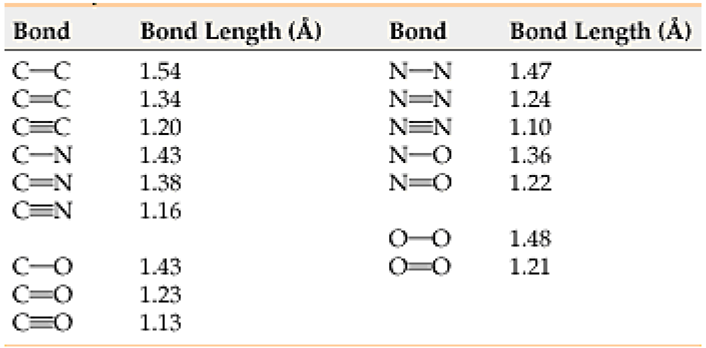
What can we easily measure in a 3D Cartesian space?





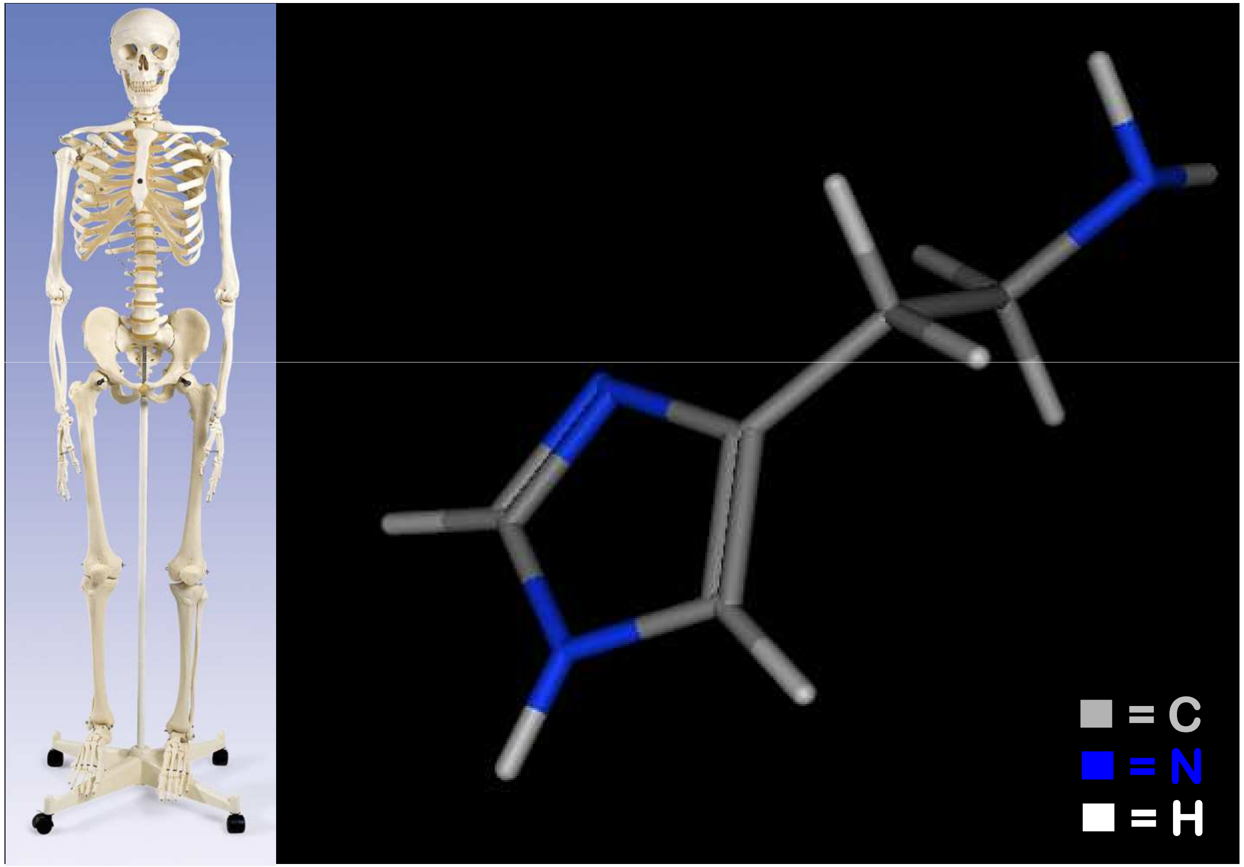
$$\overline{AB} = \sqrt{(X1 - X2)^2 + (Y1 - Y2)^2 + (Z1 - Z2)^2}$$





Bond	Bond Length (Å)	Bond	Bond Length (Å)
C—C	1.54	N—N	1.47
C=C	1.34	N=N	1.24
C≡C	1.20	N≡N	1.10
C—N	1.43	N—O	1.36
C=N	1.38	N=O	1.22
C≡N	1.16		
		O—O	1.48
C—O	1.43	O=O	1.21
C=O	1.23		
C≡O	1.13		

But how can we restore physicality to our molecular models?



Atomic radius… could be one?

The atomic radius of a chemical element is a measure of the size of its atoms, usually the mean or typical distance from the nucleus to the boundary of the surrounding cloud of electrons.

Since the boundary is not a well-d ned physical entity, there are various non-equivale definitions of atomic radius. Three widely used de itions of atomic radius are *van der Waals radius, ionic radius, and covalent radius.*

Johannes Diderik van der Waals (23 November 1837 – 8 March 1923) was a Dutch theoretical physicist and thermodynamicist famous for his work on an equation of state for gases and liquids.

His name is primarily associated with the van der Waals equation of state that describes the behavior of gases and their condensation to the liquid phase. His name is also associated with van der Waals forces (forces between stable molecules), with van der Waals molecules (small molecular clusters bound by van der Waals forces), and with van der Waals radii (sizes of molecules).

He became the first physics professor of the University of Amsterdam when it opened in 1877 and won the 1910 Nobel Prize in physics.

WIKIPEDIA
The Free Encyclopedia

van der Waals radius: in principle, half the minimum distance between the nuclei of two atoms of the element that are not bound to the same molecule.

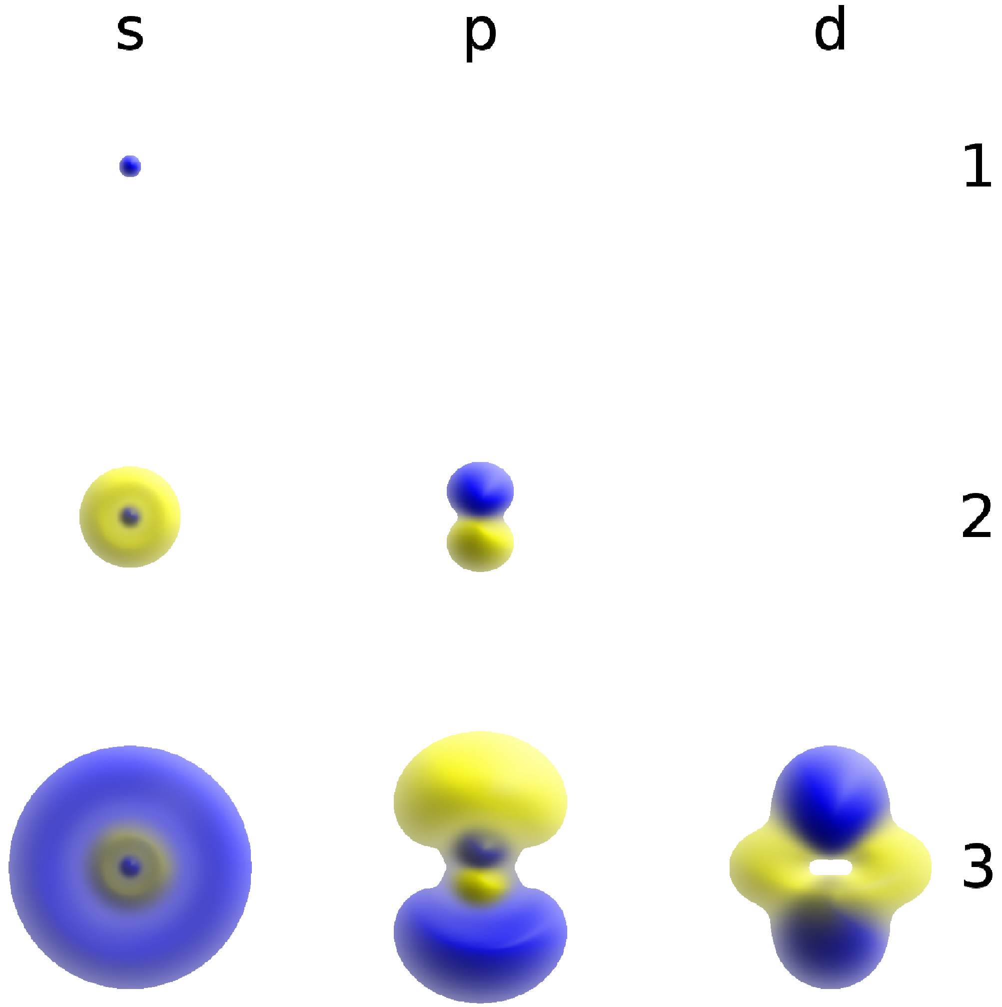
A quick refresh…

Element	radius (Å)
Hydrogen	1.20
Carbon	1.70
Nitrogen	1.55
Oxygen	1.52
Fluorine	1.47
Phosphorus	1.80
Sulfur	1.80
Chlorine	1.75
Copper	1.4

$$V_{vdW} = \frac{4}{3}\pi\, r_{vdW}^{3}$$

$$r_{vdW} = \sqrt[3]{V_{vdW}\,\frac{3}{4\pi}}$$

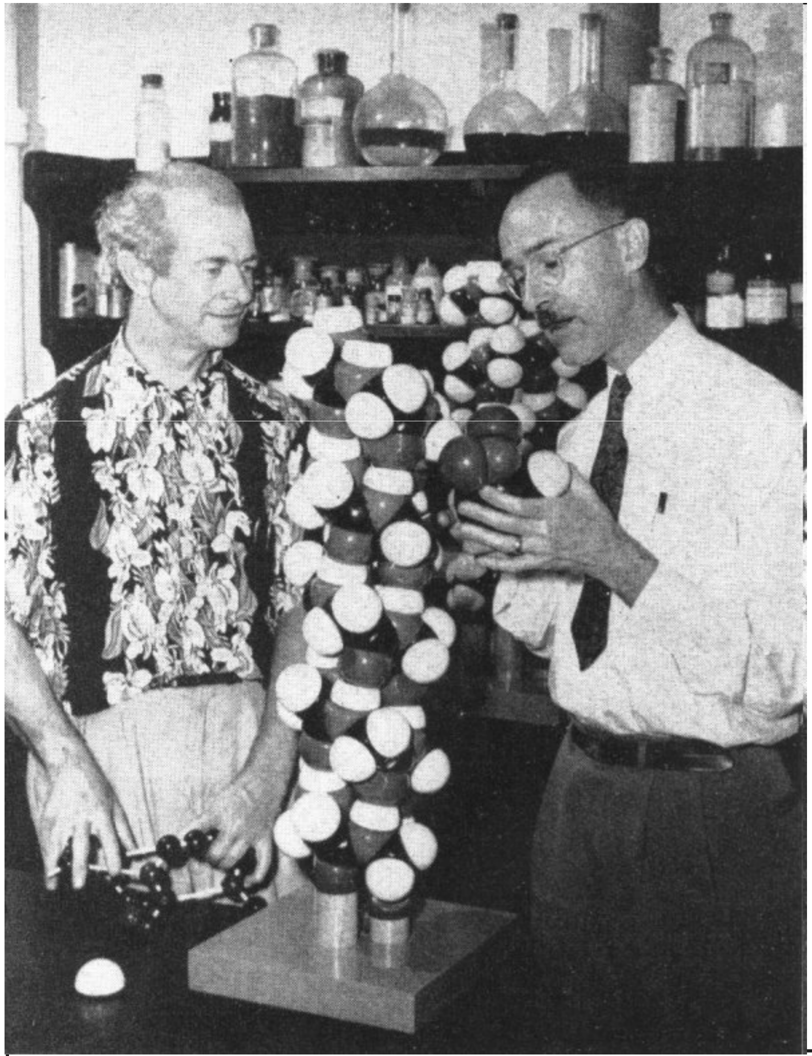
Why an atom has a spherical shape?



The images are 3D renderings of the spatial density distribution of $|\psi|^2$ with the color depicting the phase of ψ. The spatial distribution is smooth and vanishes for large radii. The cloud is a more realistic representation of an orbital than the more common solid-body approximations.

Credits: https://en.wikipedia.org/wiki/Electron#/media/File:Atomic-orbital-clouds_spd_m0.png

From Cartesian coordinates to van der Waals spheres:



Robert B. Corey and Linus Pauling “Molecular Models of Amino Acids, Peptides, and Proteins.” Rev. Sci. Instrum. 24, 621 (1953).

Abstract: A set of accurate scale models has been developed for use in studies of the structures of amino acids, peptides, and proteins. Models representing atoms or groups of atoms built from hard wood to the scale 1 in. = 1Å are connected by a clamping device which maintains desired molecular configurations. These accurate models have been used as substitutes for calculation in investigations of the probable configuration of the polypeptide chain in proteins. Analogous models constructed of rubber-like plastic to the scale 1 in. = 2Å and connected by snap fasteners are designed for qualitative studies of protein structure.

From Cartesian coordinates to van der Waals spheres:

NOW AVAILABLE FOR THE FIRST TIME-THE LONG AWAITED, INEXPENSIVE

NEW COREY-PAULING SPACE FILLING ATOMIC MODELS WITH KOLTUN CONNECTORS

Note the strength of the individual linkages. Locking connectors are used to assure the planarity of the amide group.

These new CPK Models were designed by the Atomic Models Subcommittee of the Biophysics and Biophysical Chemistry Study Section of the National Institutes of Health, Washington, D.C. Over 40 scientists from more than two dozen research centers contributed to the work of this Committee.

The CPK designs which emerged have been implemented by the American Society of Biological Chemists with financial support from the United States National Science Foundation.

A definitive history and technical analysis of the CPK Models has appeared: W. L. Koltun, *Biopolymers*, 3, (1965), pp. 665-79. Reprints are available upon request.

These new CPK Models are offered by Ealing for immediate delivery from stock:
- —individually
- —as Research and Teaching Sets, Protein Sets, Nucleic Acids Sets and Steroid Sets each in heavy wooden storage boxes
- —as 18 convenient pre-assembled structures

Get your free copy today of this new 32 page catalog.
Telephone, wire, telex, cable, or write to:

THE EALING CORPORATION
2225 MASSACHUSETTS AVE., CAMBRIDGE, MASS. 02140 • TEL. (617) 491-1515 CABLE: EALING, TELEX: 94-6378

426

SCIENCE, VOL. 152

DR. WALTER LANG KOLTUN

Don’t forget this opportunity:

2 PRIMO PIANO

Cattedre vuote

Scuole al collasso senza insegnanti Caccia ai supplenti su siti e Whatsapp

Graduatorie esaurite, si arruolano anche studenti e diplomati
Ci sono avvocati che lasciano lo studio per riciclarsi docenti

Cristiano Cadoni

PADOVA. Ora li cercano con Whatsapp, un messaggino lanciato nelle chat dei genitori con l’appello: “Se conoscete qualcuno, segnalatelo”. Prima ancora di questa mossa disperata, avevano messo un annuncio sul sito della scuola. Un manifesto pubblico della resa. Cercansi insegnanti, perché la scuola

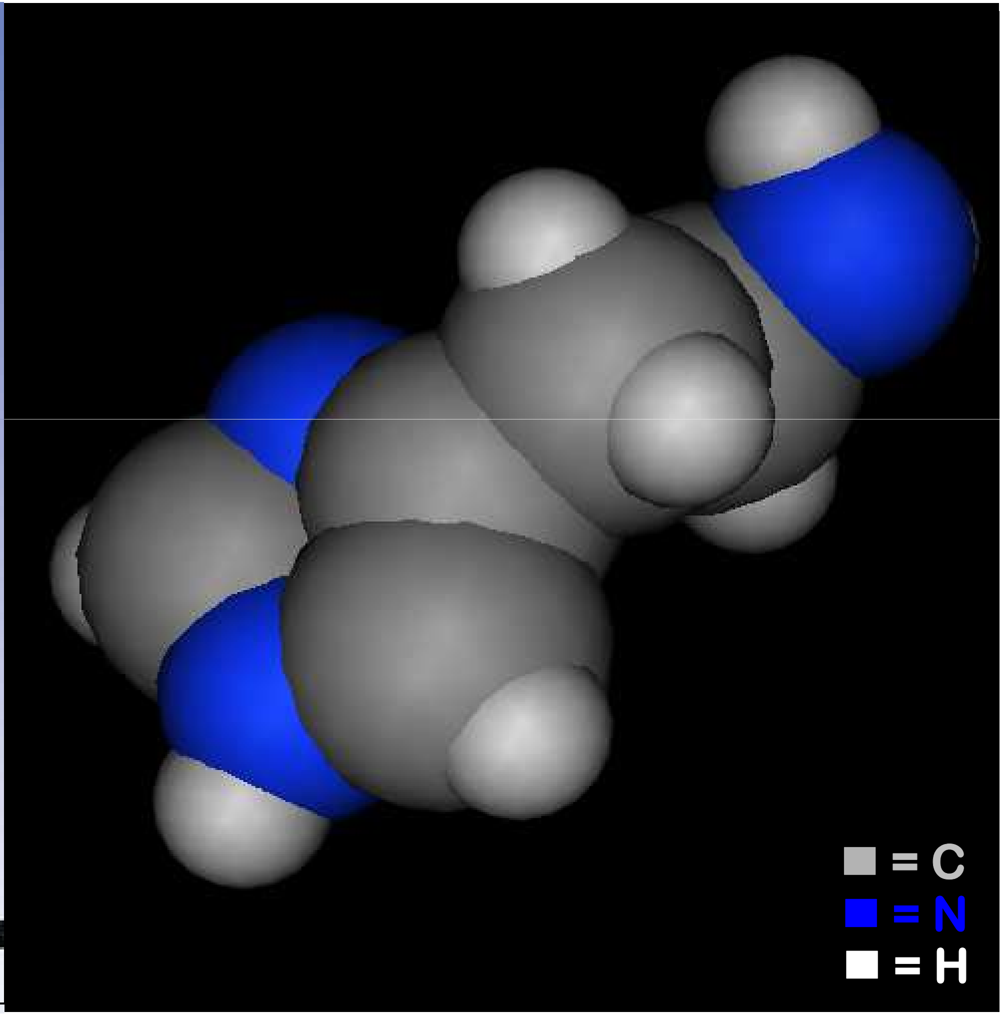
devono fare i salti mortali per trovare insegnanti a cui affidare le cattedre. E la crisi scoppia definitivamente nel momento in cui si cercano supplenti, perché non ce ne sono. Tant’è che per le supplenze brevi, le più difficili da assegnare, si “arruolano” gli insegnanti di sostegno, che a loro volta devono trascurare gli studenti di cui dovrebbero occuparsi.

I NUMERI DELLA CRISI

2200

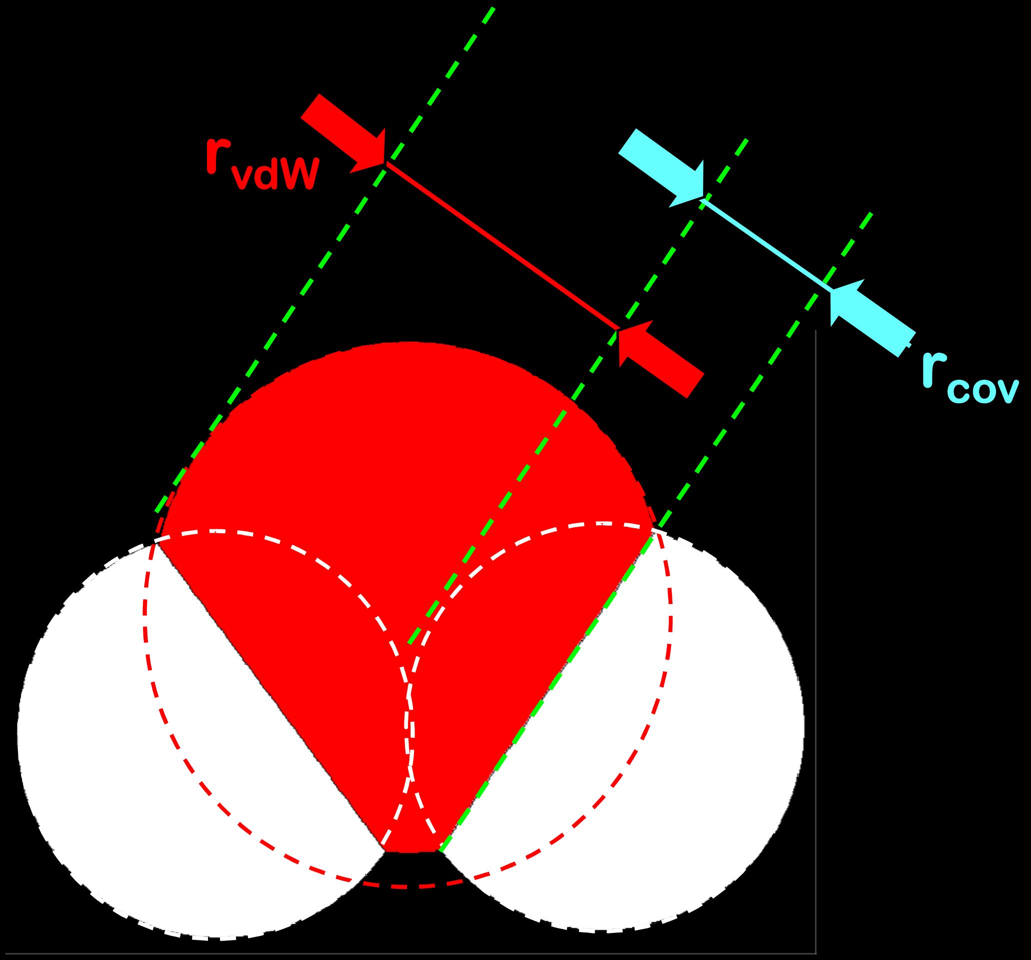
Sono i ruoli scoperti in quest'anno scolastico in tutto il Veneto. Ogni anno la scuola perde, su scala regionale, fra i 700 e gli 800 insegnanti, e ne ''recupera'' soltanto 350 dal corso di laurea delle università di Padova e Verona. Il fabbisogno è dunque soddisfatto per meno del 50

Corey and Pauling models: put the *flesh* to the *bone* !!!



CPK MODELS

(by Corey, Pauling and Kultun)

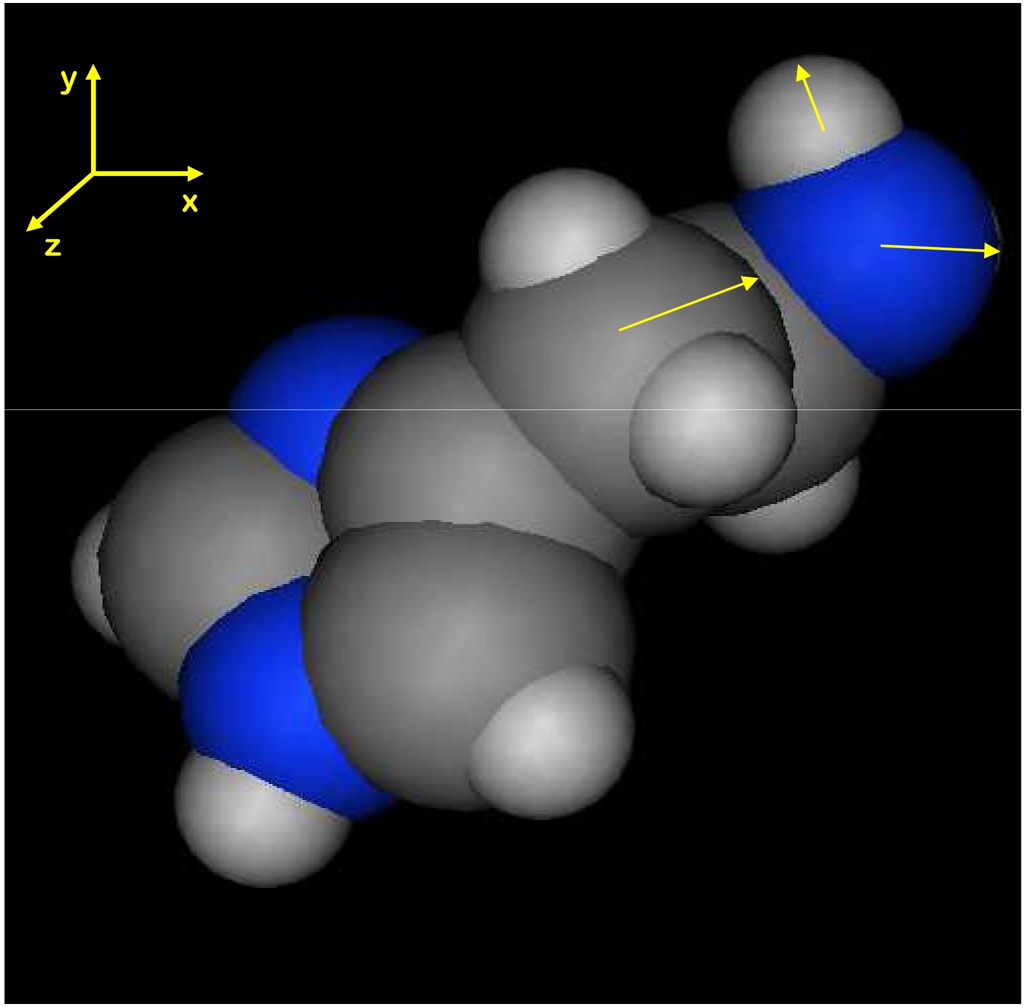


r_{vdW} = van der Waals radius

r_c = covalent radius

From an informatics point of view, is it clear how we can translate a CPK model?

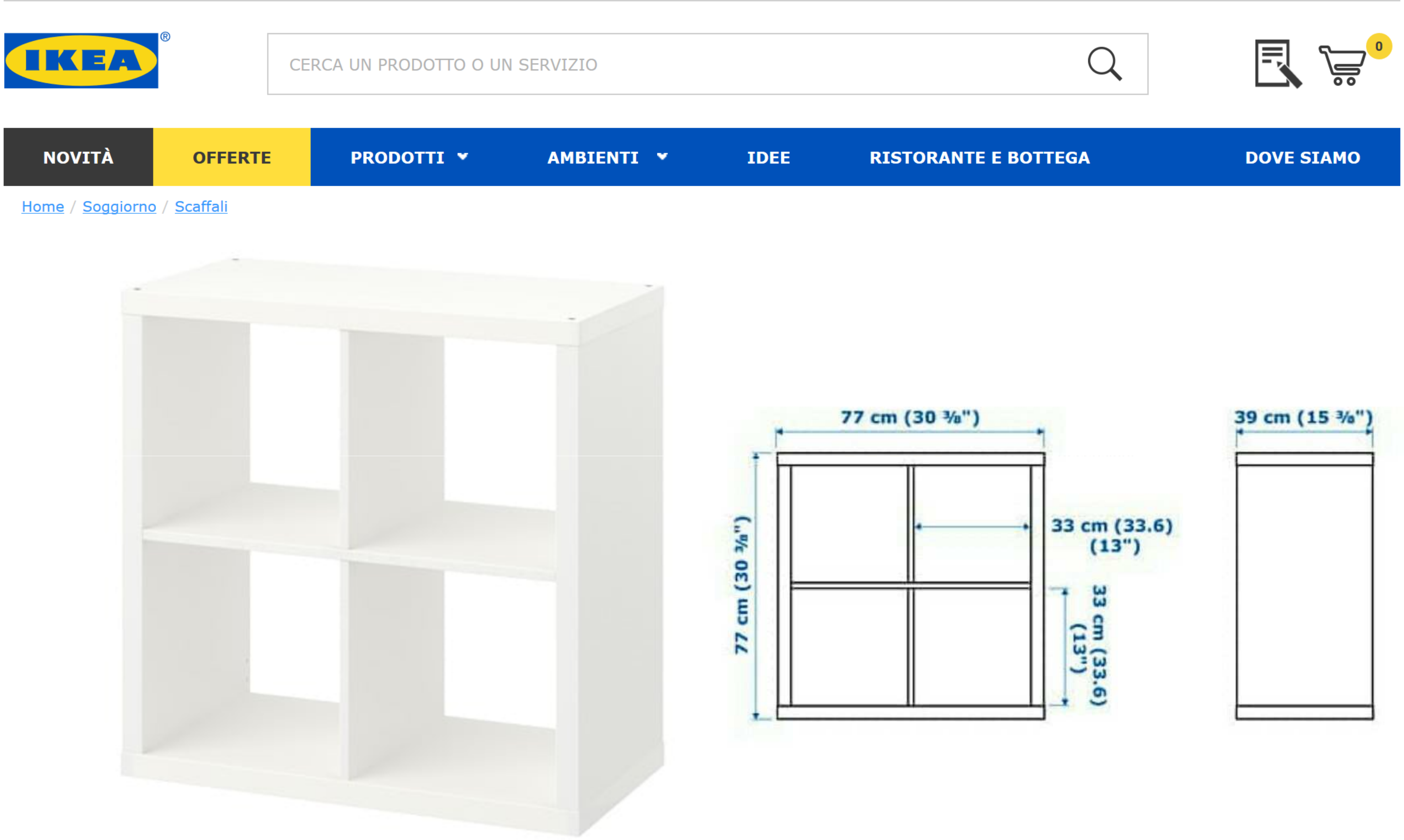
0.9760	0.5530	0.1180	C
2.1810	1.4240	0.1010	C
2.9110	1.3580	-1.2420	C
2.2460	1.6700	-2.0550	H
3.2180	0.3270	-1.4520	H
4.1000	2.2110	-1.2400	N
3.8190	3.1840	-1.1170	H
4.5460	2.1620	-2.1560	H
2.8610	1.1260	0.9100	H
1.8740	2.4560	0.3100	H
-0.1610	0.9170	-0.5670	N
-1.0220	-0.0510	-0.3510	C
-2.0310	-0.1010	-0.7380	H
-0.4920	-1.0270	0.4440	N
-0.9630	-1.8680	0.7470	H
0.7870	-0.6560	0.7530	C
1.4290	-1.2620	1.3760	H



At this point measure the molecular size is easy… or at least so it seems to Arie Verloop!

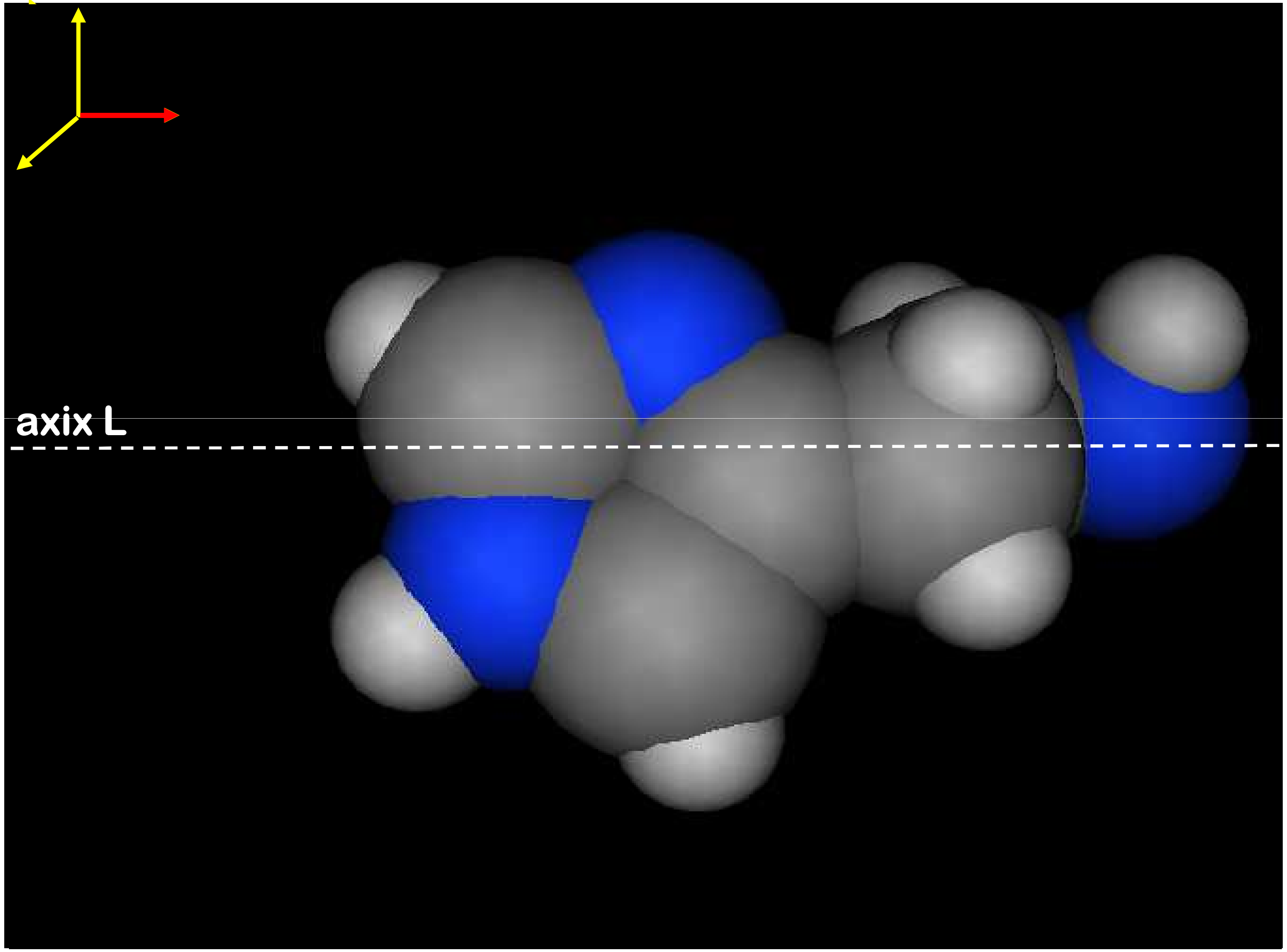
Verloop A., Hoogenstraaten W., Tipker J. “Development and application of new steric substituent parameters in drug design.” In Drug Design (Ed. Ariëns), vol.7, pp.165-207 (1976), New York: Academic Press.

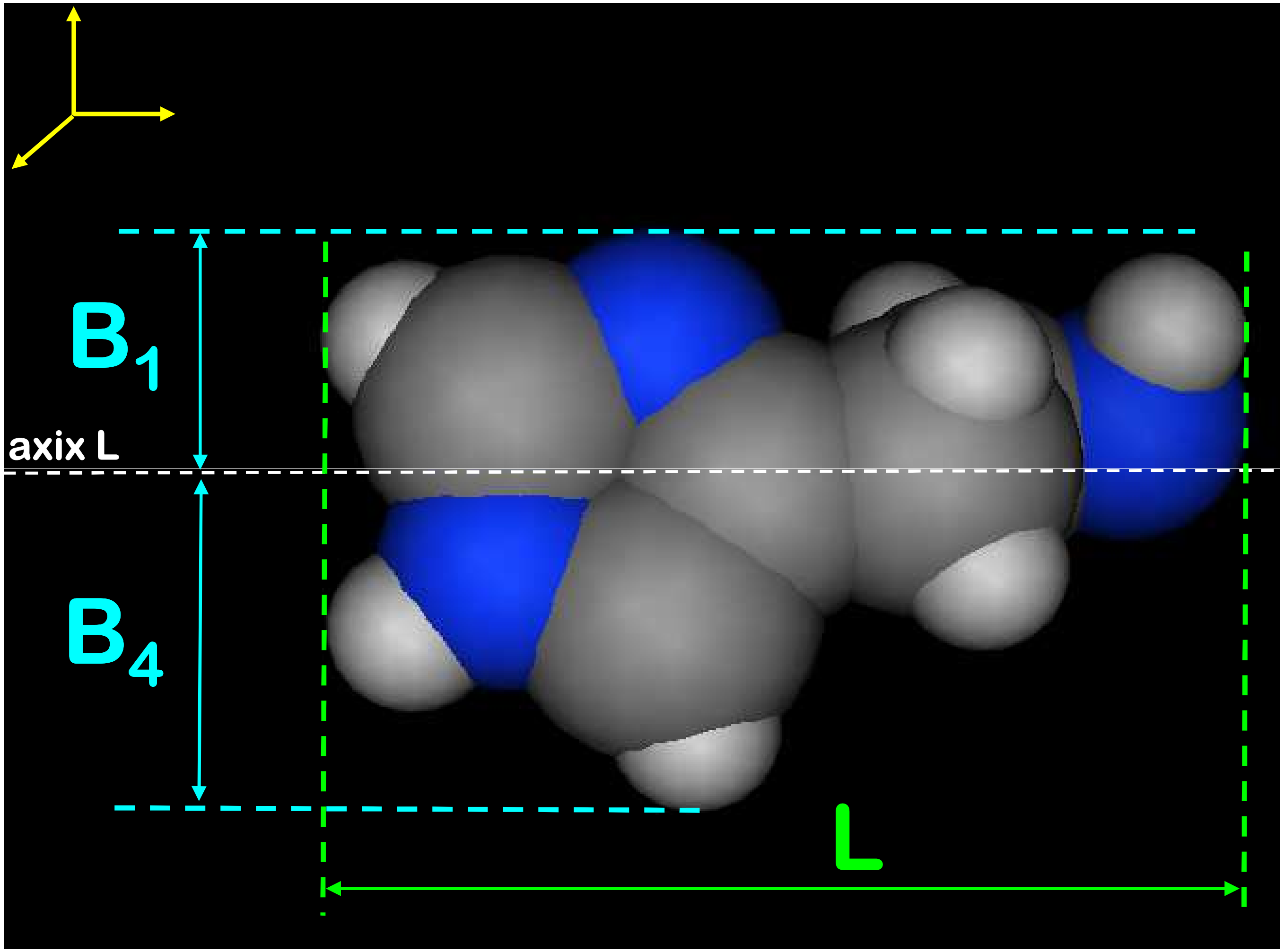
My favorite example:

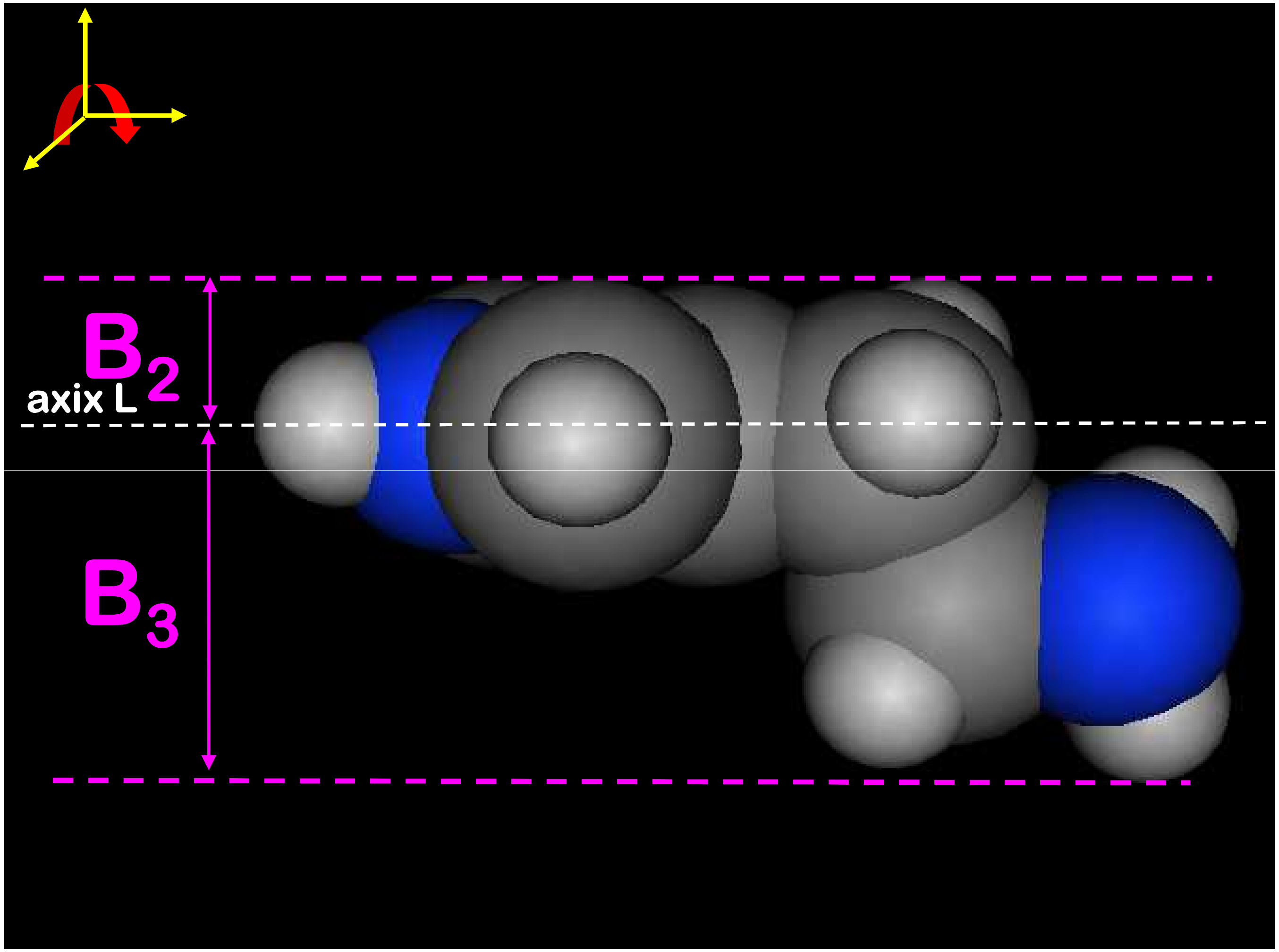


Credits: https://www.ikea.com/it/it/catalog/products/20275814/

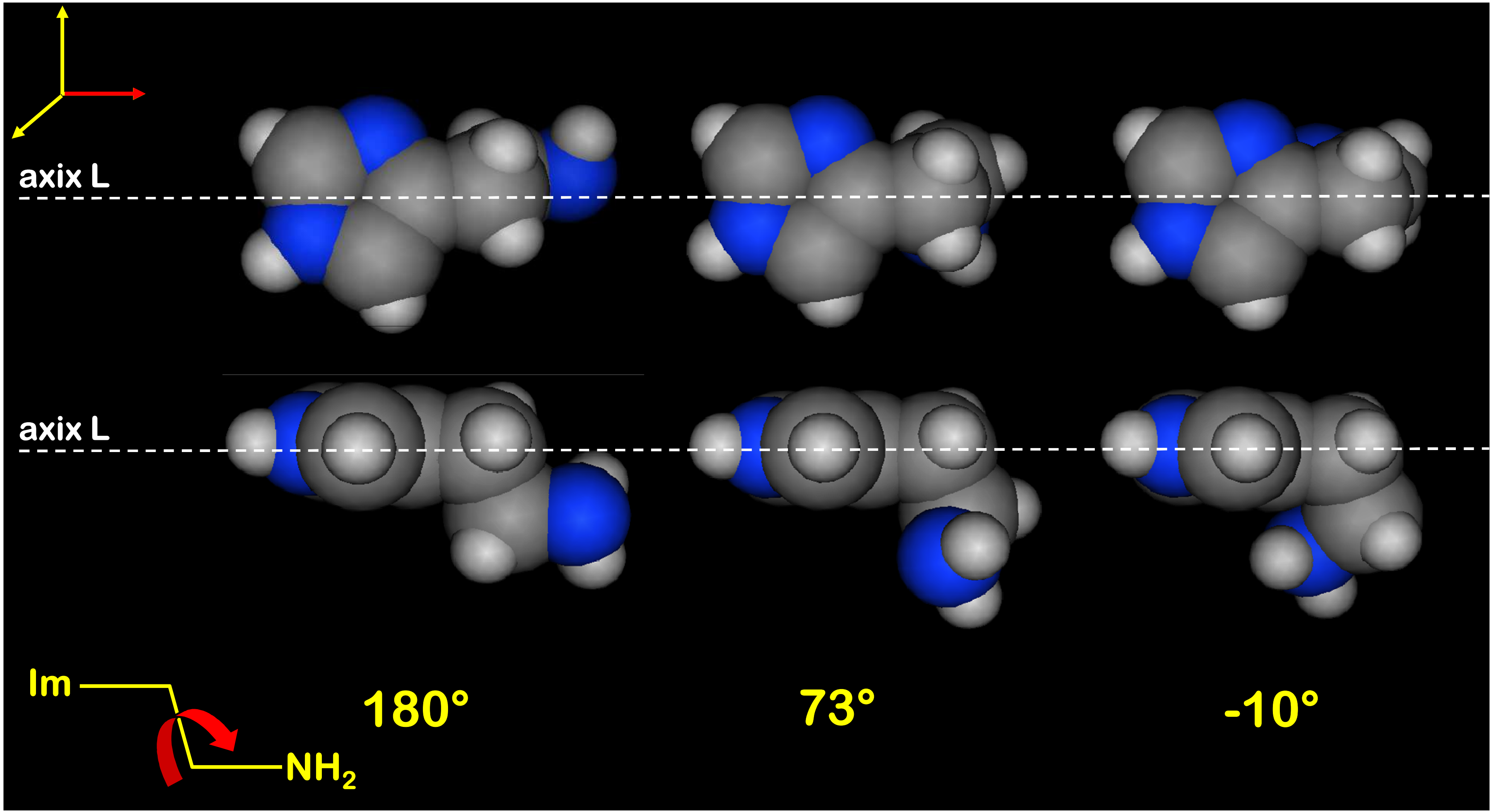
Arie Verloop and STERIMOL descriptors:



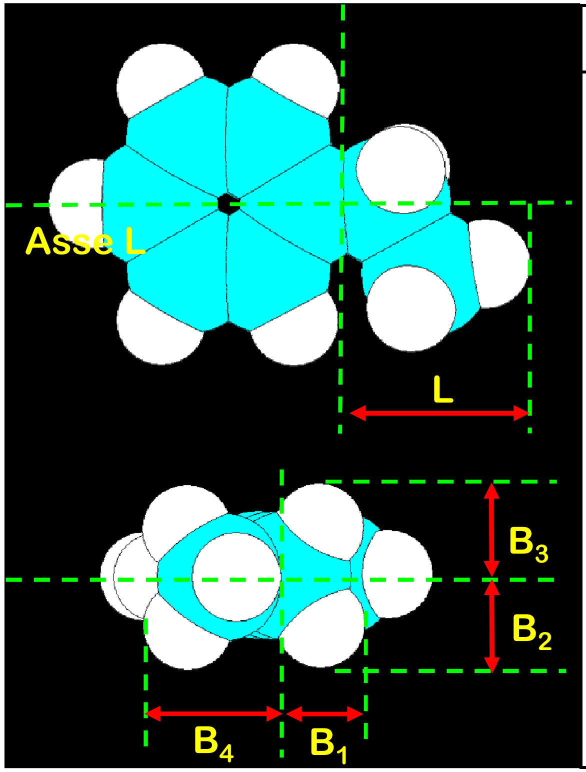

B1
B4
axix L
L


B_2
axix L
B_3

Beware: as you have already guessed the STERIMOL parameters are by definition conformation-dependent!



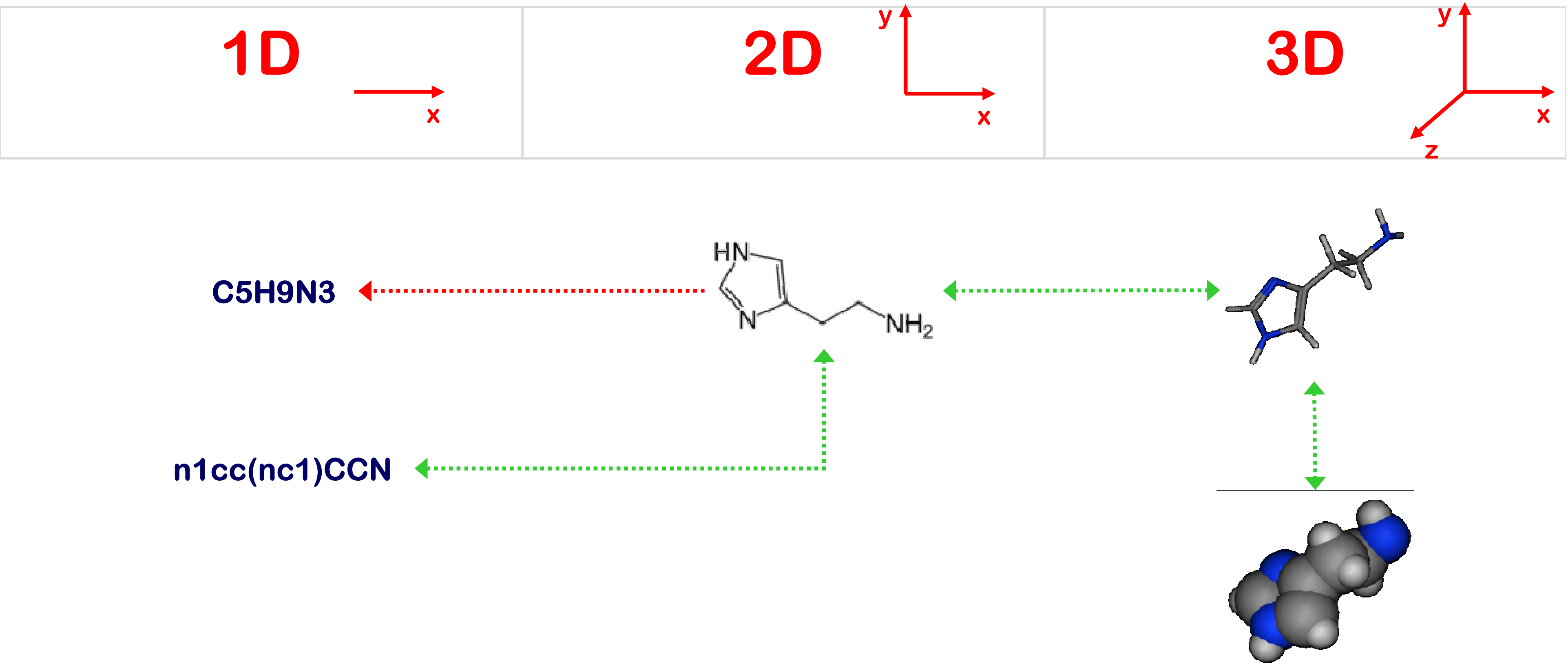
From CPK model is now very easy to measure (and collect) STERIMOL descriptors of the most common substituents:



Substituent	L	B_1	B_2	B_3	B_4
CH_3	3.00	1.52	1,90	1.90	2.04
C_2H_5	4.11	1.52	1.90	1.90	2.97
n-C_4H_9	6.17	1.52	1.52	1.90	4.42
t-C_4H_9	4.11	2.59	2.86	2.86	2.97
C_6H_5	6.28	2.30	2.30	2.83	2.83
$CH_2C_6H_5$	4.62	1.52	3.11	3.11	6.02
OH	2.74	1.35	1.35	1.35	1.93
OCH_3	3.98	1.35	1.90	1.90	2.87
$N(CH_3)_2$	3.53	1.59	2.56	2.80	2.80
NO_2	3.44	1.70	1.70	2.44	2.44

L/B_1 e B_4/B_1 ratios give as an indication of the substituent *sphericity*.

It is summary time!



1D

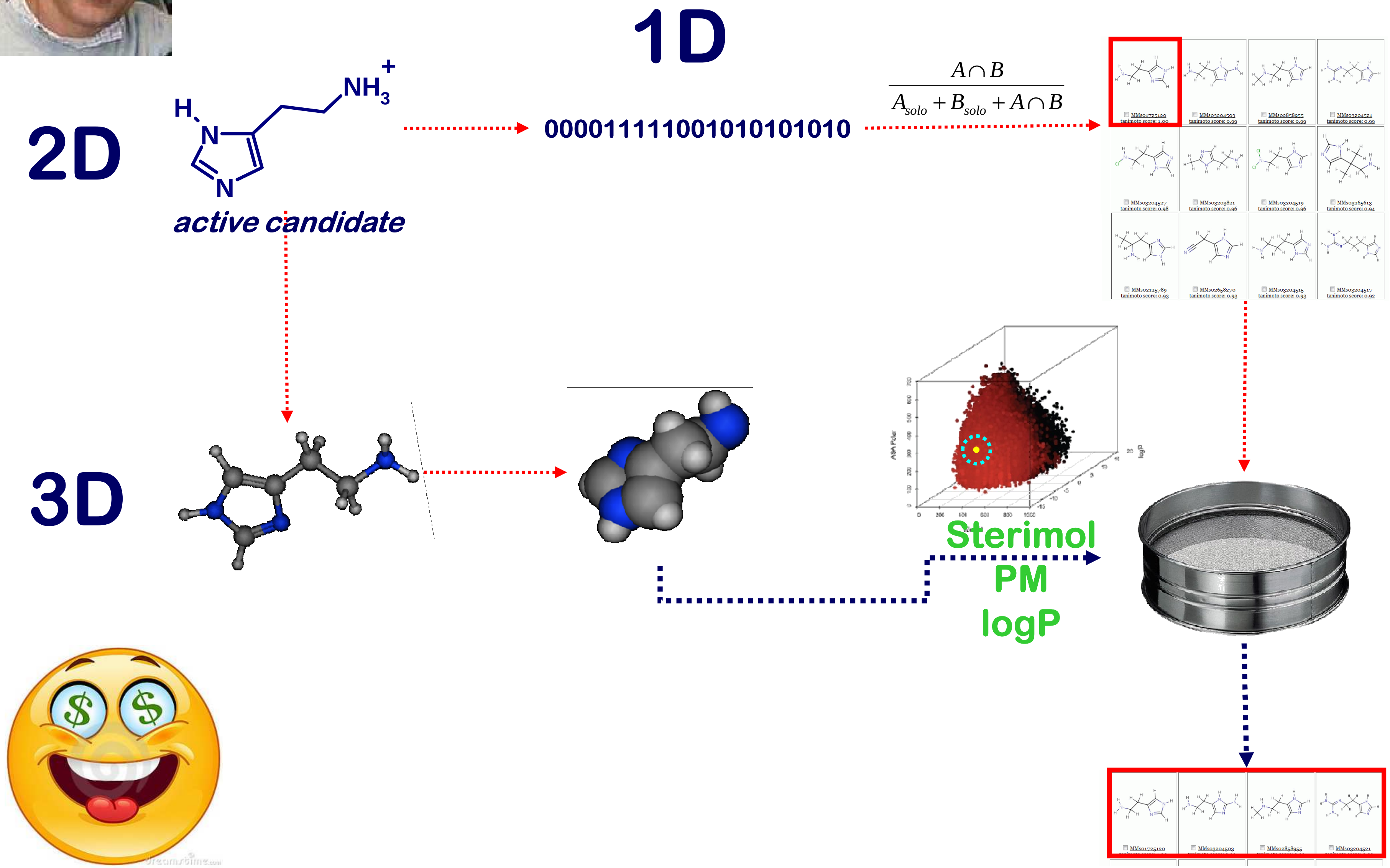
2D

3D

C5H9N3

n1cc(nc1)CCN

Do you remember:



1D

2D

3D

active candidate

NH_3^+

00001111001010101010

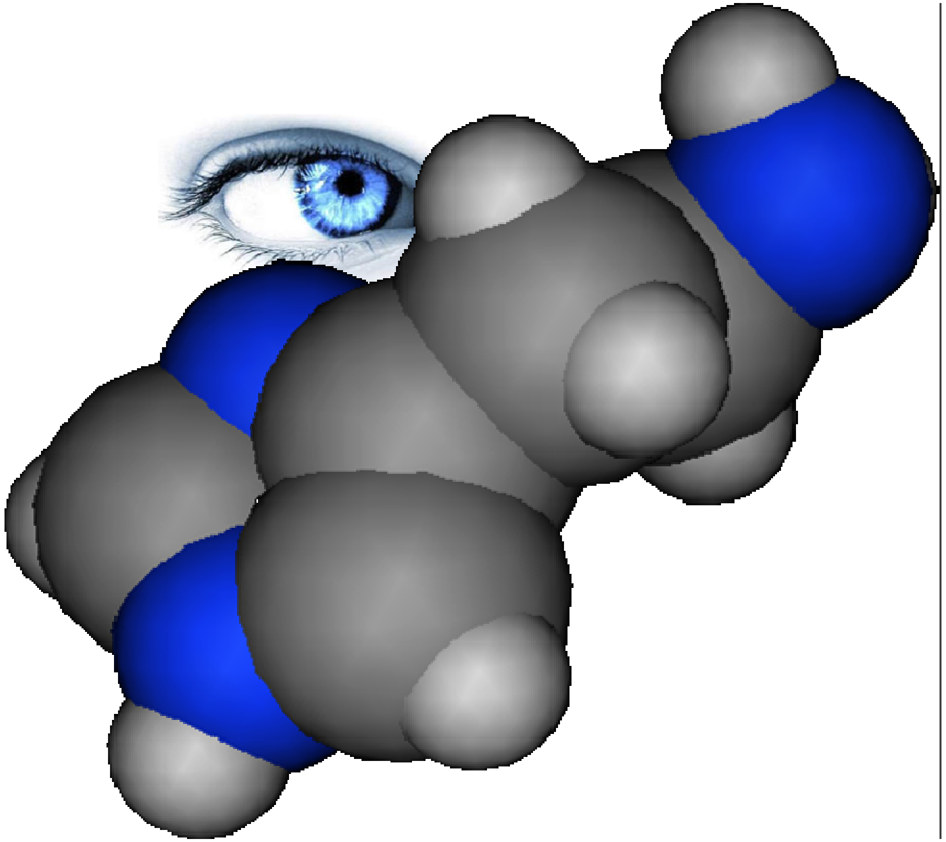
$$\frac{A\cap B}{A_{solo}+B_{solo}+A\cap B}$$

Sterimol

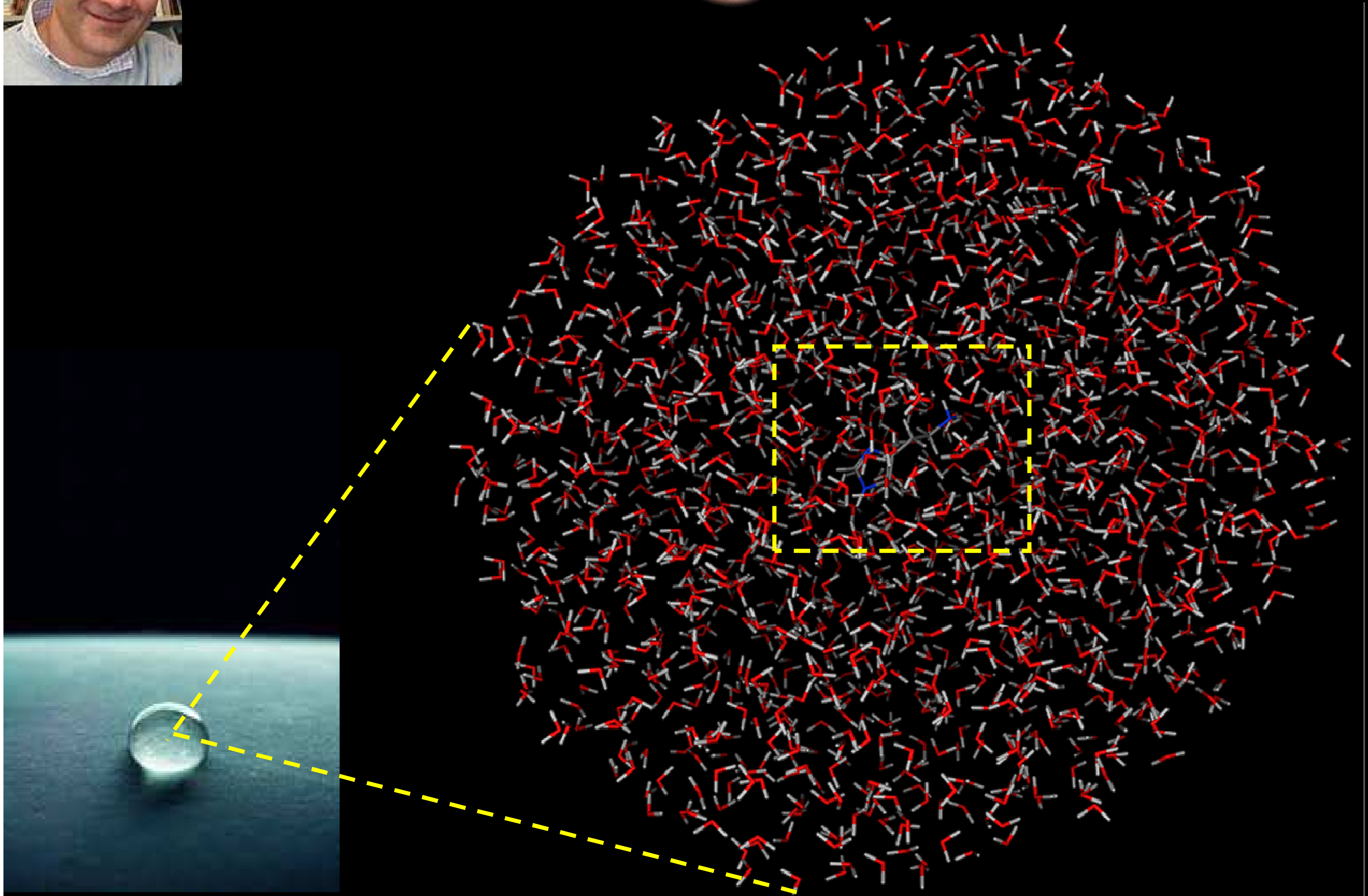
PM

logP

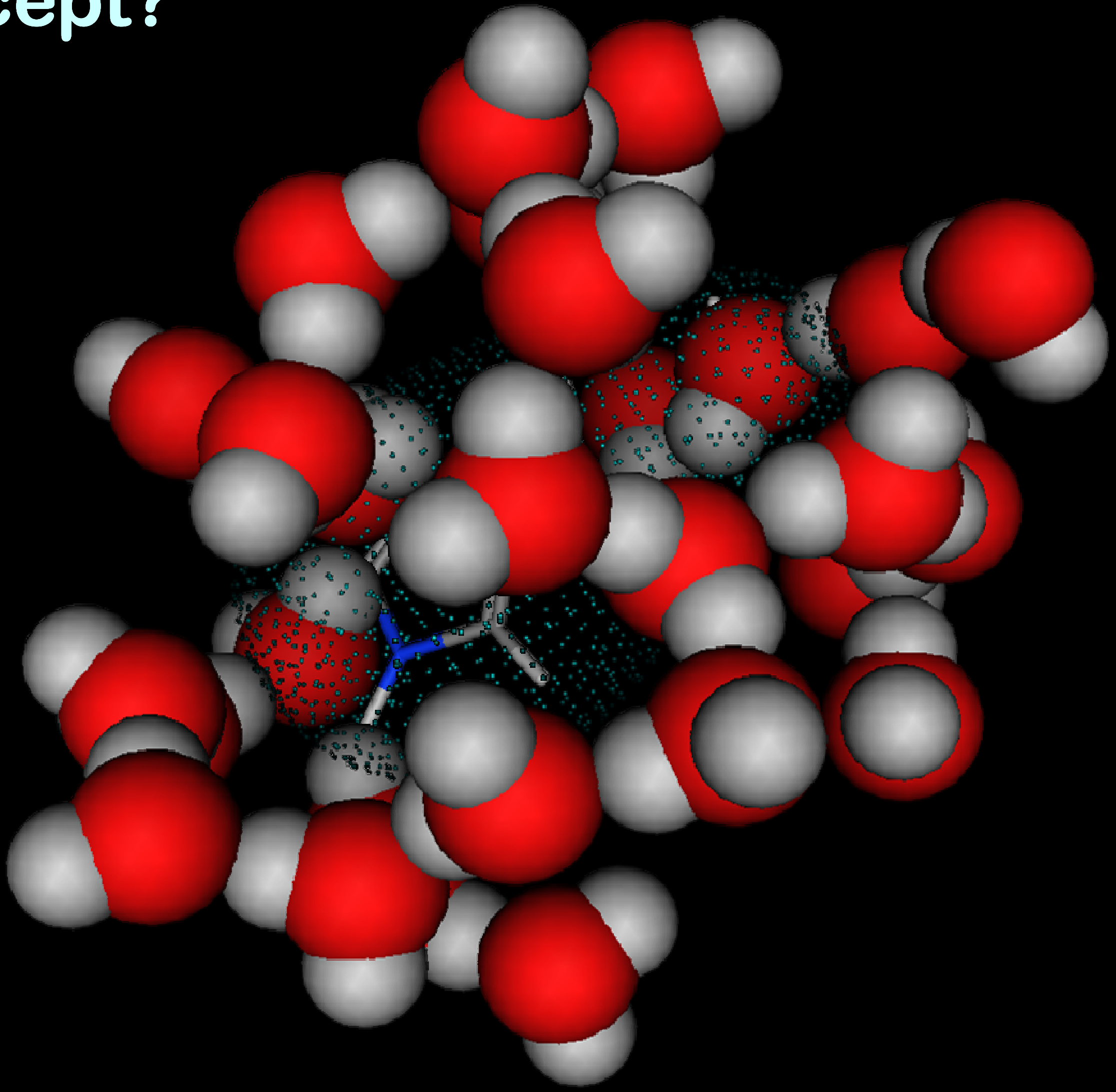
CPK is a nice 3D shape representation… but probably on for human being. But a molecule as will see another molecule?



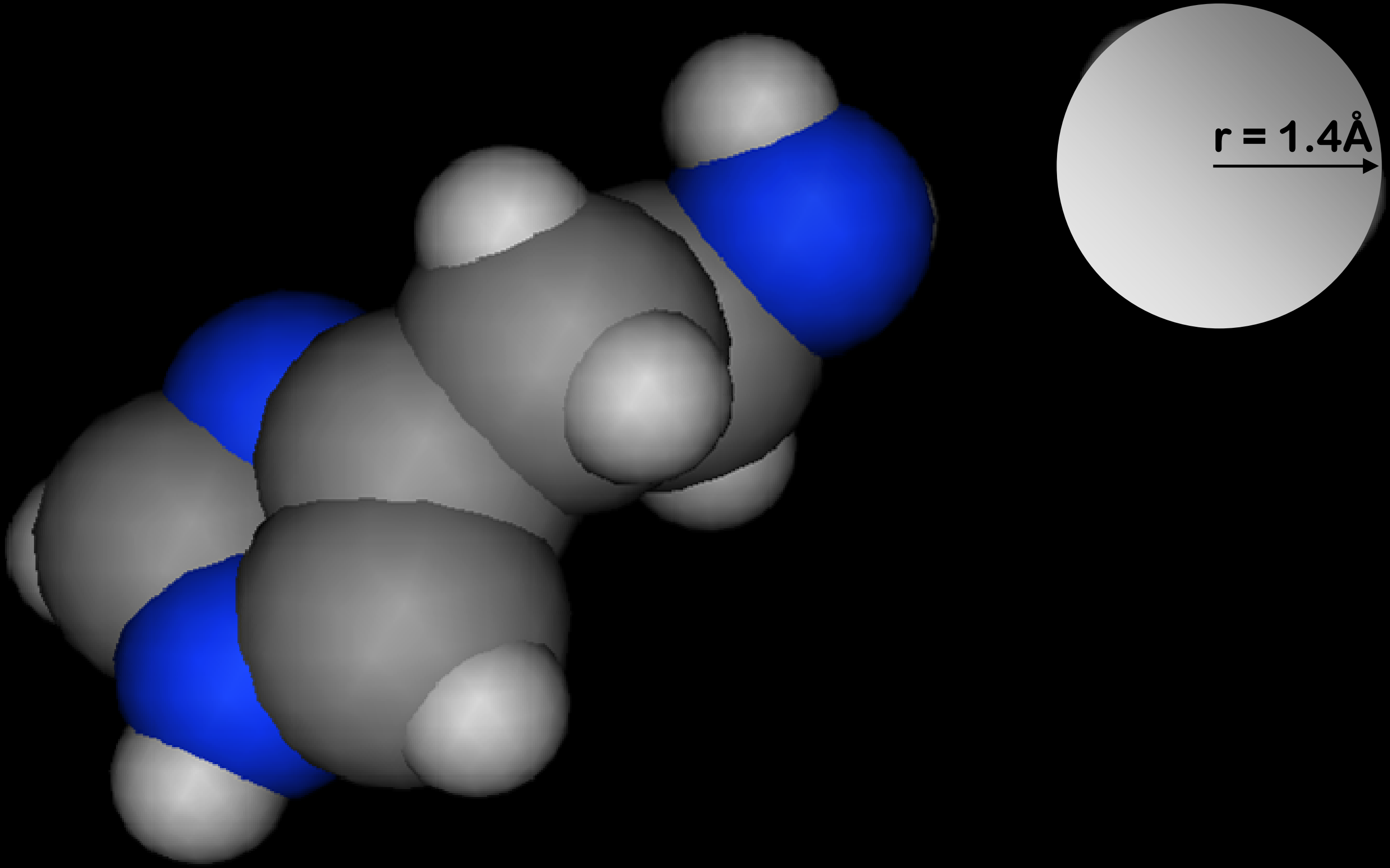
Here's who could  molecular objects !



Do you remember the *hydration shell* concept?

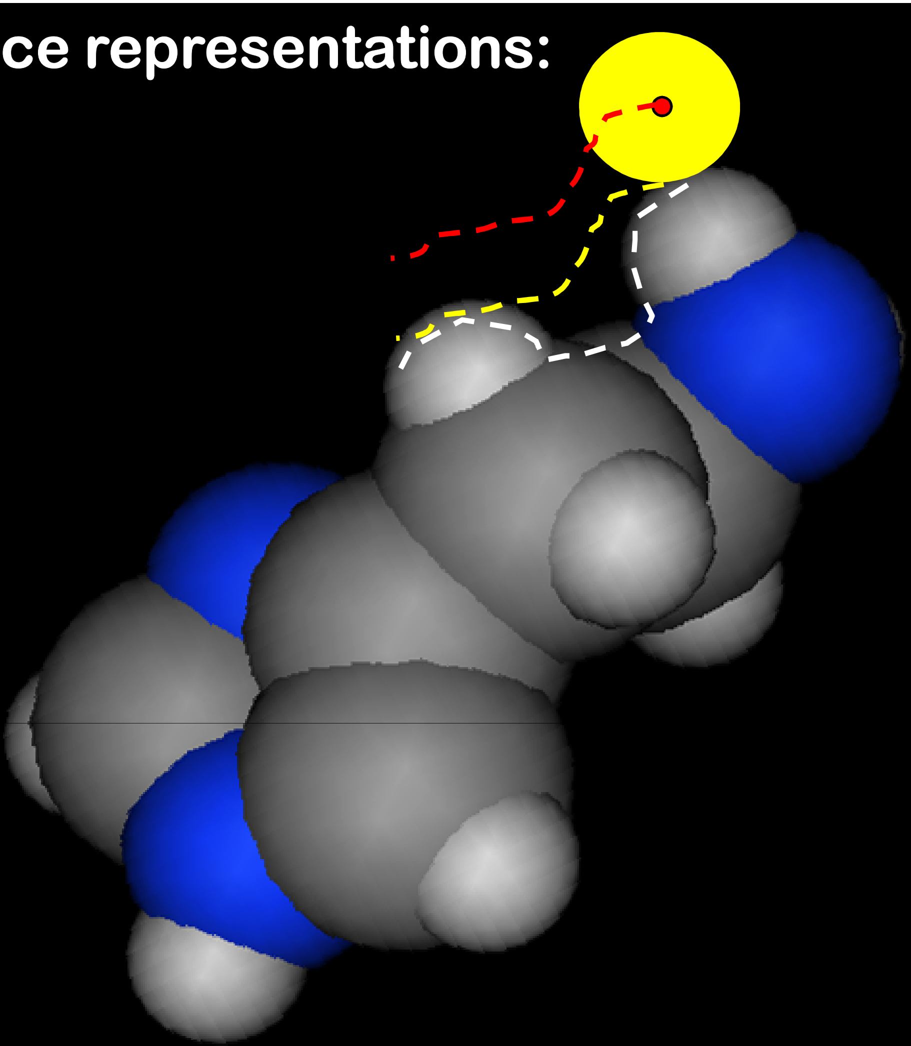


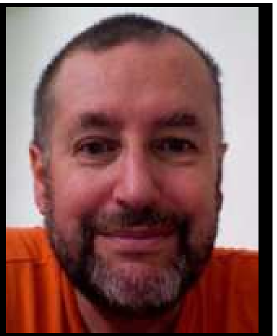
Now, the virtual experiment is pretty easy!



The *rolling ball* algorithm developed by Shrake & Rupley in 1973: *"Environment and exposure to solvent of protein atoms. Lysozyme and insulin". J Mol Biol. 79 (2): 351–71.*

Molecular surface representations:





Michael Connolly

The van der Waals Surface (vdWS) is the exterior boundary of the union of van der Waals spheres.

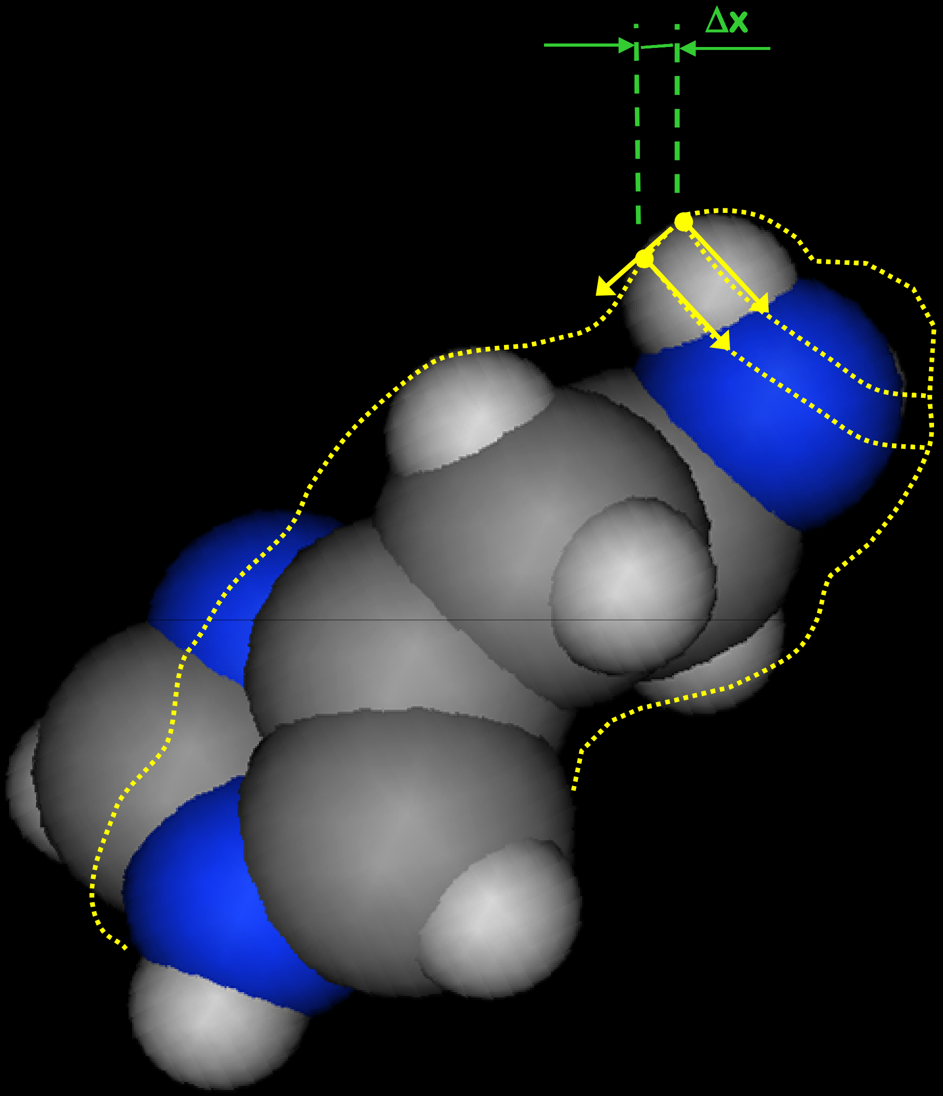
The Solvent Excluded Surface (SES, also known as *Connolly surface*) is the results of the SAS erosion by the same probe.

The Solvent Accessible Surface (SAS) is the result of the vdW Surface dilatation by a structuring element, or a probe, representing a solvent molecule, typically water.

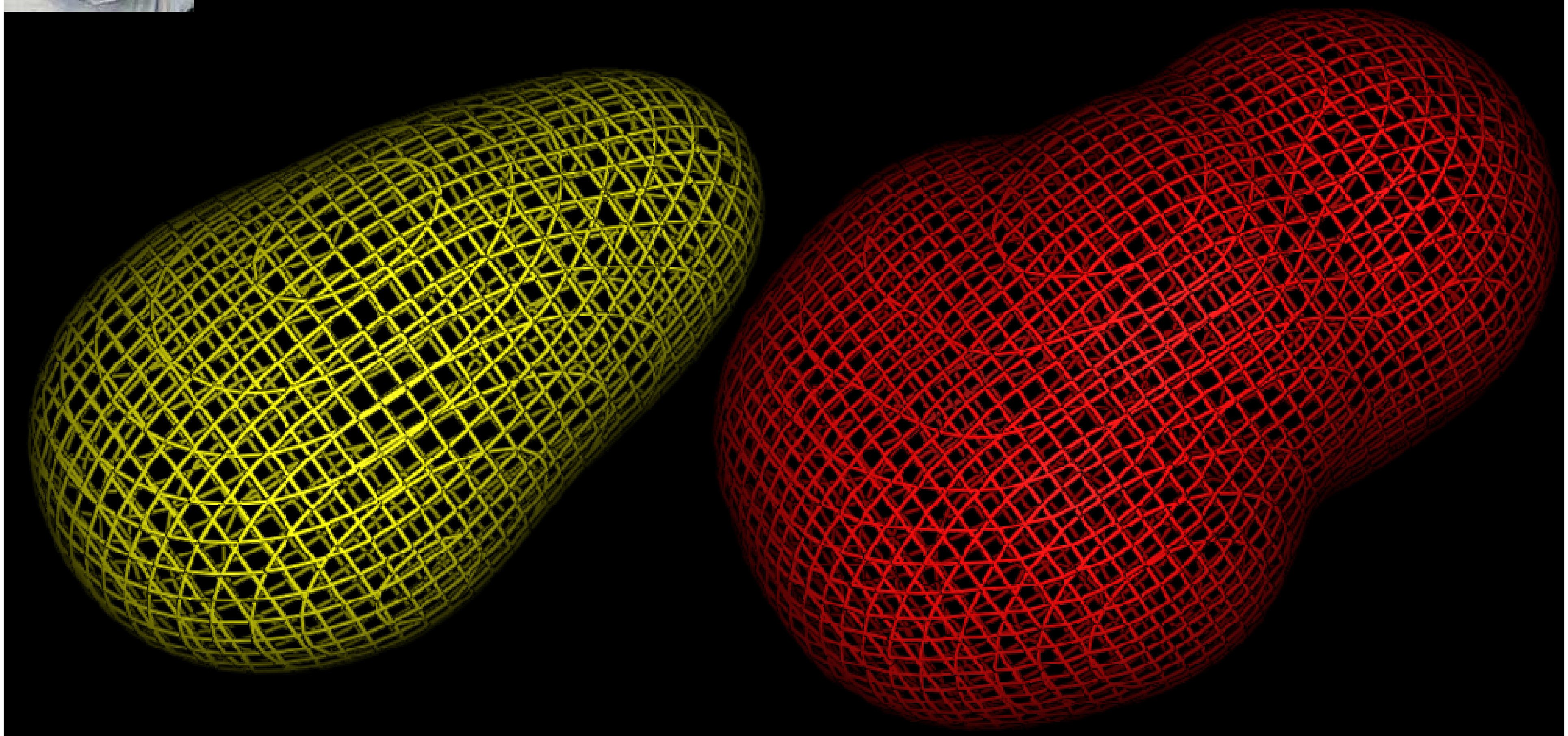
but, to transform this virtual experiment is an useful experiments we need TWO smart ideas:

1. Remember Pollicino’s fairytale;
2. and…

A bit of algorithm:



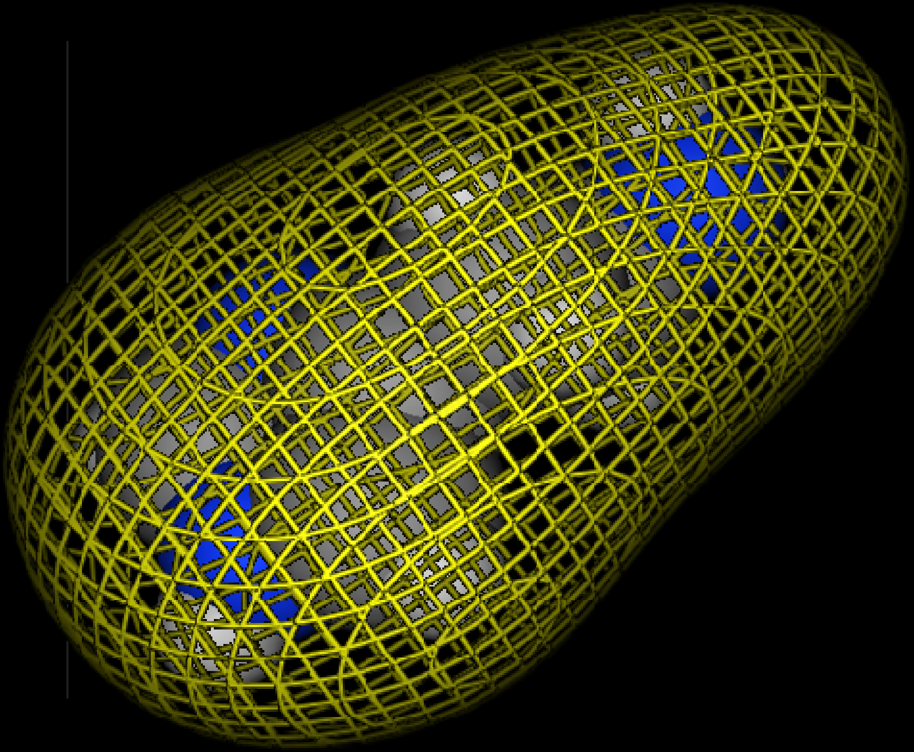
Here is the results of our virtual experiment!



SES or Connolly Surface

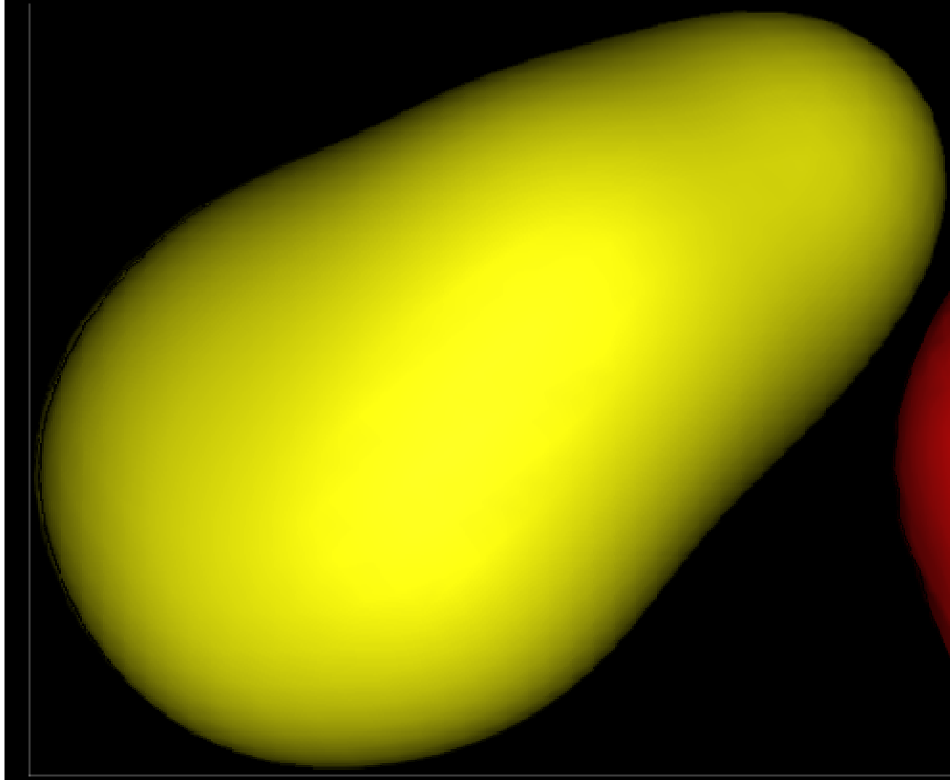
SAS

You understand the differences!

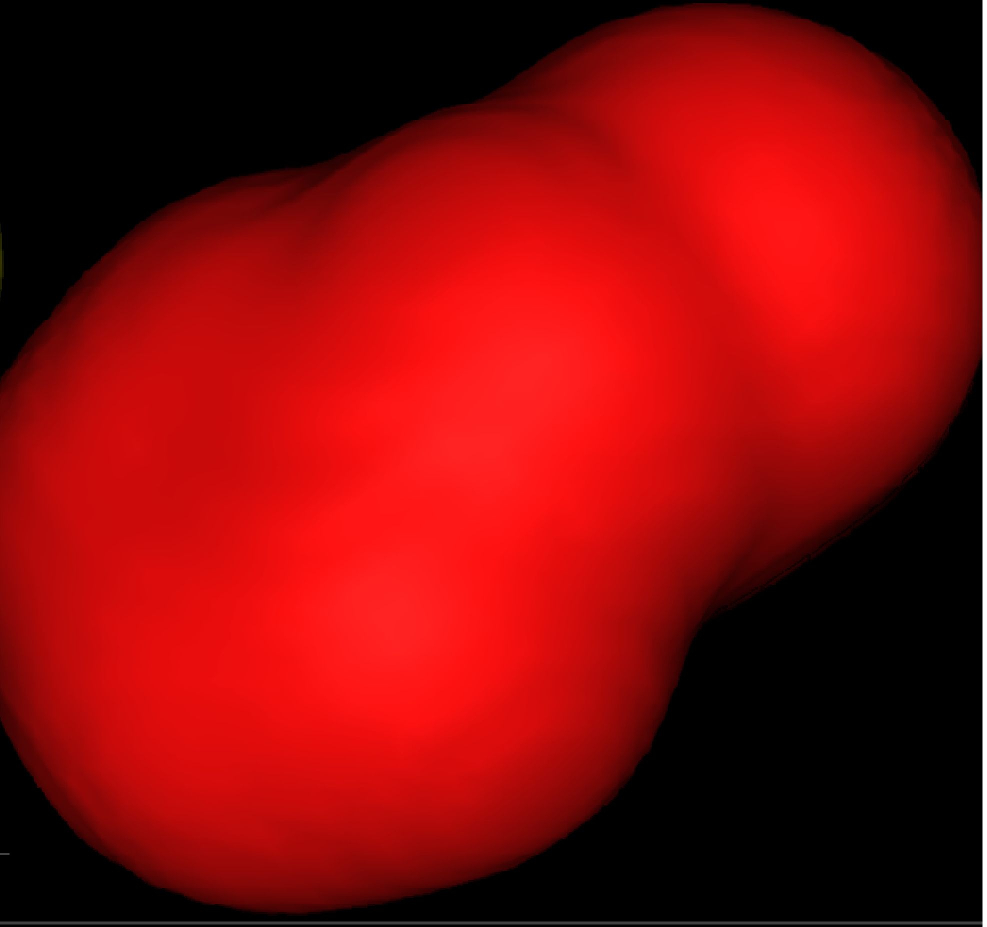


vdW *versus* Connolly Surface

This is really… PlayStation!

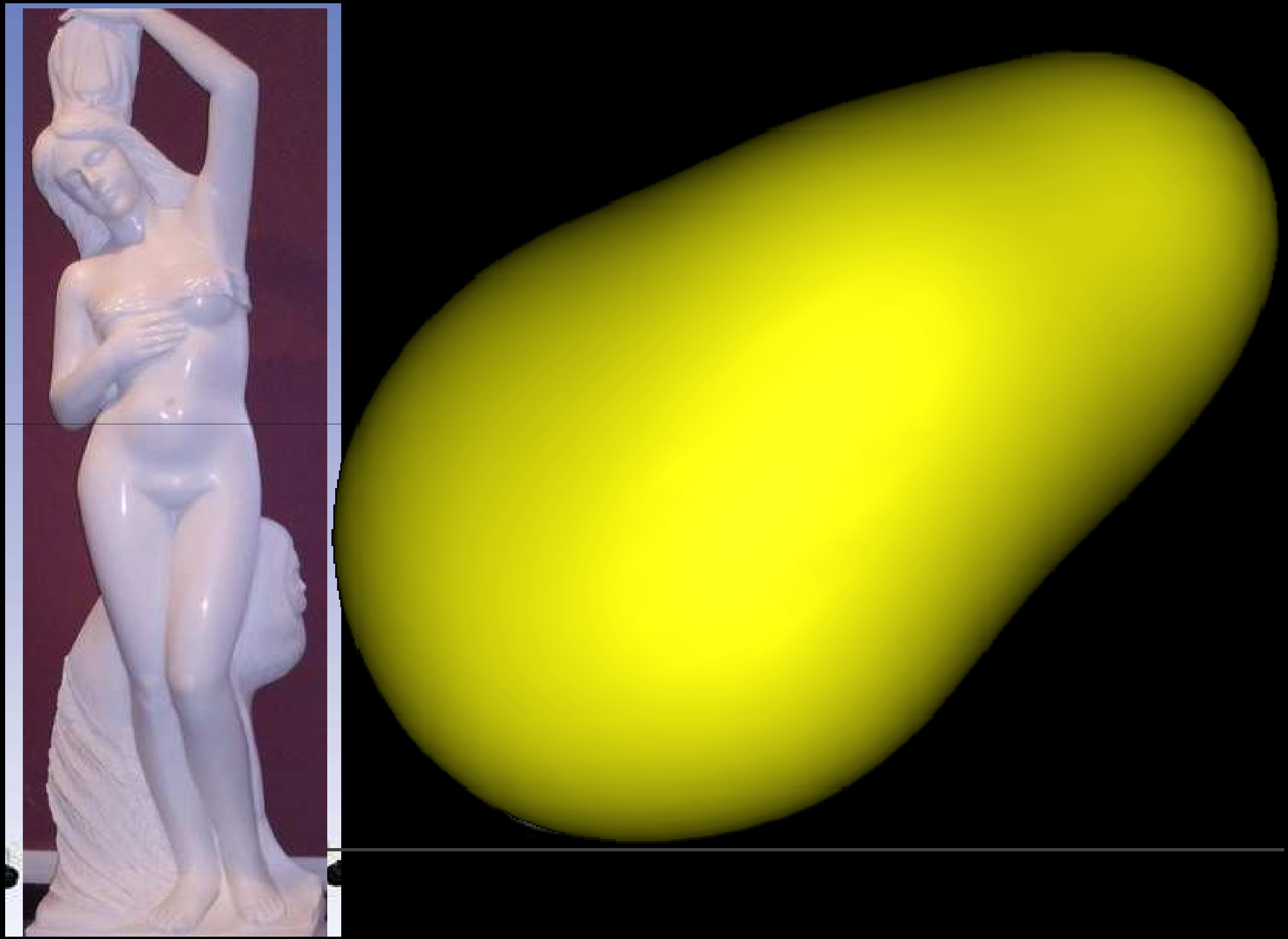


SES or Connolly Surface

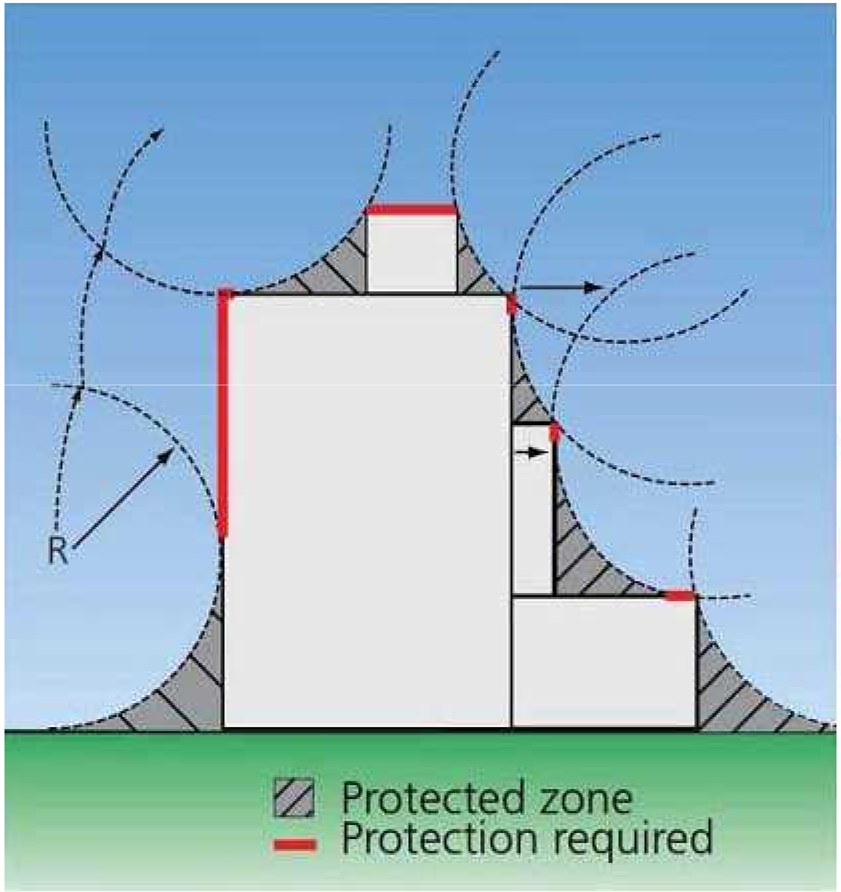


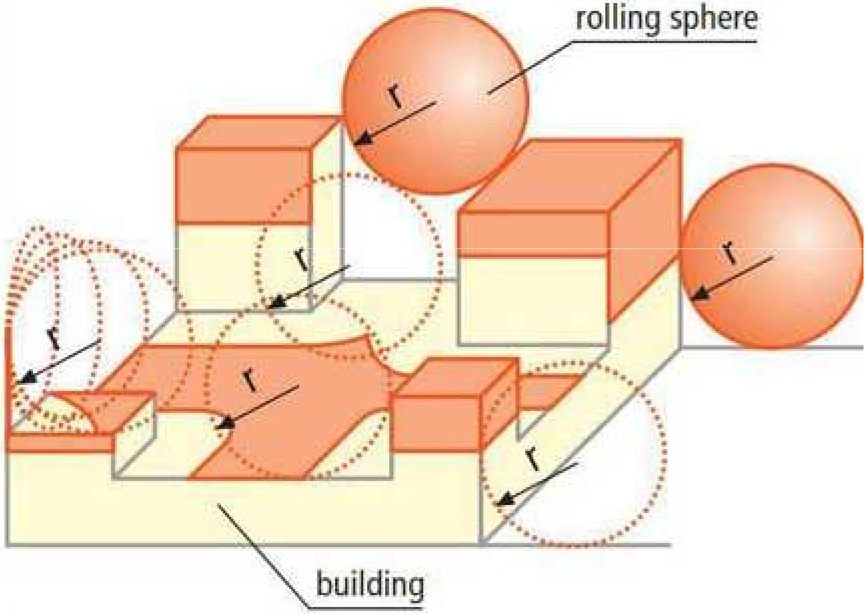
SAS

Do you understand the morphing?



Rolling Sphere method: choose a method of installation Lightning

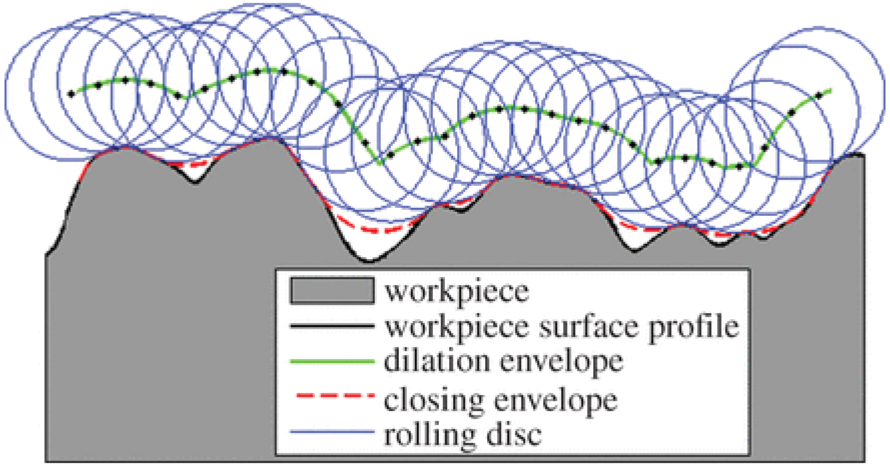




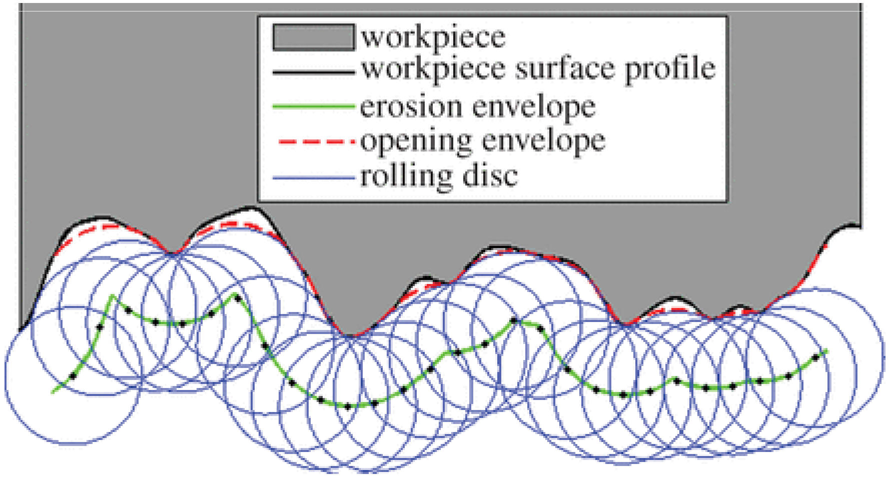
Credits: https://www.bornika.ir/en/Article/AR-52/Choose-a-method-of-installation-Lightning

Geometric computation theory for morphological filtering on freeform surfaces:

"The surface of a geometrical component is an interface limiting the body of the component and separating it from the surrounding medium. It governs the functional behaviours of the product, whether that be a mechanical, *tribological*, hydrodynamic, optical, thermal, chemical or biological property, all of which are of tremendous importance to product performance. Many emerging products and devices are based on achieving surfaces with special functionalities. Manufactured items such as micro- and nanometre-scale transistors, microelectromechanical systems and nanoelectromechanical systems, microfluidic devices, optics components with freeform geometry and structured surface products are clear evidence of products where the surface plays the functional role.



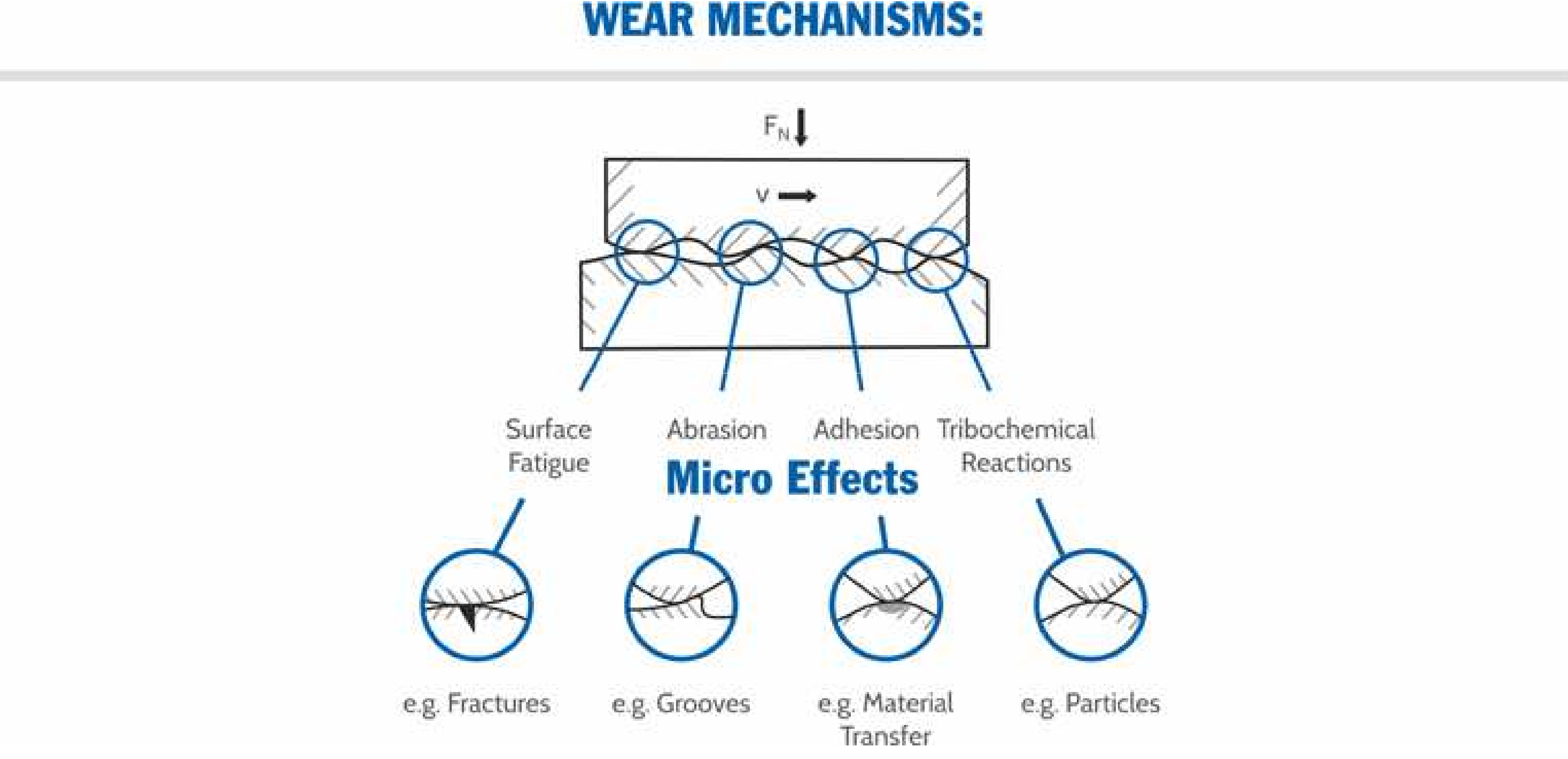
The dilation and closing envelope of an open profile by a disc.



The erosion and opening envelope of an open profile by a disc.

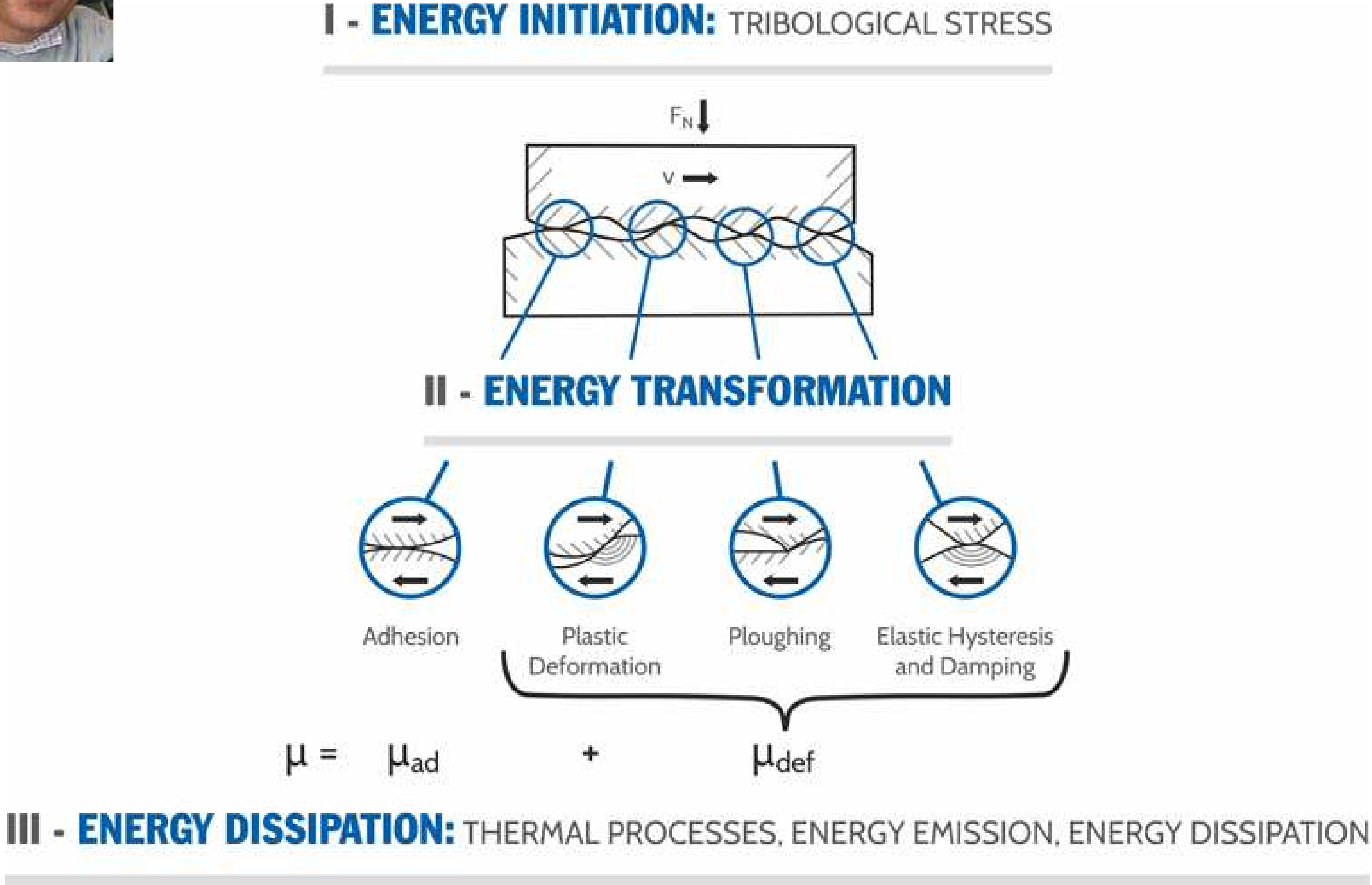
Credits: http://rspa.royalsocietypublishing.org/content/469/2159/20130150

Tribology: the Study of Interacting Surfaces in Motion



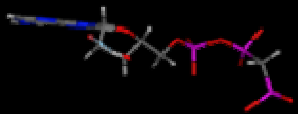
Credits: https://www.ggbearings.com/en/company/tribology

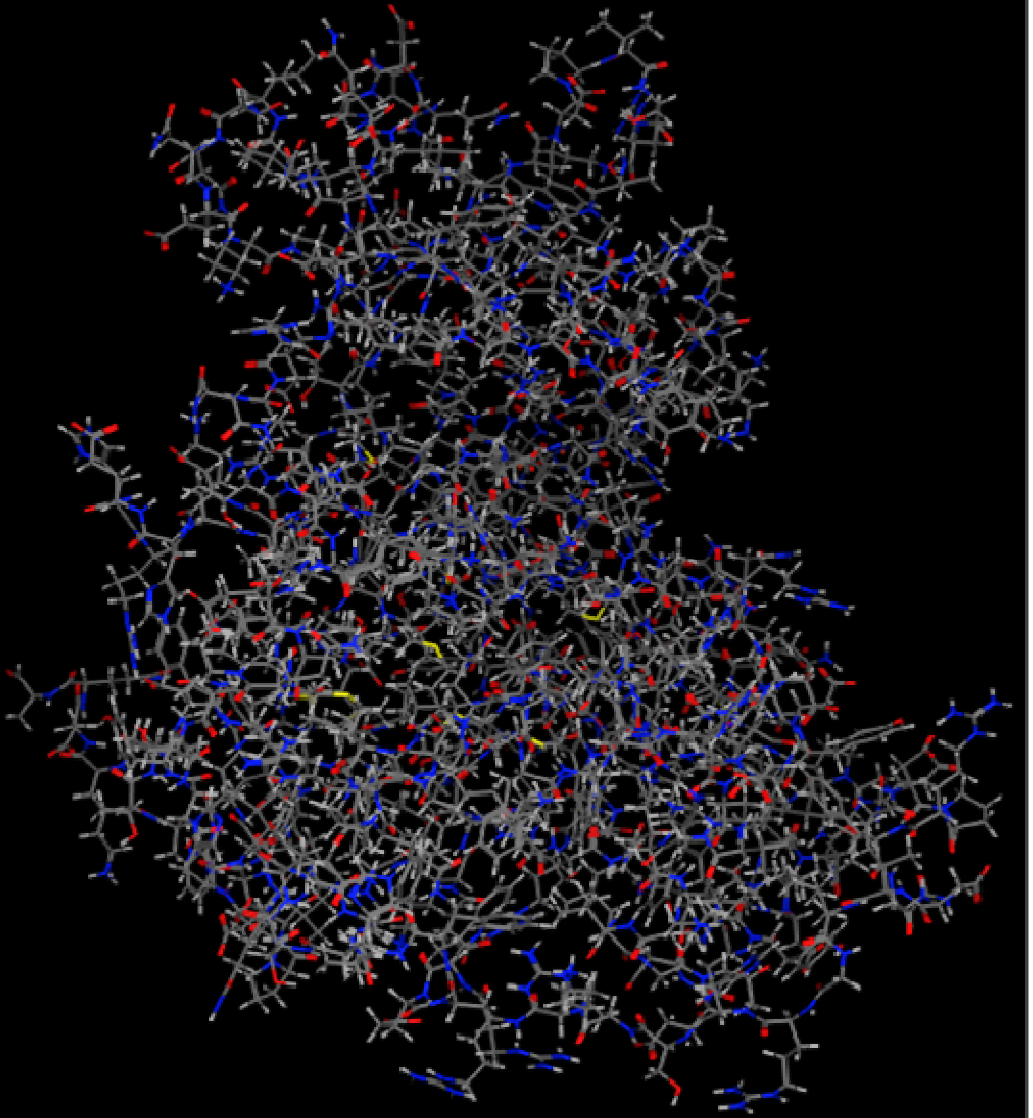
Tribology: the Study of Interacting Surfaces in Motion

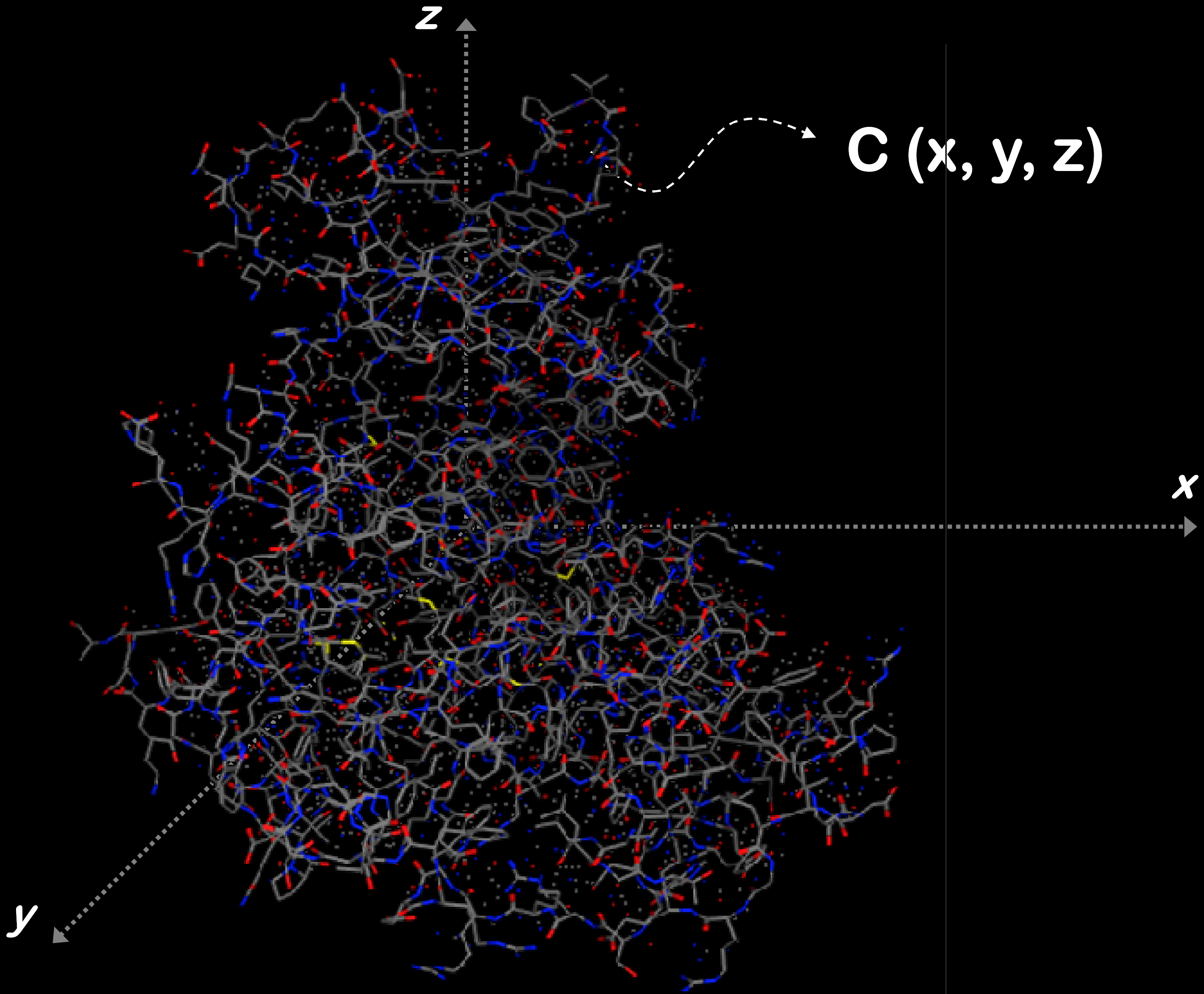


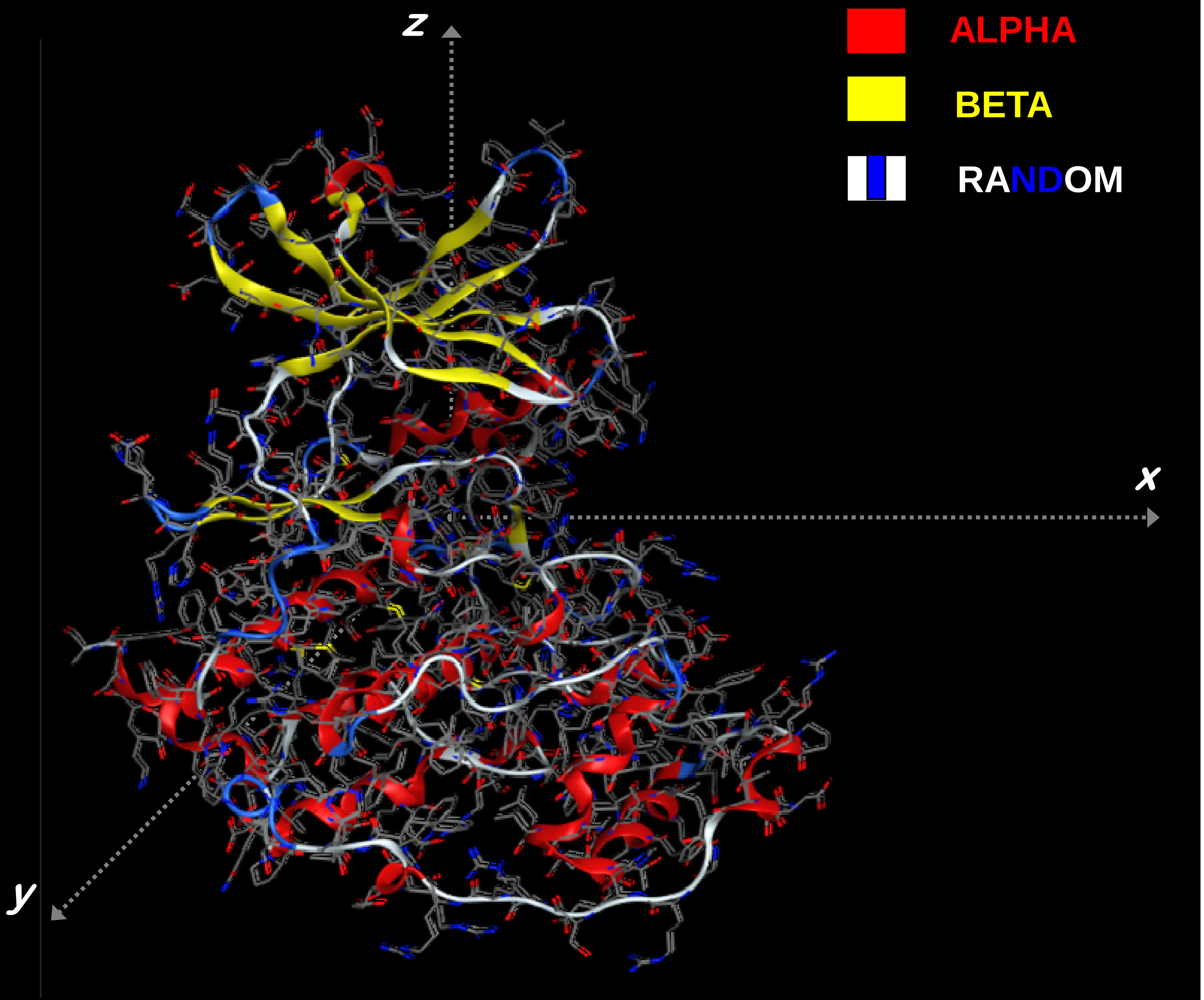
Credits: https://www.ggbearings.com/en/company/tribology

From small molecule to its biological target…

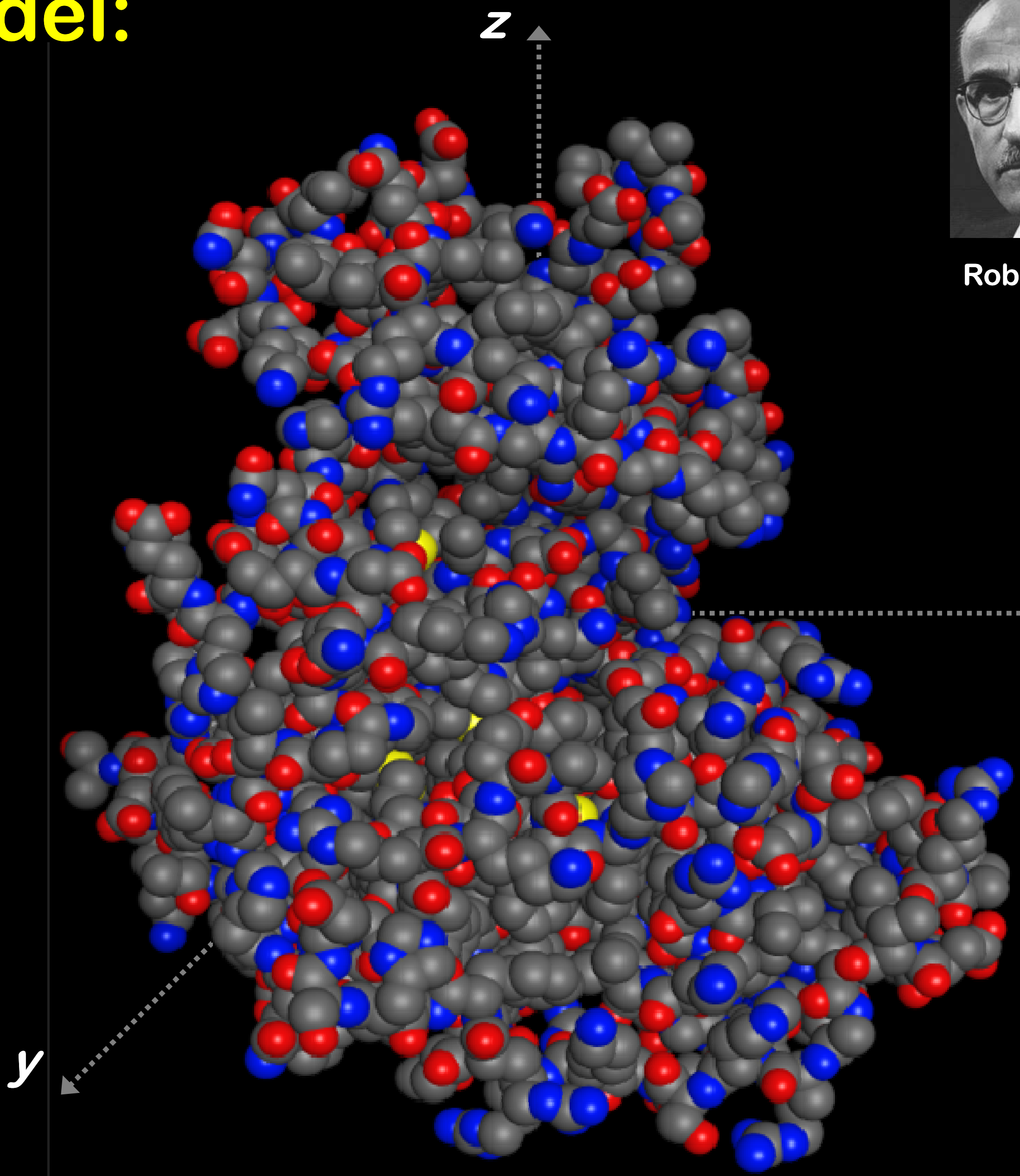


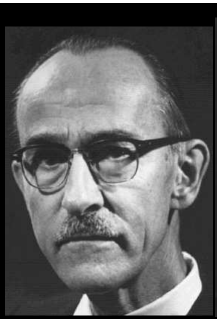



z
C (x, y, z)
x
y


z
x
y
ALPHA
BETA
RANDOM

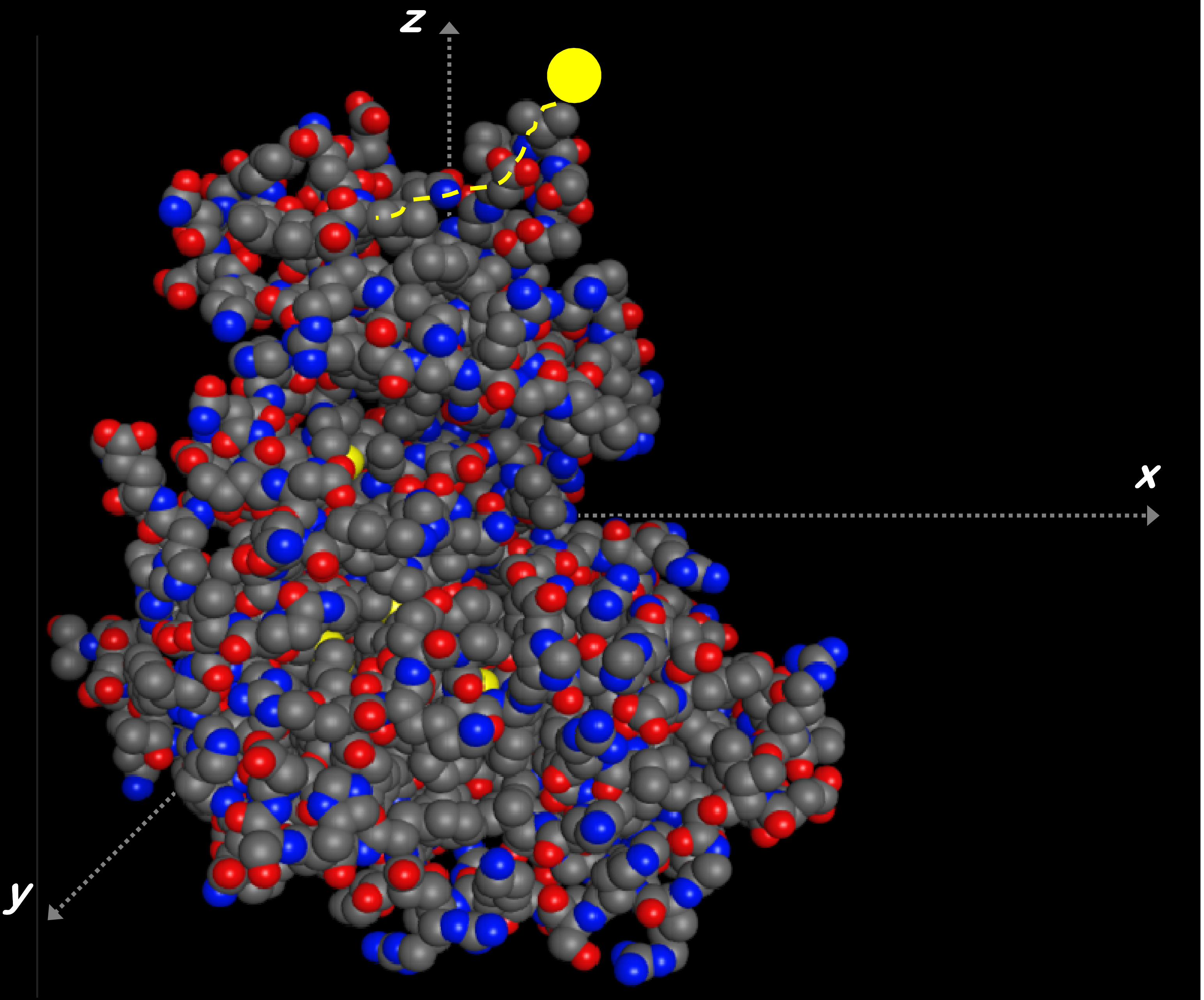
CPK model:

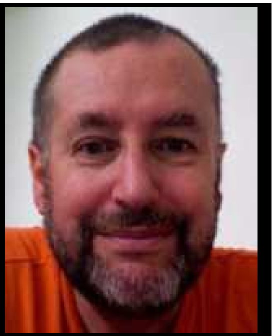



Robert Corey

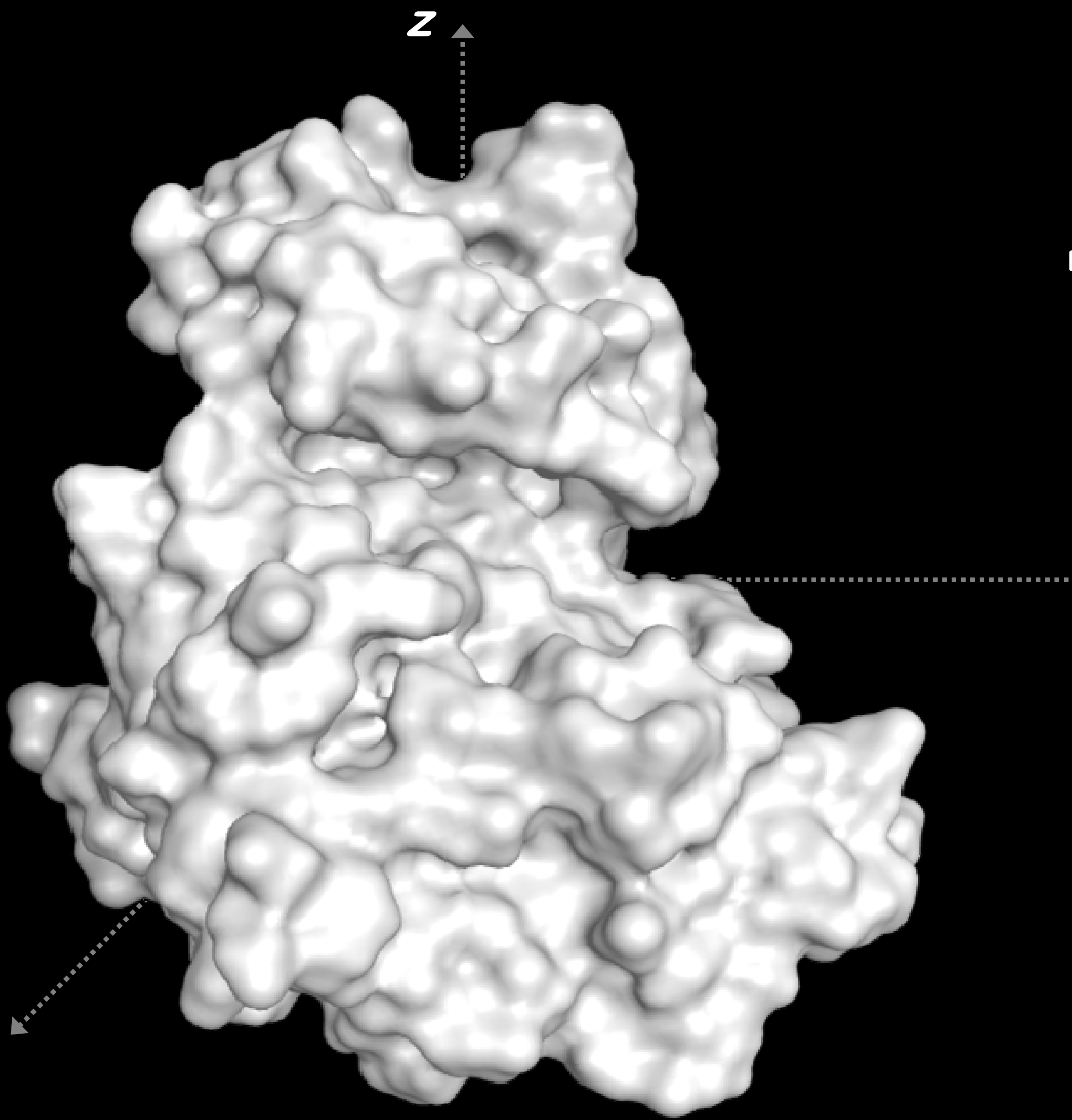

Linus Pauling


Walter Koltun

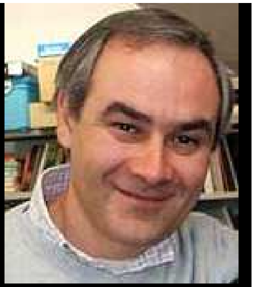

z
x
y

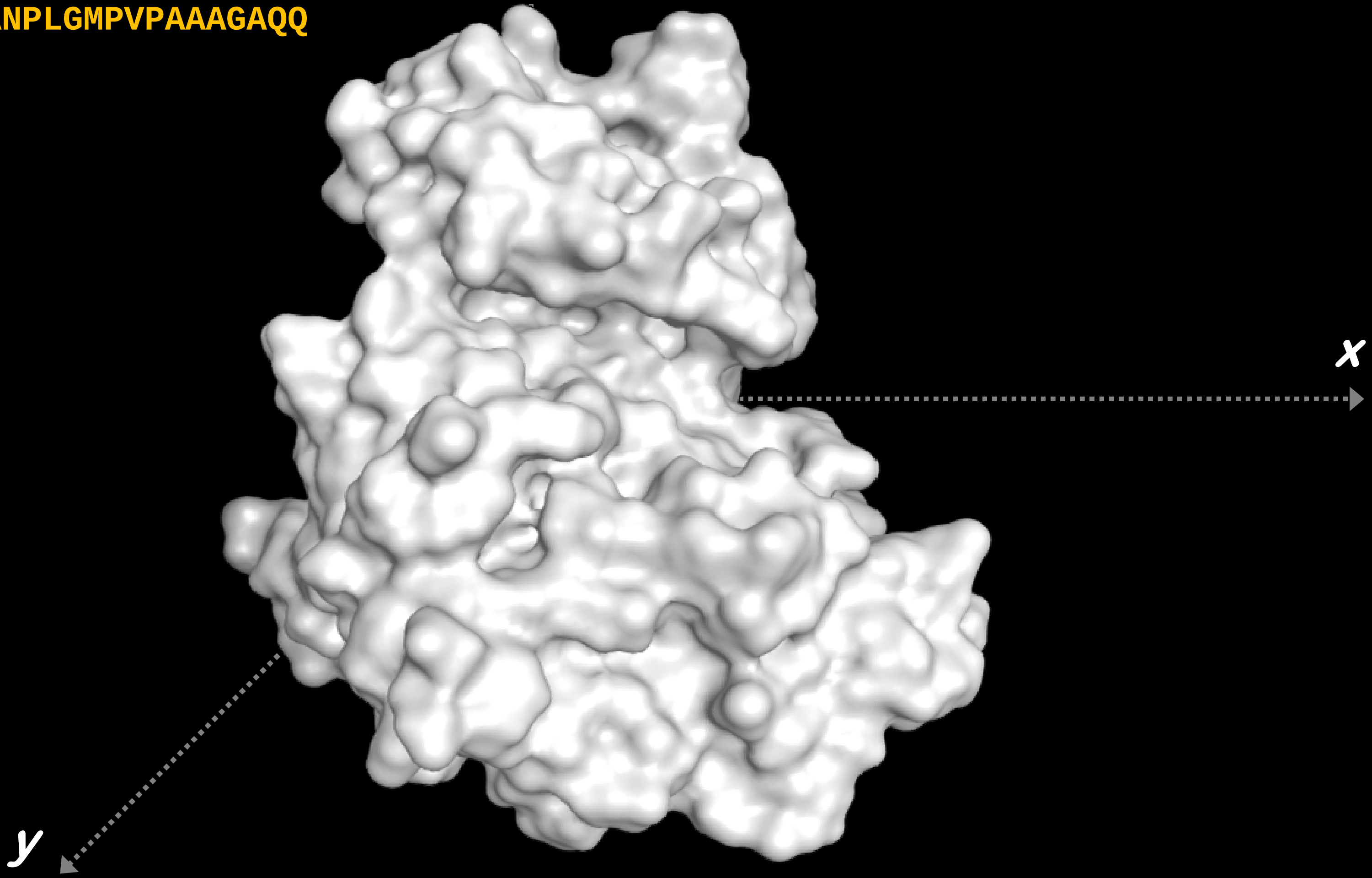


Michael Connolly



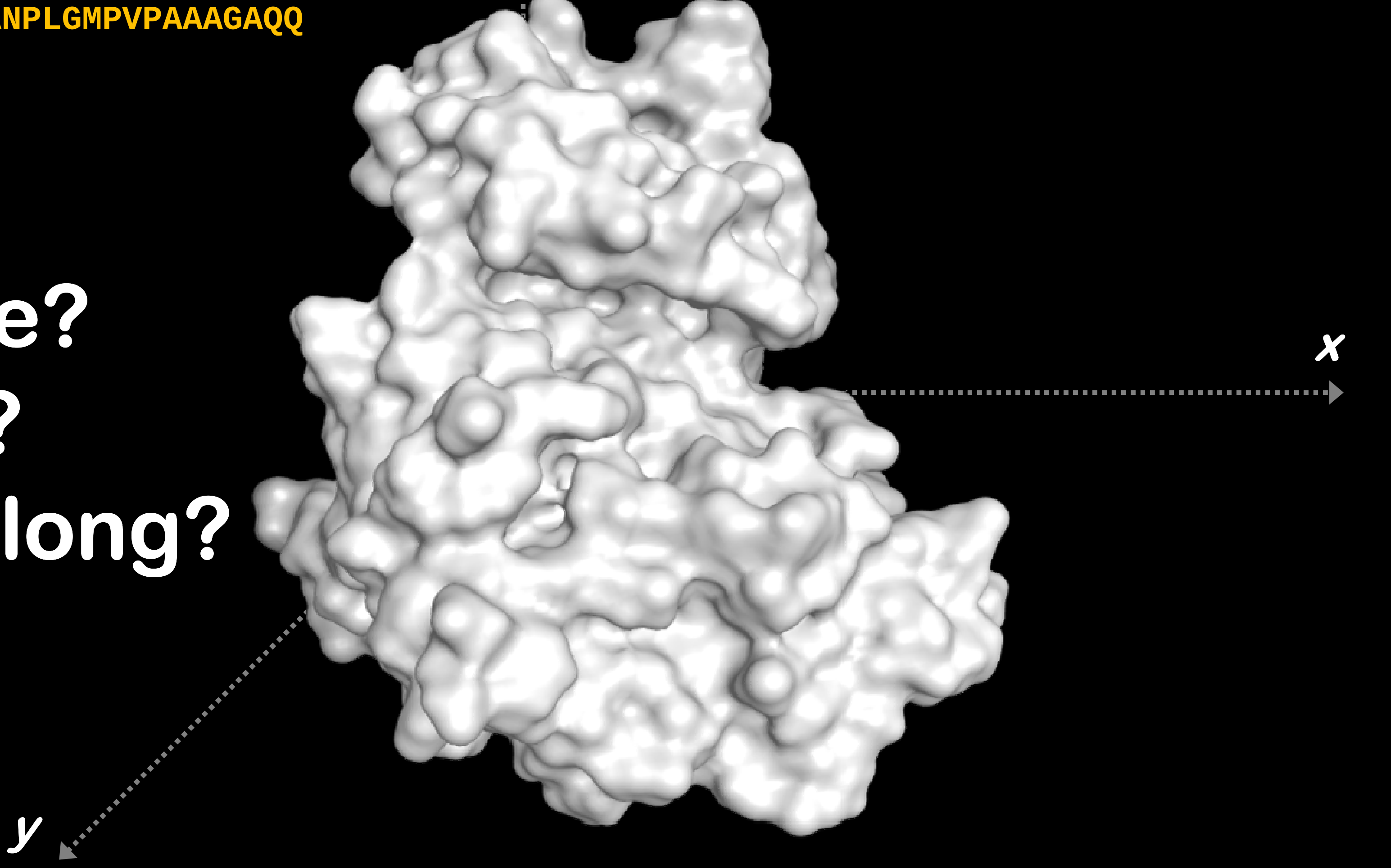
MSGPVPSRARVYTDVNTHRPREYWDYESHVVEWGNQDDYQLVRKLGRGKYSEVFEAINIT
NNEKVVVKILKPVKKKKIKREIKILENLRGGPNIITLADIVKDPVSRTPALVFEHVNNTD
FKQLYQTLTDYDIRFYMYEILKALDYCHSMGIMHRDVKPHNVMIDHEHRKLRLIDWGLAE
FYHPGQEYNVRVASRYFKGPELLVDYQMYDYSLDMWSLGCMLASMIFRKEPFFHGHDNYD
QLVRIAKVLGTEDLYDYIDKYNIELDPRFNDILGRHSRKRWERFVHSENQHLVSPEALDF
LDKLLRYDHQSRLTAREAMEHPYFYTVVKDQARMGSSSMPGGSTPVSSANMMSGISSVPT
PSPLGPLAGSPVIAAANPLGMPVPAAAGAQQ





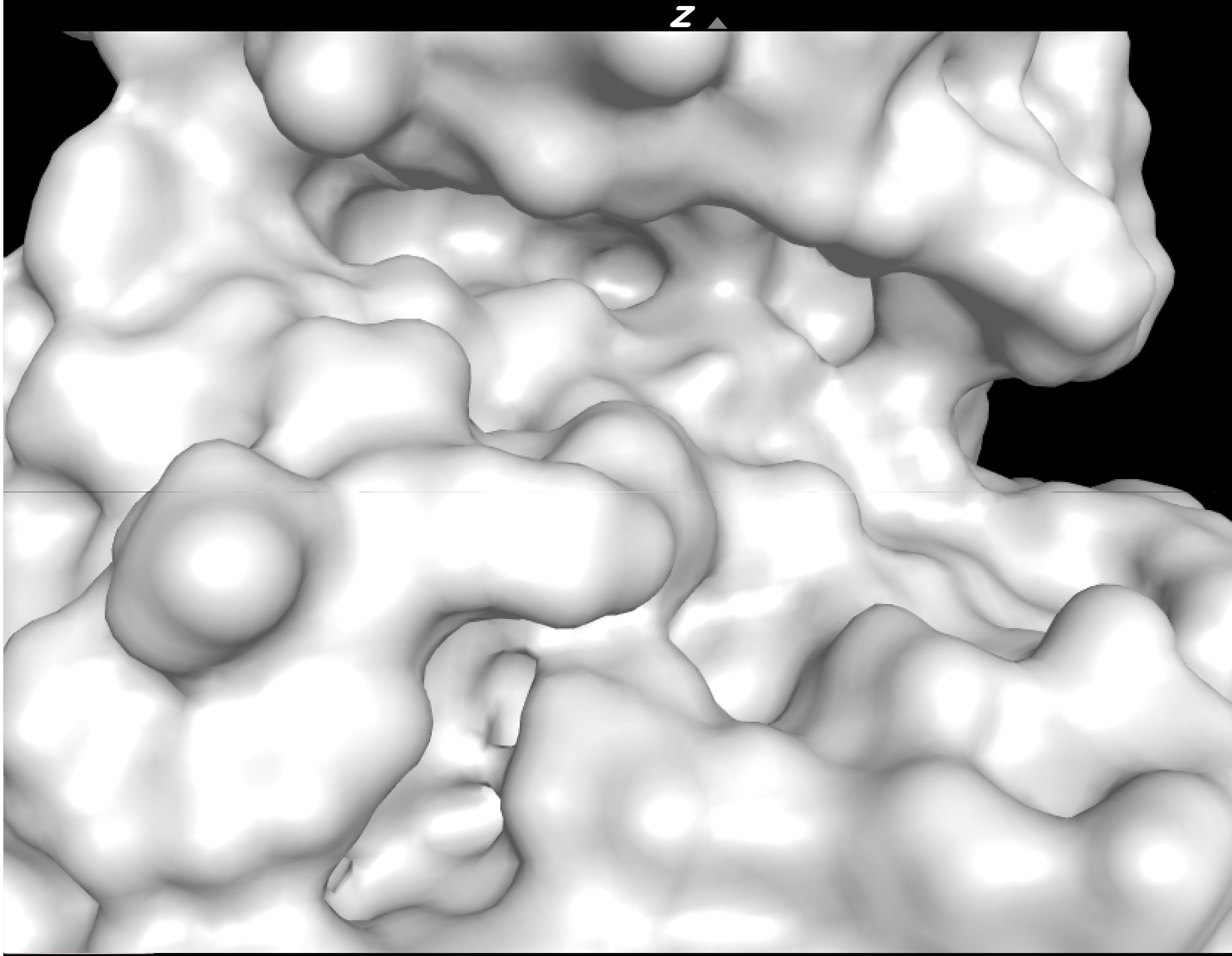
From sequence to topology… from topology to function

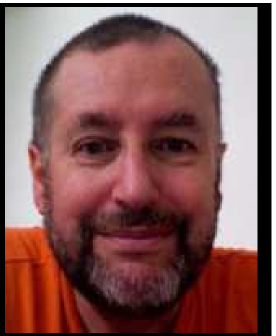
MSGPVPSRARVYTDVNTHRPREYWDYESHVVEWGNQDDYQLVRKLGRGKYSEVFEAINIT
NNEKVVVKILKPVKKKKIKREIKILENLRGGPNIITLADIVKDPVSRTPALVFEHVNNTD
FKQLYQTLTDYDIRFYMYEILKALDYCHSMGIMHRDVKPHNVMIDHEHRKLRLIDWGLAE
FYHPGQEYNVRVASRYFKGPELLVDYQMYDYSLDMWSLGCMLASMIFRKEPFFHGHDNYD
QLVRIAKVLGTEDLYDYIDKYNIELDPRFNDILGRHSRKRWERFVHSENQHLVSPEALDF
LDKLLRYDHQSRLTAREAMEHPYFYTVVKDQARMGSSSMPGGSTPVSSANMMSGISSVPT
PSPLGPLAGSPVIAAANPLGMPVPAAAGAQQ

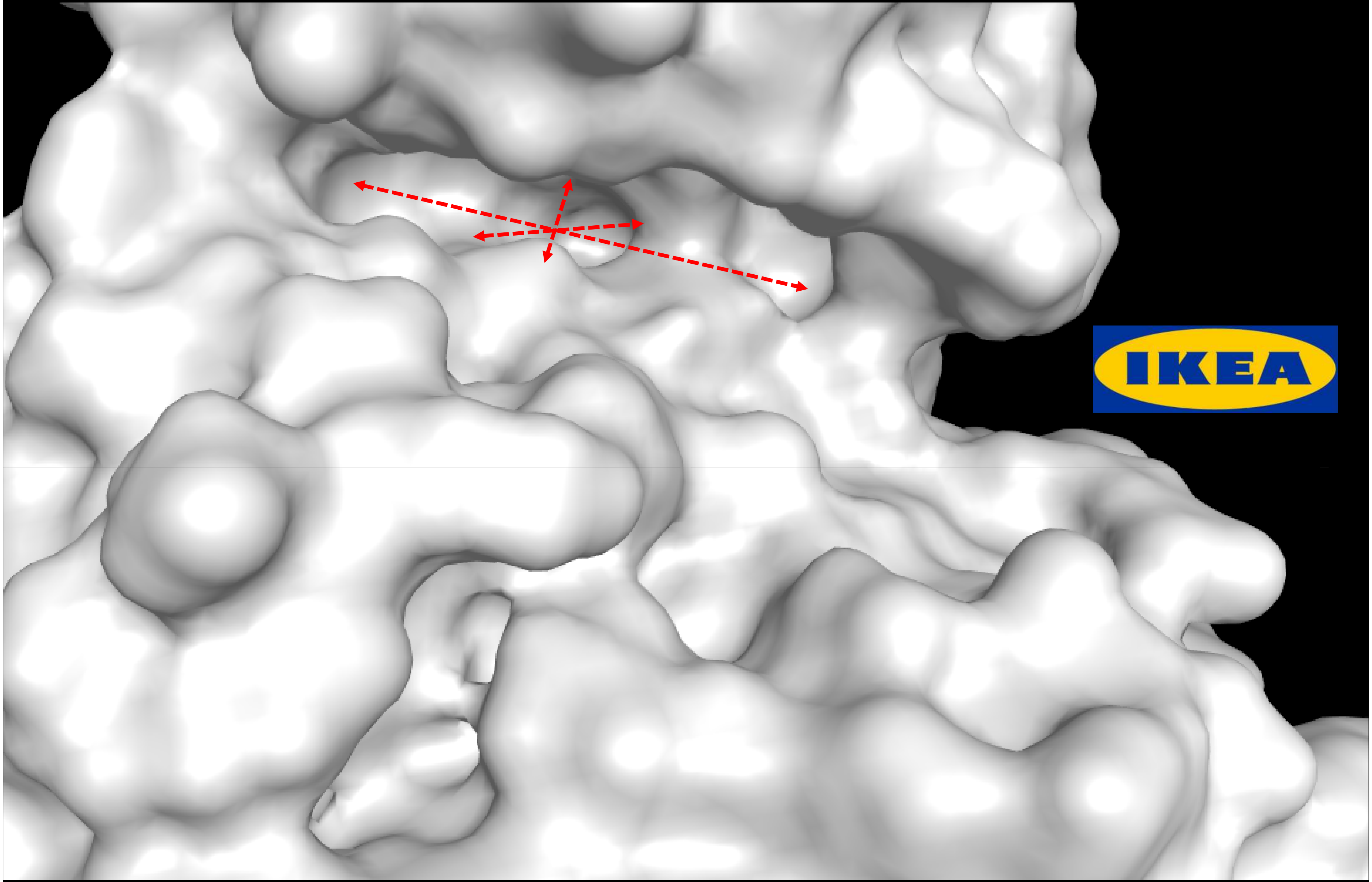


1. where?
2. how?
3. how long?

From sequence to topology… from topology to recognition

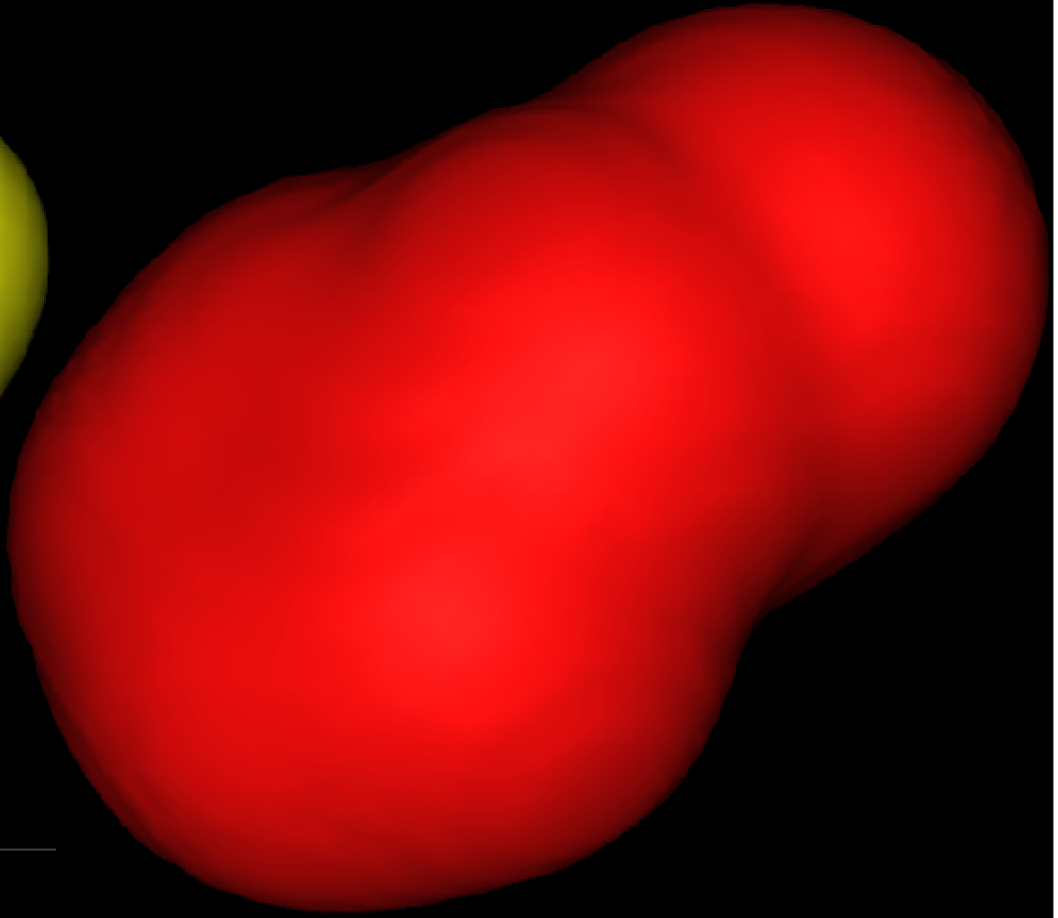



Michael Connolly


IKEA

Come back to surface representations: are you able to calculate the volume of these objects?

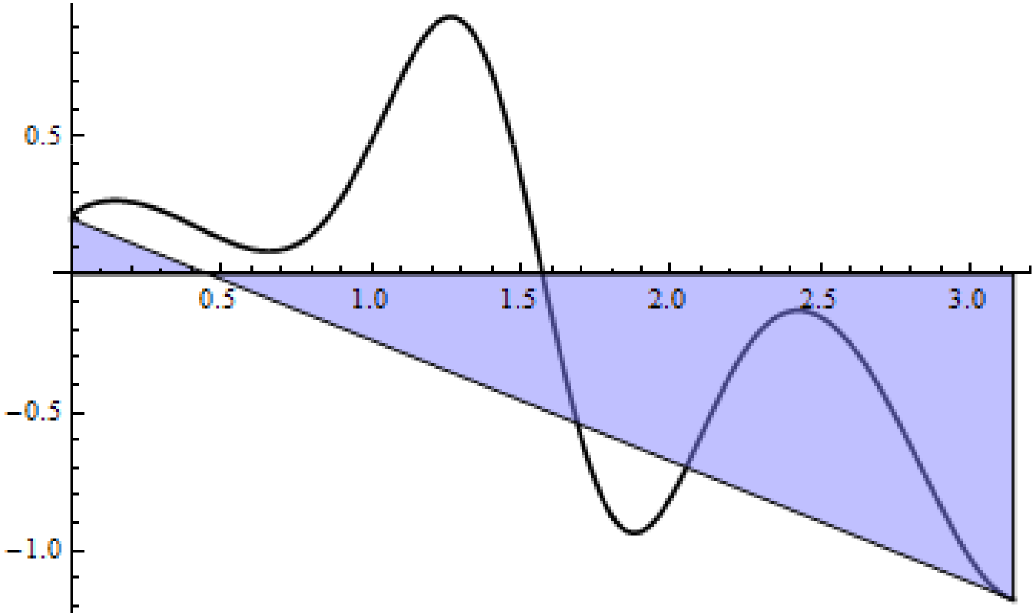
SES or Connolly Surface



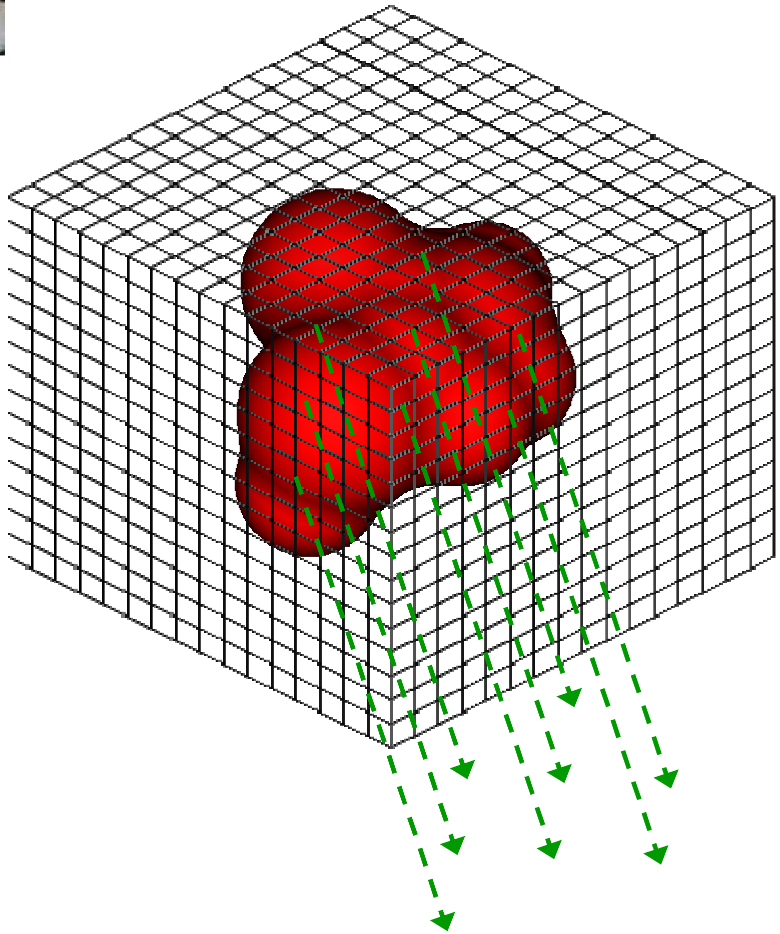
SAS

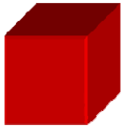
do you remember the *trapezoidal rule*?

$$\int_a^b f(x)dx \approx (b-a)\frac{f(a)+f(b)}{2}$$

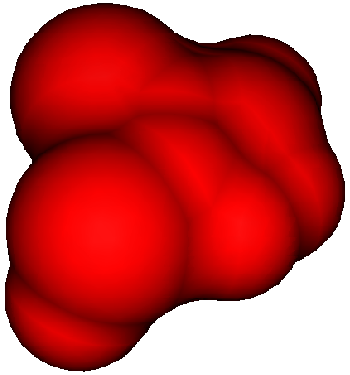


a 3D version of our *trapezoidal rule*

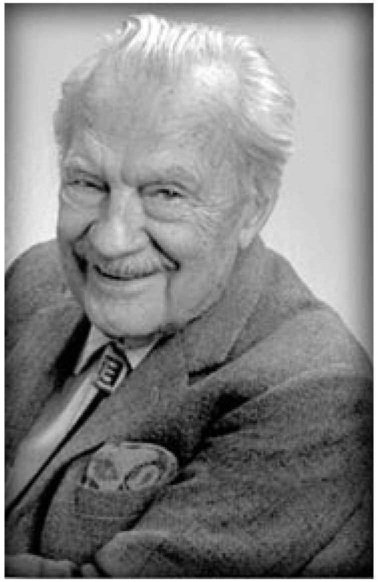




$= V_i$



$$V_{mol} \cong n\, V_i$$



Nicholas Constantine Metropolis and the Monte Carlo method

(N. Metropolis and S. Ulam. 1949. The Monte Carlo method. *Journal of the American Statistical Association 44:335-341.)*

The Monte Carlo method is one of the ten algorithms that have had major influence in the development of science and engineering in the 20th century.

Here is the list!

1946: The Metropolis Algorithm for Monte Carlo. Through the use of random processes, this algorithm offers an efficient way to stumble toward answers to problems that are too complicated to solve exactly.

1947: Simplex Method for Linear Programming. An elegant solution to a common problem in planning and decision-making.

1950: Krylov Subspace Iteration Method. A technique for rapidly solving the linear equations that abound in scientific computation.

1951: The Decompositional Approach to Matrix Computations. A suite of techniques for numerical linear algebra.

1957: The Fortran Optimizing Compiler. Turns high-level code into efficient computer-readable code.

1959: QR Algorithm for Computing Eigenvalues. Another crucial matrix operation made swift and practical.

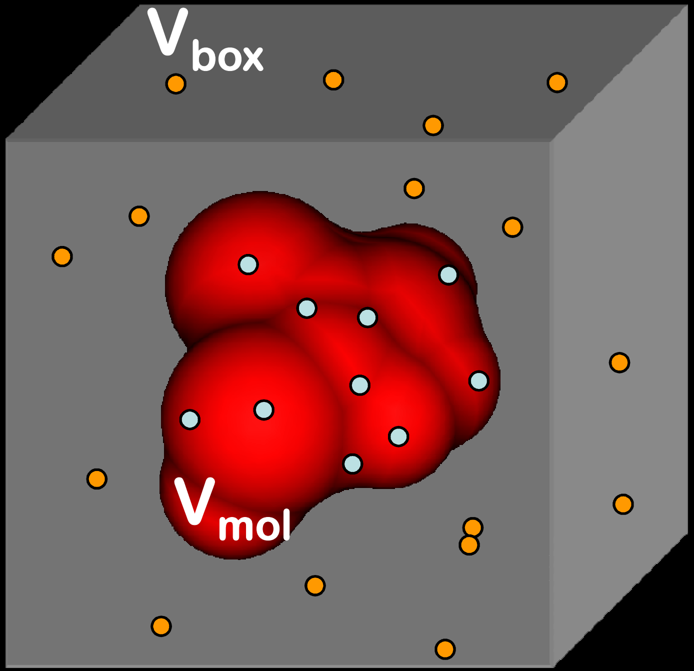
1962: Quicksort Algorithms for Sorting. For the efficient handling of large databases.

1965: Fast Fourier Transform. Perhaps the most ubiquitous algorithm in use today, it breaks down waveforms (like sound) into periodic components.

1977: Integer Relation Detection. A fast method for spotting simple equations satisfied by collections of seemingly unrelated numbers.

1987: Fast Multipole Method. A breakthrough in dealing with the complexity of n-body calculations, applied in problems ranging from celestial mechanics to protein folding.

Monte Carlo Integration Method



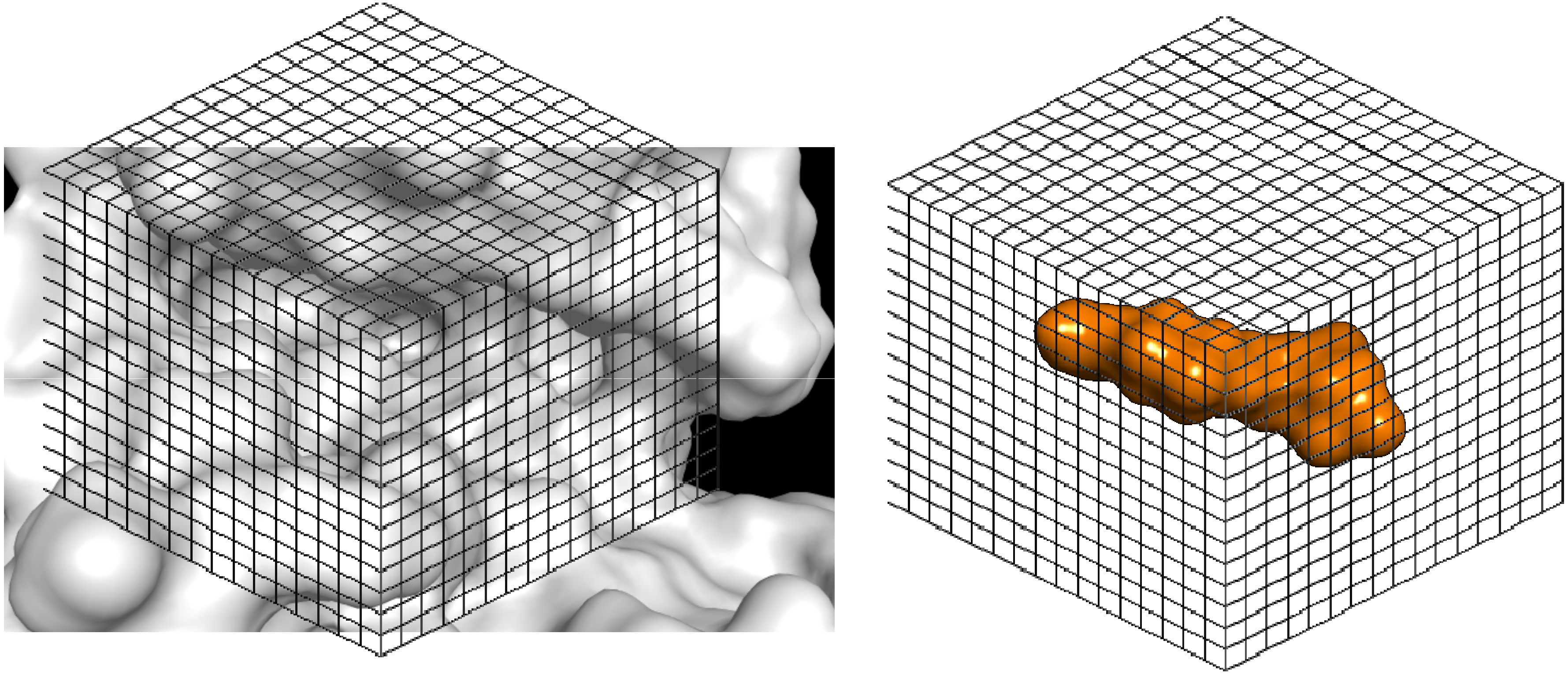
$$V_{mol} : V_{box} \cong n_{mol} : n_{tot}$$

$$V_{mol} \cong V_{box} \times n_{mol} / n_{tot}$$

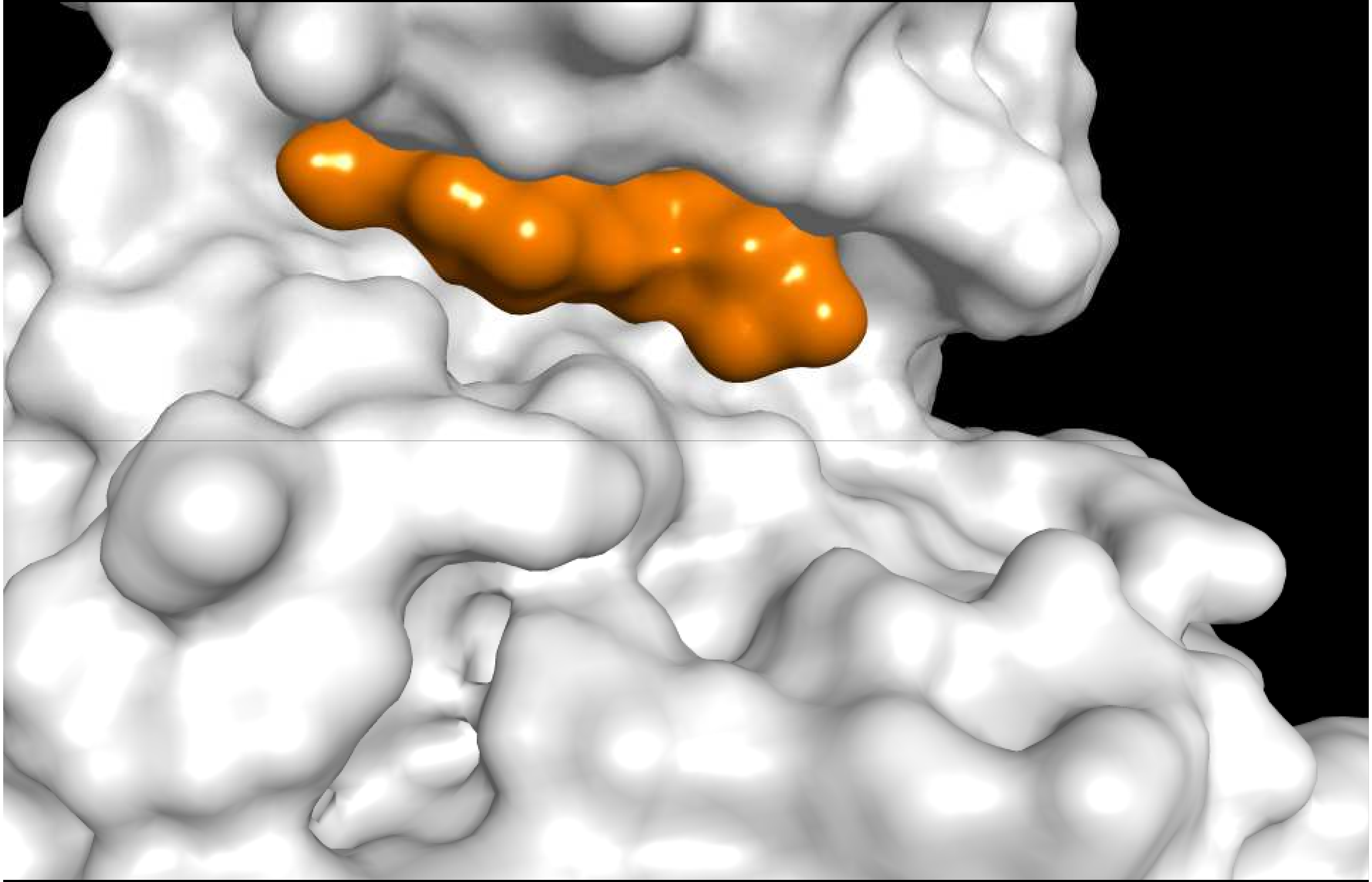
Monte Carlo Integration Method

by Giulio Mattedi (2015/2016)

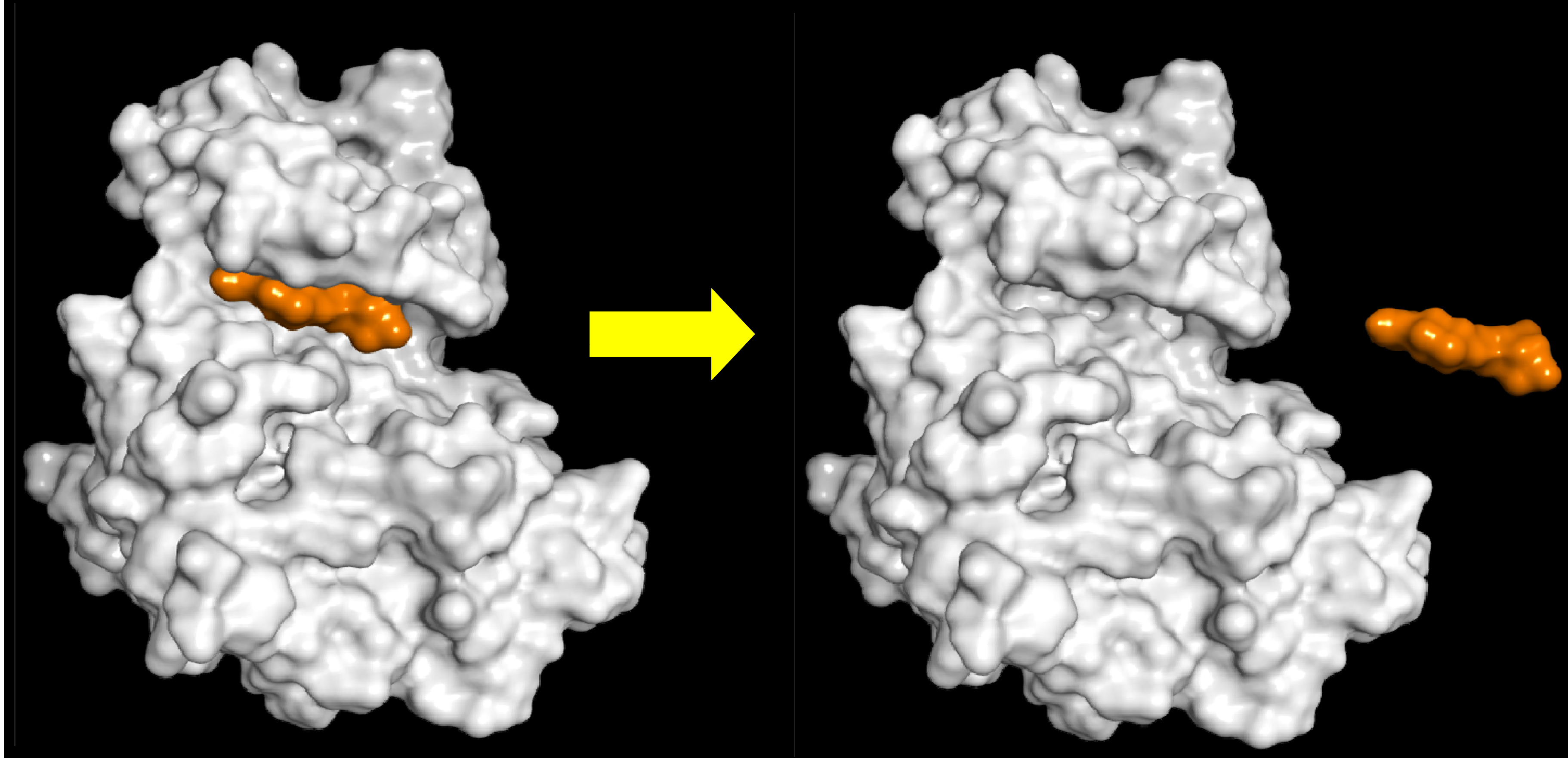
The measure of a *full* (space) equal to the measure of an *empty* (space)?



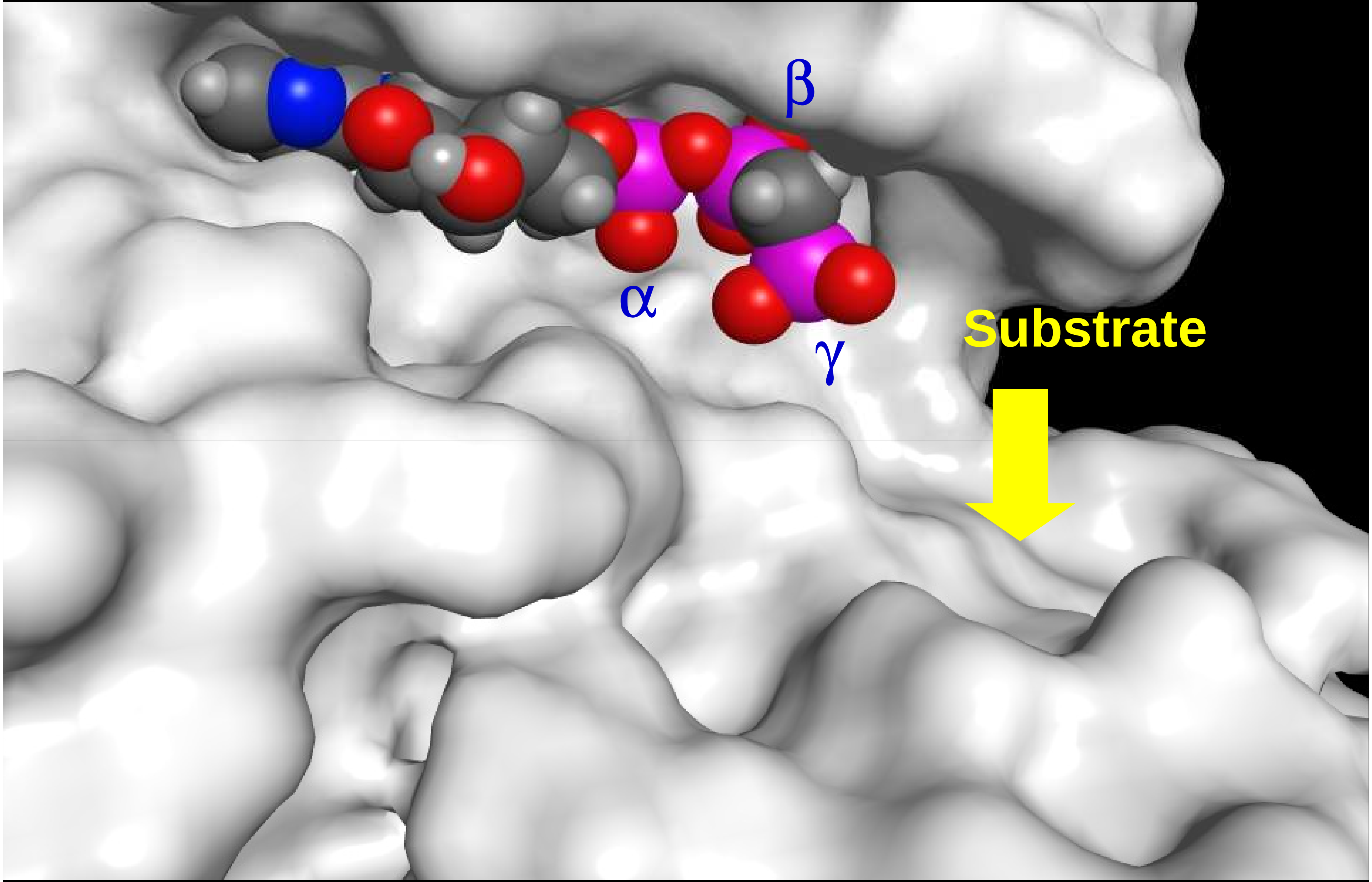
Do you note where the difference is?



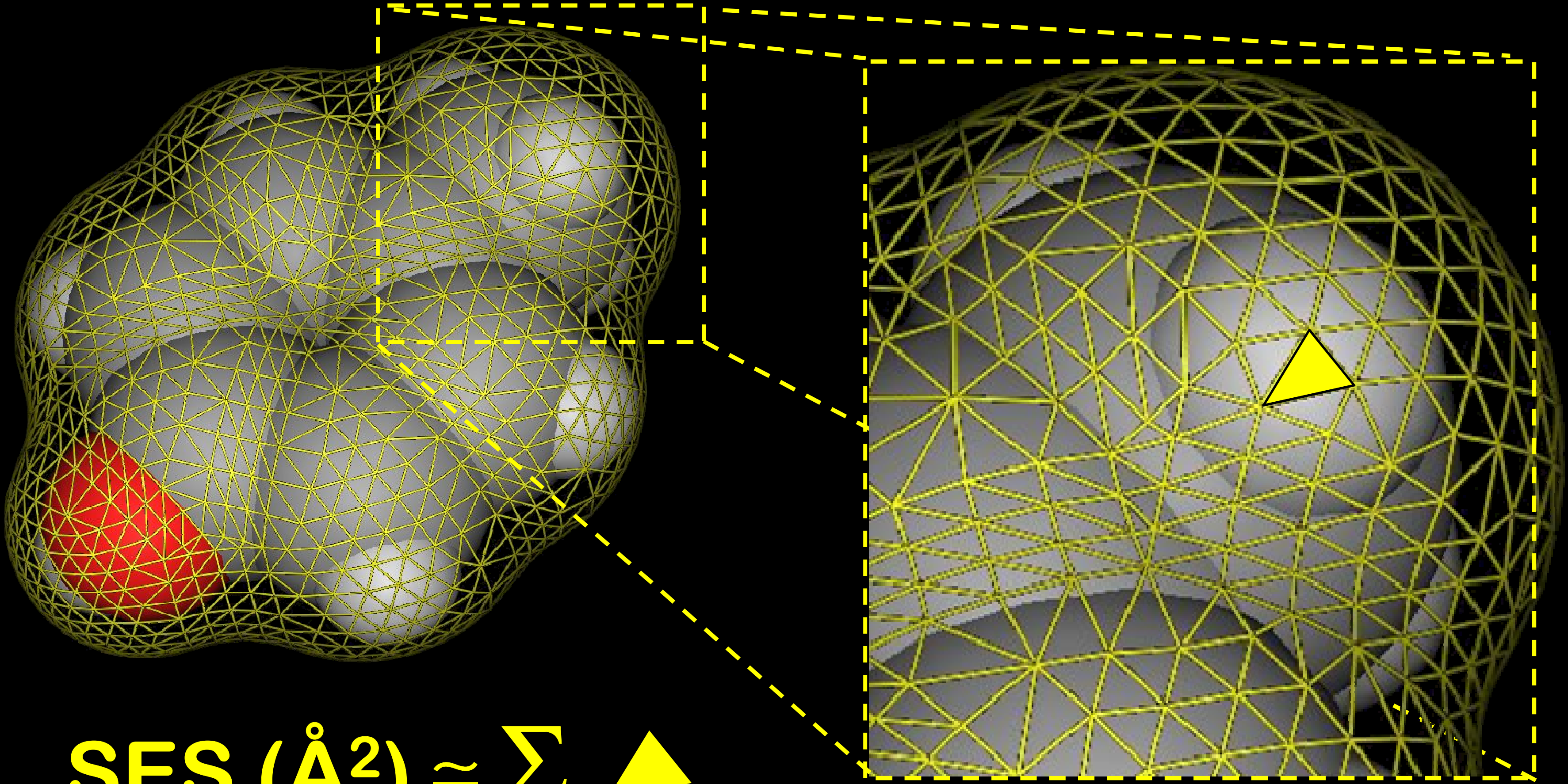
You have now a wonderful tool to estimate the topological complementarity between a cavity and its ligand:



$$\text{Complementarity} \propto \text{Vol}_{\text{cavity}} - \text{Vol}_{\text{ligand}}$$

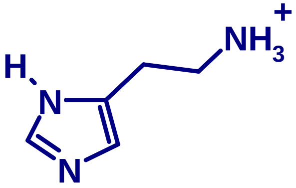

β
α
γ
Substrate

Now also surface extension is very easy to calculate... approximately!



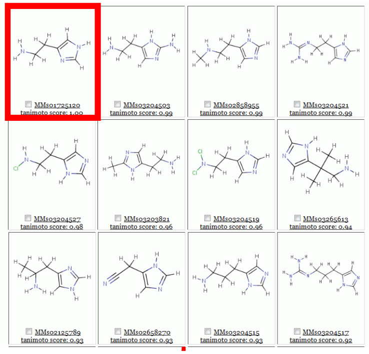
$\text{SES}\ (\text{Å}^2) \cong \Sigma$ ▲

Do you remember:



00001111001010101010

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



☐ MMs01725120 tanimoto score: 1.00

☐ MMs03204503 tanimoto score: 0.99

☐ MMs02858955 tanimoto score: 0.99

☐ MMs03204521 tanimoto score: 0.99

☐ MMs03204527 tanimoto score: 0.98

☐ MMs03203821 tanimoto score: 0.96

☐ MMs03204519 tanimoto score: 0.96

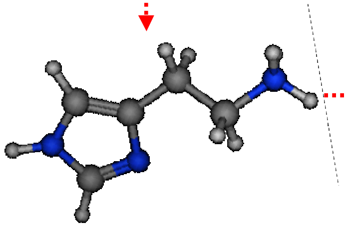
☐ MMs03265613 tanimoto score: 0.94

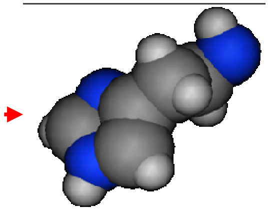
☐ MMs02125789 tanimoto score: 0.93

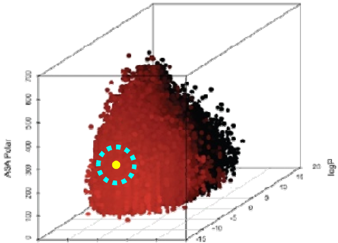
☐ MMs02658270 tanimoto score: 0.93

☐ MMs03204515 tanimoto score: 0.93

☐ MMs03204517 tanimoto score: 0.92



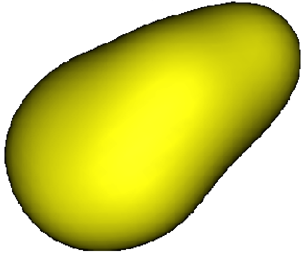


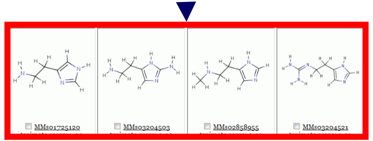


Sterimol

Vol (Å^3)

SES (Å^2)





☐ MMs01725120

☐ MMs03204503

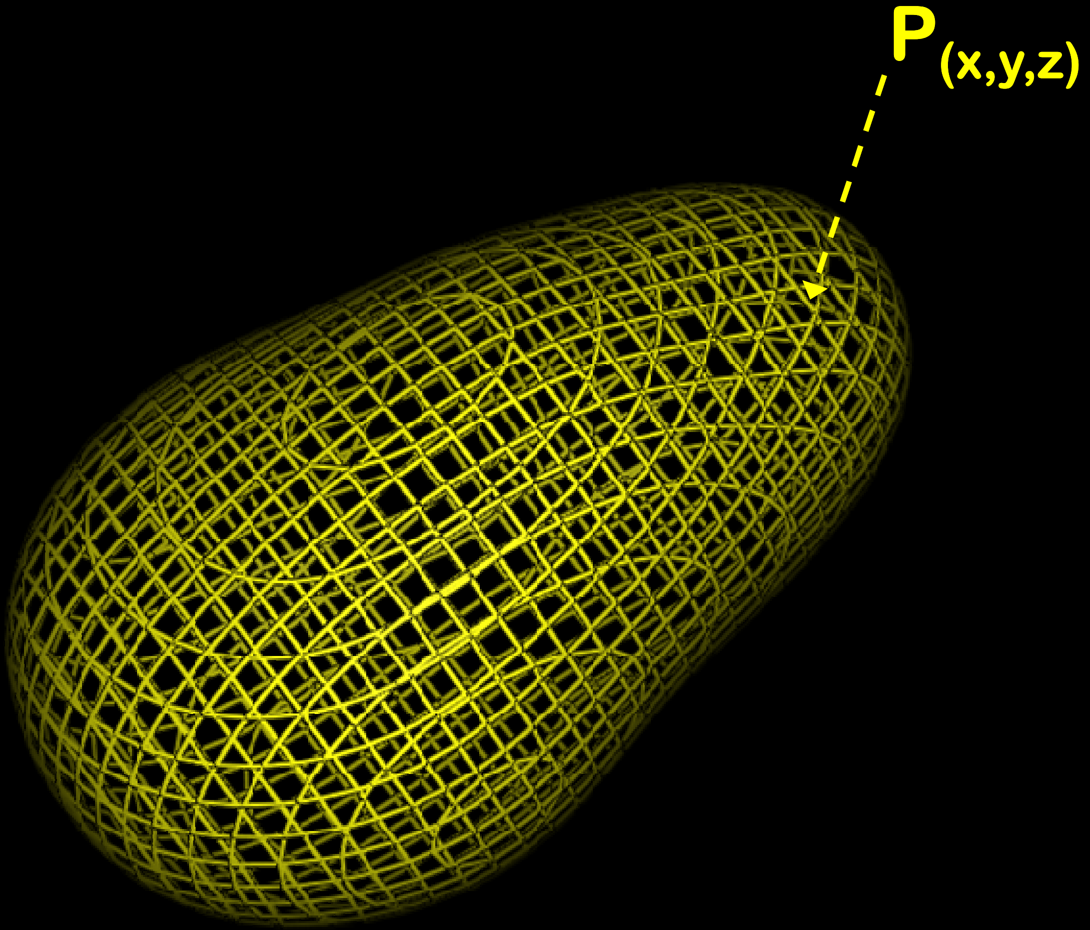
☐ MMs02858955

☐ MMs03204521

And now I would like to morph this…

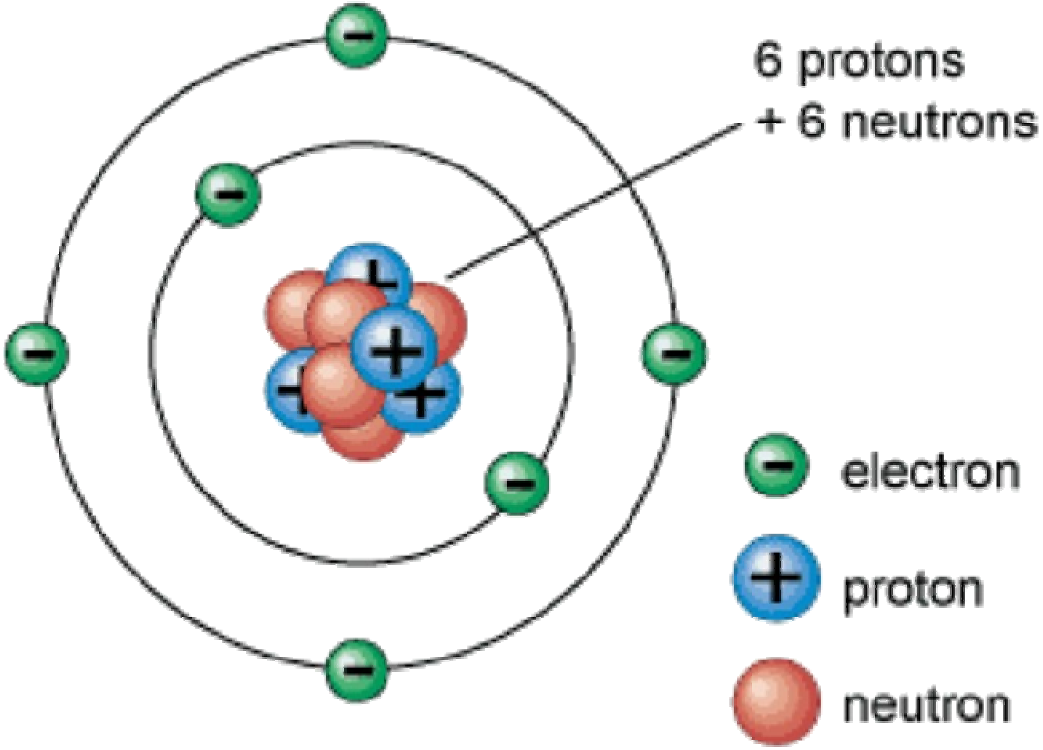
in this!

Molecular surfaces as a locus for the projection of molecular properties:



As chemists, what is the first property that you would like to see projected ?

We cannot forget that under the real surface we surely find... charges!



Carbon atom

Back when I was young…

"It is generally accepted that receptor and substrate molecules recognize each other at their molecular surfaces. Therefore, the binding strength of a receptor-drug complex depends on the shape of the substrate surface and on the **distribution of certain properties on this surface**. Any method attempting to model biological activity should take into account this information and try to correlate it to biological activity…"

by Johann Gasteiger *et al J.A.C.S.* 1995, **117**, 7769-7775

Considering that molecules are an ensemble of electric charges… probably the distribution of their *electrostatic potentials*?

How we can virtualize the projection of the electrostatic potential on a molecular surface?

We can surely start with him!

Charles Augustin de Coulomb (Angoulême, 14 giugno 1736 – Parigi, 23 agosto 1806).

$$F_{el} = k_0 \frac{q_1 q_2}{r^2}$$

$$\text{Coulomb's costant}: k_0 = 9 \times 10^9 \frac{Nm^2}{C^2}$$

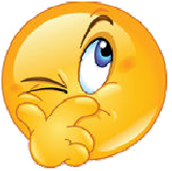
$$k_0 = \frac{1}{4\pi\varepsilon_0}$$

where ε_0 is the electric permittivity of free space

$$\varepsilon_0 = 8{,}9 \times 10^{-12} \frac{C^2}{Nm^2}$$

I try to recite it:

"The magnitude of the electrostatic force of interaction between two **point charges** is directly proportional to the scalar multiplication of the magnitudes of charges and inversely proportional to the square of the distance between them."



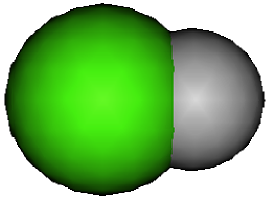
How we can deal with the *point charges* concept when, in the real world, we usually deal with charge distributions?

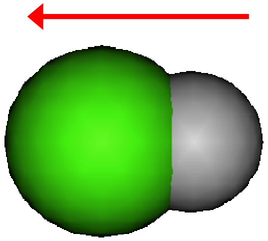
Probably, using a very interesting physical trick:

$\Delta\chi = 1.9$

$\mu = 1.82$ D

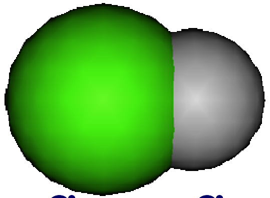
F - H





$\chi_F = 4.0$ $\chi_H = 2.1$

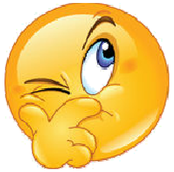
$-\delta$ $+\delta$



q, partial atomic charge q_F q_H

$\Delta\chi$ f $(q_F; q_H)$

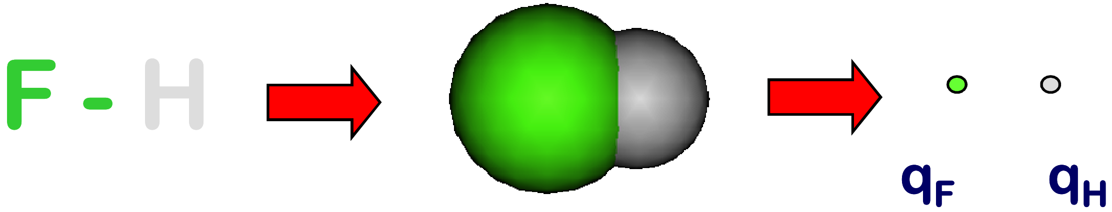
μ f $(q_F; q_H)$



Probably, using a very interesting physical trick:

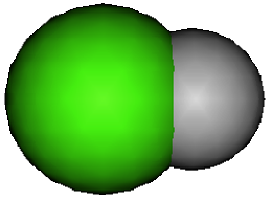
Introducing an atomic property, definite after the formation of a chemical bond, called *partial atomic charge*.

It is very important to underline that this property is not a real physical atomic property but it is very useful to use the Coulombic's law!!



The formalism of Johann Gasteiger:

$\Delta\chi = 1.9$



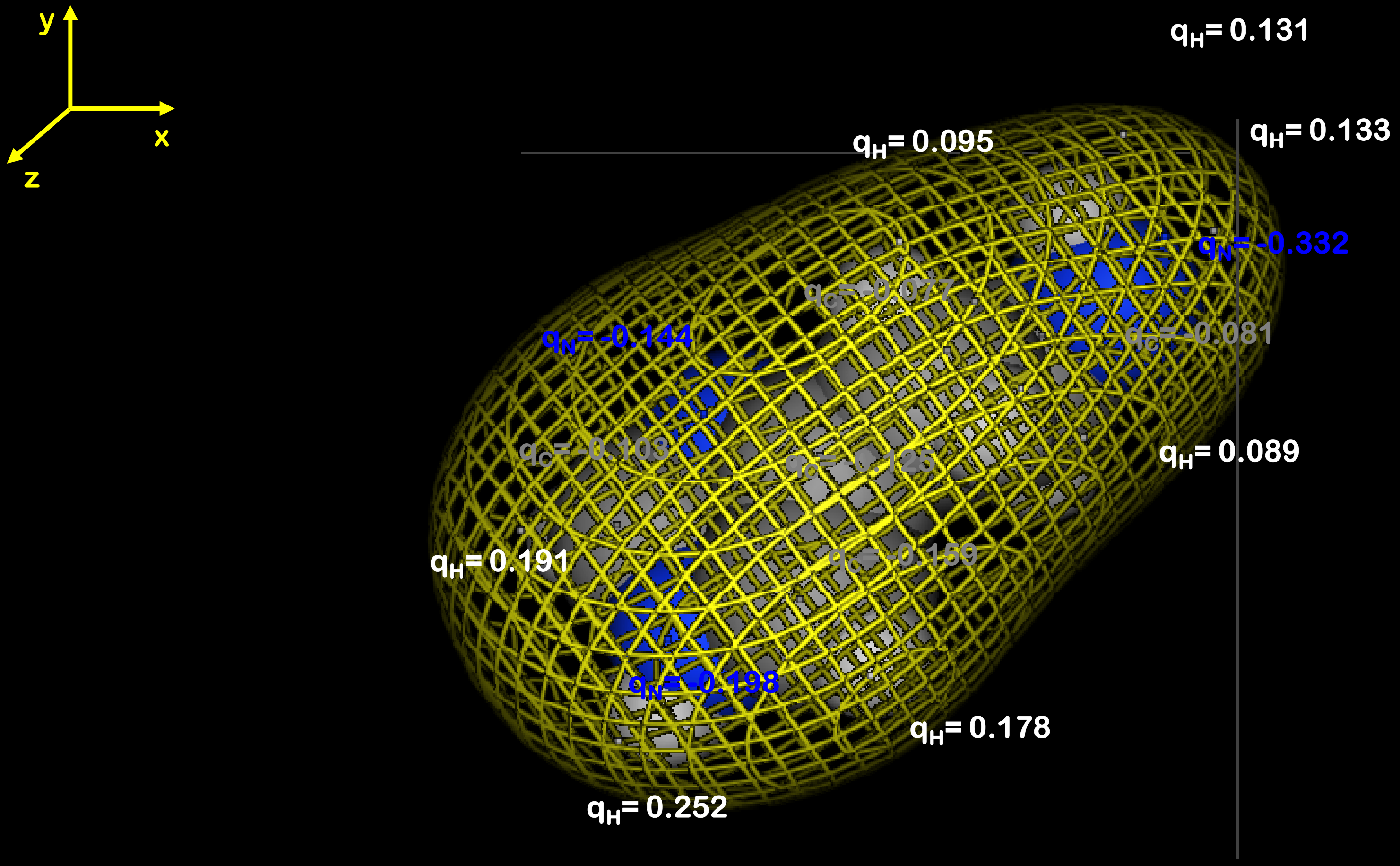
q_F q_H

$\chi_F = 4.0$ $\chi_H = 2.1$

$\Delta\chi \; f\,(q_F; q_H)$

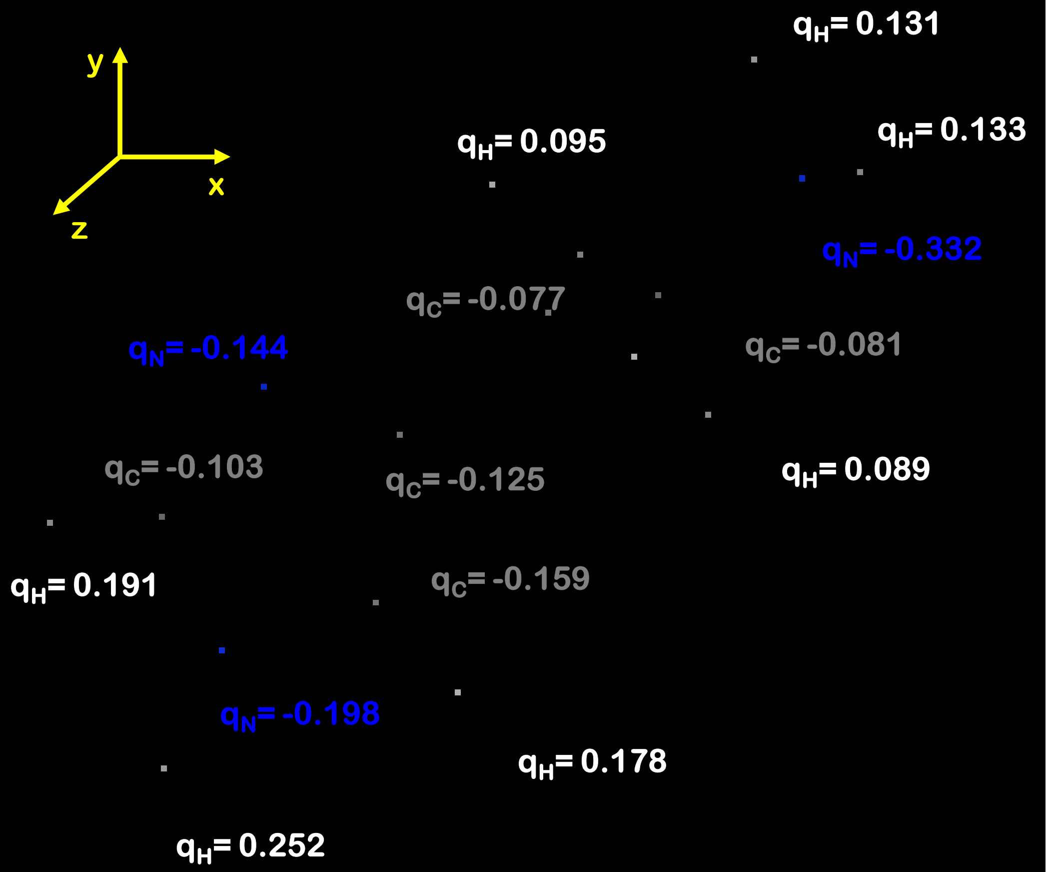
$$\chi_A = a_A q_A + b_A q_A^2$$

Informatically speaking, it is very clear…



vdW *versus* Connolly Surface

The partial atomic charges are *punctiform ambassadors* of charge distribution!



$q_H = 0.131$

$q_H = 0.095$

$q_H = 0.133$

$q_N = -0.332$

$q_C = -0.077$

$q_N = -0.144$

$q_C = -0.081$

$q_C = -0.103$

$q_C = -0.125$

$q_H = 0.089$

$q_H = 0.191$

$q_C = -0.159$

$q_N = -0.198$

$q_H = 0.178$

$q_H = 0.252$

According to Cramer (2002), this earlier classification of methods involves the following four classes:

Class I: are those that are not determined from quantum mechanics, but from some intuitive or empiric approach. These approaches can be based on experimental data such as dipoles and electronegativities.

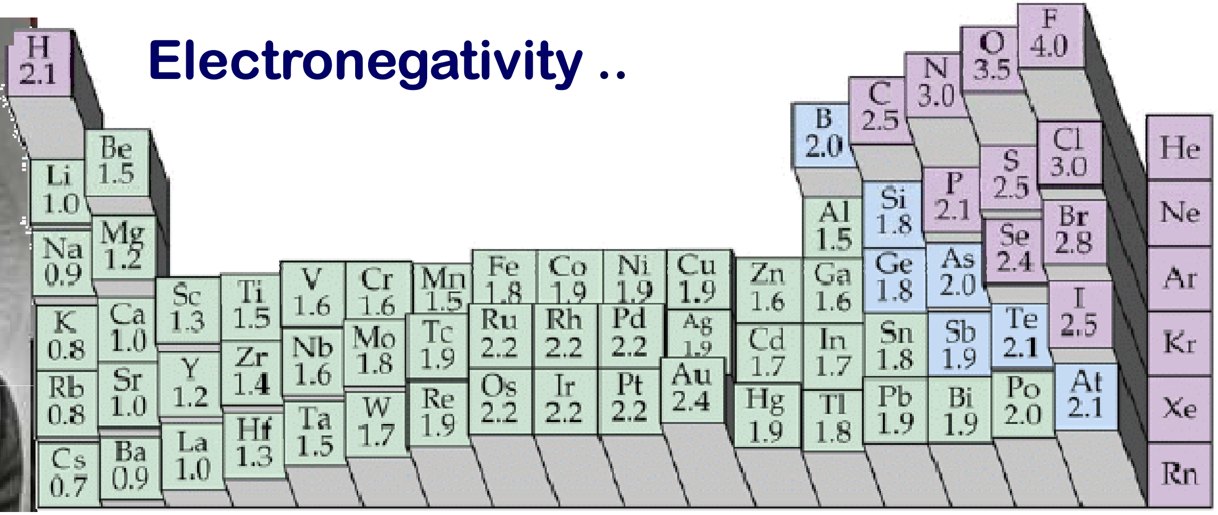
Class II: are derived from partitioning the molecular wave function using some arbitrary, orbital based scheme.

Class III: are based on a partitioning of a physical observable derived from the wave function, such as electron density.

Class IV: are derived from a semiempirical mapping of a precursor charge of type II or III to reproduce experimentally determined observables such as dipole moments.

C. J. Cramer (2002). *Essentials of Computational Chemistry: Theories and Methods*. Wiley. pp. 278–289.

Partial atomic charges derived from electronegativity values (Class I):



Electronegativity ..

Pauling (1932):

$$\chi_{\mathrm{A}} - \chi_{\mathrm{B}} = (\mathrm{eV})^{-1/2}\sqrt{E_{\mathrm{d}}(\mathrm{AB}) - [E_{\mathrm{d}}(\mathrm{AA}) + E_{\mathrm{d}}(\mathrm{BB})]/2}$$

Mulliken (1934):

$$\chi = 0.187(E_{\mathrm{i}} + E_{\mathrm{ea}}) + 0.17$$

Allred-Rochow (1958):

$$\chi = 0.359\frac{Z^*}{r_{\mathrm{cov}}^2} + 0.744$$

Sanderson (1983):

Allen (1989):

$$\chi = \frac{n_{\mathrm{s}}\varepsilon_{\mathrm{s}} + n_{\mathrm{p}}\varepsilon_{\mathrm{p}}}{n_{\mathrm{s}} + n_{\mathrm{p}}}$$

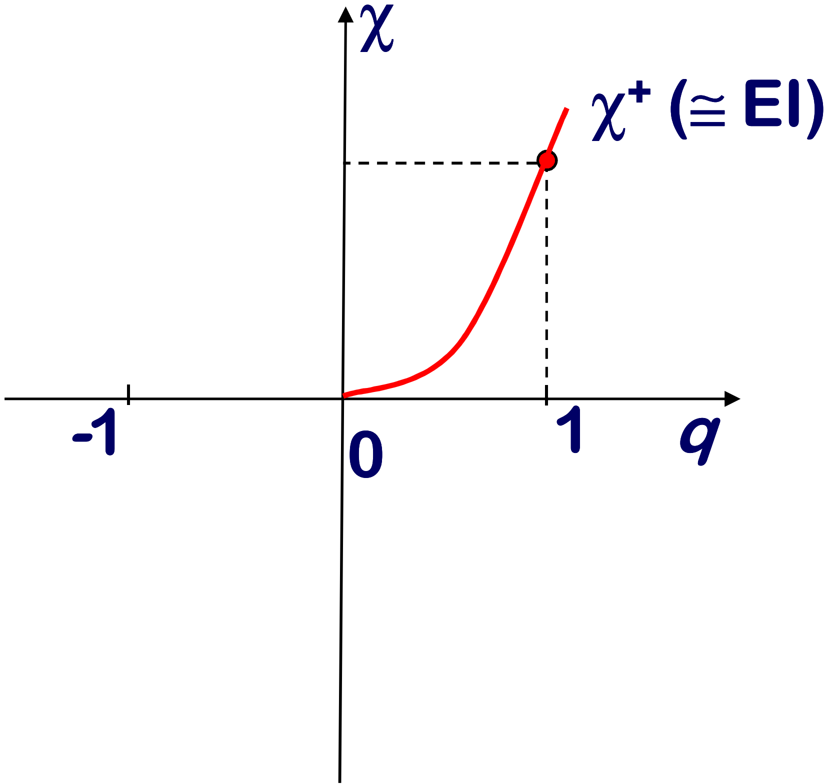
Partial atomic charges derived from electronegativity values (Class I):

H-H	H-F	H-Cl	H-Br
$\Delta\chi = 0$	$\Delta\chi = 1.9$	$\Delta\chi = 0.9$	$\Delta\chi = 0.7$
$q_H\,(H_2)$	$q_H\,(HF)$ $q_F\,(HF)$	$q_H\,(HCl)$ $q_{Cl}\,(HCl)$	$q_H\,(HBr)$ $q_{Br}\,(HBr)$

The mission is very clear: find the optimal values of q to accurately calculate values of $\Delta\chi$!

PEOE formalism (Partial Equalization of Orbital Electronegativities), Gastaiger 1980:

In this formalism the electronegativity χ of an atom A, is described as a quadratic function of its partial atomic charge *q*, where *a* e *b* parameters have being calculated:



$$\chi_A = a_A q_A + b_A q_A^2$$

PEOE formalism (Partial Equalization of Orbital Electronegativities), Gastaiger 1980:

For the determination of *a* and *b* parameters, we can follow this consideration:

$$\chi_A = a_A q_A + b_A q_A^2$$

When the value of the *partial atomic charge q* its equal to its possible *formal charge* (+1 or -1) then the electronegative value χ must be coincident with the *ionization energy* (I.E. for a formal charge equal to +1) or to the *electron affinity* (E.A. for a formal charge equal to -1)

$$q_A \to +1 \ \text{ allora } \ \chi_A \to I.E.$$

$$q_A \to -1 \ \text{ allora } \ \chi_A \to E.A.$$

Beware of units of measurement !!

PEOE formalism (Partial Equalization of Orbital Electronegativities), Gastaiger 1980:

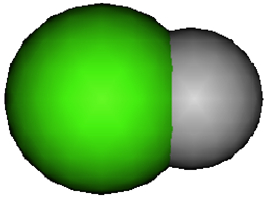
Now could be much easier to calculate them:

$$I.E._{(A)} \cong a_A + b_A$$

$$E.A._{(A)} \cong -a_A + b_A$$

PEOE formalism (Partial Equalization of Orbital Electronegativities), Gastaiger 1980:

Considering two atoms A e B:



$q_F = -\delta$; $q_H = \delta$

$$q_A + q_B = 0$$

$$\Delta\chi = \chi_A - \chi_B$$

$$\chi_A = a_A q_A + b_A q_A^2 \qquad \chi_B = a_B q_B + b_B q_B^2$$

$$\Delta\chi = a_A q_A + b_A q_A^2 - a_B q_B - b_B q_B^2$$

$$\Delta\chi = a_A q_A + b_A q_A^2 + a_B q_A + b_B q_A^2$$

$$\Delta\chi = (a_A + a_B) q_A + (b_A + b_B) q_A^2$$

Do you remember how to solve it?

PEOE formalism (Partial Equalization of Orbital Electronegativities), Gastaiger 1980:

H-H	H-F	H-Cl	H-Br
$\Delta\chi = 0$	$\Delta\chi = 1.9$	$\Delta\chi = 0.9$	$\Delta\chi = 0.7$
$q_H\,(H_2)$	$q_H\,(HF)$ $q_F\,(HF)$	$q_H\,(HCl)$ $q_{Cl}\,(HCl)$	$q_H\,(HBr)$ $q_{Br}\,(HBr)$
	±0.267	±0.145	±0.113

PEOE formalism (Partial Equalization of Orbital Electronegativities), Gastaiger 1980:

C-H (CH_4)

$\Delta\chi = 0.3$

$q_C = 4q_H$

C-H (CH_3F)

$\Delta\chi_{C\text{-}H} = 0.3$

$\Delta\chi_{C\text{-}F} = -1.6$

$q_C = 3q_H + q_F$

$q_C = -0.078$

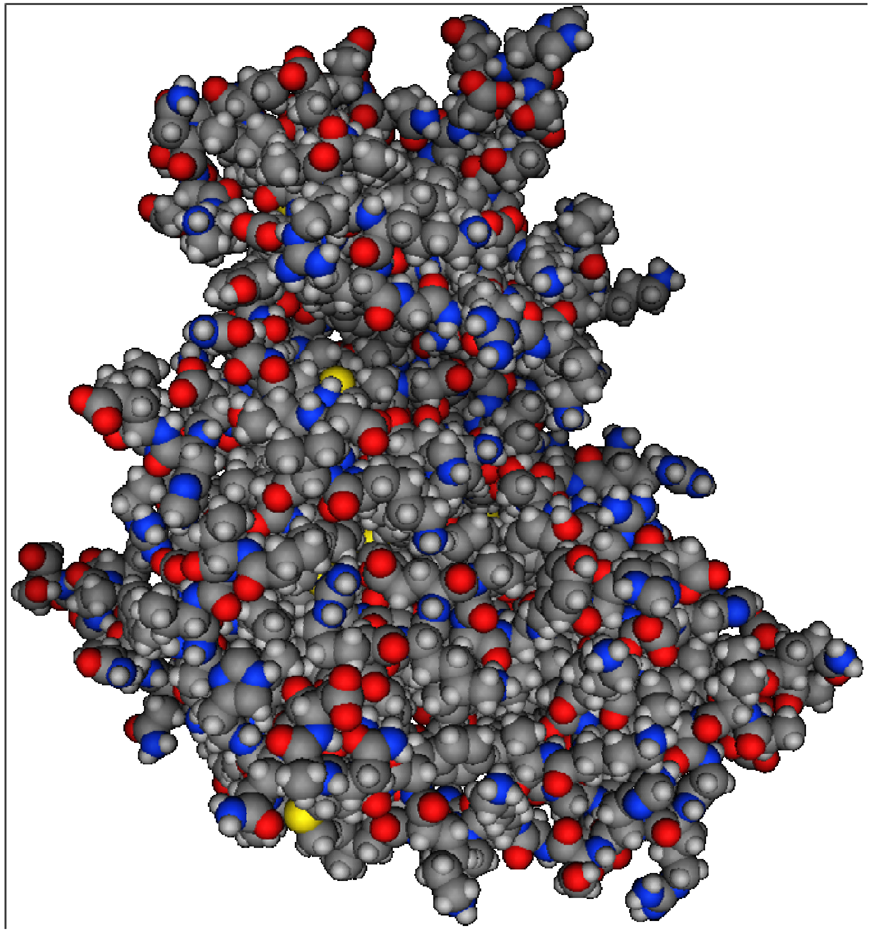
$q_H = 0.019$

$q_C = 0.079$

$q_H = 0.058$

$q_F = -0.253$

Why the Class I methods are very popular in calculating partial atomic charges… even if sometimes are not so accurate? Here is the answer:

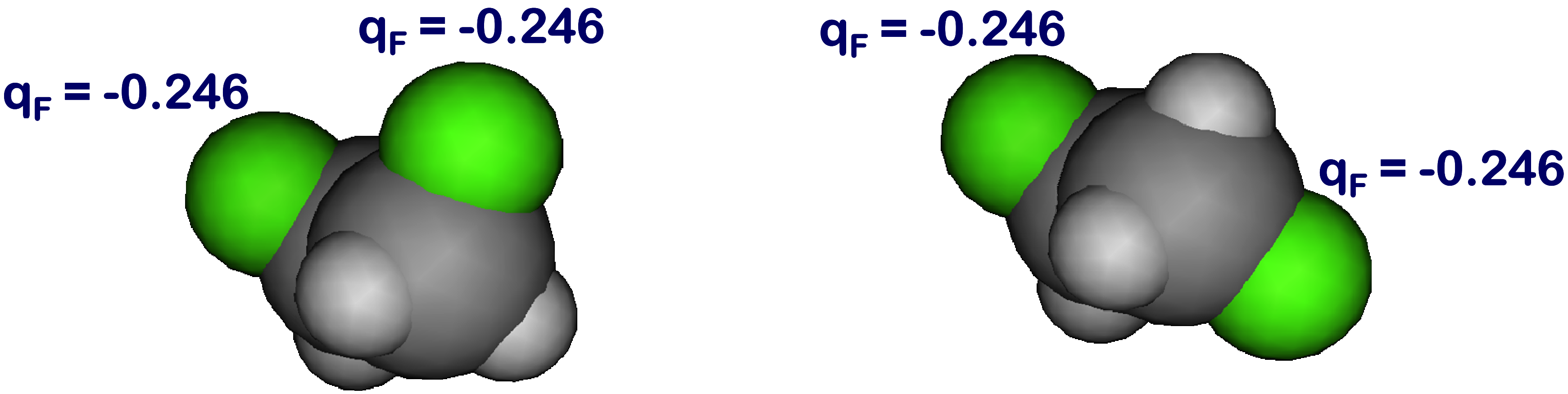


The calculations of partial atomic charges are very easy to compute for small and big molecular structures!!!

But also in this case there is the other side of the medal …

C-F ($C_2H_4F_2$)

$\Delta\chi_{C-F} = -1.6$

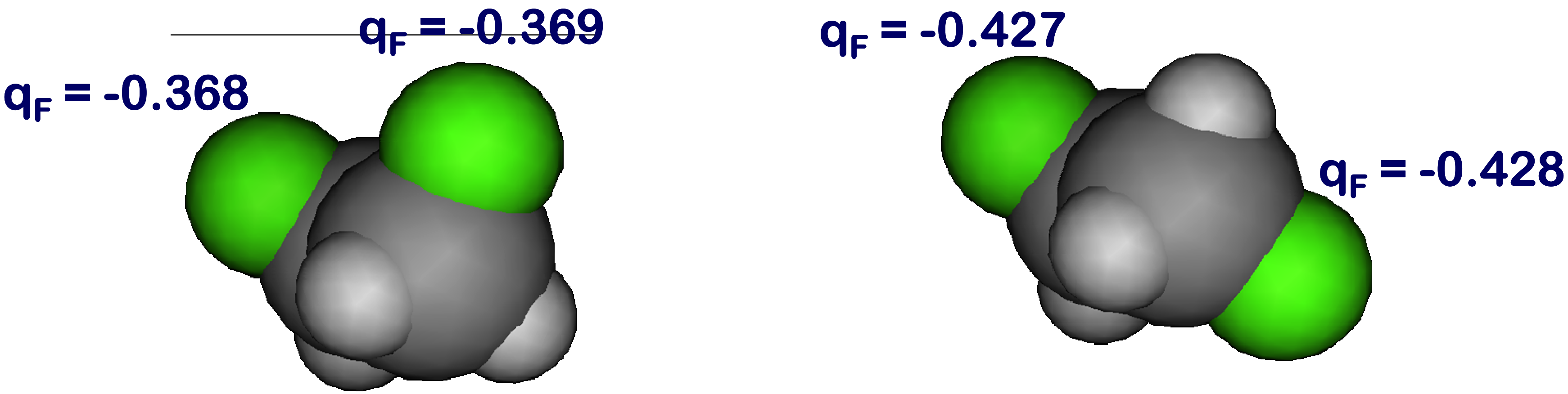


As you suspected, PEOE-driven partial atomic charges are conformation independent… this is a pity!!!

Look what's happened using type II or III…

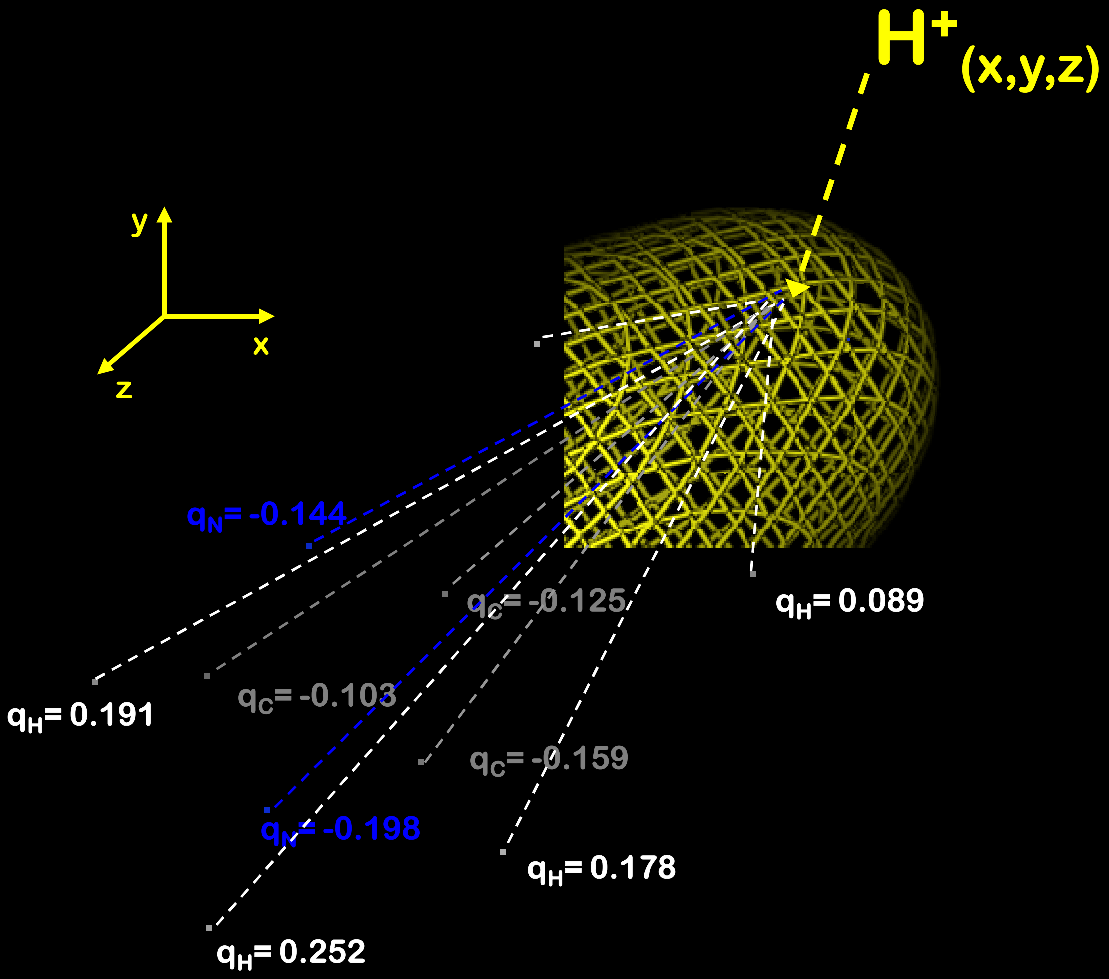
$C\text{-}F\ (C_2H_4F_2)$

$\Delta\chi_{C\text{-}F} = -1.6$

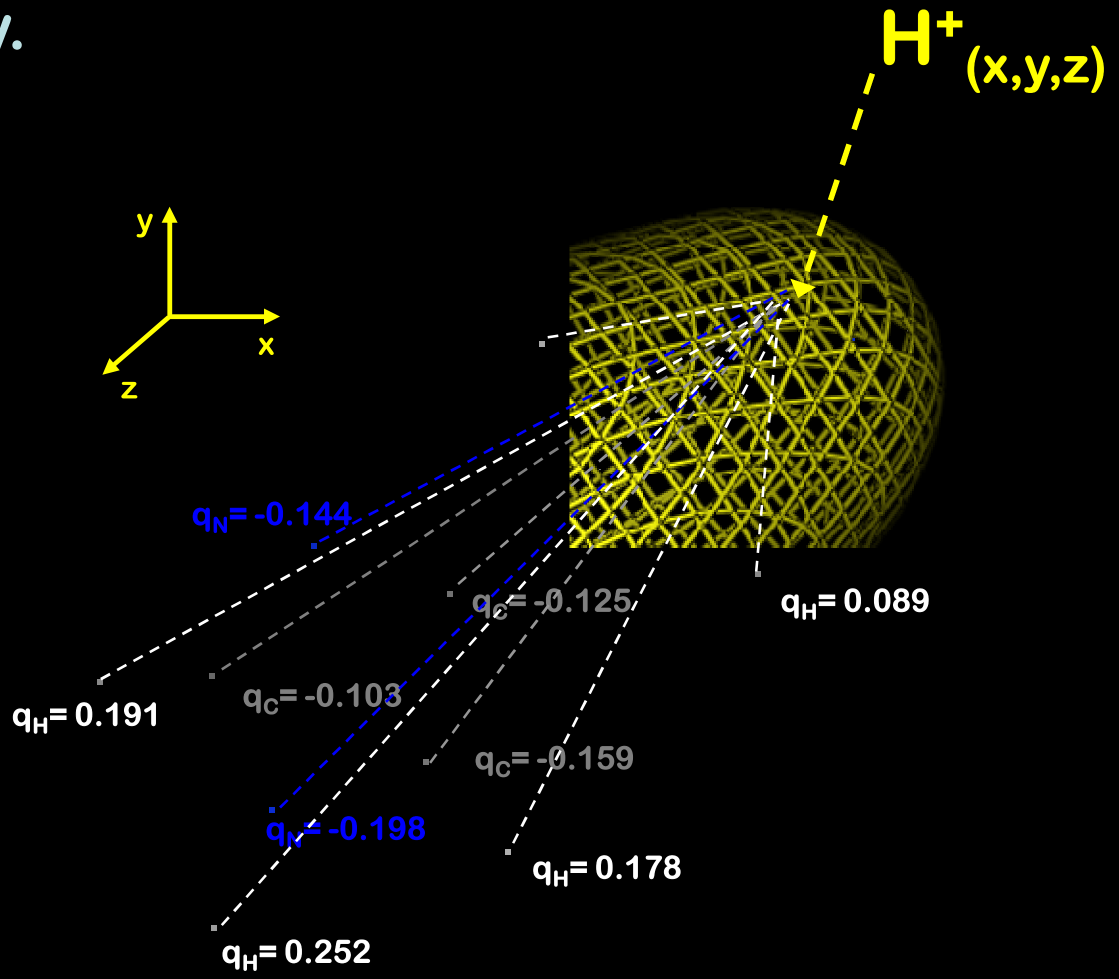


EPS charges (RHF/AM1)

Back to our dear friend, Charles…



The field produced by multiple point charges : the electric field of two or more point charges is equal to the sum vector of the electric fields of each of these taken separately.



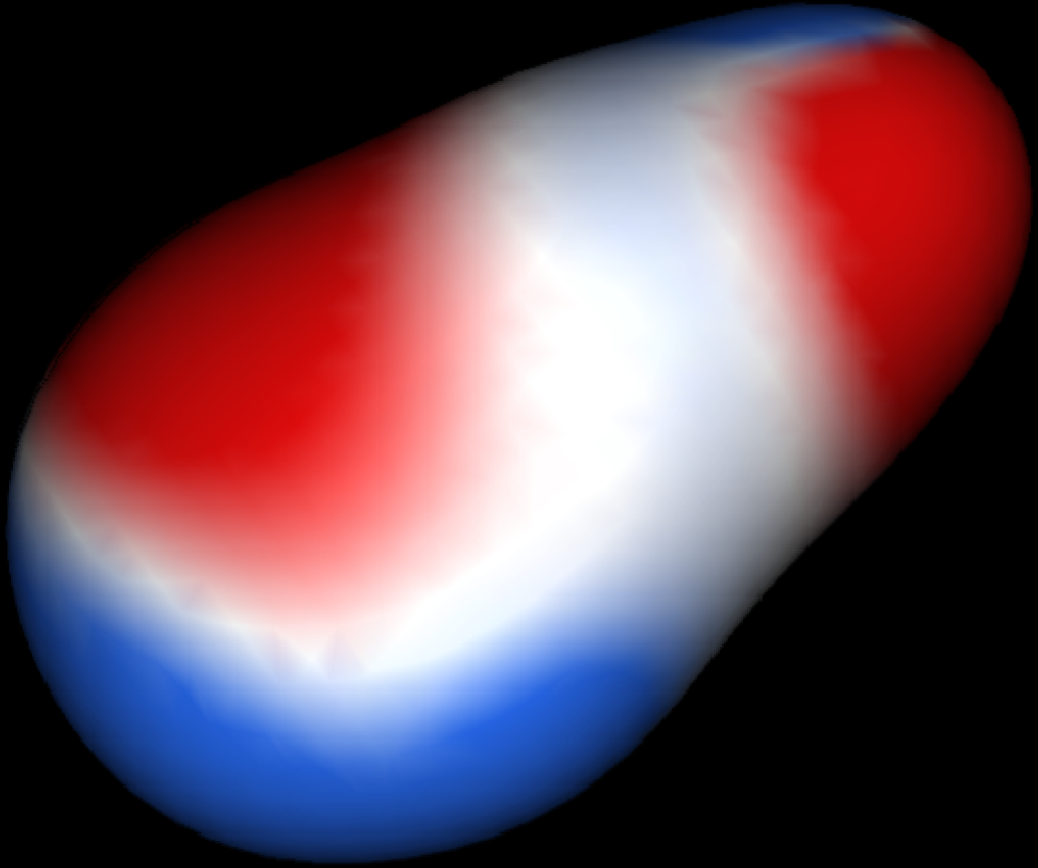
do you remember?

The work to move a particle from A to B is, in the case of conservative forces, equal to the potential energy change sign changed.

$$V_{el} = -\int_A^B F_{el} \times \mathrm{ds} = -\int_A^B \frac{1}{4\pi\varepsilon_0} \frac{q}{r^2} \mathrm{d}r =$$

$$= -\frac{q}{4\pi\varepsilon_0} \int_A^B \frac{1}{r^2} \mathrm{d}r = -\frac{q}{4\pi\varepsilon_0} \left[-\frac{1}{r} \right]_A^B = \frac{q}{4\pi\varepsilon_0 r}$$

Charming!

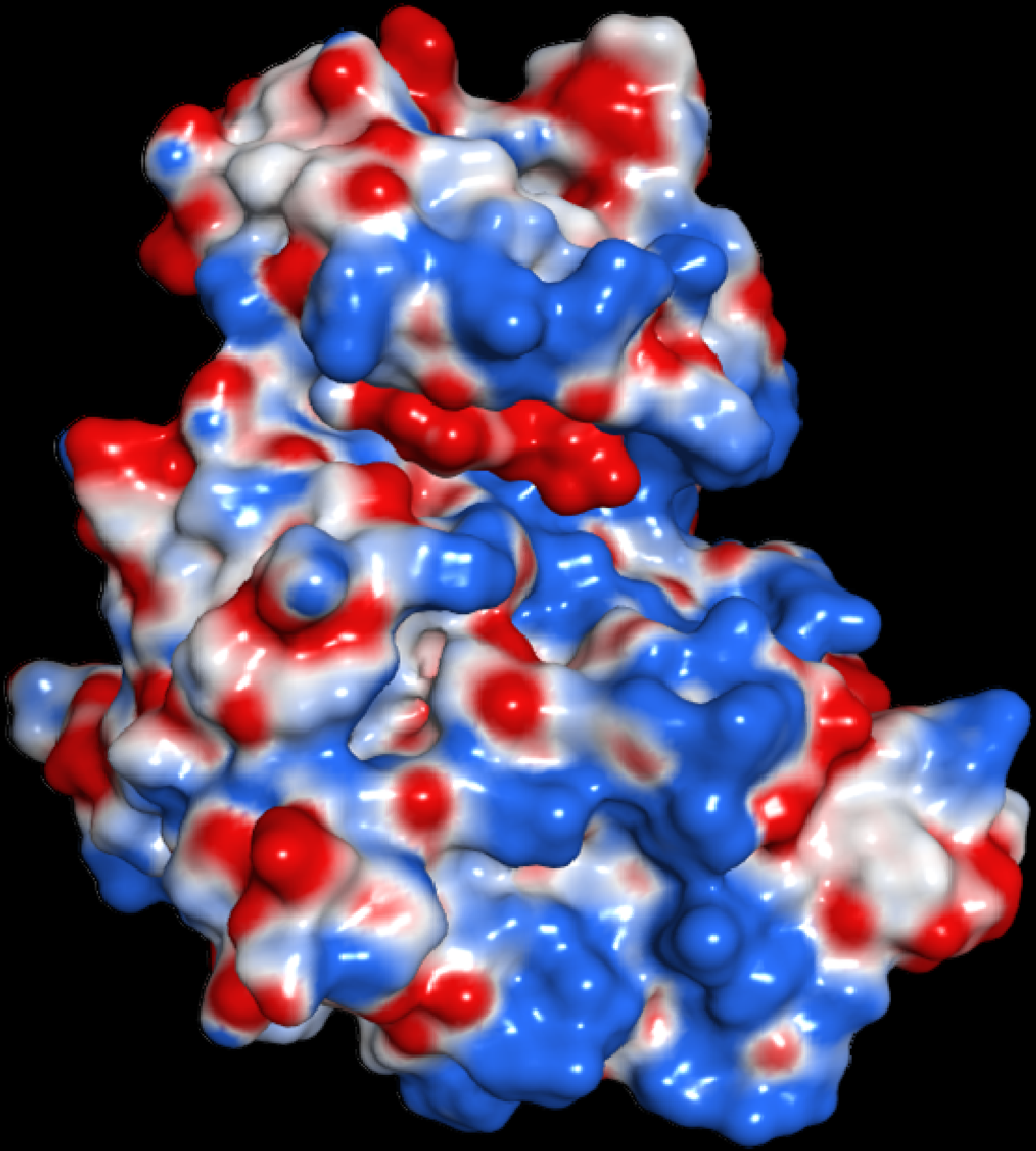


🟥 $V_{el} < 0$

⬜ $V_{el} = 0$

🟦 $V_{el} > 0$

… very charming!

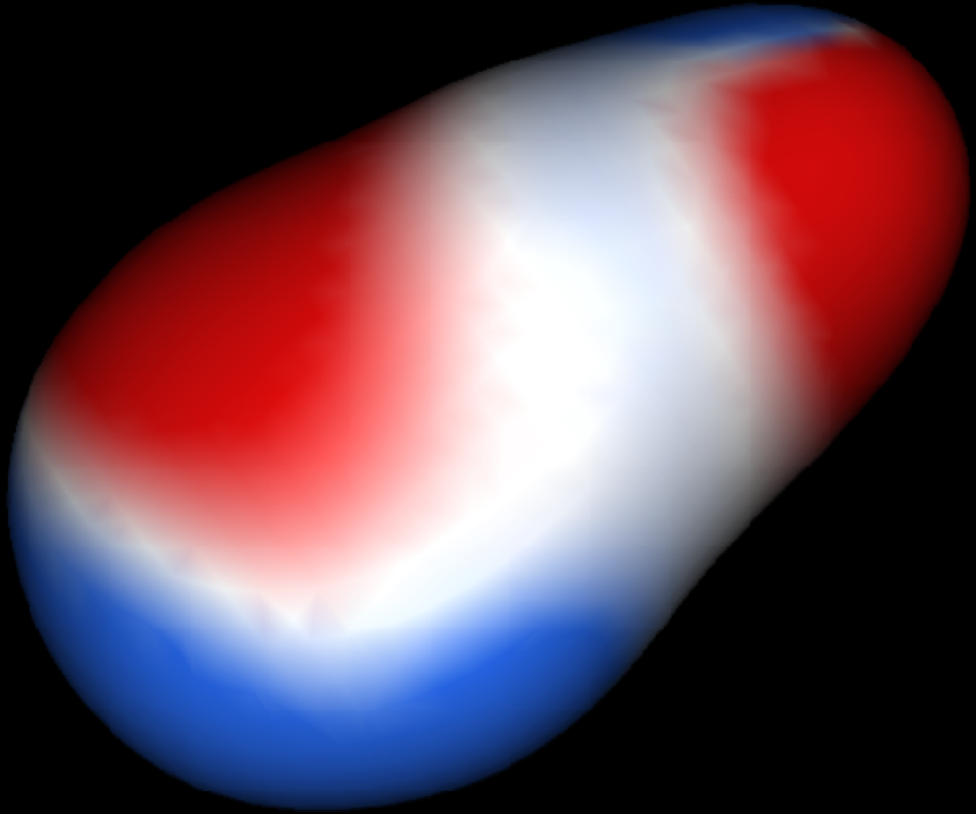


🟥 $V_{el} < 0$

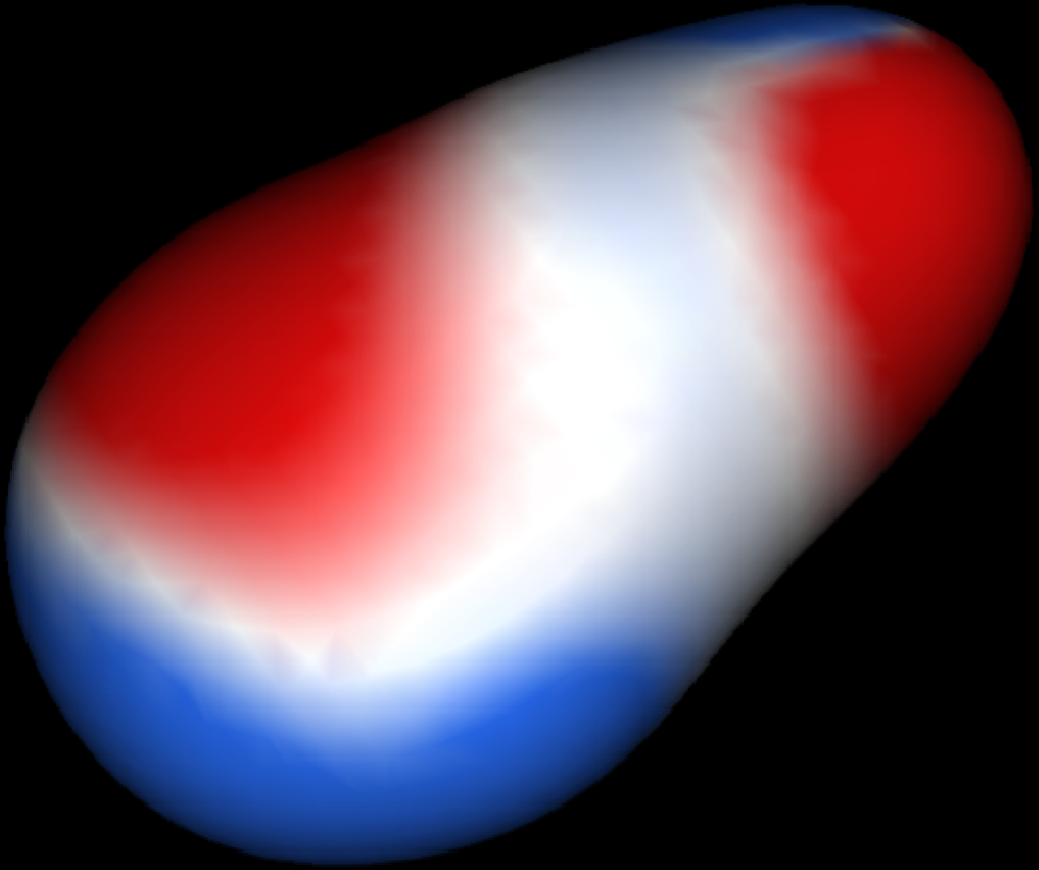
⬜ $V_{el} = 0$

🟦 $V_{el} > 0$

Do you note the similarity?



… and do you appreciate the meaning of topological and electrostatic complementarities?

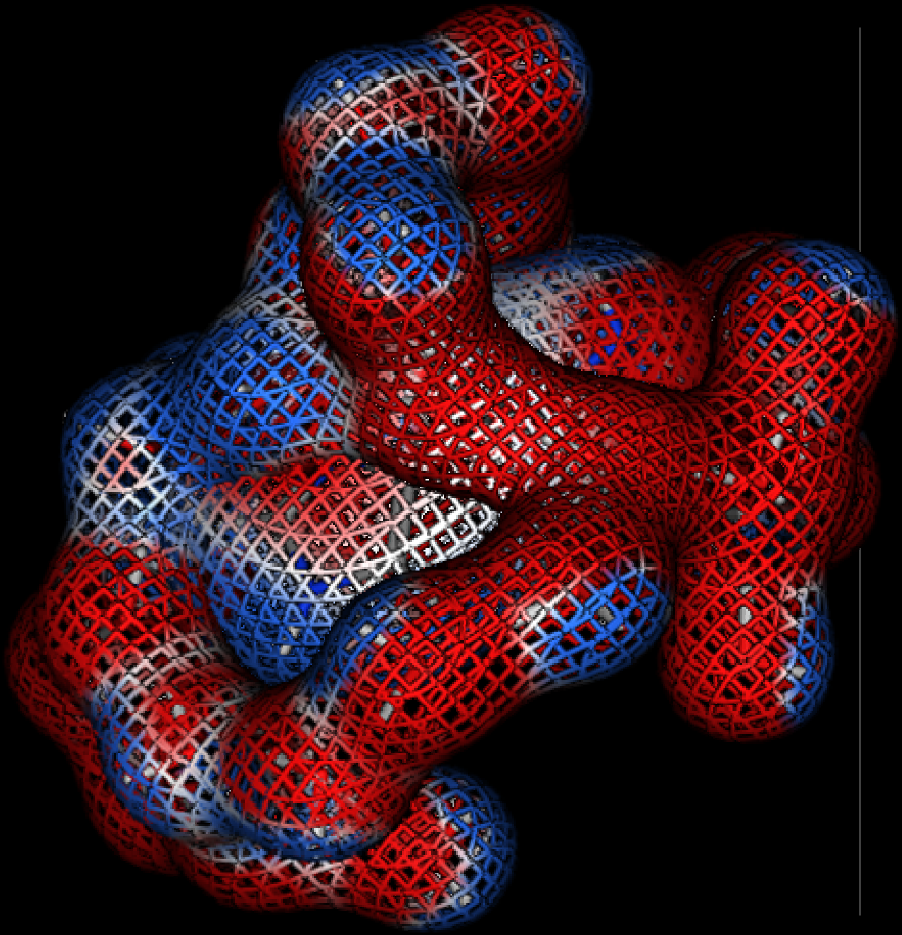


🟥 $V_{el} < 0$

⬜ $V_{el} = 0$

🟦 $V_{el} > 0$

from the simple concept of “solvation”…

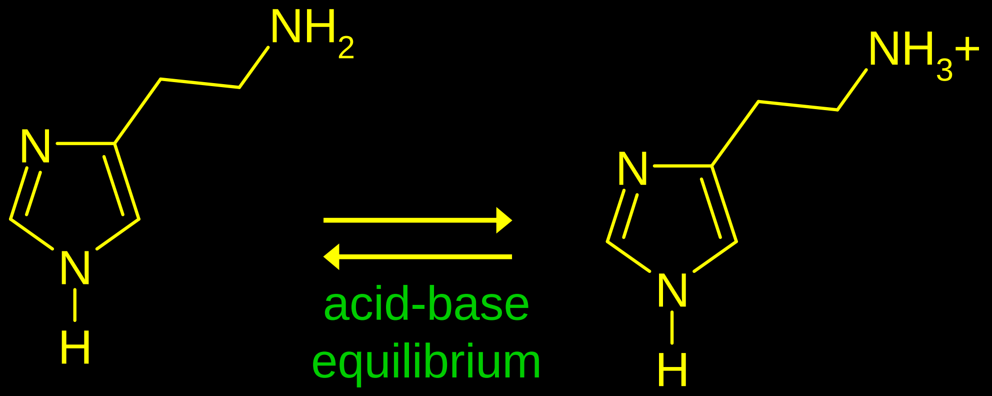


🟥 $V_{el} < 0$

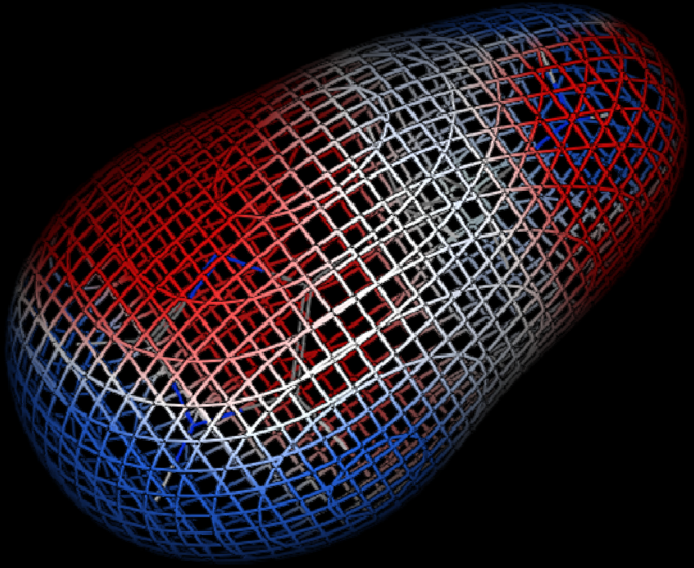
⬜ $V_{el} = 0$

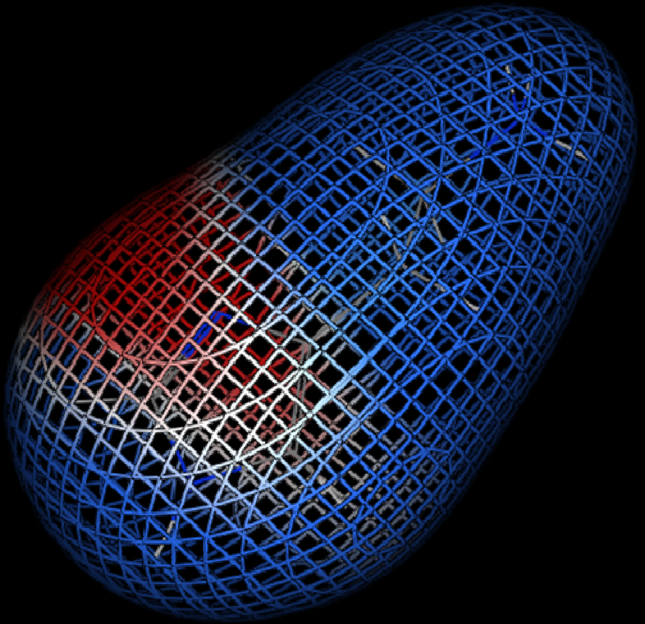
🟦 $V_{el} > 0$

...up to the most intriguing from a chemical point of view!

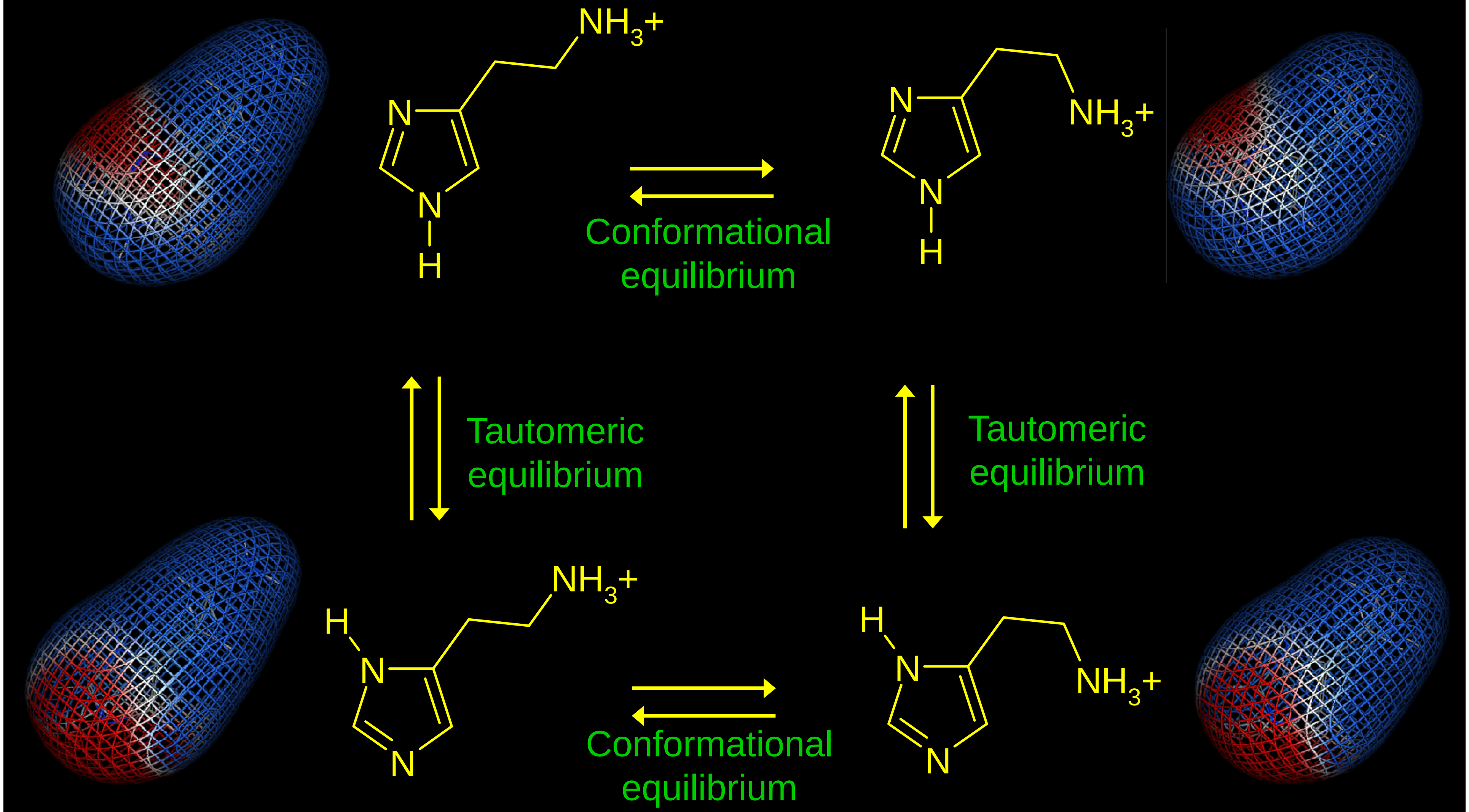


acid-base equilibrium





and these…



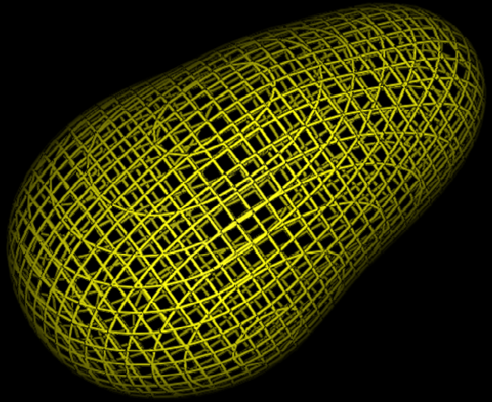
Conformational equilibrium

Tautomeric equilibrium

Tautomeric equilibrium

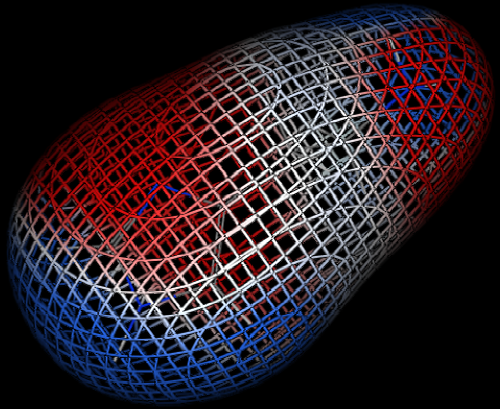
Conformational equilibrium

Back to this example: new molecular descriptors:



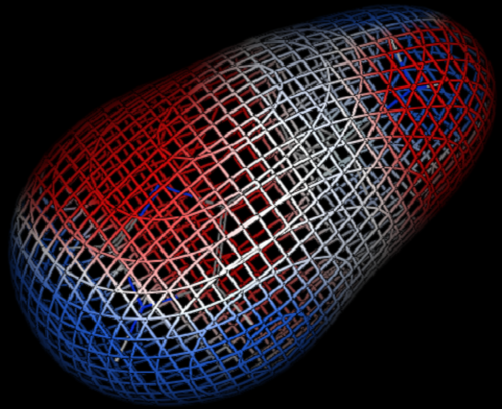
SES (Å^{2})

$\cong \sum$ ▲



SES+ (Å^{2})

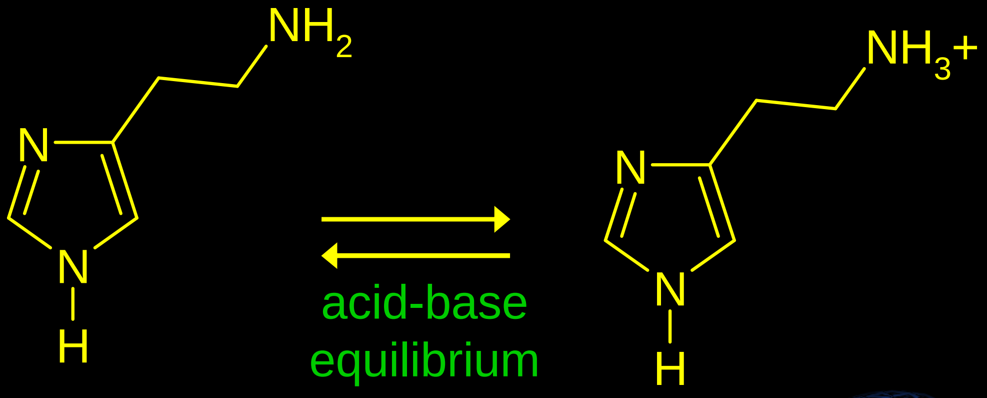
$\cong \sum$ ▲



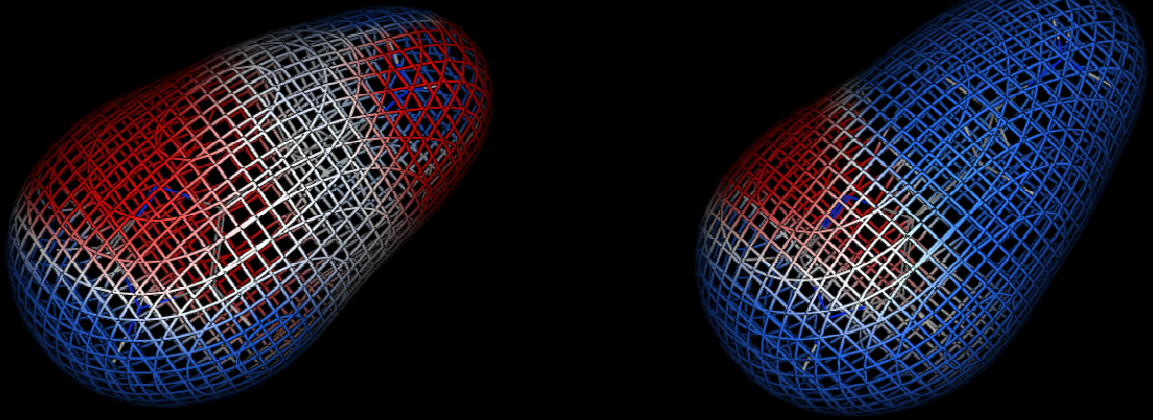
SES- (Å^{2})

$\cong \sum$ ▲

Back to this example: new molecular descriptors:

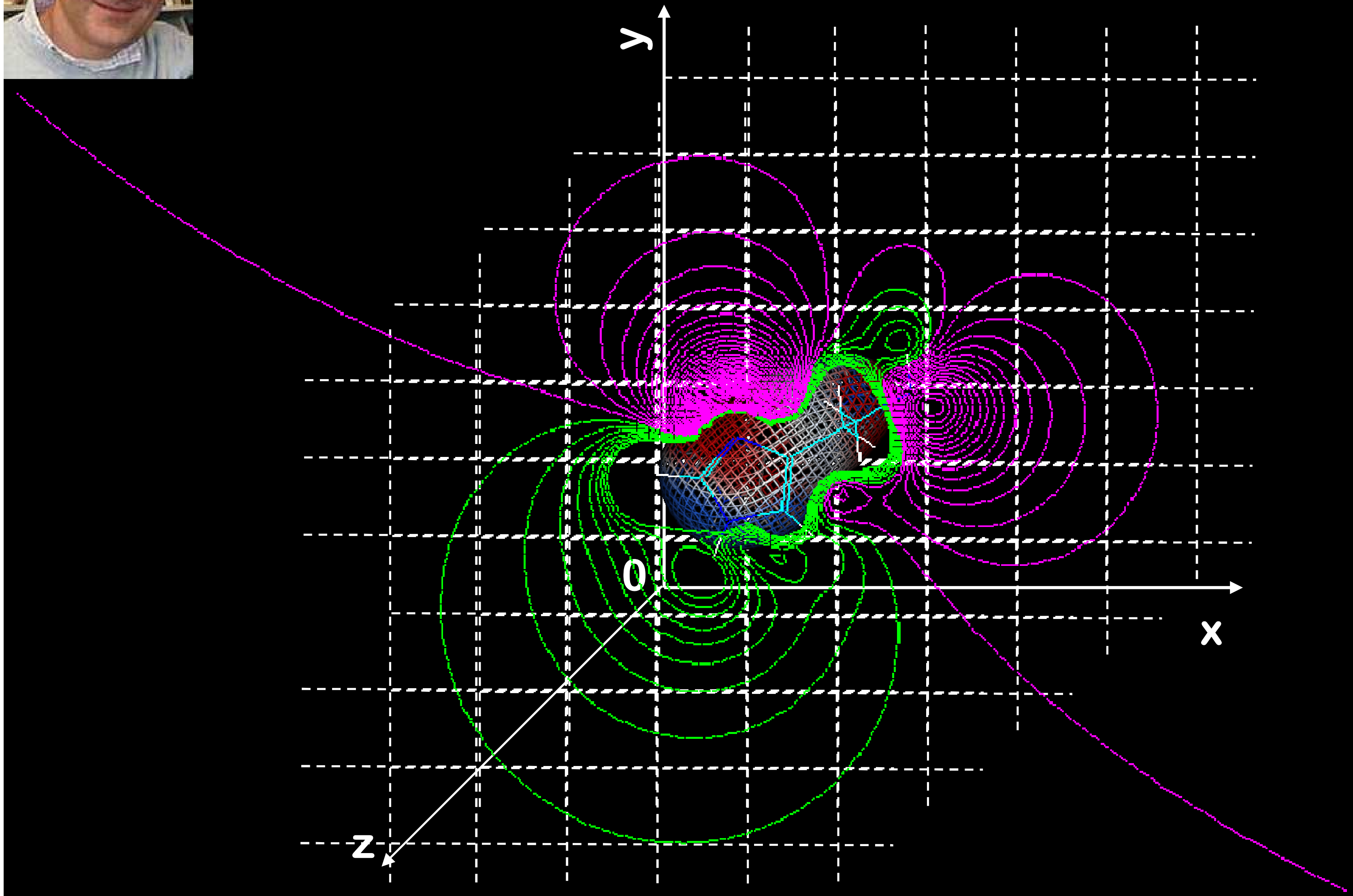


acid-base equilibrium



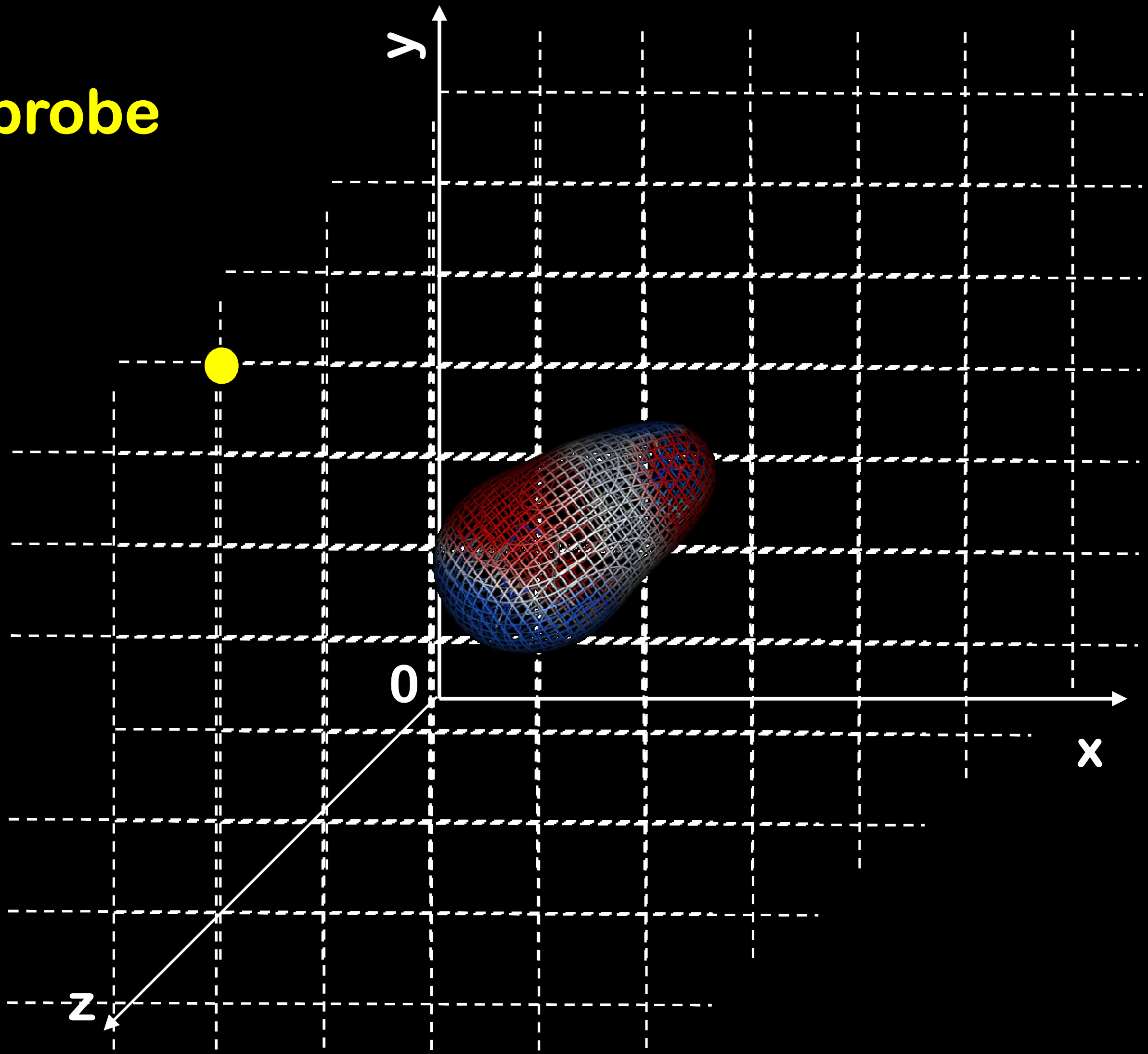
Volume (Å^3)	117.0	119.9
ASA (Å^2)	304.2	306.6
ASA+ (Å^2)	204.5	263.8
ASA- (Å^2)	99.7	42.8

Beyond the molecular surface:



Molecular Fields: a possible molecular language?

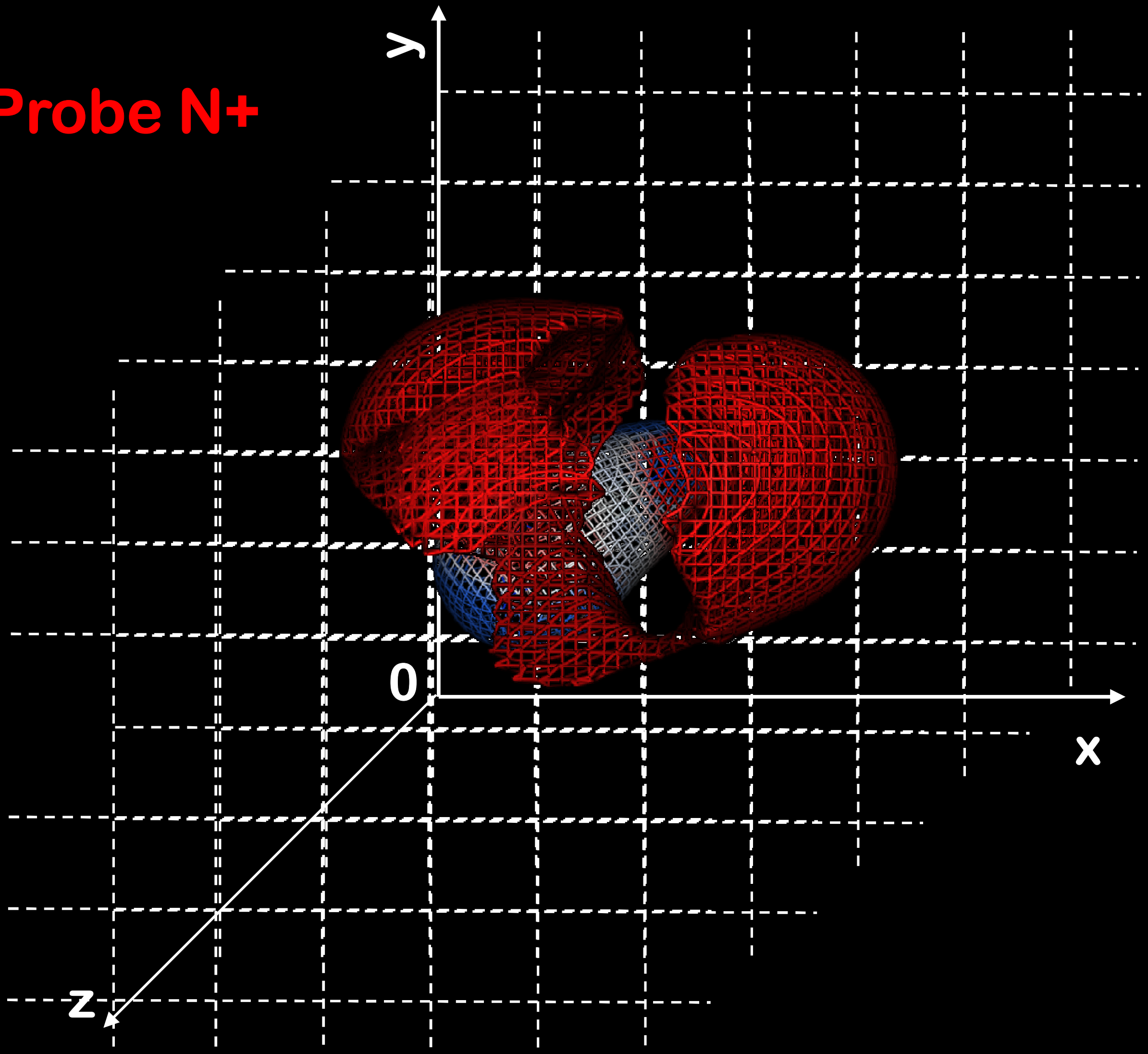
Molecular probe



Molecular Fields: a possible molecular language?

Molecular Probe N+

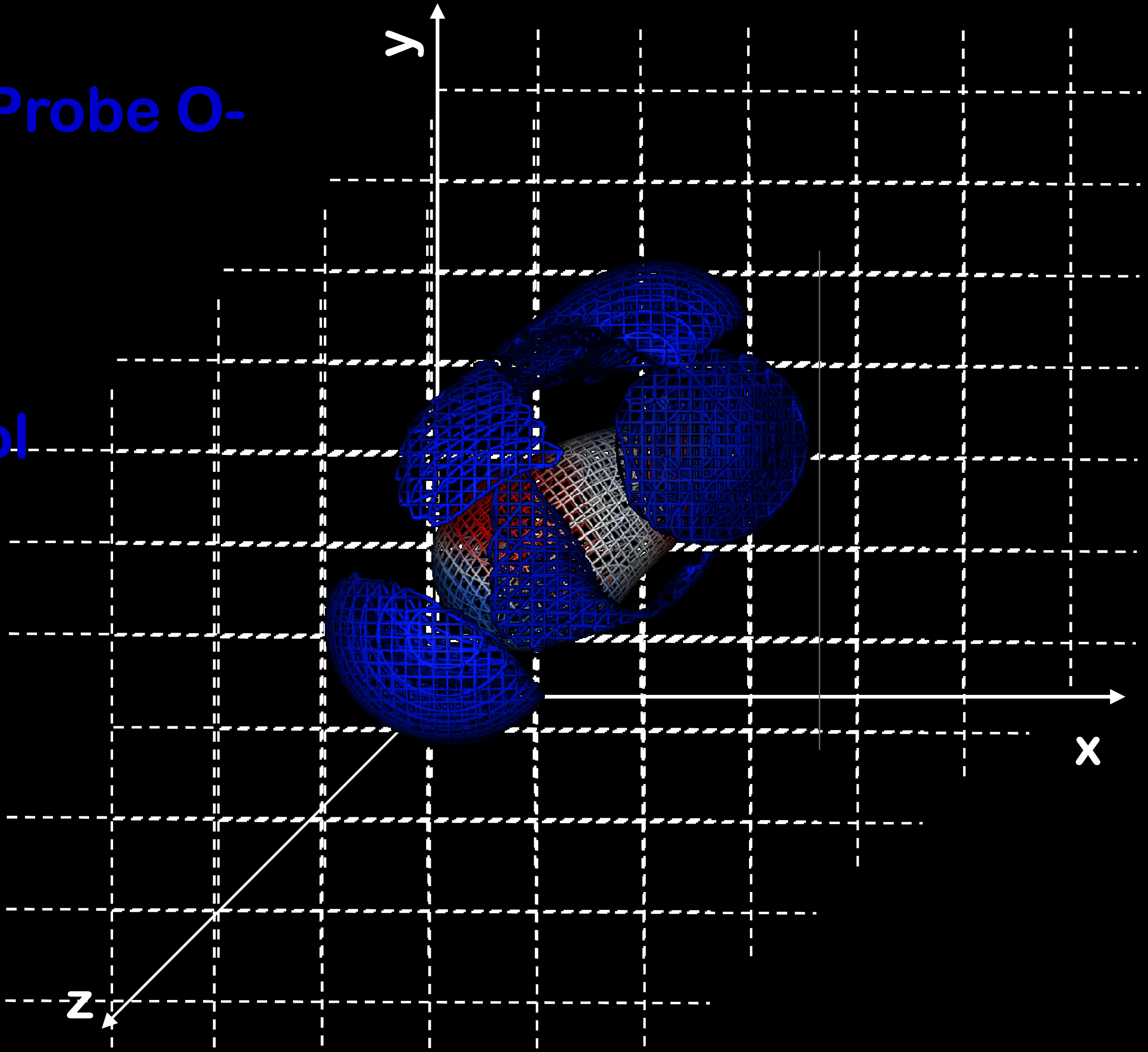
3 kcal/mol
2 kcal/mol
1 kcal/mol



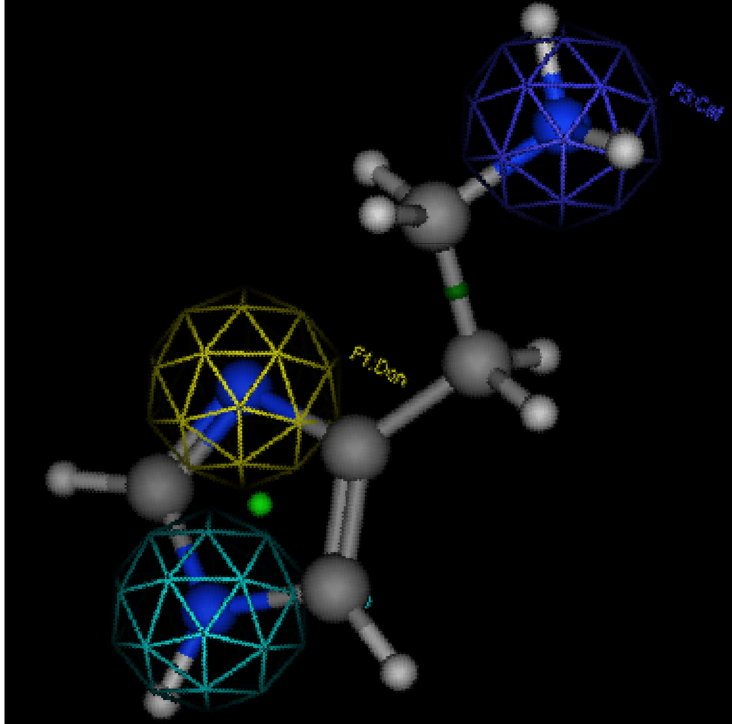
Molecular Fields: a possible molecular language?

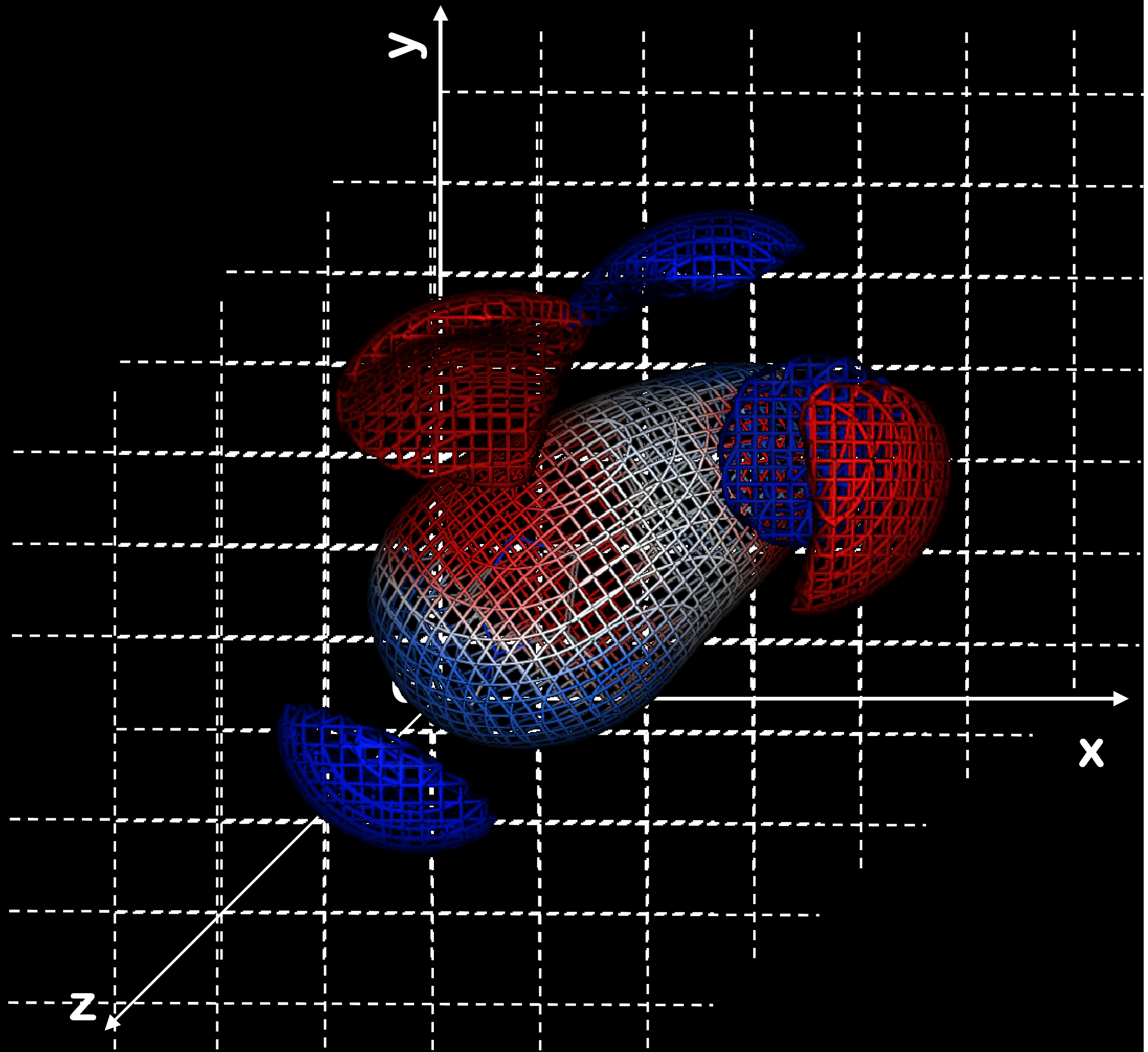
Molecular Probe O-

3 kcal/mol
2 kcal/mol
1.5 kcal/mol

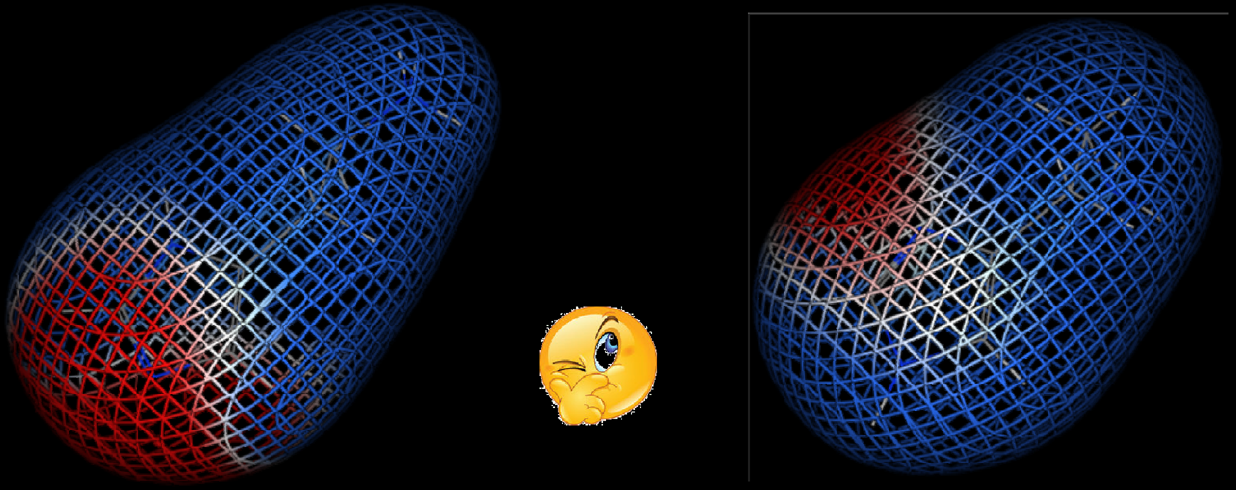


Molecular Fields: a new age view of pharmacophoric hypothesis ?



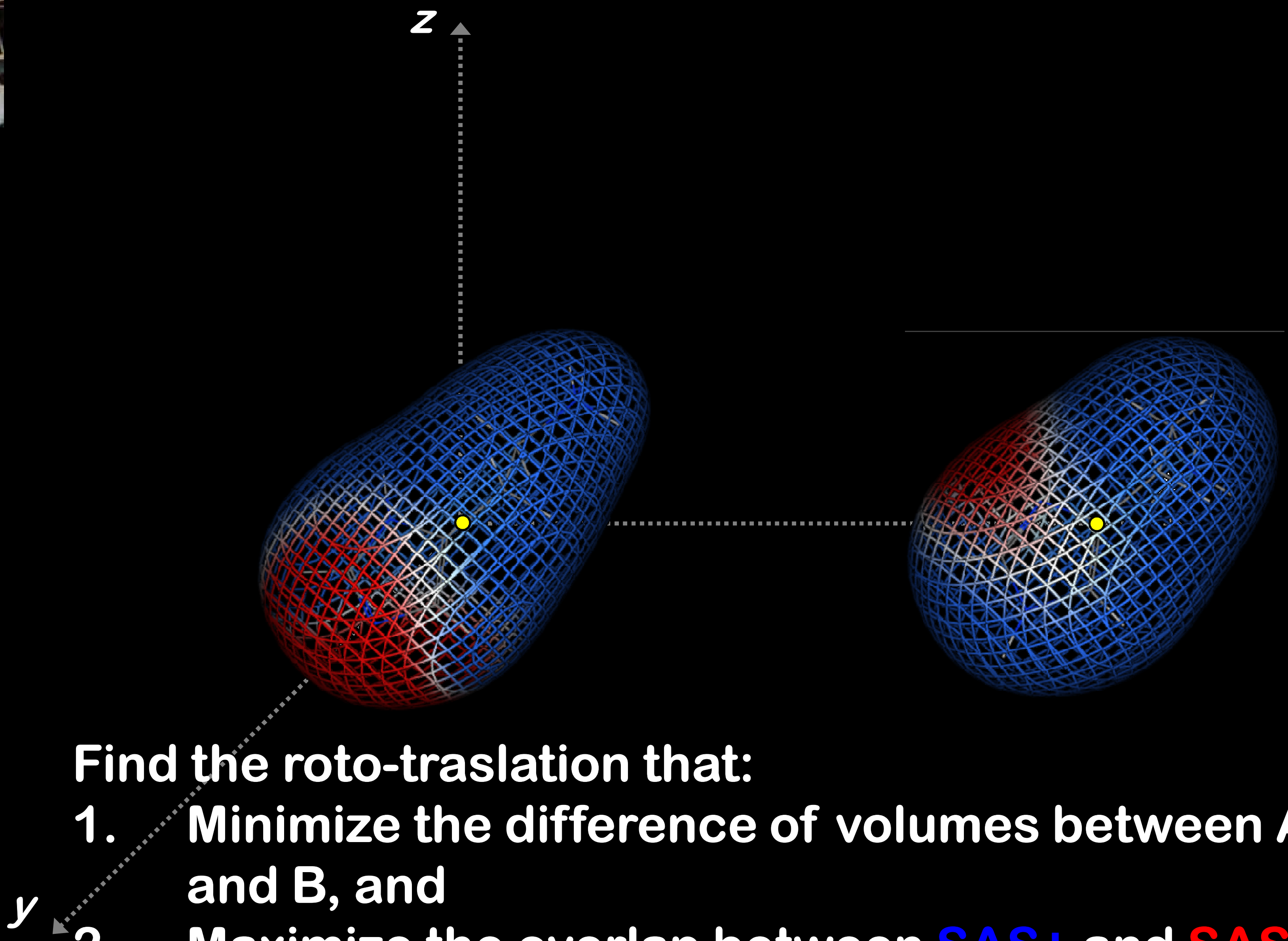


Before moving on, I need a good idea… how can we evaluate/measure the *similarity* between these two objects?



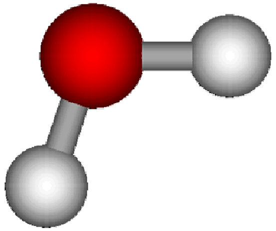
Try to answer: ***when are they identical?***

How you can do it?

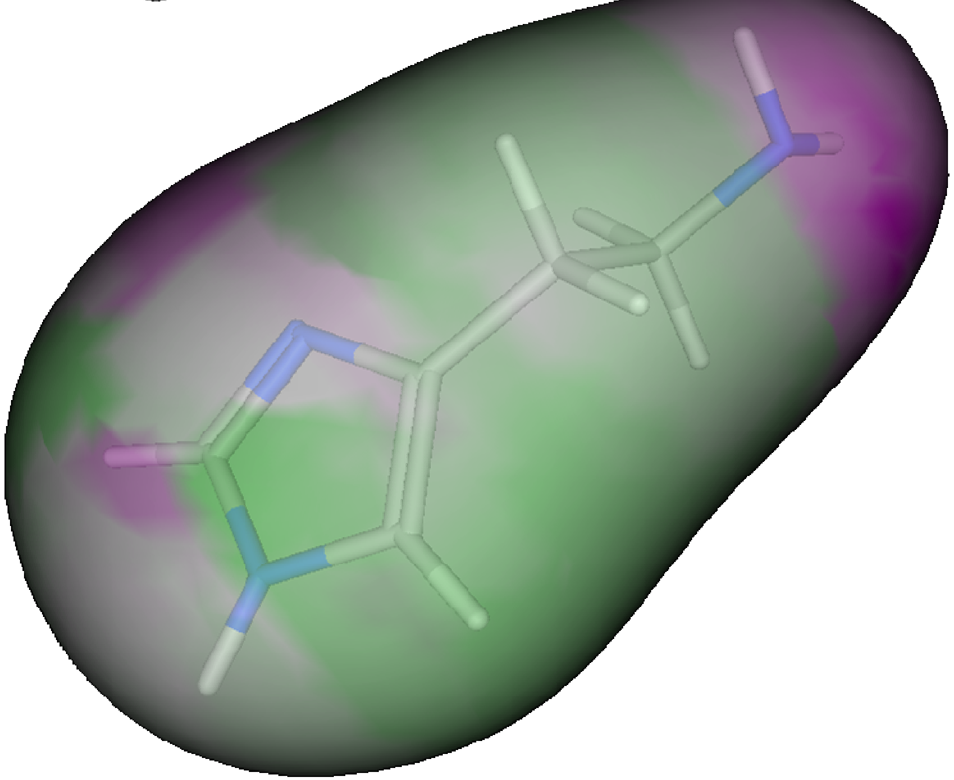


Find the roto-traslation that:

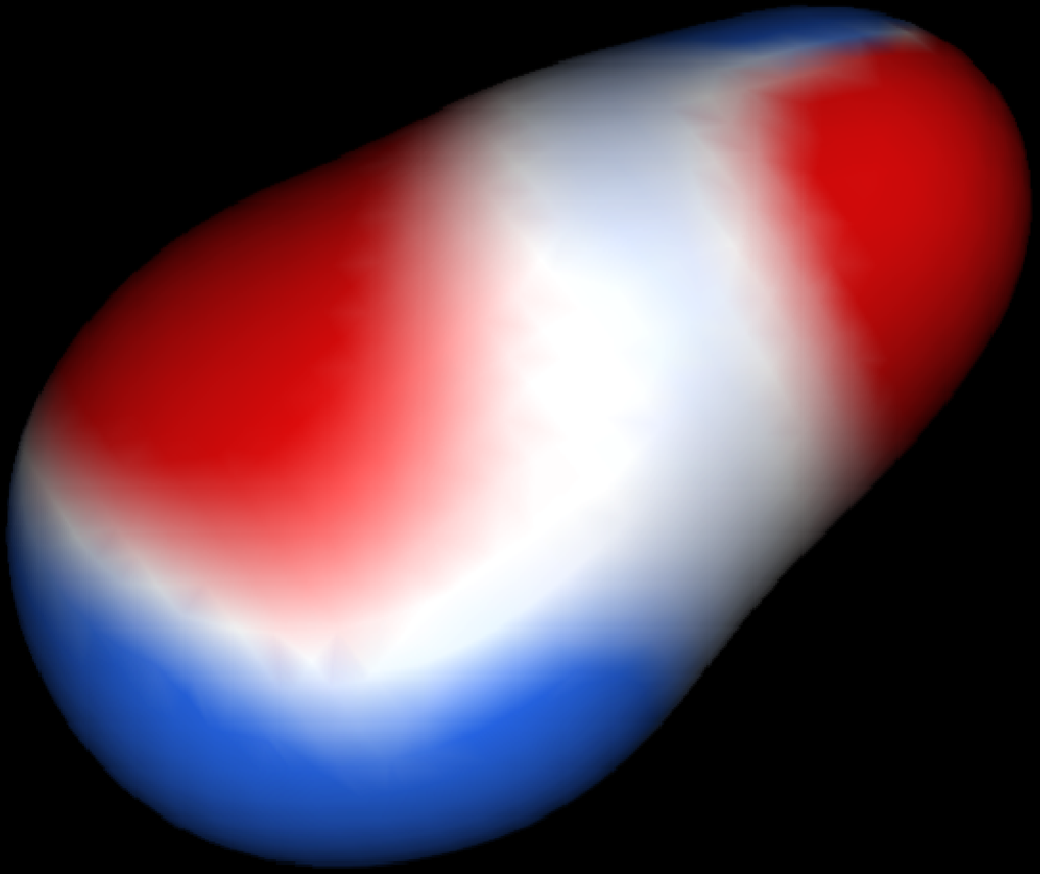
1. Minimize the difference of volumes between A and B, and
2. Maximize the overlap between SAS+ and SAS-



phobicity



What is a plausible meaning of these “white” regions ?

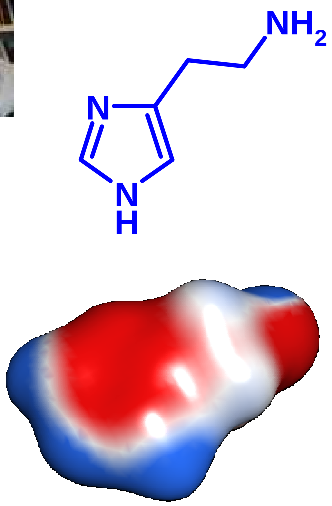


🟥 $V_{el} < 0$

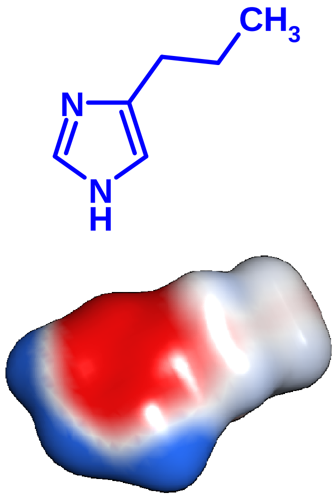
⬜ $V_{el} = 0$

🟦 $V_{el} > 0$

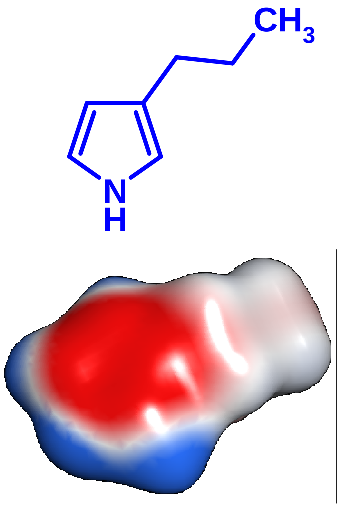
and this is understandable!



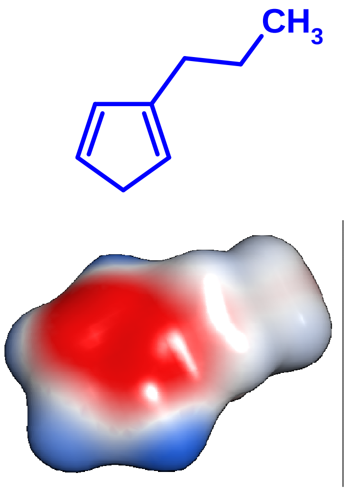
log P = -0.26



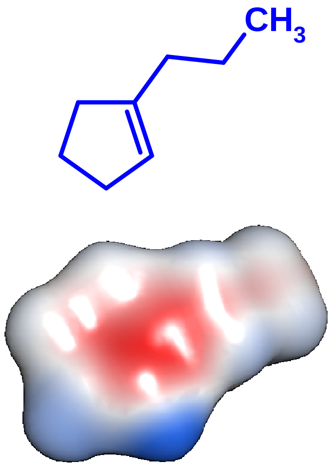
log P = 0.90



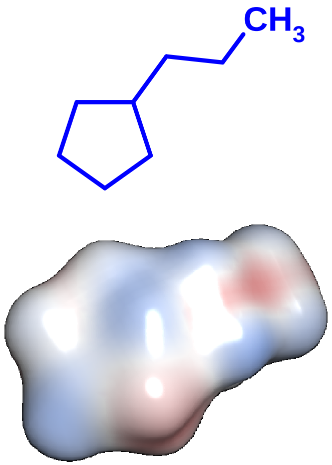
log P = 2.17



log P = 2.94

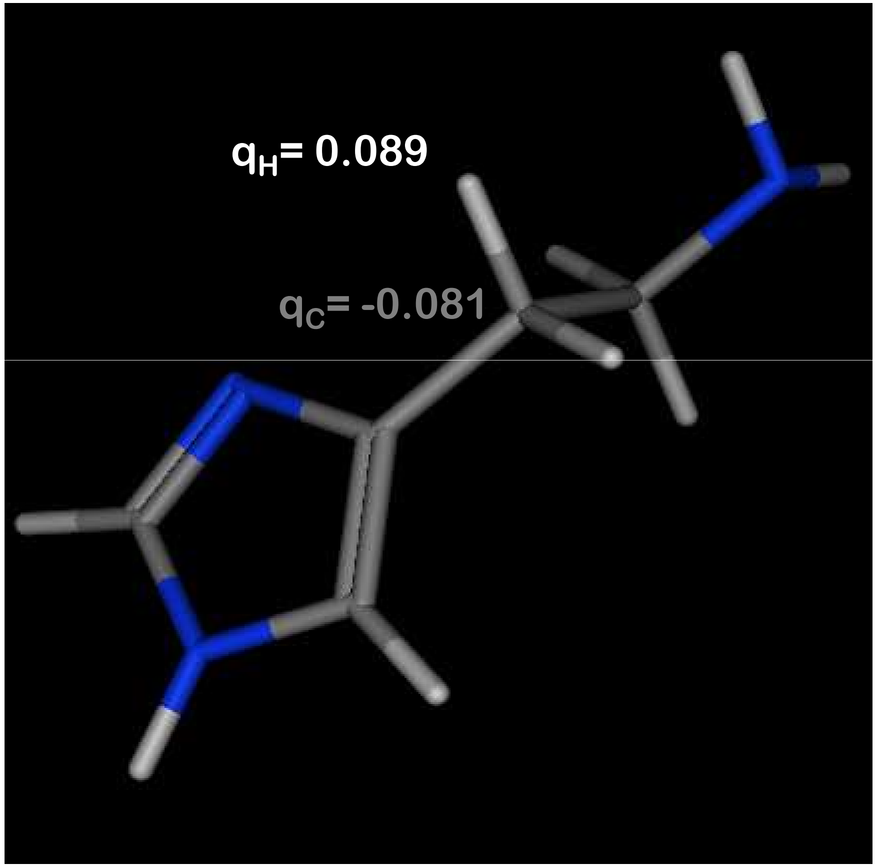


log P = 3.31



log P = 3.84

In first approximation we can make this simple assumption:

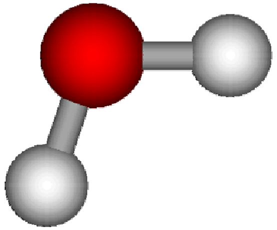


Regions of the molecular surface in proximity to bonds with low polarity (ΔEN $\cong$ 0) can mediate only weak interactions with our charged probe!

Back to our “white” potential surface:

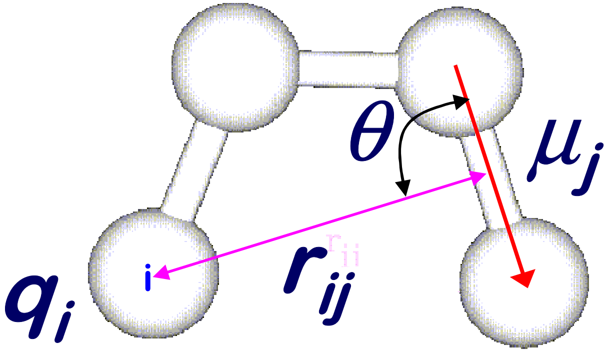
In organic chemistry, we have a family of compounds characterized by bonds with ΔEN ≅ 0

HYDROCARBONS

and they are surely  phobics!!!

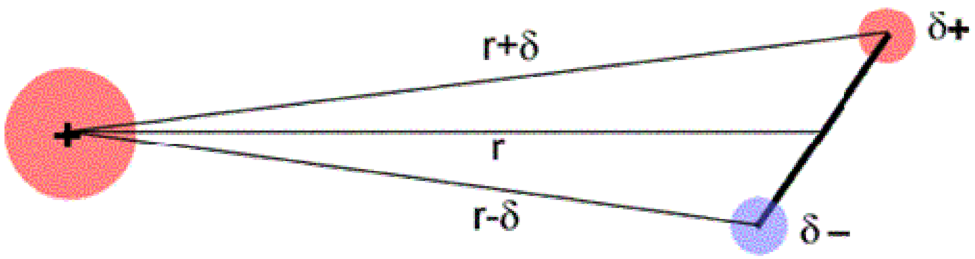
Back for a moment to the real world: beyond to electrostatic interaction…

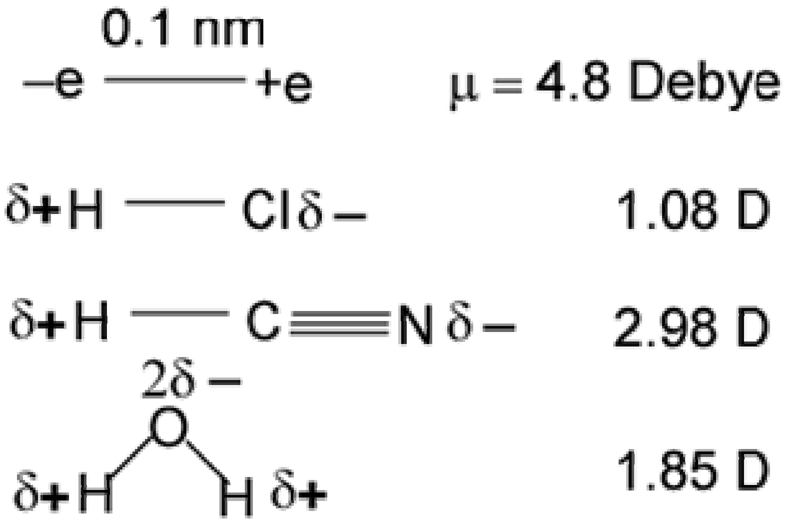
Charge-dipole interactions:



$$V(r,\theta) = k\frac{q_i\mu_j}{r_{ij}^{\ 2}}\cdot\cos\theta$$

Charge-dipole interactions:

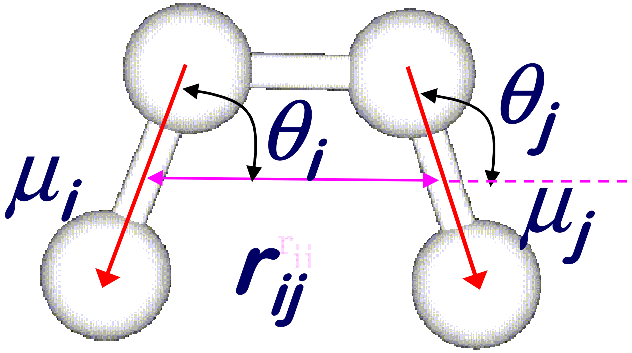


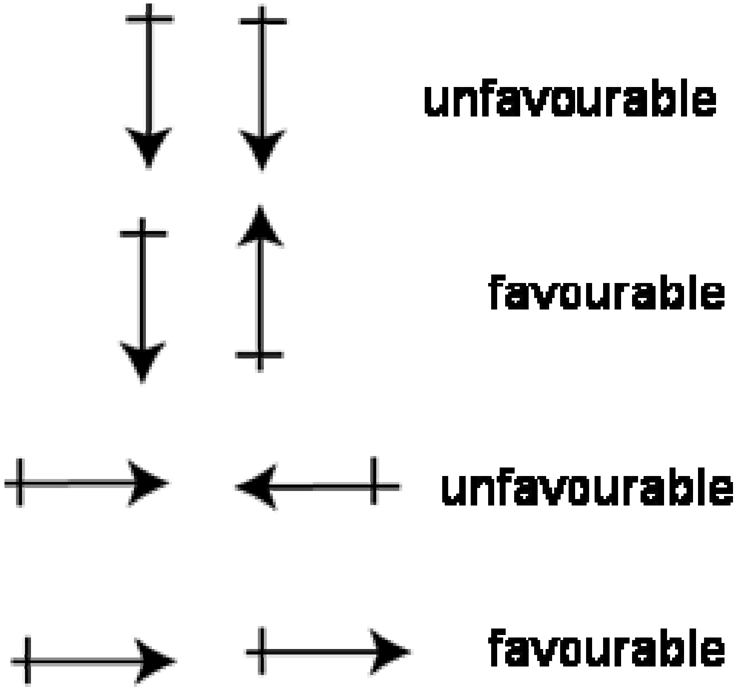


Formally, we can decompose the charge-dipole interaction in two charge-charge interactions…

And even worst:

dipole-dipole interactions





unfavourable

favourable

unfavourable

favourable

$$V(r,\theta) = k\frac{\mu_i \mu_j}{{r_{ij}}^3} \cdot \left(2\cos\theta_i \cos\theta_j - \sin\theta_i \sin\theta_j\right)$$

And even even worst:

ion-dipole and dipole-dipole interactions are temperature dependent!!!

charge-dipole interactions:

$$V(r) = -\frac{1}{3}k^2 \frac{q_i{}^2 \mu_j{}^2}{k_B T r_{ij}{}^4}$$

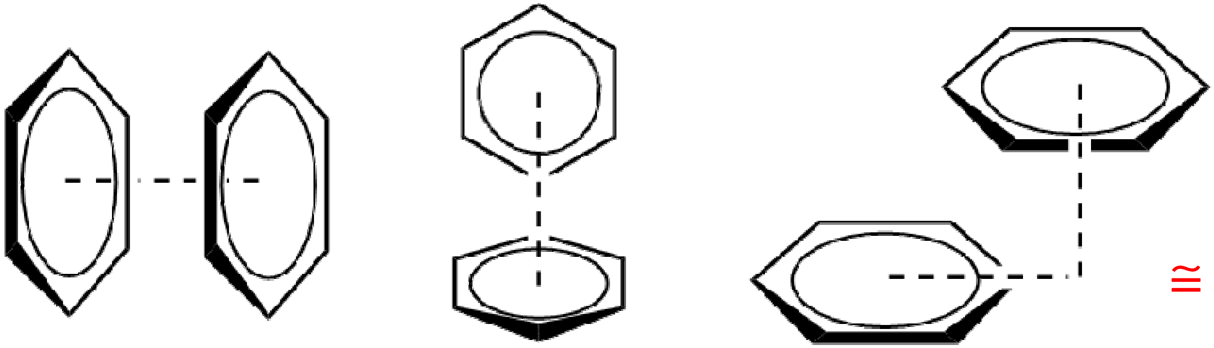
dipole-dipole interactions:

$$V(r) = -\frac{2}{3}k^2 \frac{\mu_i{}^2 \mu_j{}^2}{k_B T r_{ij}{}^6}$$

remember:

no temperature, no time, …no dipoles interactions honest computation!!!

The mystery of π-π interactions:

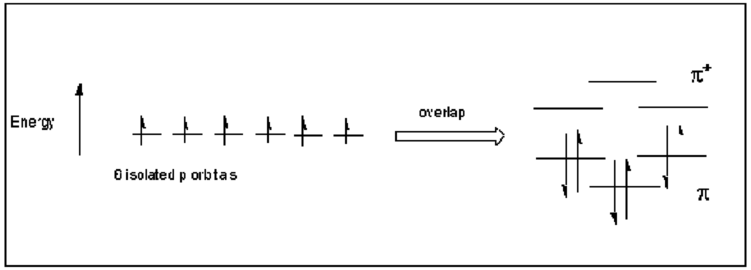


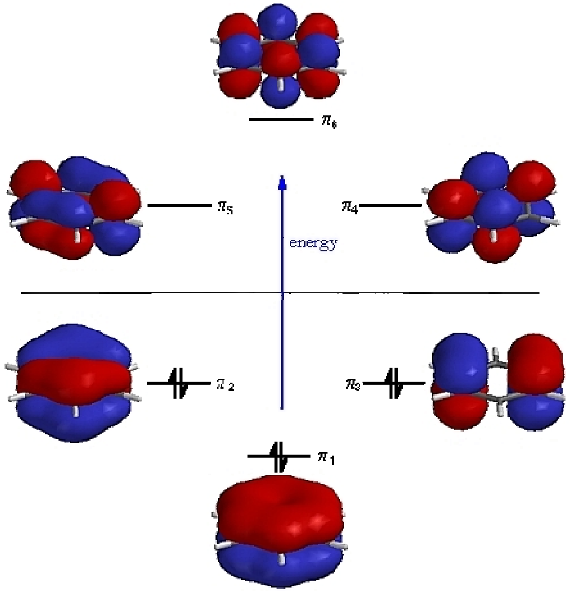
Sandwich T-shaped Parallel-dispaced

≅ 4.9 Å

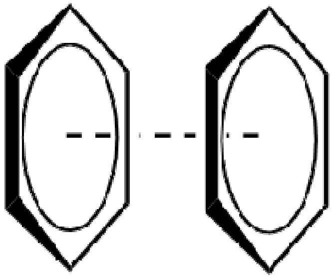
≅ -2.5 kcal/mol

≅ -1.5 kcal/mol



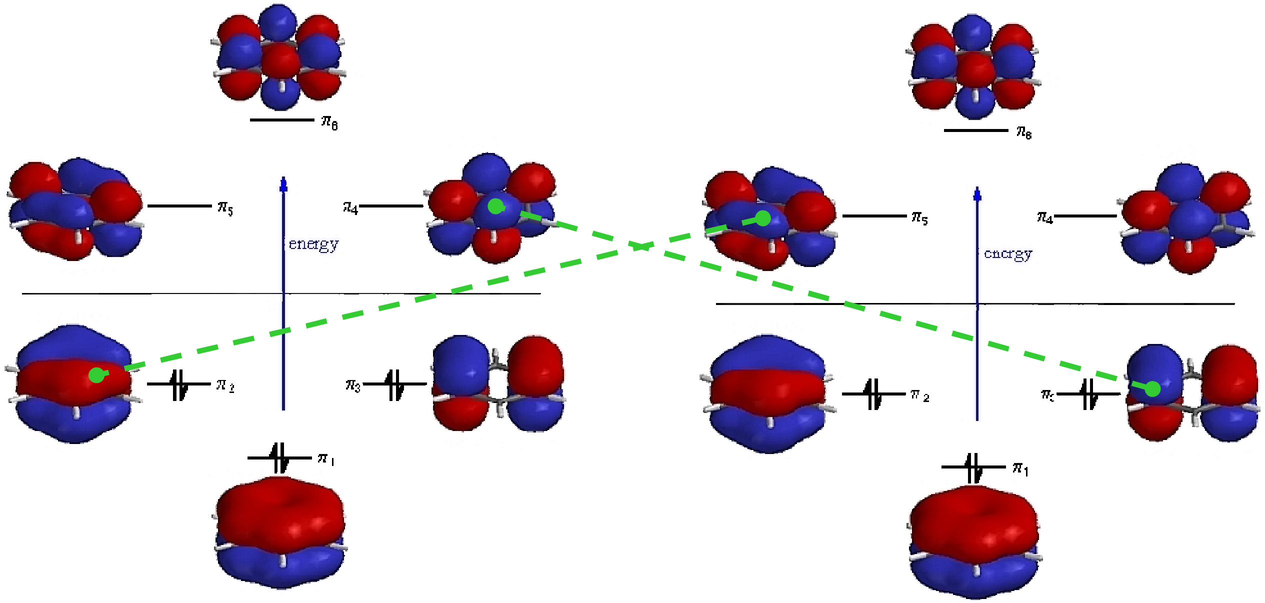


The mystery of π-π interactions:



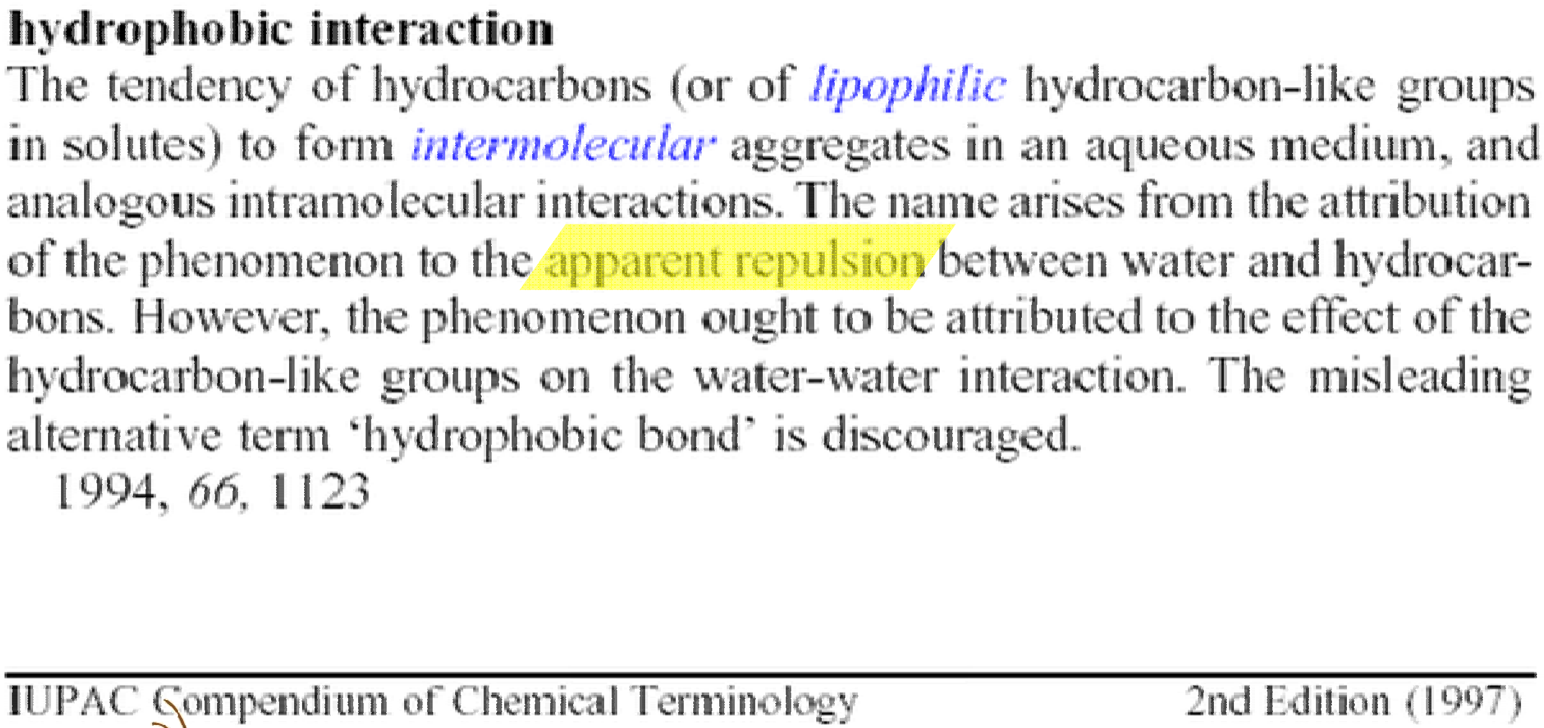
Sandwich

≅ -1.5 kcal/mol



Not only electrostatic moves the molecular world!!

… *in primis*, ask to IUPAC!



hydrophobic interaction

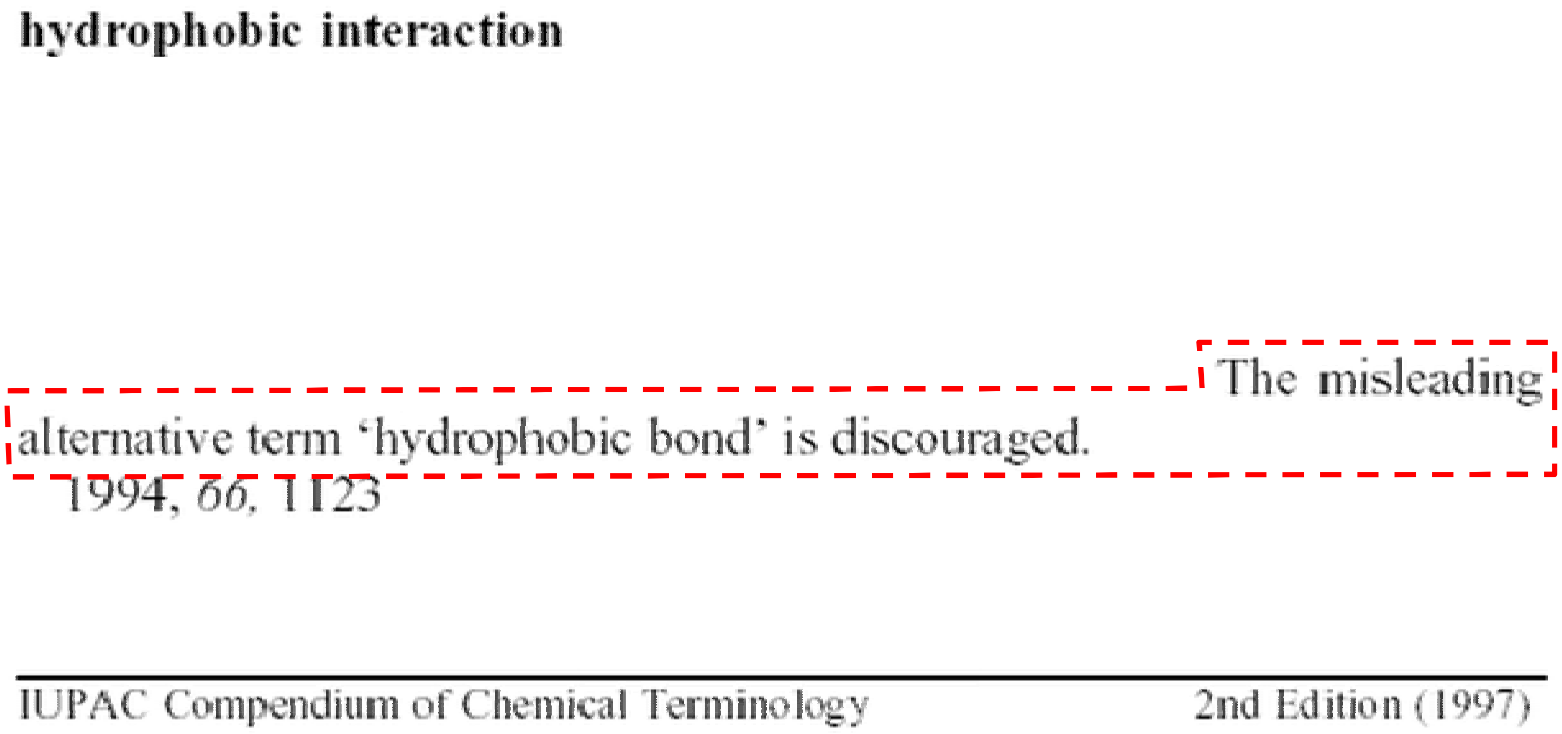
The tendency of hydrocarbons (or of *lipophilic* hydrocarbon-like groups in solutes) to form *intermolecular* aggregates in an aqueous medium, and analogous intramolecular interactions. The name arises from the attribution of the phenomenon to the apparent repulsion between water and hydrocarbons. However, the phenomenon ought to be attributed to the effect of the hydrocarbon-like groups on the water-water interaction. The misleading alternative term ‘hydrophobic bond’ is discouraged.

1994, *66*, 1123

IUPAC Compendium of Chemical Terminology — 2nd Edition (1997)

Yes, embarrassing!!!

.. this is the only sentence that we salve!

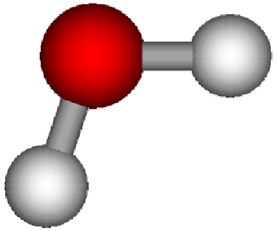


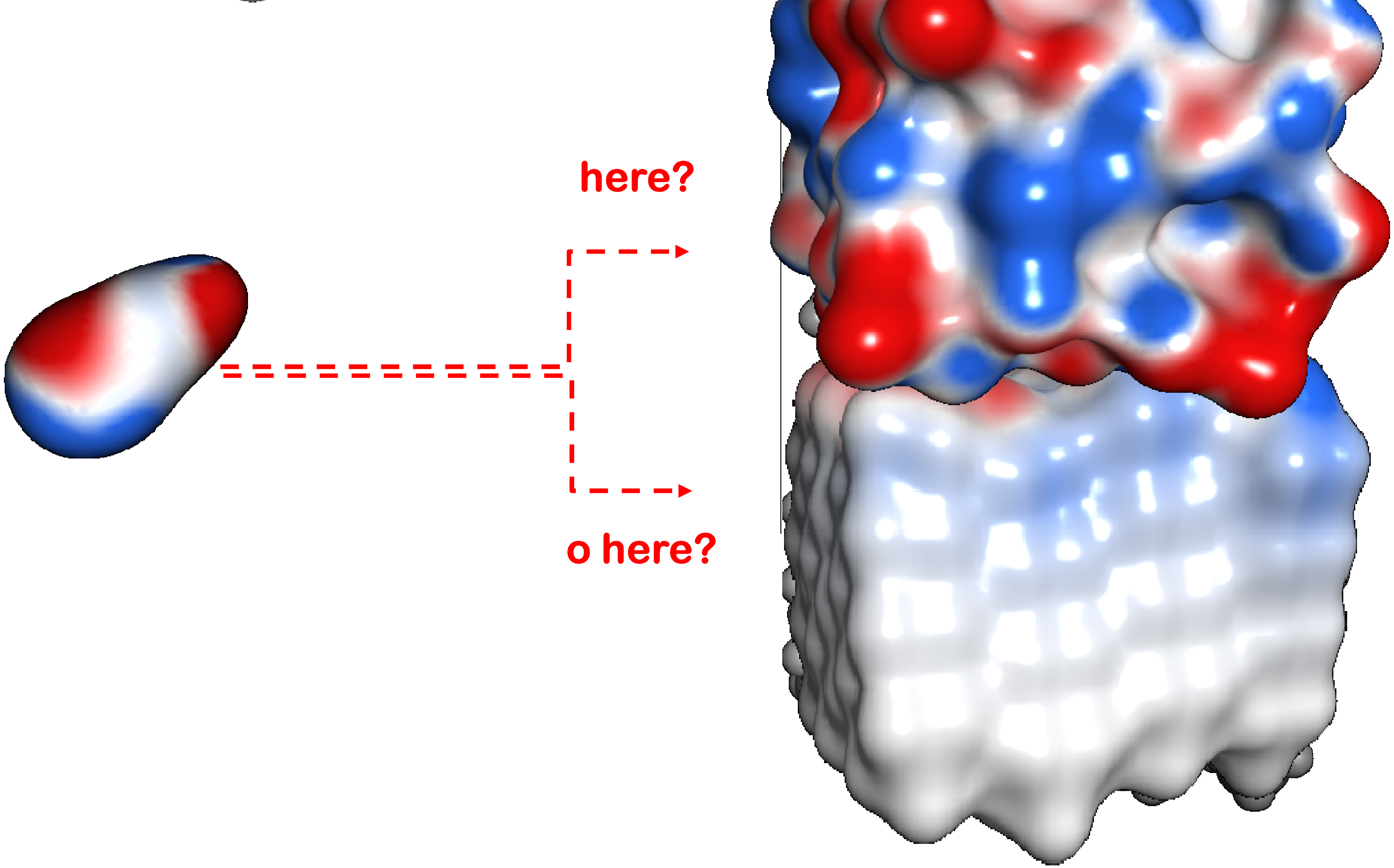
hydrophobic interaction

The misleading alternative term ‘hydrophobic bond’ is discouraged.

1994, *66*, 1123

IUPAC Compendium of Chemical Terminology 2nd Edition (1997)

How we can address in the real world about  phobicity ?

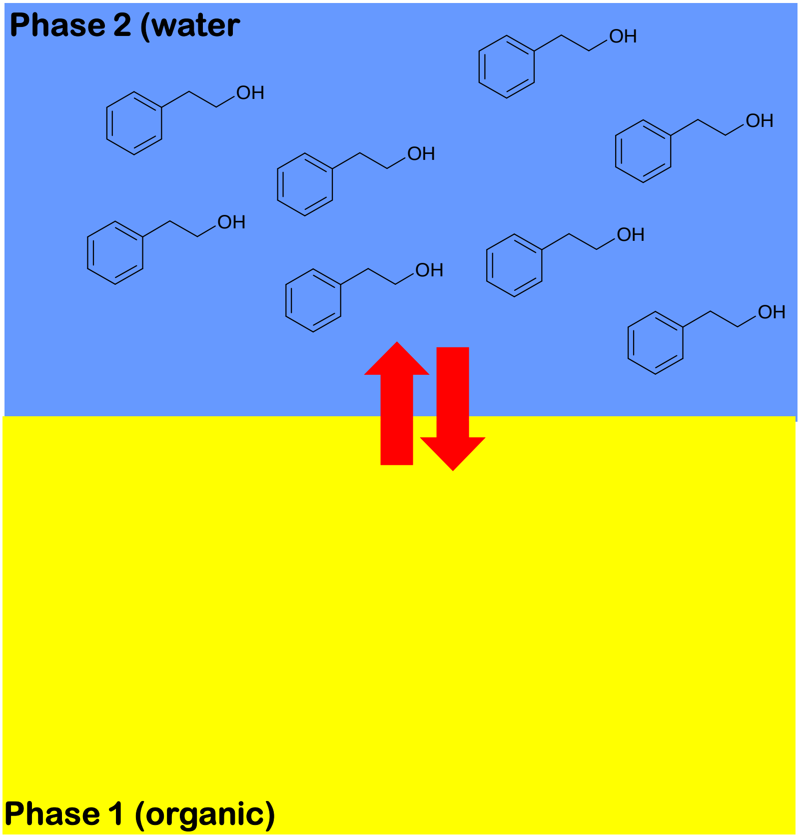


here?

o here?

Hydrophobicity and partition coefficient

$$\mu_{(phase2)} = \mu^0_{(phase2)} - RT\ln\left[C_{(phase2)}\right]$$



$$\mu_{(phase1)} = \mu^0_{(phase1)} - RT\ln\left[C_{(phase1)}\right]$$

At the equilibrium:

$$\mu_{(phase1)} = \mu_{(phase2)}$$

$$\mu^0_{(phase1)} - RT\ln\left[C_{(phase1)}\right] = \mu^0_{(phase2)} - RT\ln\left[C_{(phase2)}\right]$$

$$\mu^0_{(phase1)} - \mu^0_{(phase2)} = RT\ln\left[C_{(phase1)}\right] - RT\ln\left[C_{(phase2)}\right]$$

$$\mu^0_{(phase1)} - \mu^0_{(phase2)} = RT\ln\left[\frac{C_{(phase1)}}{C_{(phase2)}}\right]$$

$$\left[\frac{C_{(phase1)}}{C_{(phase2)}}\right] = P$$ **Partition coefficient**

Hydrophobicity and partition coefficient

$$\left[\frac{C_{(phase1)}}{C_{(H_2O)}}\right] = P$$

Partition coefficient

We can define: "*hydrophobic*" a compound with $P > 1$;
"*hydrophilic*" a compound with $P < 1$.

$$\log\left[\frac{C_{(phase1)}}{C_{(H_2O)}}\right] = \log P$$

We can define: "*hydrophobic*" a compound with $\log P > 0$;
"*hydrophilic*" a compound with $\log P < 0$.

Just a bit of thermodynamics:

from the van't Hoff isotherm:

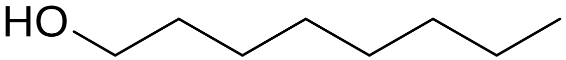
$$\Delta G = 2.303RT \log P$$

$$\Delta G = \Delta H - T\Delta S$$

$$\log P = -2.303\frac{\Delta H}{RT} + 2.303\frac{\Delta S}{R}$$

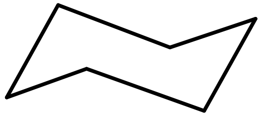
Hydrophobicity and partition coefficient

How we can choose the second phase:



***n*-octanol**

- “Similarity” with biological membrane;
- immiscible in water even if 27% of water dissolves in it… so the first region of hydration of the solute is preserved;
- UV transparent;
- Low vapor pressure.



cyclohexane

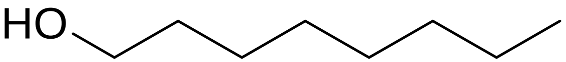
Immiscible in water and very low amount of water dissolves in it… so also the first region of hydration of the solute is lost.

The differences between the logP values in *n*-octanol and cyclohexane is a measure of the de-hydration energy of a solute.

Do you see any similarity?

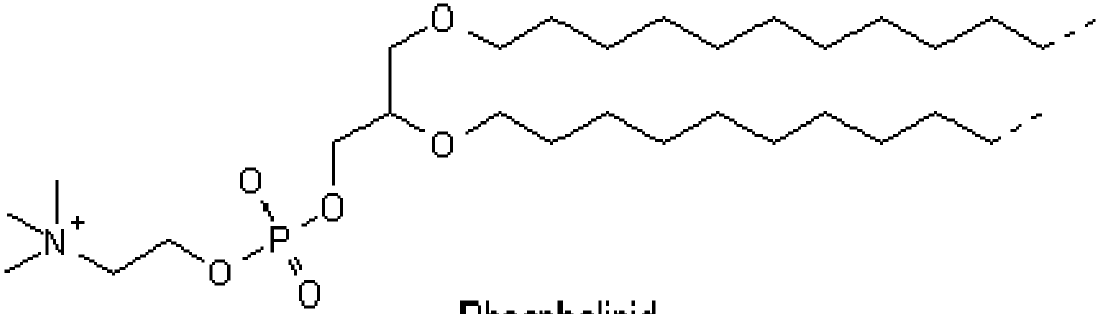
…in addition to this?

How we can choose the second phase:

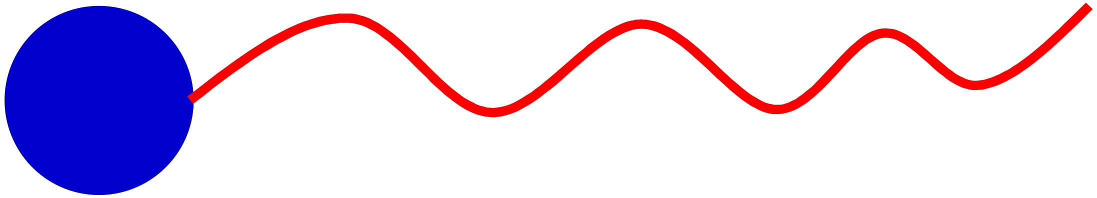


***n*-octanol**

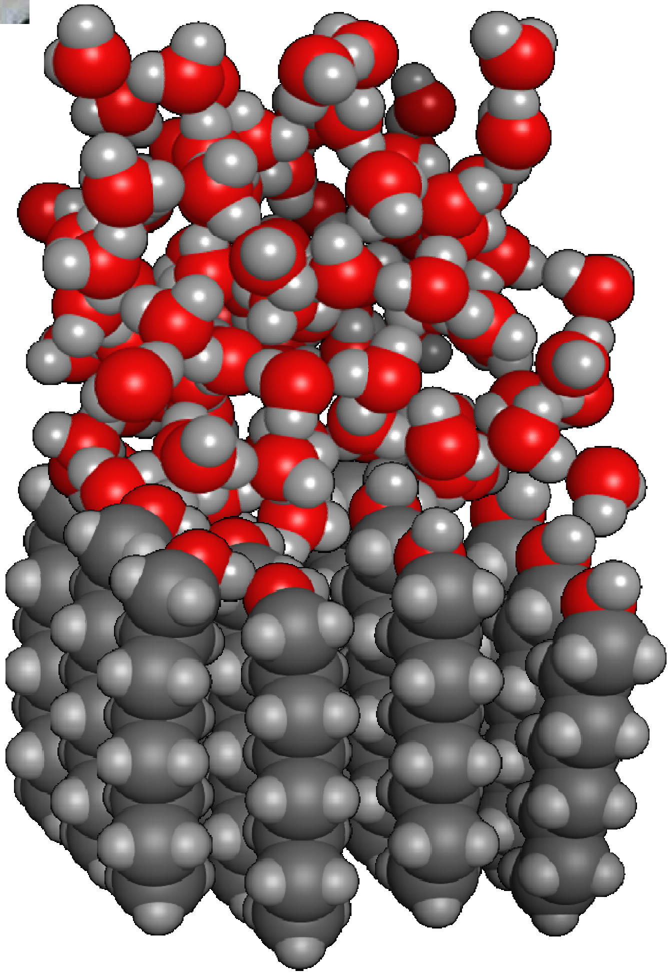
- "Similarity" with biological membrane;

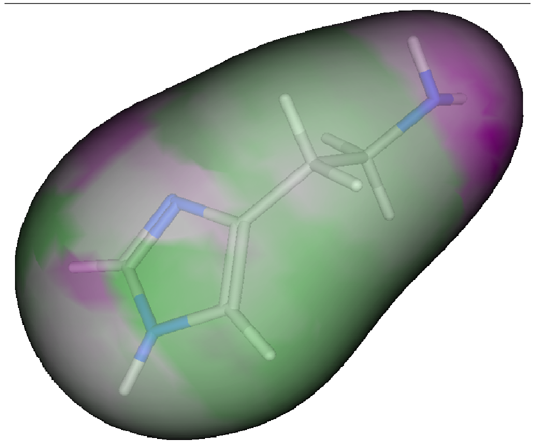


Phospholipid



From thermodynamics to molecule:

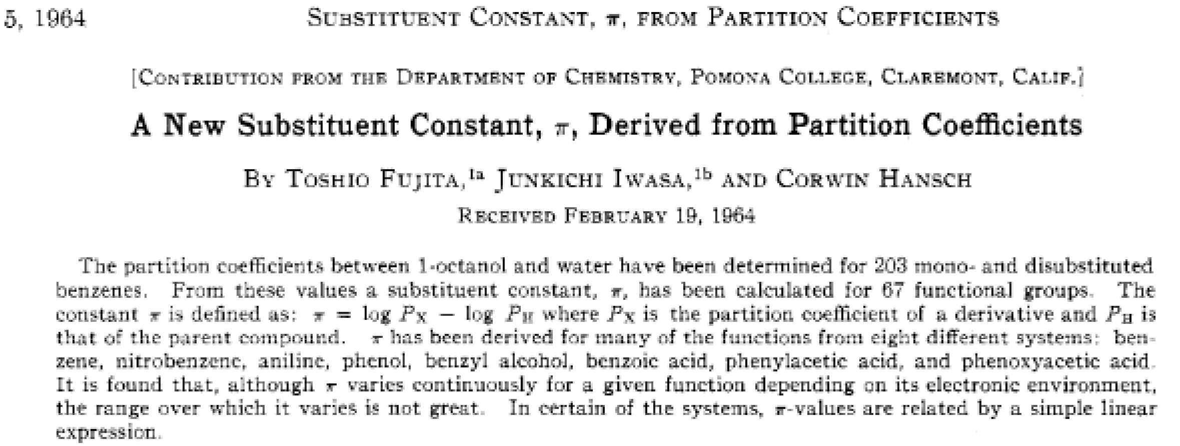




The brilliant idea of Corwin Hansch!

The fragments method of Hansch-Fujita.

(*J. Am. Chem. Soc.* 1964, 86, 5175)



Dec. 5, 1964 SUBSTITUENT CONSTANT, π, FROM PARTITION COEFFICIENTS 5175

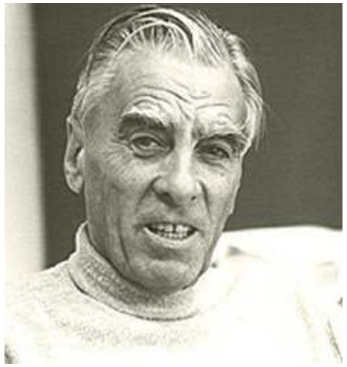
[CONTRIBUTION FROM THE DEPARTMENT OF CHEMISTRY, POMONA COLLEGE, CLAREMONT, CALIF.]

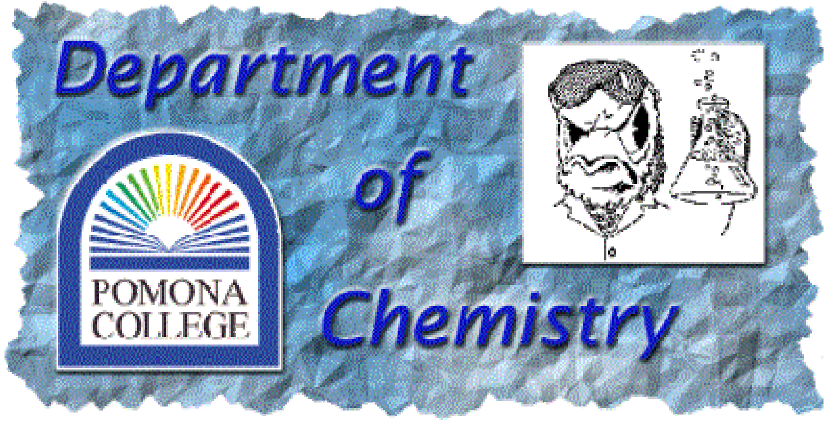
A New Substituent Constant, π, Derived from Partition Coefficients

BY TOSHIO FUJITA,[1a] JUNKICHI IWASA,[1b] AND CORWIN HANSCH

RECEIVED FEBRUARY 19, 1964

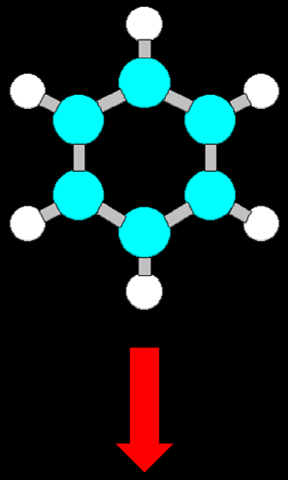
The partition coefficients between 1-octanol and water have been determined for 203 mono- and disubstituted benzenes. From these values a substituent constant, π, has been calculated for 67 functional groups. The constant π is defined as: $\pi = \log P_X - \log P_H$ where P_X is the partition coefficient of a derivative and P_H is that of the parent compound. π has been derived for many of the functions from eight different systems: benzene, nitrobenzene, aniline, phenol, benzyl alcohol, benzoic acid, phenylacetic acid, and phenoxyacetic acid. It is found that, although π varies continuously for a given function depending on its electronic environment, the range over which it varies is not great. In certain of the systems, π-values are related by a simple linear expression.



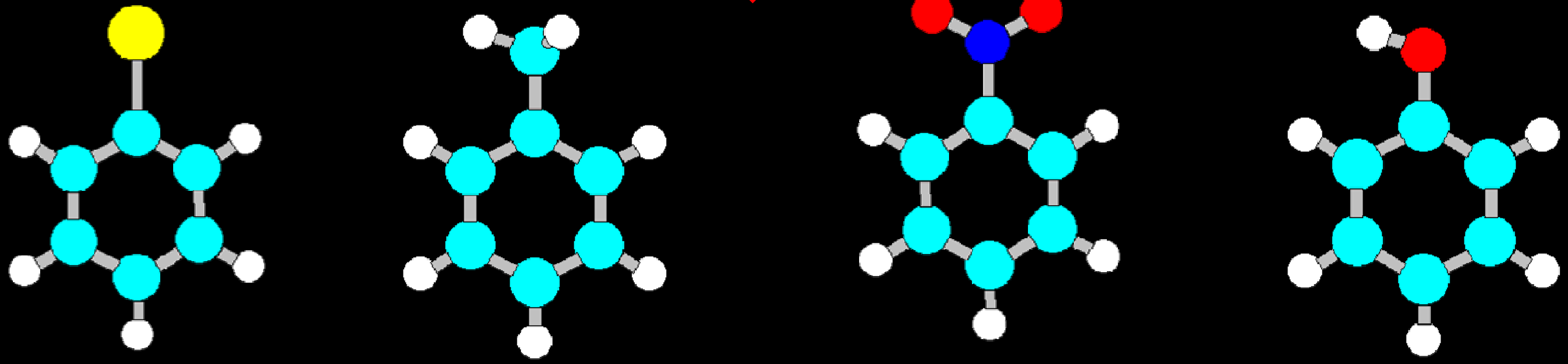


POMONA COLLEGE (Claremont, California)

The hydrophobic constant of substituent, π



log P = 2.13

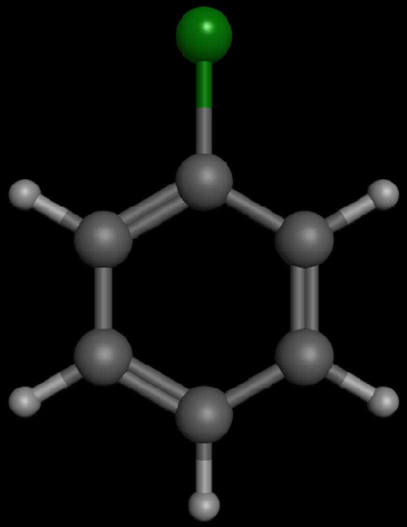


log P = 2.99   log P = 2.69   log P = 1.85   log P = 1.48

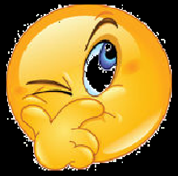
$$\pi_X = \log P_{C_6H_5X} - \log P_{C_6H_6}$$

$\pi_{Br} = 0.86$ $\pi_{CH_3} = 0.56$ $\pi_{NO_2} = -0.28$ $\pi_{OH} = -0.67$

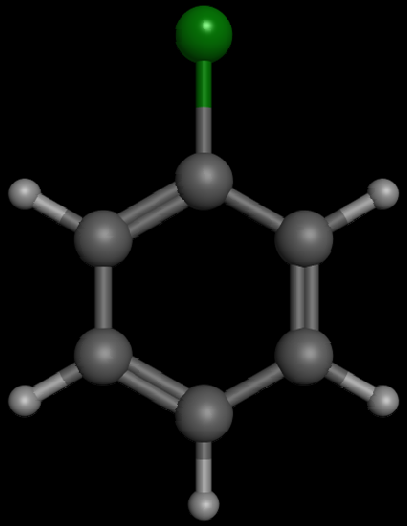
The hydrophobic constant of substituent, π



Substituent	π aromatic
F	0.14
Cl	0.71
Br	0.86
I	1.12
OH	-0.67
OCH_3	-0.02
SCH_3	0.61
CN	-0.57
COOH	-0.28
$COOCH_3$	-0.01
$COCH_3$	-0.55
NH_2	-1.23
$N(CH_3)_2$	-0.28
NO_2	-0.28
CH_3	0.56

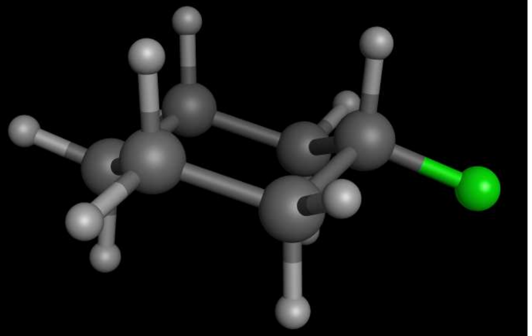


The hydrophobic constant of substituent, π



Ref. benzene
log P = 2.13

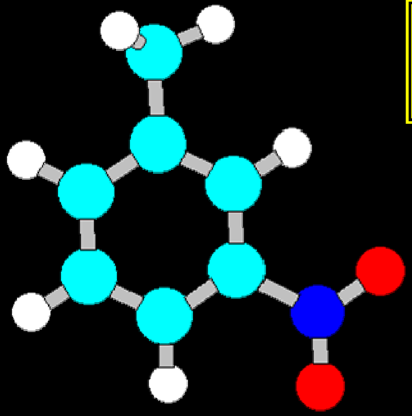
Substituent	π aromatic
F	0.14
Cl	0.71
Br	0.86
I	1.12
OH	-0.67
OCH_3	-0.02
SCH_3	0.61
CN	-0.57
COOH	-0.28
$COOCH_3$	-0.01
$COCH_3$	-0.55
NH_2	-1.23
$N(CH_3)_2$	-0.28
NO_2	-0.28
CH_3	0.56



Ref. cyclohexane
log P = 3.44

The additive rule:

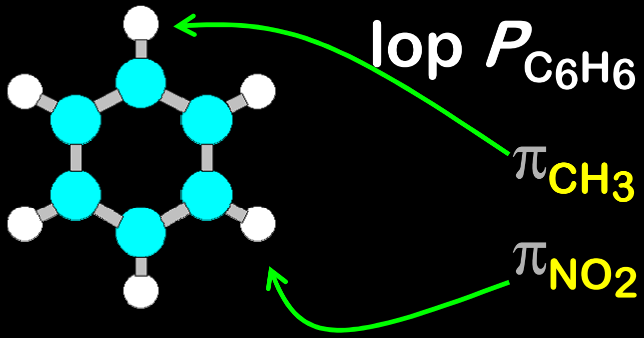
It works? Just check together:



log *P*: 2.45

IMPRESSIVE!!!!!

log *P* $_{calculated}$ (ClogP):



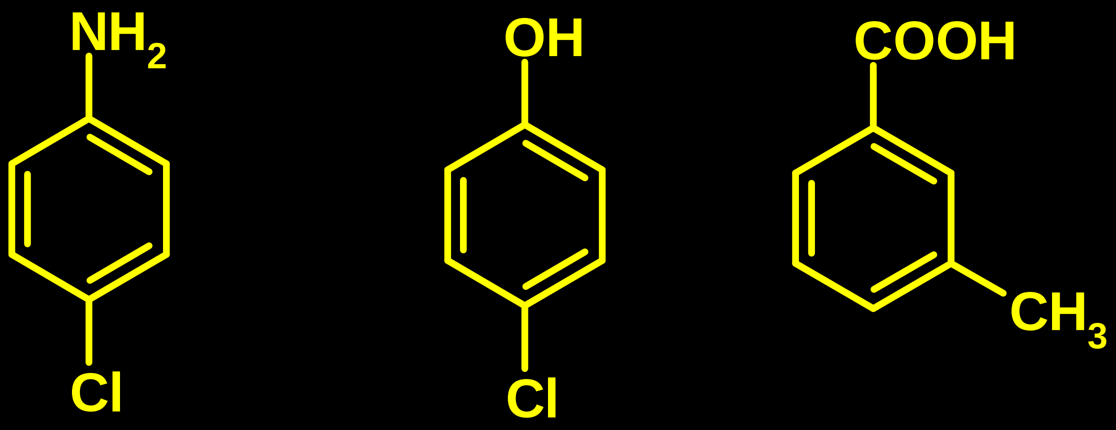
lop P_{C6H6} = 2.13

π_{CH_3} = 0.56

π_{NO_2} = -0.28

2.41

The hydrophobic constant of substituent, π



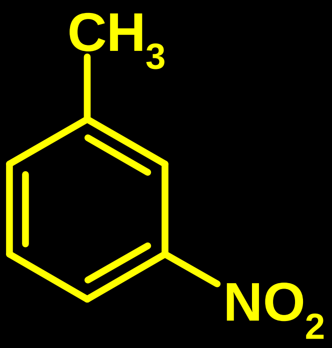
…

log *P* 1.80	log *P* 2.30	log *P* 2.40
Clog *P* 1.61	Clog *P* 2.14	Clog *P* 2.38

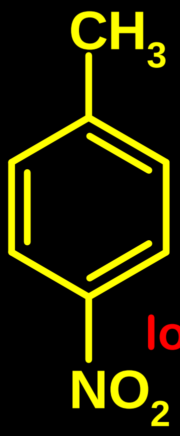
…not too bad!!!

The hydrophobic constant of substituent, π

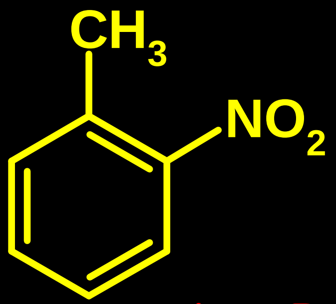
But not so good…



log *P* 2.45

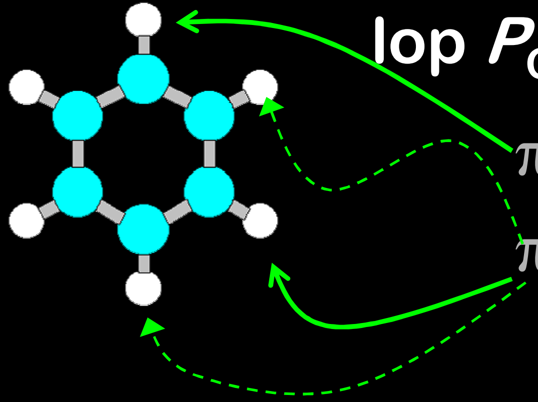


log *P* 2.42



log *P* 2.30

log *P* calculated (ClogP):



$$\text{lop } P_{C6H6} = 2.13$$

$$\pi_{CH3} = 0.56$$

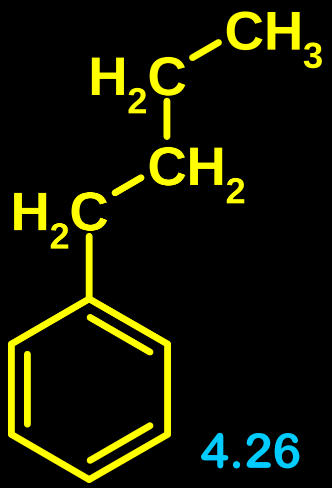
$$\pi_{NO2} = -0.28$$

$$2.41$$

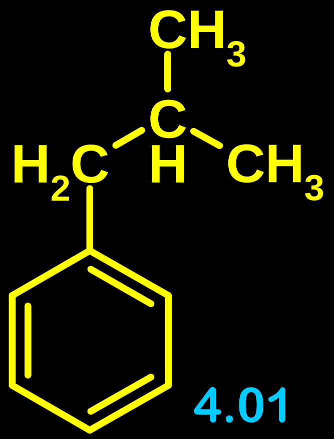
?

The hydrophobic constant of substituent, π

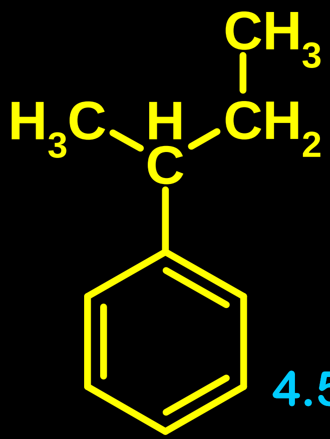
again…



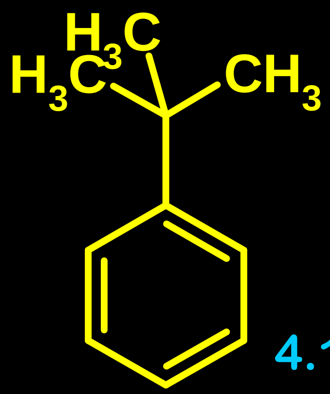
4.26



4.01

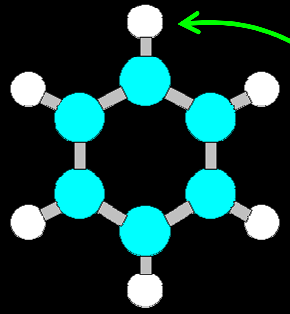


4.57



4.11

log $P_{\text{calculated}}$ (ClogP):



$\text{lop } P_{C6H6} = 2.13$

$4 \times \pi_{CH3} = 2.24$

$\overline{4.37}$?

Corwin… we have a problem!

Here is how he fixed the problem… as an engineer usually done introduction the magic *correction factors* !!!

Remember : a *correction factor* is any mathematical adjustment made to a calculation to account for deviations in either the sample or the method of measurement.

The hydrophobic correction factors, $\Delta\pi$

- Branched carbon chain ($\Delta\pi$ = -0.20);
- Double bond ($\Delta\pi$ = -0.30);
- Intra-molecular H-bond ($\Delta\pi$ = 0.65);
- Ring condensation ($\Delta\pi$ = -0.20)
- …

Finally the log *P* calculated by Hansch-Fujita:

$$\text{Clog P} = \log P_{ref} + \Sigma\,\pi_{Xi} + \Sigma\,\Delta\pi$$

Corwin… where is your problem?

$$\pi_H = 0$$

Is this true?

But how we can demonstrate this is not true?

From Hansch's π to Rekker's *f*…

A second nasty problem arose between Hansch and me on the simplest conceivable molecule: H_2.

We rejected the Hansch approach of the *substituent* constant (the lipophilicity value of a group or atom when it *substitutes* an H atom in a structure) and introduced our new concept, the hydrophobic *fragmental* constant: the *proper* lipophilicity contribution of a *constituent part of* a structure to the total lipophilicity.

Rekker, R., The Hydrophobic Fragmental Constant, Elsevier, Amsterdam (1977)

From Hansch's π to Rekker's *f*...

The idea:

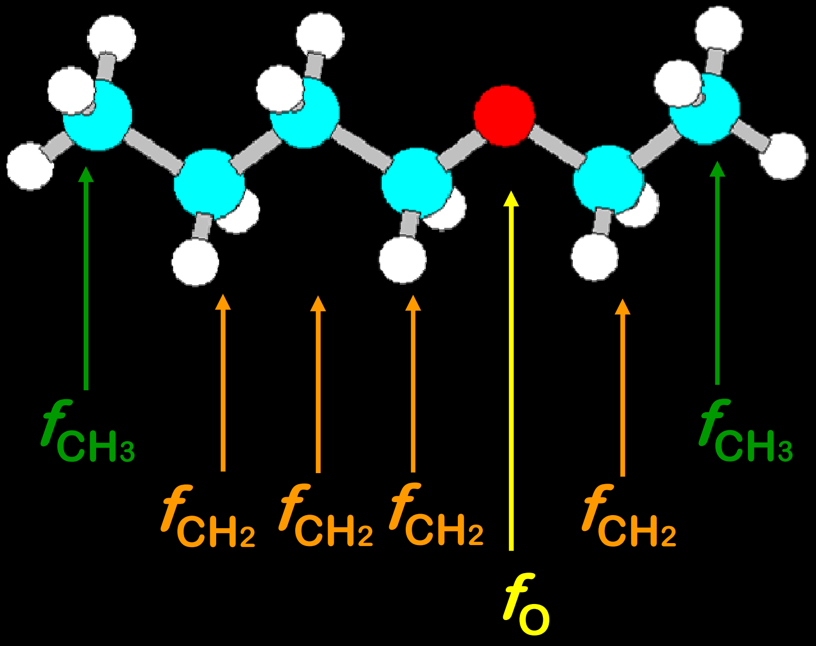
$$f_H = ½ \log P_{H2} = ½ \; 0.45 = 0.225$$

Clog *P* recalculated by Rekker:

$$\text{Clog P} = \Sigma\, a_X\, f_X + \Sigma\, \Delta\pi$$

Rekker, R., The Hydrophobic Fragmental Constant, Elsevier, Amsterdam (1977)

Substituent	f Aliphatic		f Aromatic
H	0.21		0.18
C		0.15	
CH		0.24	
CH_2		0.53	
CH_3		0.70	
C_6H_3		1.43	
C_6H_4		1.69	
C_6H_5		1.89	
F	-0.46		0.40
Cl	0.06		0.92
Br	0.27		1.13
I	0.59		1.45
O	-1.51		-0.43
OH	-1.27		-0.31
SH	0.00		0.62
CN	-1.01		-0.21
COOH	-0.95		-0.03
$CONH_2$	-1.97		-1.11
NH_2	-1.43		-0.85
NO_2	-0.28		-0.85



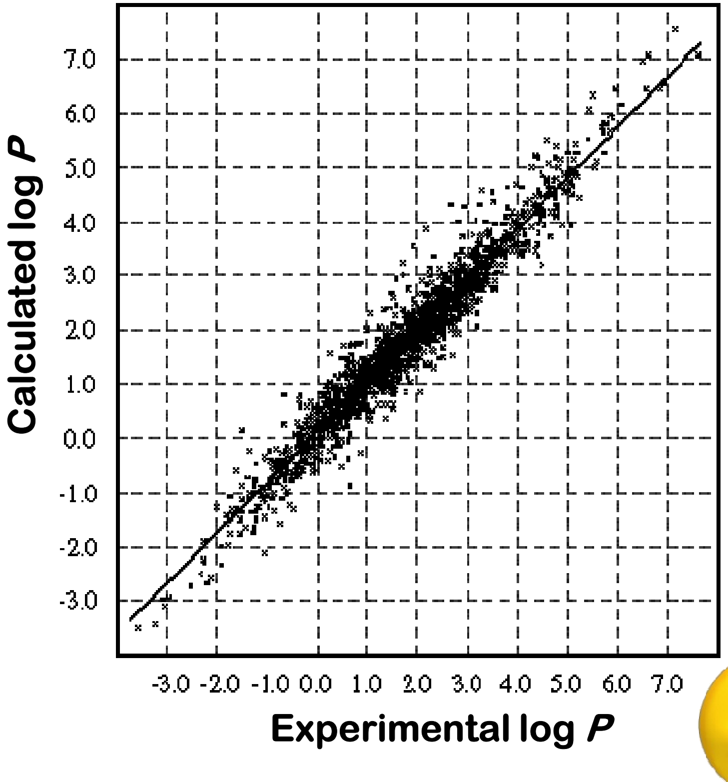
Clog *P*: 2.01

log *P*: 2.02

Nowadays, are these methods reliable?

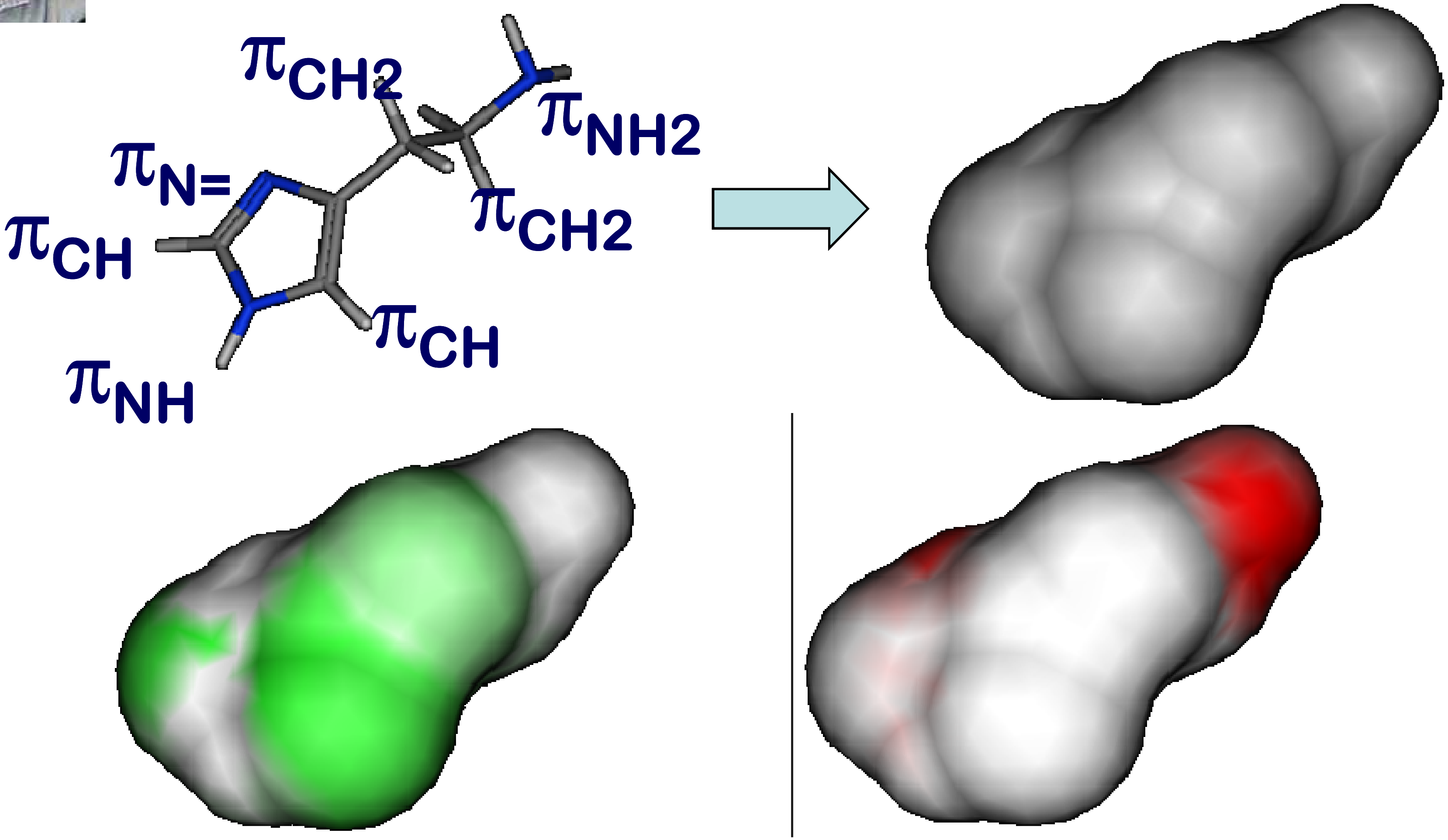
… you decide!!!

Training set referring 3600 log *P* values



ClogP… a wonderfully precious molecular descriptor!

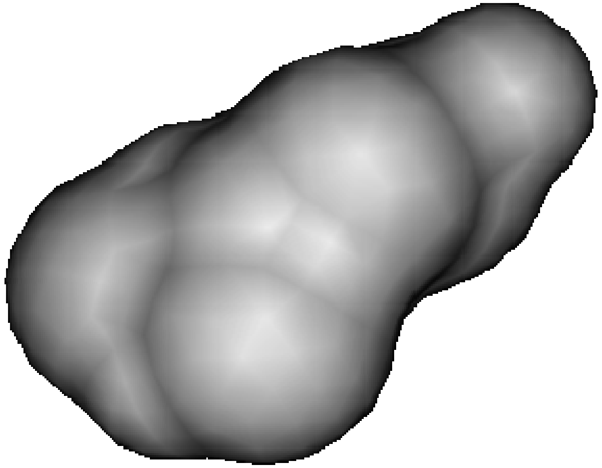
Back to our molecular surface:



$\pi > 0$

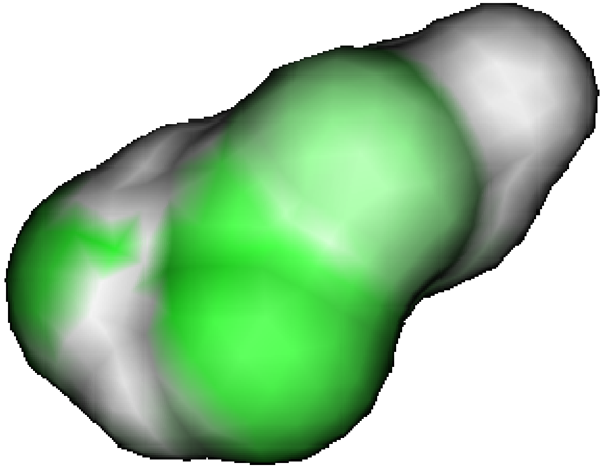
$\pi < 0$

Back to our molecular surface:



ASA ≅ 304 Å^2

$\pi > 0$

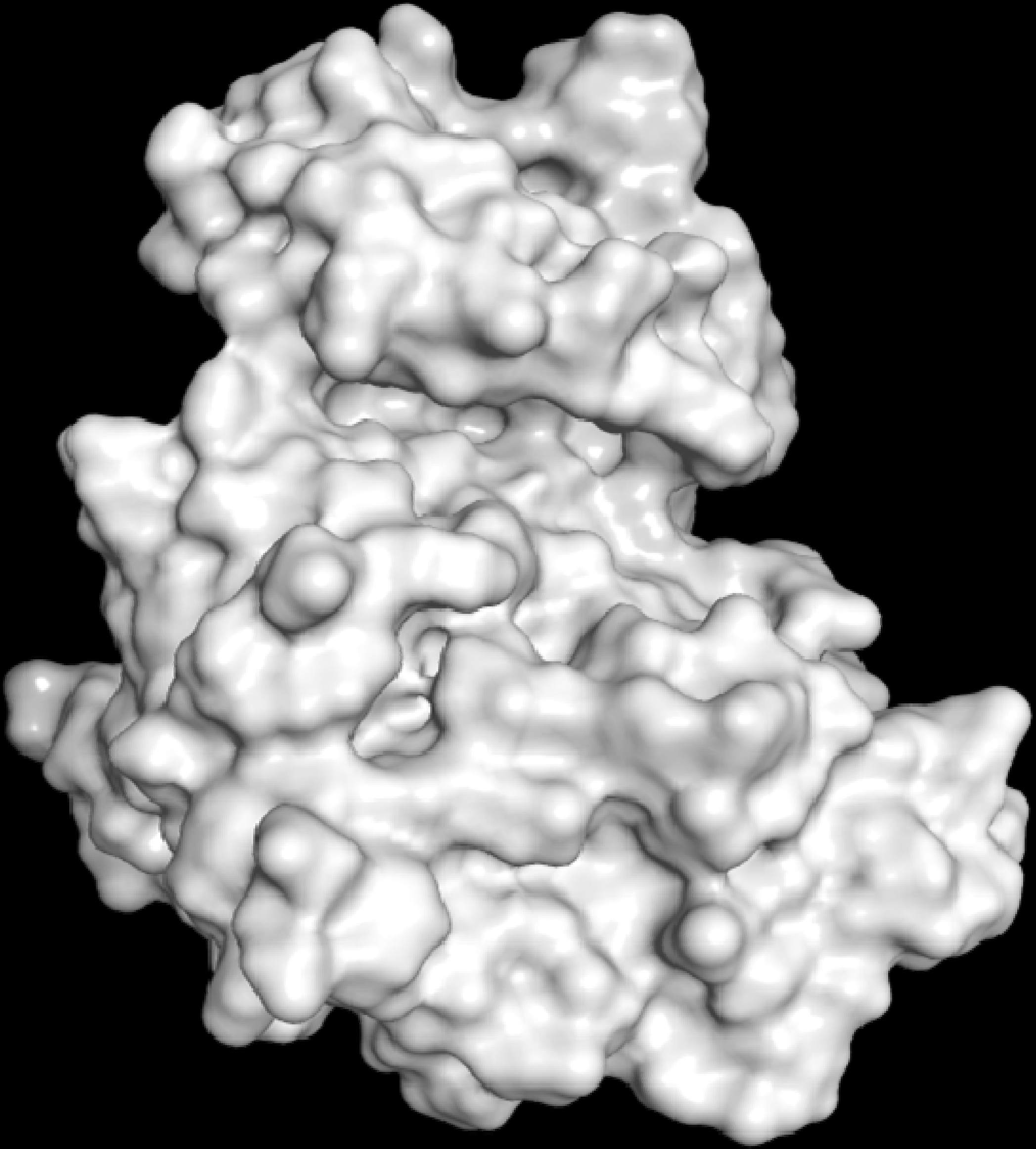


ASA_Hydrophobic ≅ 160 Å^2

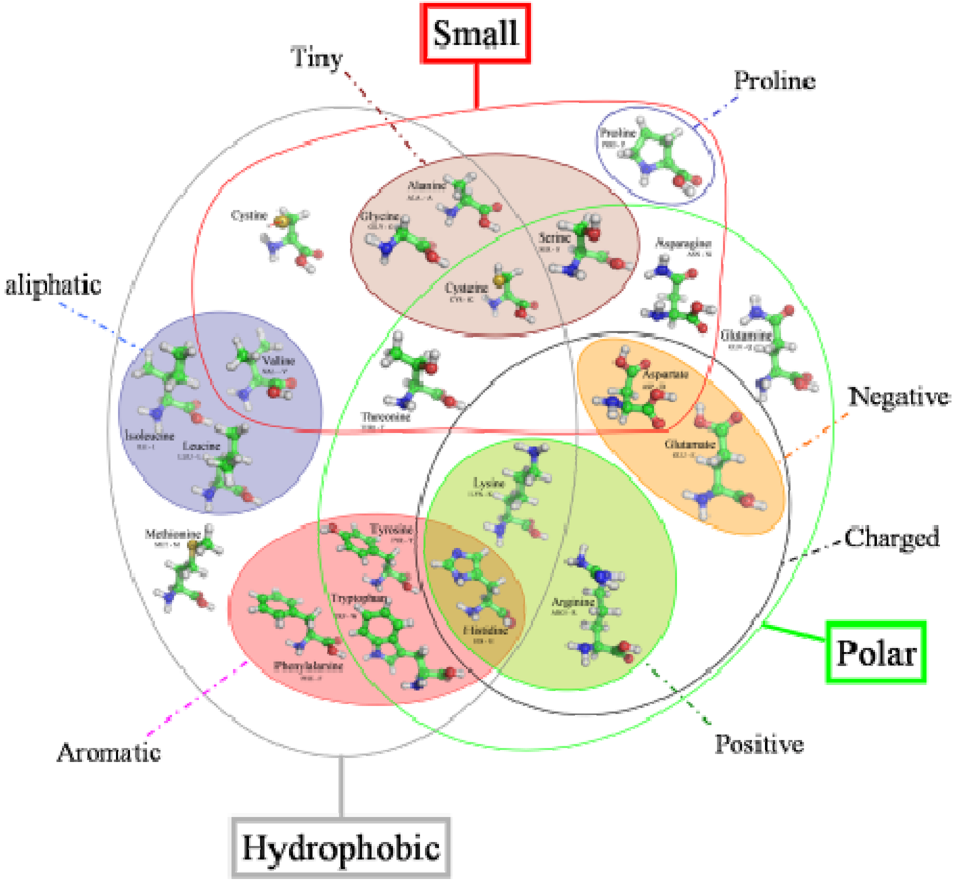
ASA_Polar ≅ 144 Å^2

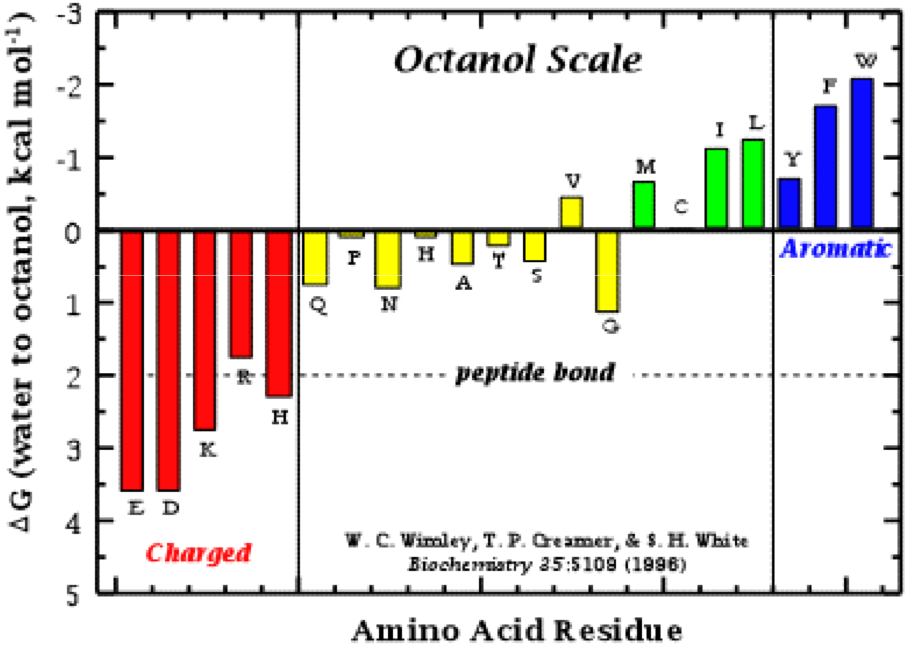
… another interesting molecular descriptor!

Back to our molecular surface:



Amino acids and hydrophobicity:



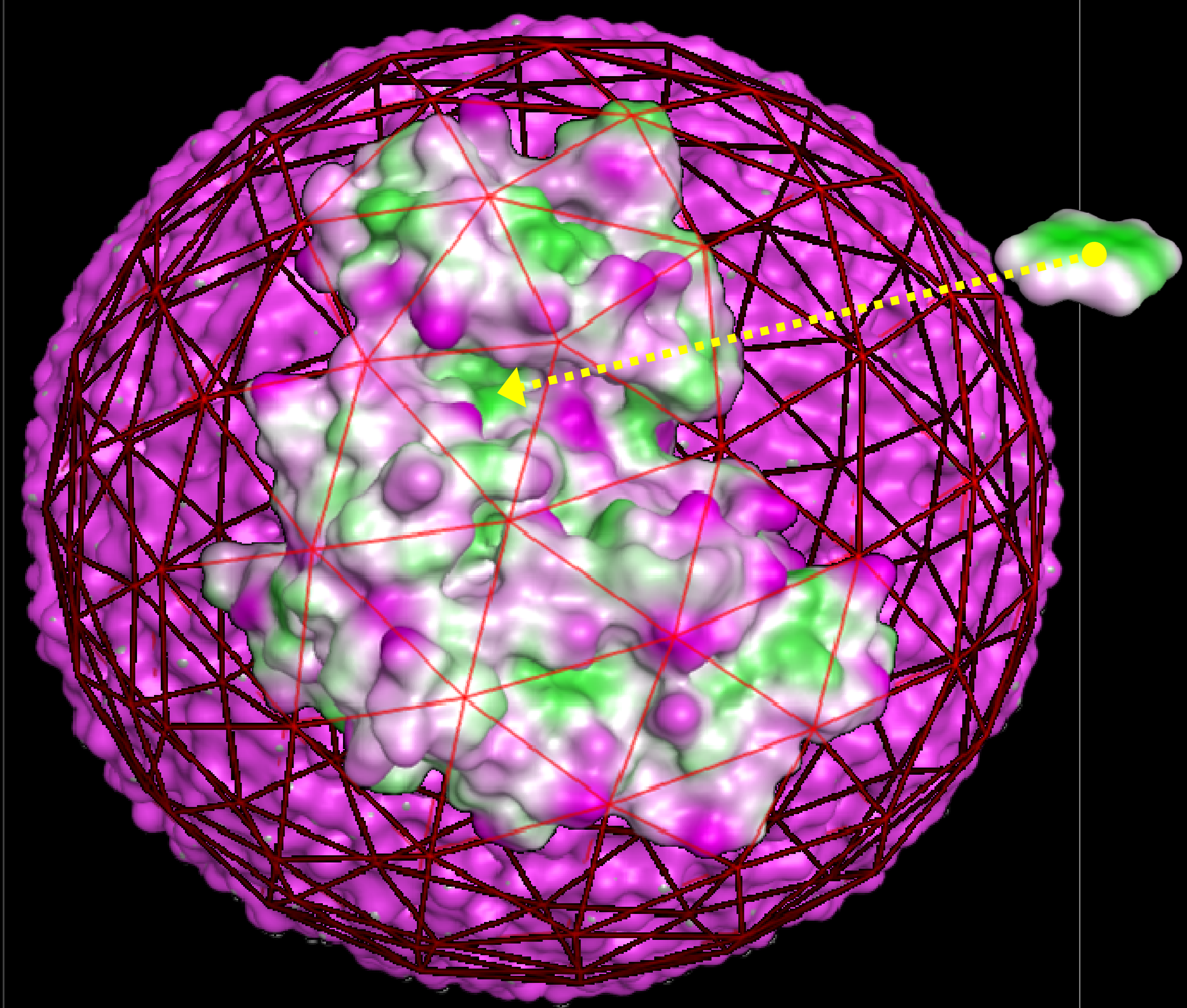


Back to our molecular surface:

🟩 $\pi > 0$

🟪 $\pi < 0$

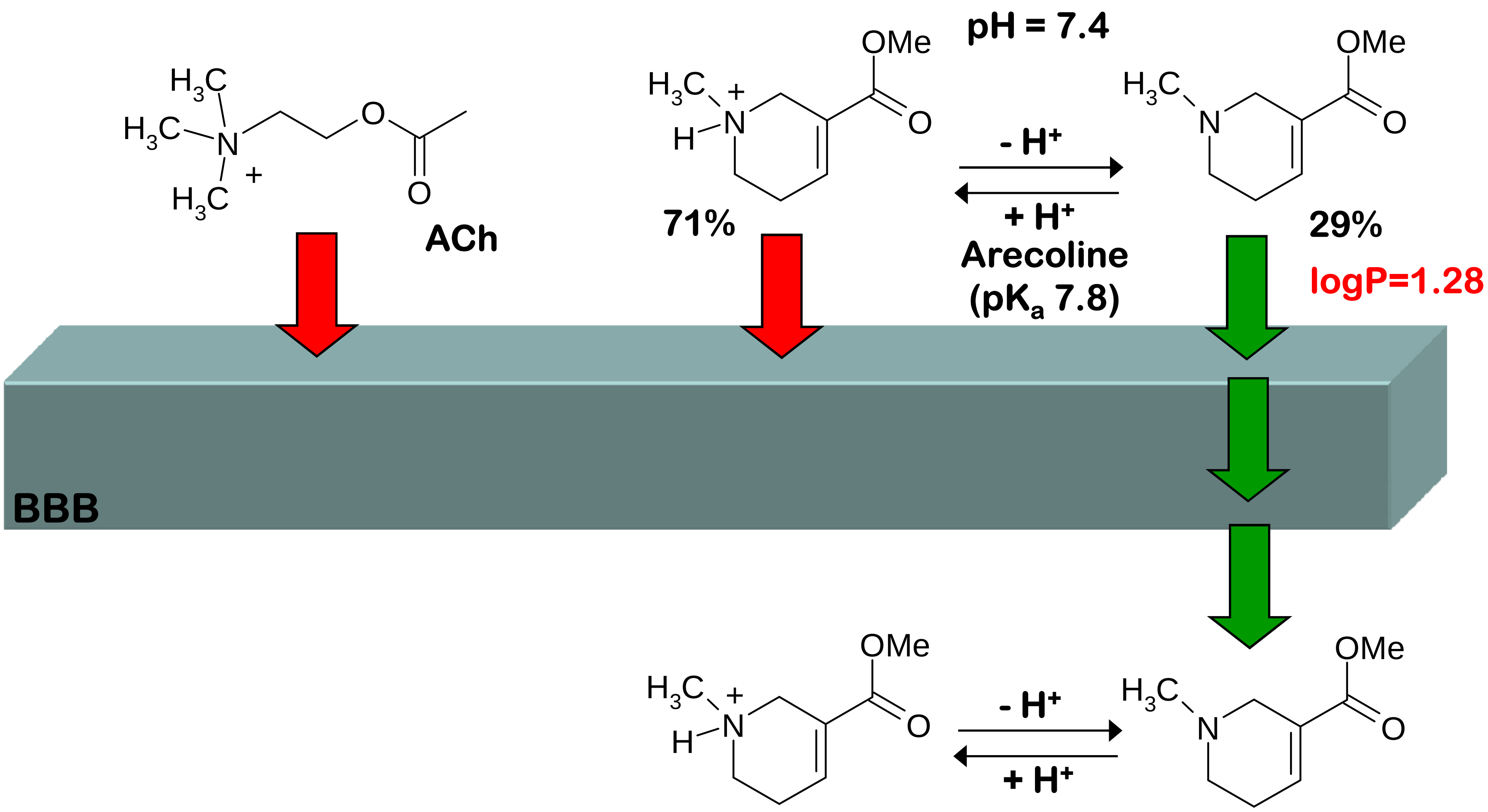
A wonderful strategy to tune the residence time of water molecules on the protein surface:



🟩 $\pi > 0$

🟪 $\pi < 0$

Don’t forget this old scheme:



Partition (logP) *vs.* distribution (logD):

We can define "*distribution coefficient* (D)":

$$\left[\frac{C_{(phase1)}}{C^{neutral}_{(H_2O)} + C^{ionized}_{(H_2O)}}\right] = D$$

and consequently the logD:

$$\log\left[\frac{C_{(phase1)}}{C^{neutral}_{(H_2O)} + C^{ionized}_{(H_2O)}}\right] = \log D$$

$$\log C_{(phase1)} - \log\left(C^{neutral}_{(H_2O)} + C^{ionized}_{(H_2O)}\right) = \log D$$

Partition (logP) *vs.* distribution (logD):

for a mono-protic acid:

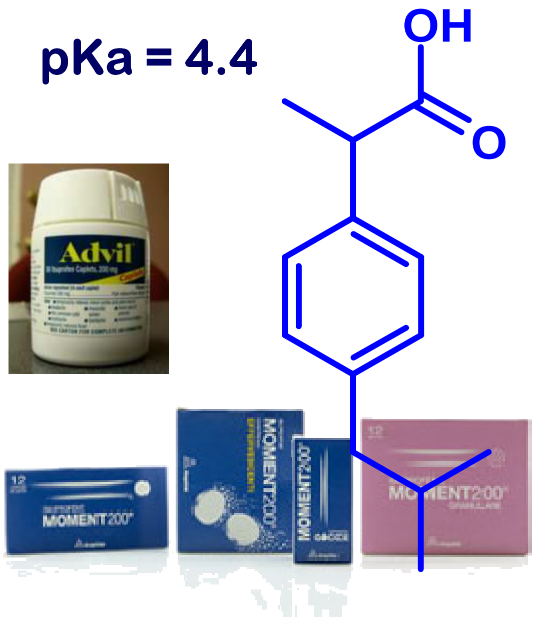
$$K_a = \frac{C^{ionized}_{(H_2O)} \times C^{H^+}_{(H_2O)}}{C^{neutral}_{(H_2O)}} \quad \text{e} \quad \frac{C_{(phase1)}}{C^{neutral}_{(H_2O)} + C^{ionized}_{(H_2O)}} = D$$

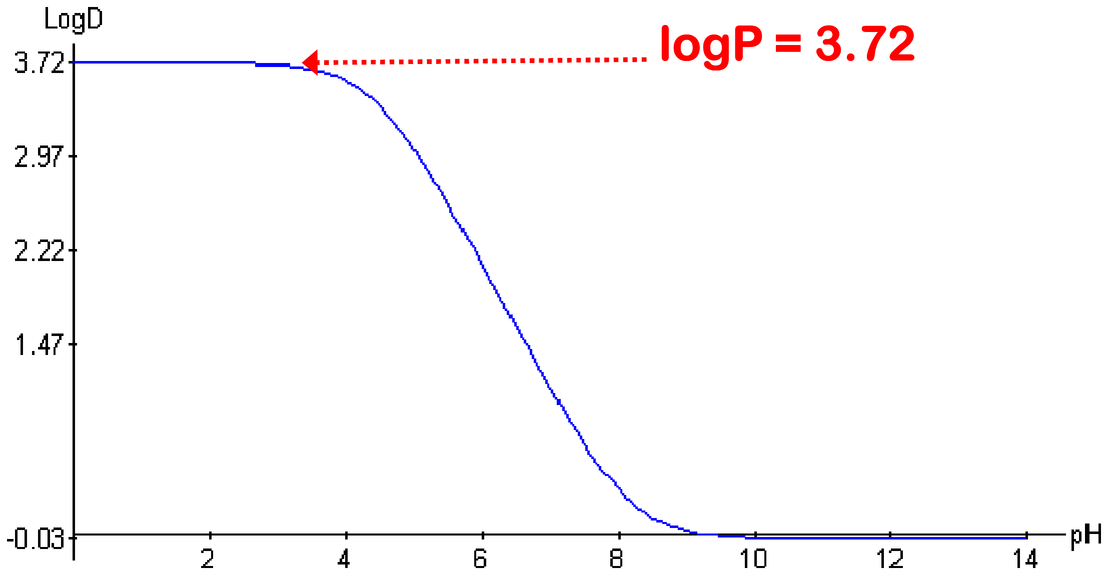
$$\log P - \log\left[1 + 10^{(pH - pK_a)}\right] = \log D$$

and for a mono-protic base:

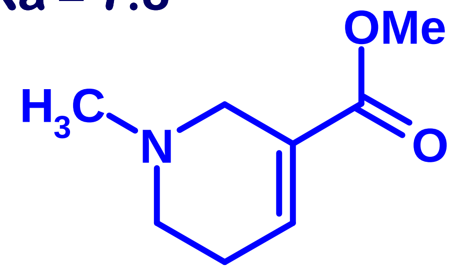
$$\log P - \log\left[1 + 10^{(pK_a - pH)}\right] = \log D$$

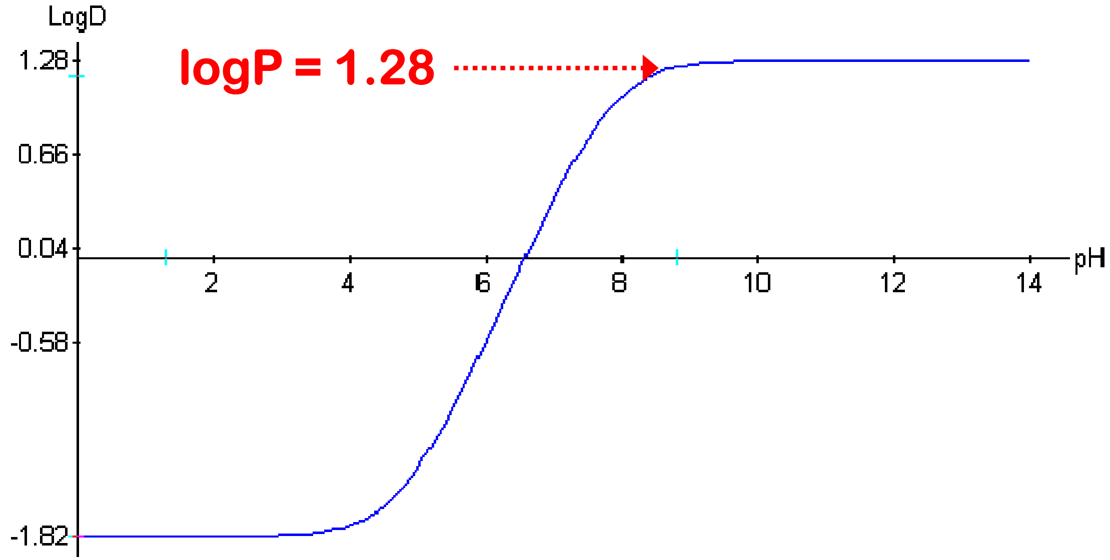
pKa = 4.4

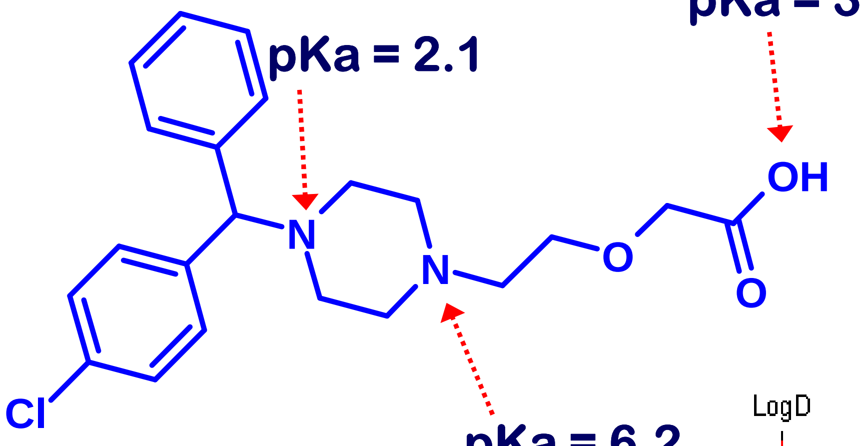

OH
O
Advil
MOMENT200
MOMENT200
MOMENT200
MOMENT200


LogD
3.72
2.97
2.22
1.47
-0.03
2
4
6
8
10
12
14
pH
logP = 3.72

pKa = 7.8


OMe
H3C
N
O


LogD
1.28
0.66
0.04
-0.58
-1.82
2
4
6
8
10
12
14
pH
logP = 1.28

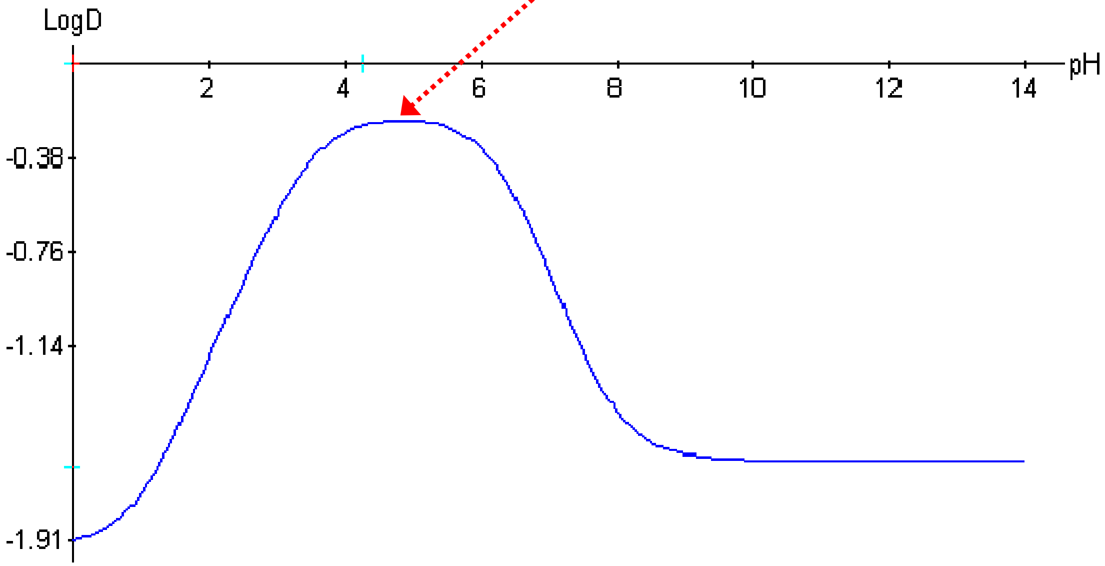


pKa = 3.5

pKa = 2.1

pKa = 6.2

logP$_{(zwitterione)}$ = -0.33



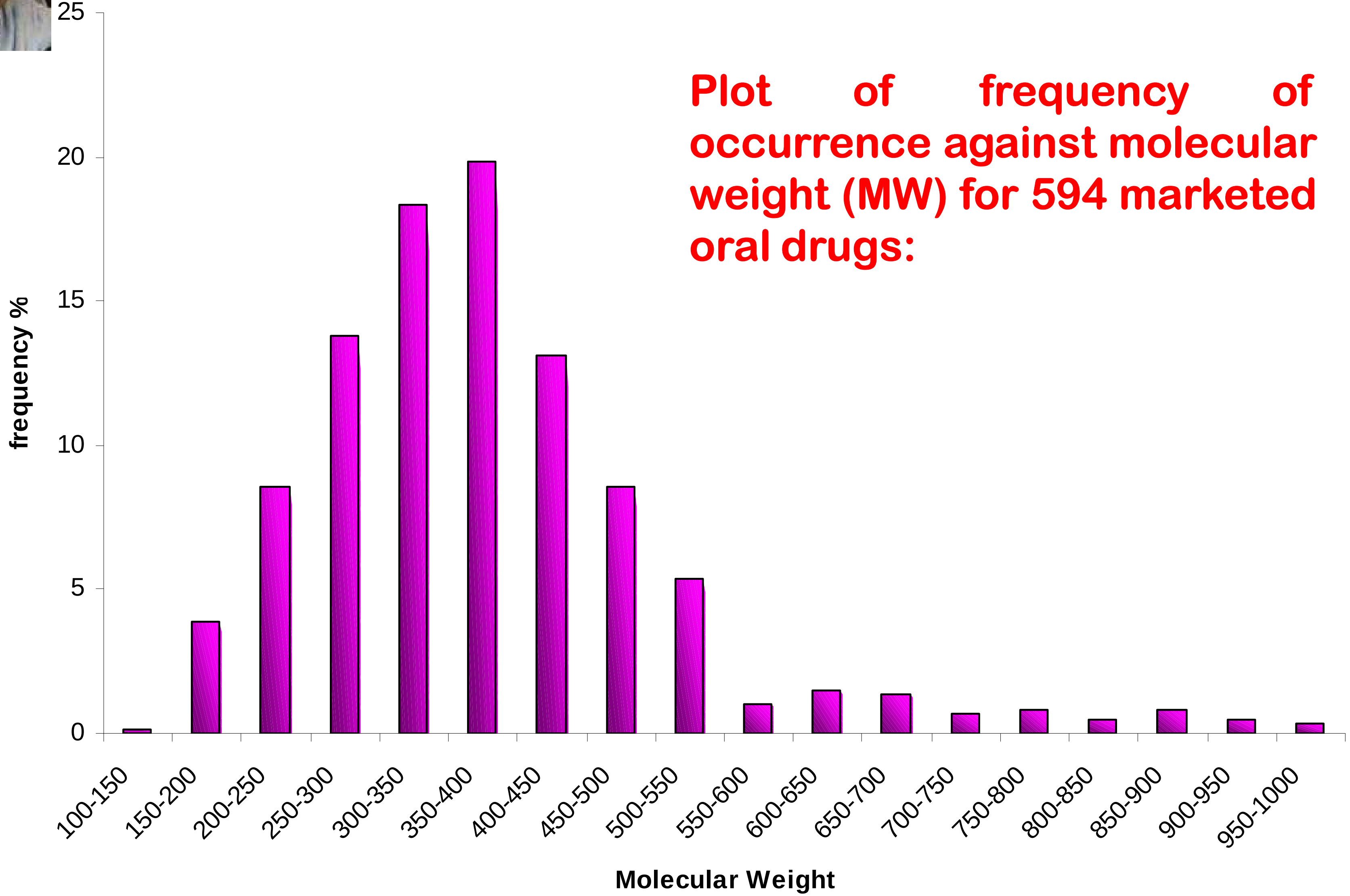
Citirizine (Zirtec®)

A short summary of useful molecular descriptors described until now:

Sterimol
Molecular Volume
Accessible Surface Area
Accessible Surface Area – Negatively Charged
Accessible Surface Area – Positively Charged
Accessible Surface Area – Hydrophobic
Accessible Surface Area – Polar

Hydrophobic substituent parameter (π)
ClogP
ClogD plot

ClogP… the largely used molecular descriptor together with MW:



Plot of frequency of occurrence against molecular weight (MW) for 594 marketed oral drugs:

Lipinski's "Rule of Five"

Poor absorption or permeation, if …

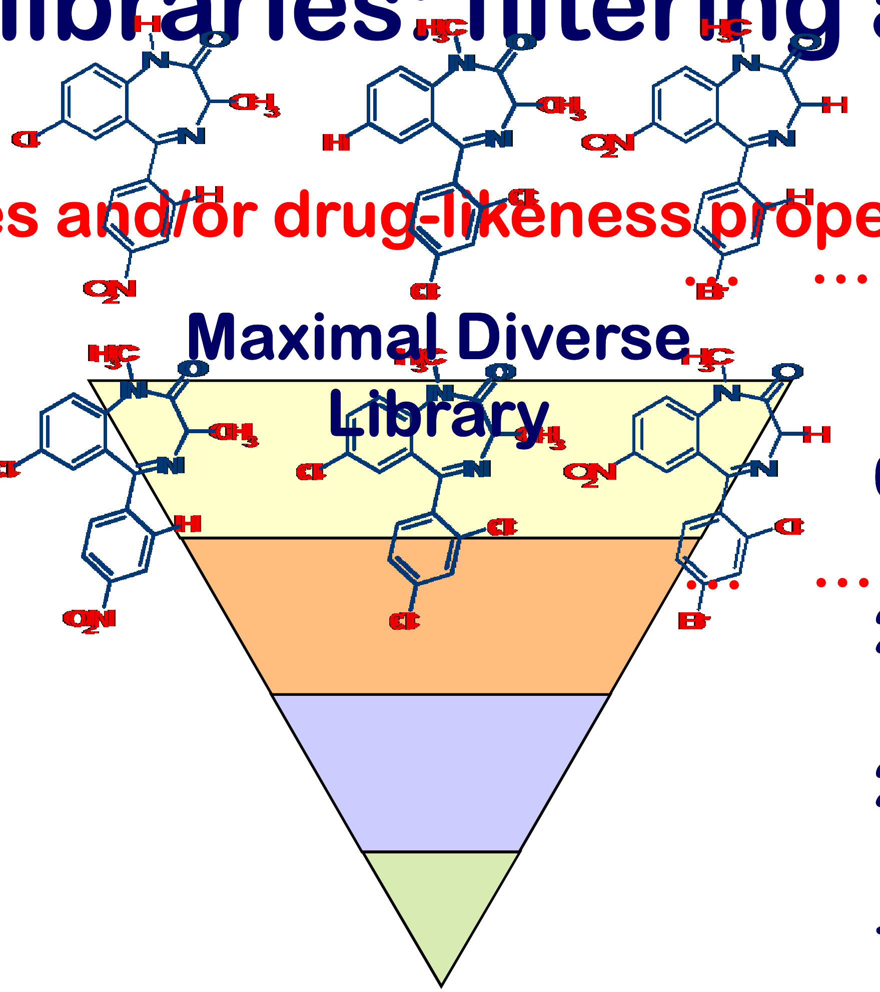
- **Number of H-bond donors > 5 (OH & NH)**
- **Number of H-bond acceptors > 10 (O & N)**
- **Molecular weight > 500**
- **log *P* > 5**

Lipinski et al., Adv. Drug Deliv. 23 (1997) 3.

N.B. Like all rules they are there to be broken and a number of exceptions exist. I have personally worked on a couple of well-absorbed drugs which broke this rule but as a general guide it works well. Remember that you may have charge in your molecule so that LogD(7.4) or LogD(5.5) is really the important parameter rather than Log P. Keeping LogD(7.4) around 2 seem generally good advice. Manipulating the pKa can be a way of improving a molecule.

Focused libraries: filtering approach

Lipinski rules and/or drug-likeness properties funnel



Maximal Diverse Library

ClogP < 5

MW < 500

HBD < 5

HBA < 10

$0 \leq \text{HBD} \leq 2$

$2 \leq \text{HBA} \leq 9$

$2 \leq \text{RTB} \leq 8$

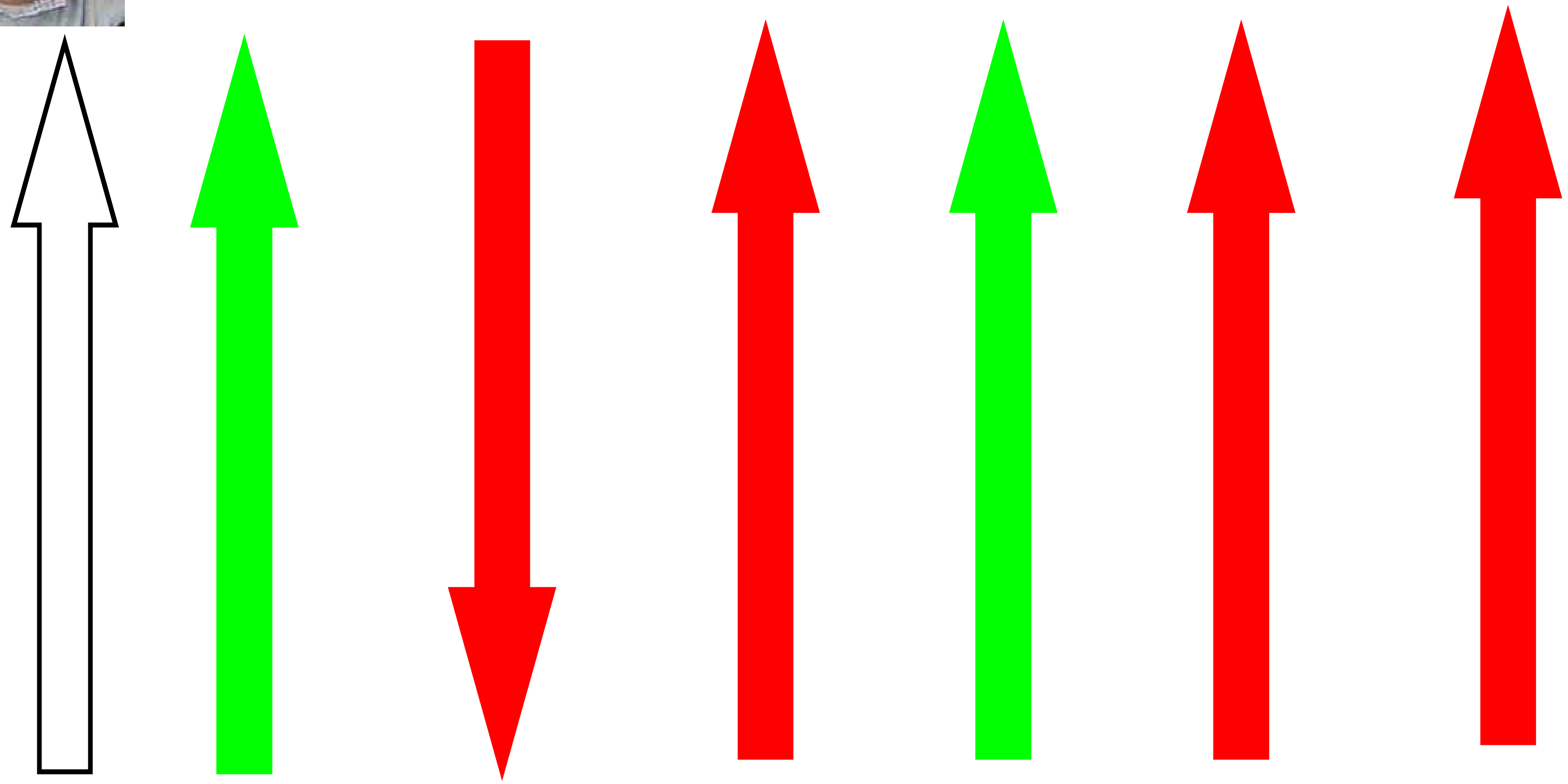
$1 < \text{RNG} \leq 4$

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

Oprea TI. Property distribution of drug-related chemical databases.
J Comput Aided Mol Des 2000 Mar;14(3):251-64

Focused Library

But remember: logP needs to be optimized!!!



logP

Binding to enzyme / receptor

Aqueous solubility

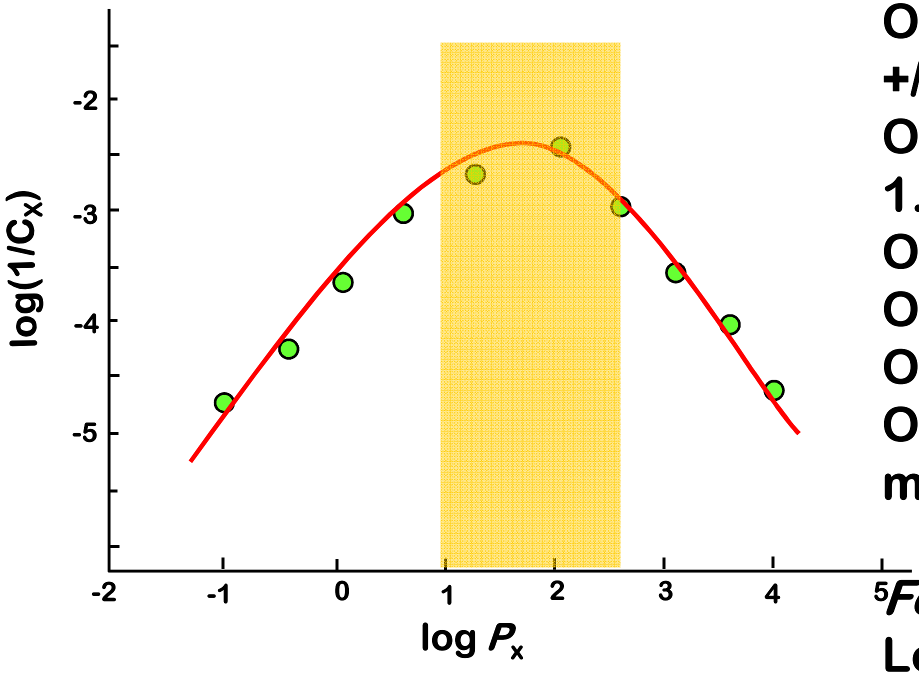
Binding to P_{450} metabolising enzymes

Absorption through membrane

Binding to blood / tissue proteins – less drug free to act

Binding to hERG heart ion channel - cardiotoxicity risk

Optimum logP (o ClogP!!)...



Studies have found: (bear in mind these may not apply to your class of chemicals)

Optimum CNS penetration around logP = 2 +/- 0.7 (Hansch)

Optimum Oral absorption around logP = 1.8

Optimum Intestinal absorption logP =1.35

Optimum Colonic absorption logP = 1.32

Optimum Sub lingual absorption logP = 1-5

Optimum Percutaneous logP = 2.6 (& low mw)

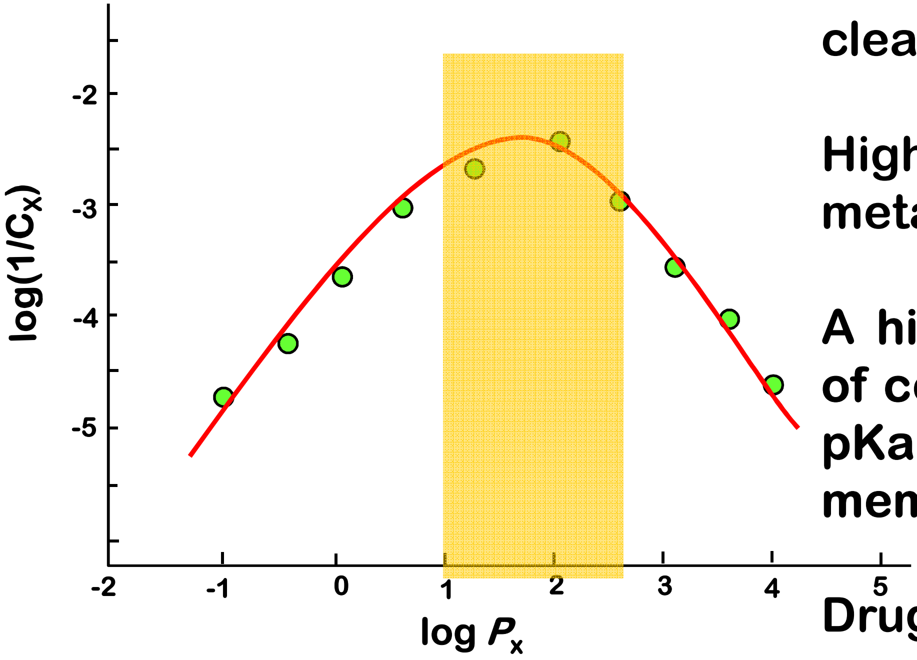
Formulation and dosing forms:

Low Log P (below 0) Injectable

Medium (0-3) Oral

High (3-4) Transdermal

Very High (4-7) Toxic build up in fatty tissues



Drug Clearance and Toxicity

Increasing LogD(7.4) above 0 will decrease renal clearance and increase metabolic clearance.

High LogD(7.4) compounds will tend to be metabolised by P450 enzymes in the liver.

A high degree of ionization keeps drugs out of cells and decreases systemic toxicity.
pKa in range 6 to 8 is advantageous for membrane penetration.

Drugs should be designed with the lowest possible logP, to reduce toxicity, non-specific binding, increase ease of formulation and bioavailability. Drugs should also be as low mw as possible to lower the risk of allergic reactions. (See principle of minimum hydrophobicity)

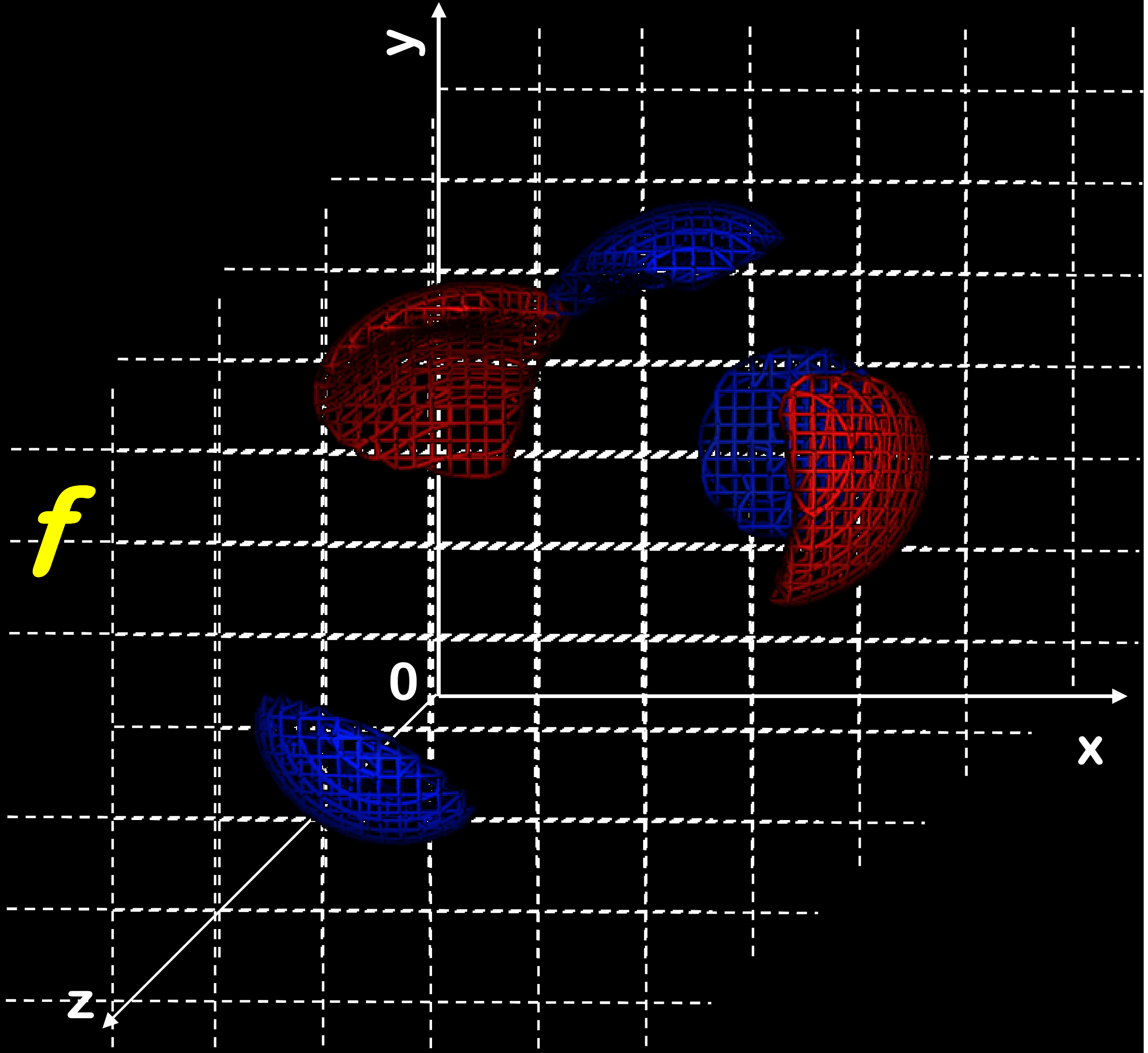
“Progettazione e Sviluppo di un Farmaco”

Stefano Moro

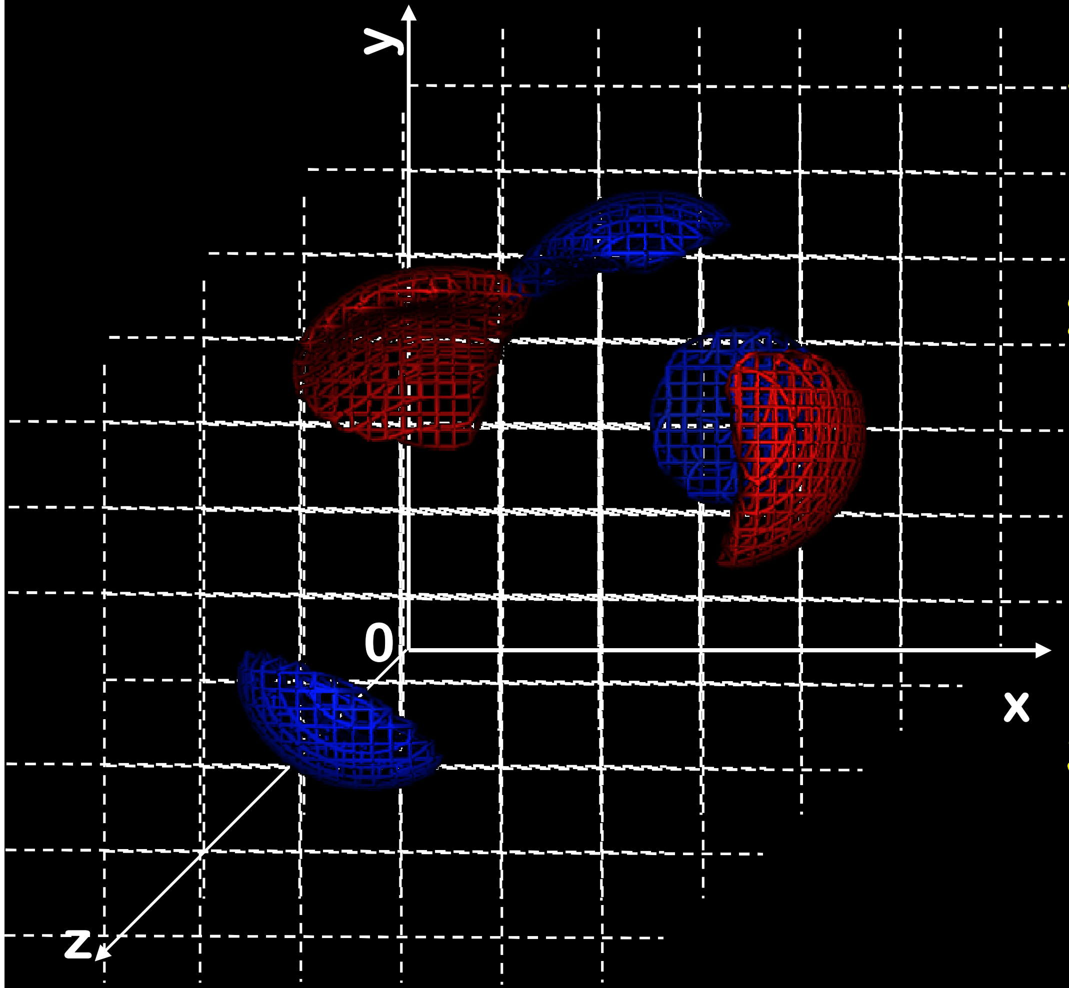
Chimica e Tecnologia Farmaceutiche

e quindi i *campi molecolari* come una nuova collezione di descrittori molecolari:

$\log 1/[D_{eff}] = f$



CoMFA (Comparative Molecular Fields Analisys): un interessante strumento QSAR… da maneggiare con cautela!



1. I campi molecolari sono conformazione dipendenti;
2. I campi molecolari sono descrittori 3D esterni, quindi per una comparazione è necessario pre-allineare le strutture molecolari;
3. i campi molecolari sono costituiti da migliaia di valori di potenziale di campo

quel bel ragazzo di Hammett… e le sue σ

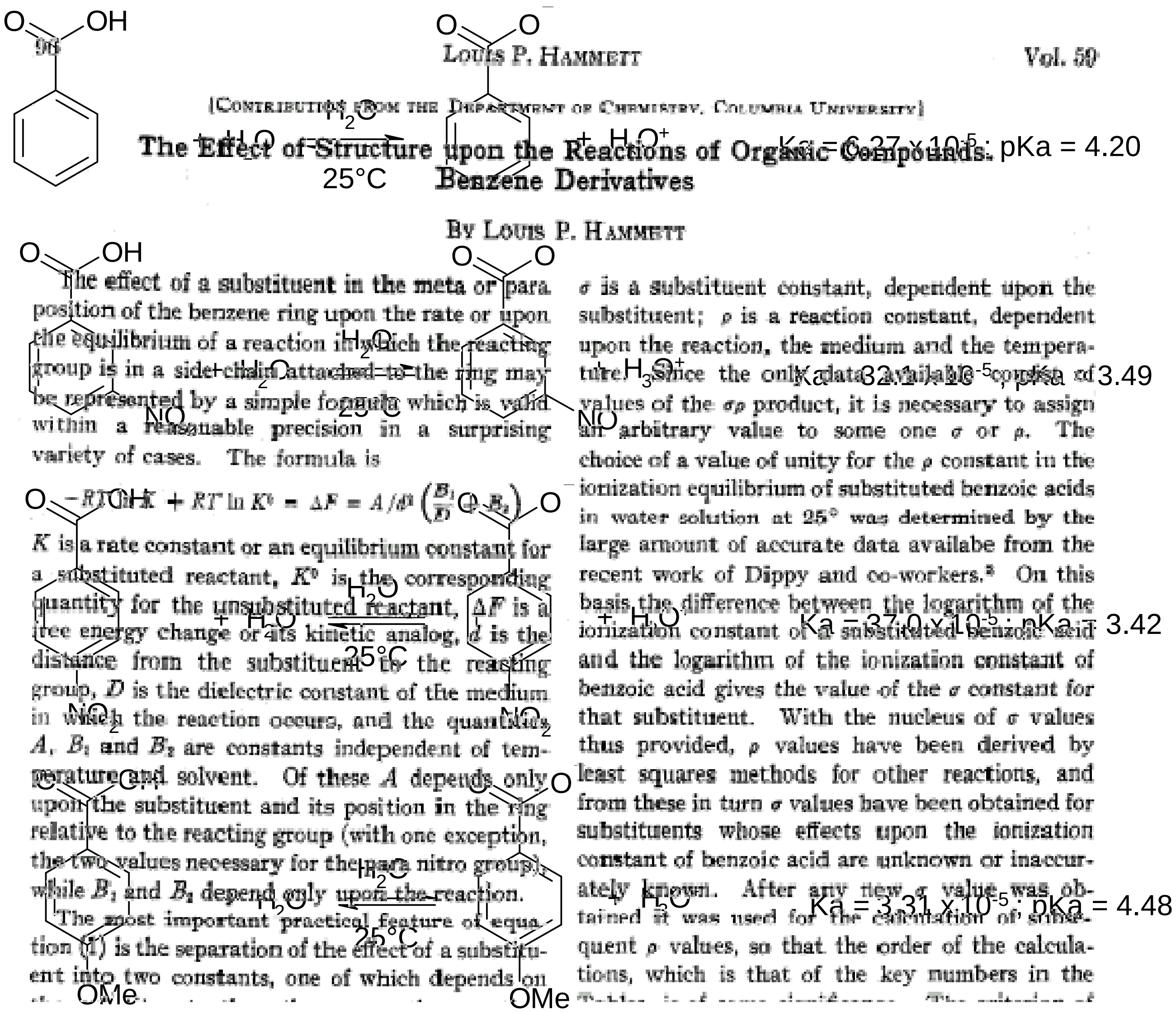
Louis Plack Hammett (April 7, 1894 — February 9, 1987)

1940

Physical Organic Chemistry.

New York; McGraw-Hill.

Dagli arbori della chimica fisica organica… ad una interessante applicazione in chimica farmaceutica.



96 LOUIS P. HAMMETT Vol. 59

[CONTRIBUTION FROM THE DEPARTMENT OF CHEMISTRY, COLUMBIA UNIVERSITY]

The Effect of Structure upon the Reactions of Organic Compounds. Benzene Derivatives

BY LOUIS P. HAMMETT

The effect of a substituent in the meta or para position of the benzene ring upon the rate or upon the equilibrium of a reaction in which the reacting group is in a side chain attached to the ring may be represented by a simple formula which is valid within a reasonable precision in a surprising variety of cases. The formula is

$$-RT \ln K + RT \ln K^0 = \Delta F = A/d^2 \left(\frac{B_1}{D} + B_2\right) \quad (1)$$

K is a rate constant or an equilibrium constant for a substituted reactant, K^0 is the corresponding quantity for the unsubstituted reactant, ΔF is a free energy change or its kinetic analog, d is the distance from the substituent to the reacting group, D is the dielectric constant of the medium in which the reaction occurs, and the quantities A, B_1 and B_2 are constants independent of temperature and solvent. Of these A depends only upon the substituent and its position in the ring relative to the reacting group (with one exception, the two values necessary for the para nitro group), while B_1 and B_2 depend only upon the reaction.

The most important practical feature of equation (1) is the separation of the effect of a substituent into two constants, one of which depends on

σ is a substituent constant, dependent upon the substituent; ρ is a reaction constant, dependent upon the reaction, the medium and the temperature. Since the only [illegible] of values of the $\sigma\rho$ product, it is necessary to assign an arbitrary value to some one σ or ρ. The choice of a value of unity for the ρ constant in the ionization equilibrium of substituted benzoic acids in water solution at 25° was determined by the large amount of accurate data availabe from the recent work of Dippy and co-workers.[3] On this basis the difference between the logarithm of the ionization constant of a substituted benzoic acid and the logarithm of the ionization constant of benzoic acid gives the value of the σ constant for that substituent. With the nucleus of σ values thus provided, ρ values have been derived by least squares methods for other reactions, and from these in turn σ values have been obtained for substituents whose effects upon the ionization constant of benzoic acid are unknown or inaccurately known. After any new σ value was obtained it was used for the calculation of subsequent ρ values, so that the order of the calculations, which is that of the key numbers in the Tables, is of some significance.

L.P. Hammett, *The Effect of Structure upon the Reactions of Organic Compounds. Benzene Derivatives.* J. Am. Chem. Soc. **1937, 59, 96-103** .

Effetto elettronico del sostituente e costante σ:

pKa per alcuni acidi benzoici sostituiti (H_2O, 25°C) e corrispondenti valori di σ dei sostituenti.

$$\sigma_X = pKa(C_6H_5COOH) - pKa(XC_6H_4COOH)$$

$$\sigma_x = \log \frac{Ka(XC_6H_4COOH)}{Ka(C_6H_5COOH)}$$

Sostituente X	pKa (XC_6H_4COOH)	σ_X
H	4.20	0
m-OCH_3	4.09	0.11
m-F	3.86	0.34
m-NO_2	3.49	0.71
p-NO_2	3.42	0.78
p-CH_3	4.37	-0.17
p-OCH_3	4.48	-0.28

Teoria dell'effetto dei sostituenti

Sostituente	σ_m	σ_p
NO_2	0.71	0.78
CN	0.61	0.70
CF_3	0.43	0.54
CH_3COO	0.39	0.31
Br	0.39	0.23
CH_3CO	0.38	0.48
$CO_2C_2H_5$	0.37	0.45
Cl	0.37	0.22
CHO	0.36	0.44
CO_2H	0.35	0.44
I	0.35	0.28
F	0.34	0.06
$C\equiv CH$	0.20	0.23
SCH_3	0.15	≈ 0
OH	0.13	-0.38
OCH_3	0.11	-0.28
C_6H_5	0.05	≈ 0
H	0	0
CH_3	-0.06	-0.17
C_2H_5	-0.07	-0.15
$CH(CH_3)_2$	-0.07	-0.15
$C(CH_3)_3$	-0.10	-0.20
$N(CH_3)_2$	-0.15	-0.63
NH_2	-0.16	-0.57

Effetto di risonanza (π-p o π-π):

+R donazione elettronica verso il centro di reazione

($-OCH_3$, $-OH$, $-NR_2$)

-R richiamo elettronico centro di reazione

($-NO_2$, $-CN$, $-COCH_3$)

orto, para > meta

Effetto induttivo

(polarizzazione ed effetto di campo)

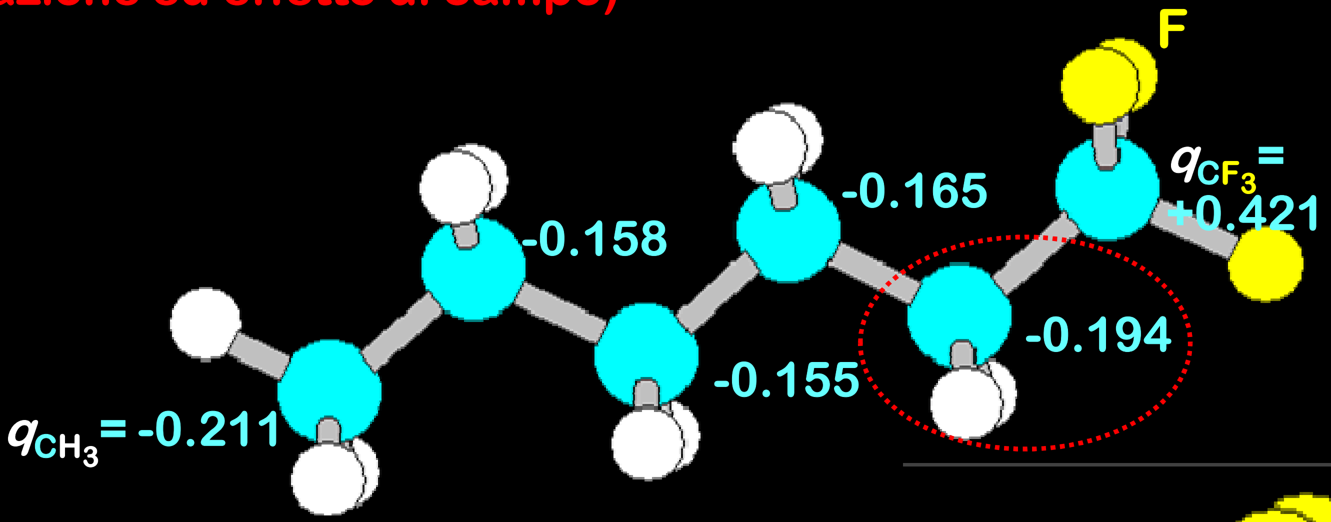
+I donazione elettronica verso il centro di reazione

-I richiamo elettronico centro di reazione

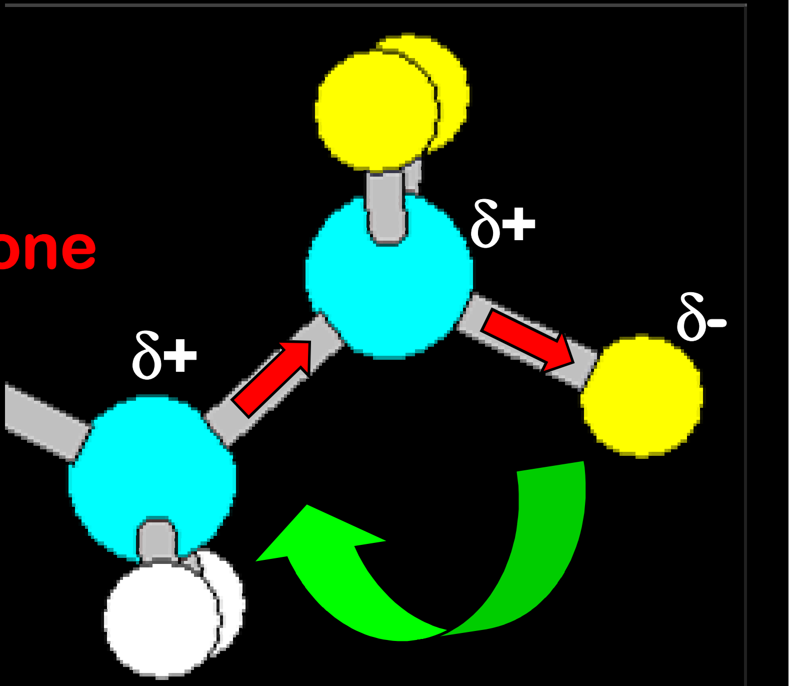
orto > meta >para

Effetto induttivo

(polarizzazione ed effetto di campo)

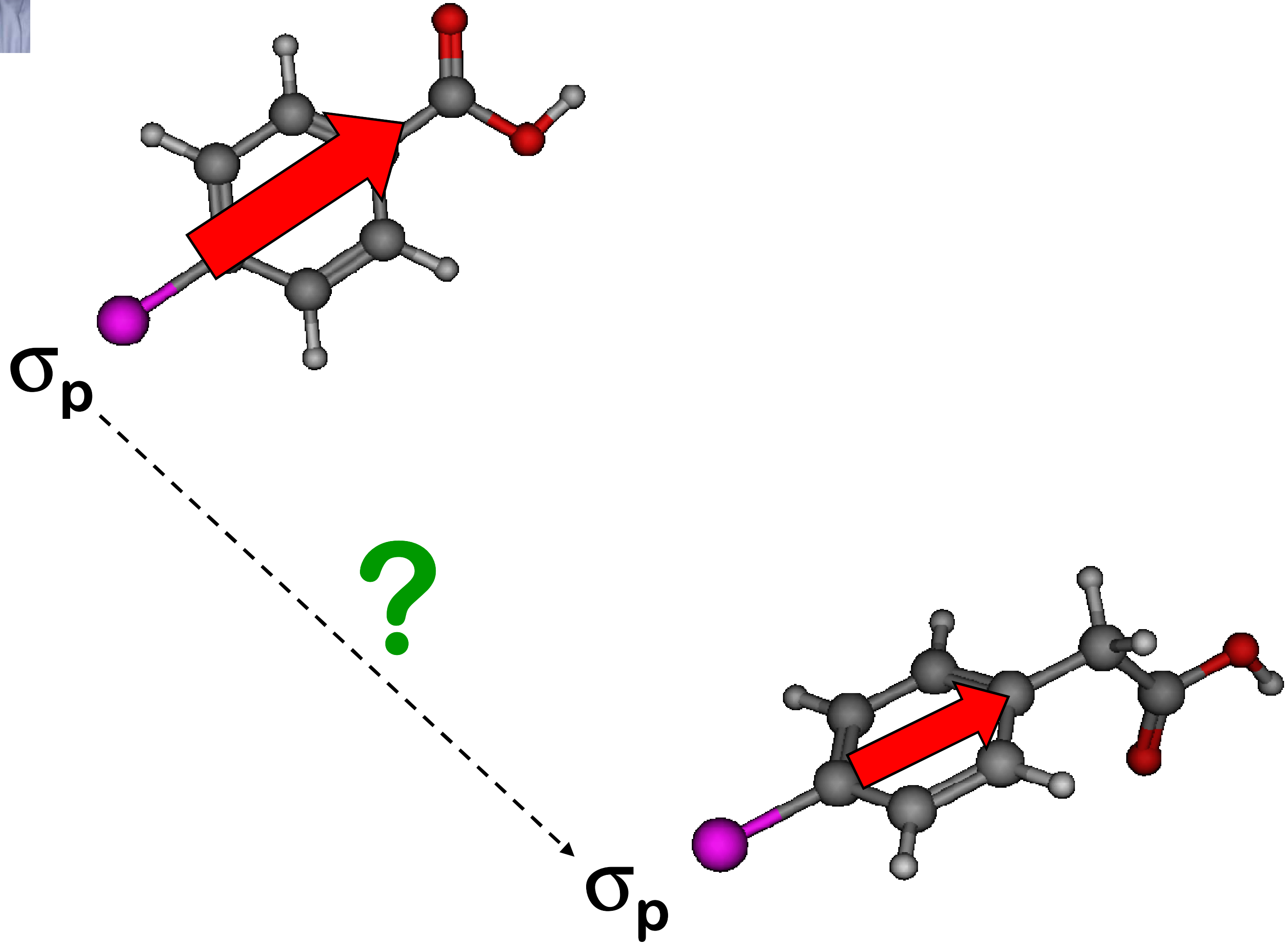


➡ Effetto di polarizzazione

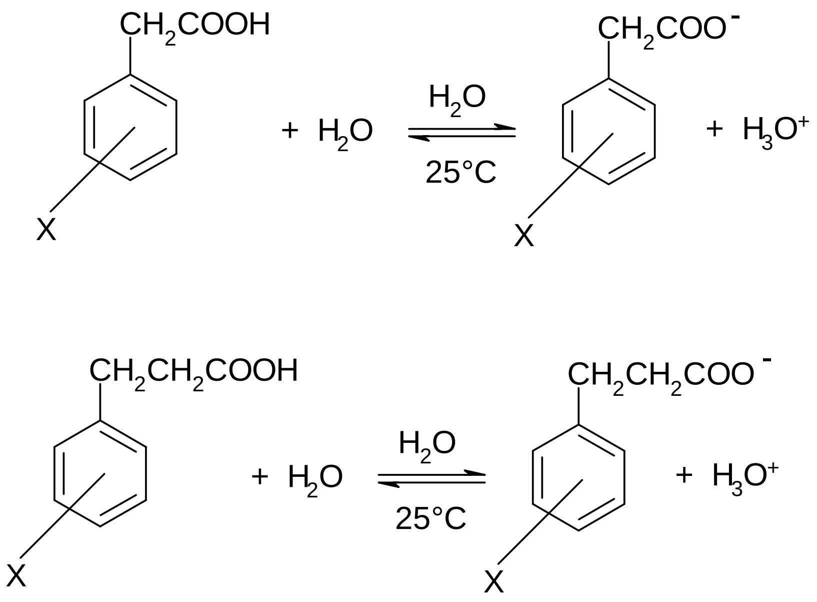


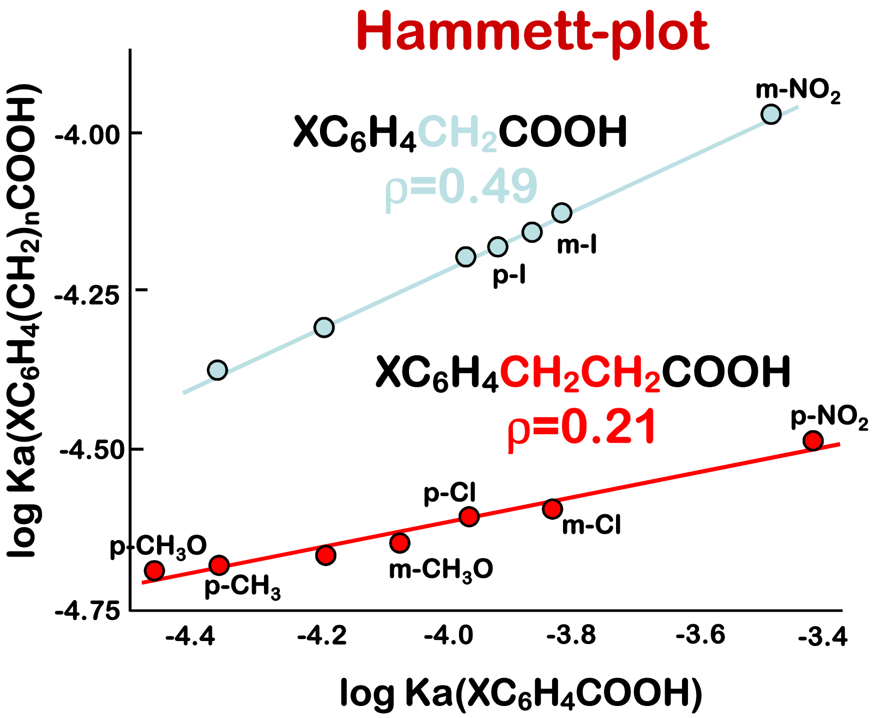
➡ Effetto di campo

ma sigma (σ) è davvero una proprietà invariante del sostituente?



Applicazione della costante σ: dissociazione di acidi *omo*-benzoici




Hammett-plot

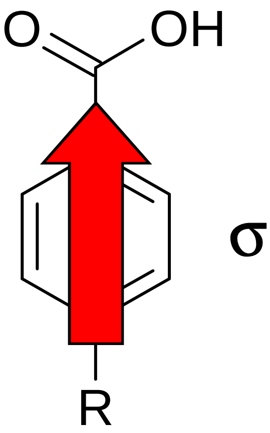
$$\log \frac{Ka(XC_6H_4CH_2COOH)}{Ka(C_6H_5CH_2COOH)} = \rho \; \log \frac{Ka(XC_6H_4COOH)}{Ka(C_6H_5COOH)}$$

$$\log \frac{Ka(XC_6H_4CH_2COOH)}{Ka(C_6H_5CH_2COOH)} = \rho \; \sigma_X$$

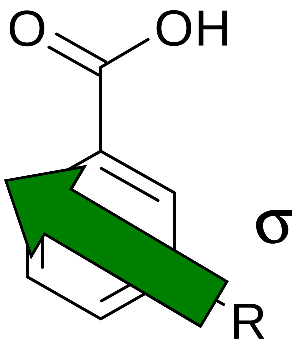
Effetto elettronico del sostituente e costante σ:

Ma Hammett, di quale effetto elettronico stava disquisendo mentre definiva la sua costante σ?

$$\sigma_x = \log \frac{Ka(XC_6H_4COOH)}{Ka(C_6H_5COOH)}$$



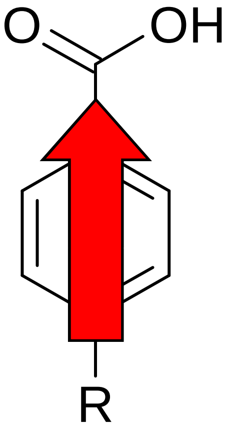
$\sigma_p = \sigma_I + \sigma_R$



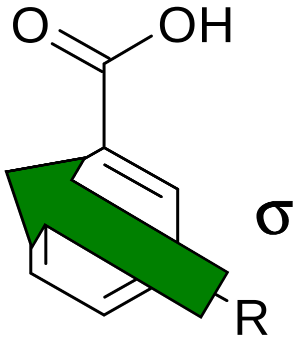
$\sigma_m = \sigma_I + \alpha\sigma_R$

Effetto elettronico del sostituente e costante σ:

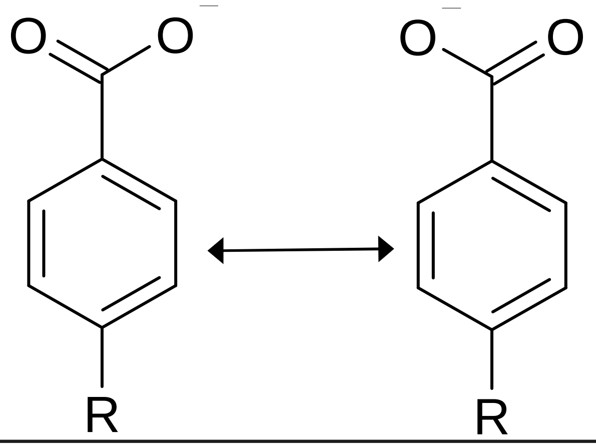
Ma è vera risonanza quella pesata dalla costante σ di Hammett?



$\sigma_p = \sigma_I + \sigma_R$

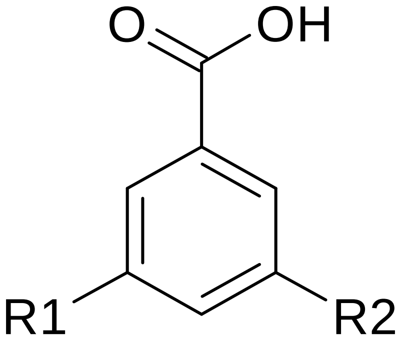


$\sigma_m = \sigma_I + \alpha\sigma_R$



Virtualizzazione del calcolo delle costanti di dissociazione acido-base:

Regola dell'additività:

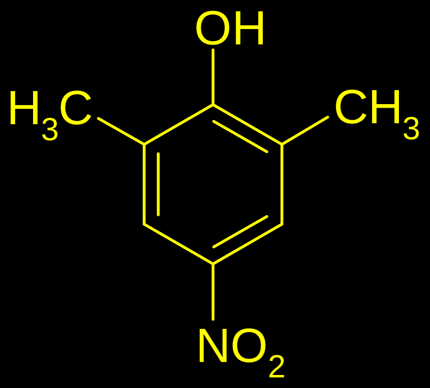


$$C \log Ka = \log Ka_{rif} + \Sigma \sigma_{Ri}$$

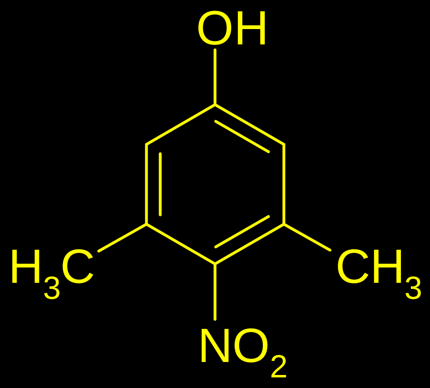
Regola della costitutività:

$$C \log Ka = \log Ka_{rif} + \Sigma \sigma_{Ri} + \Sigma \Delta\sigma_{Ri}$$

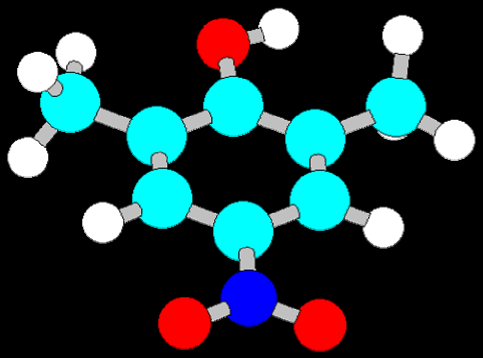
Inibizione Sterica della Risonanza

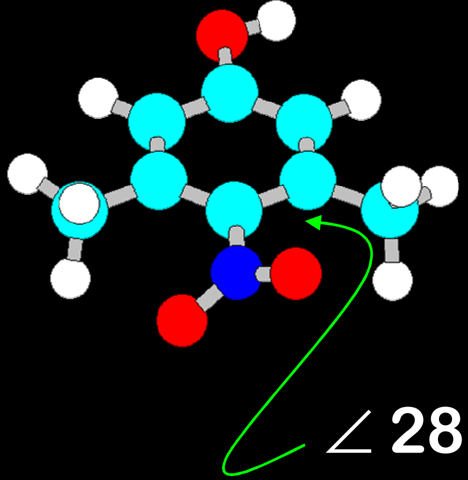


$pK_a = 7.16$



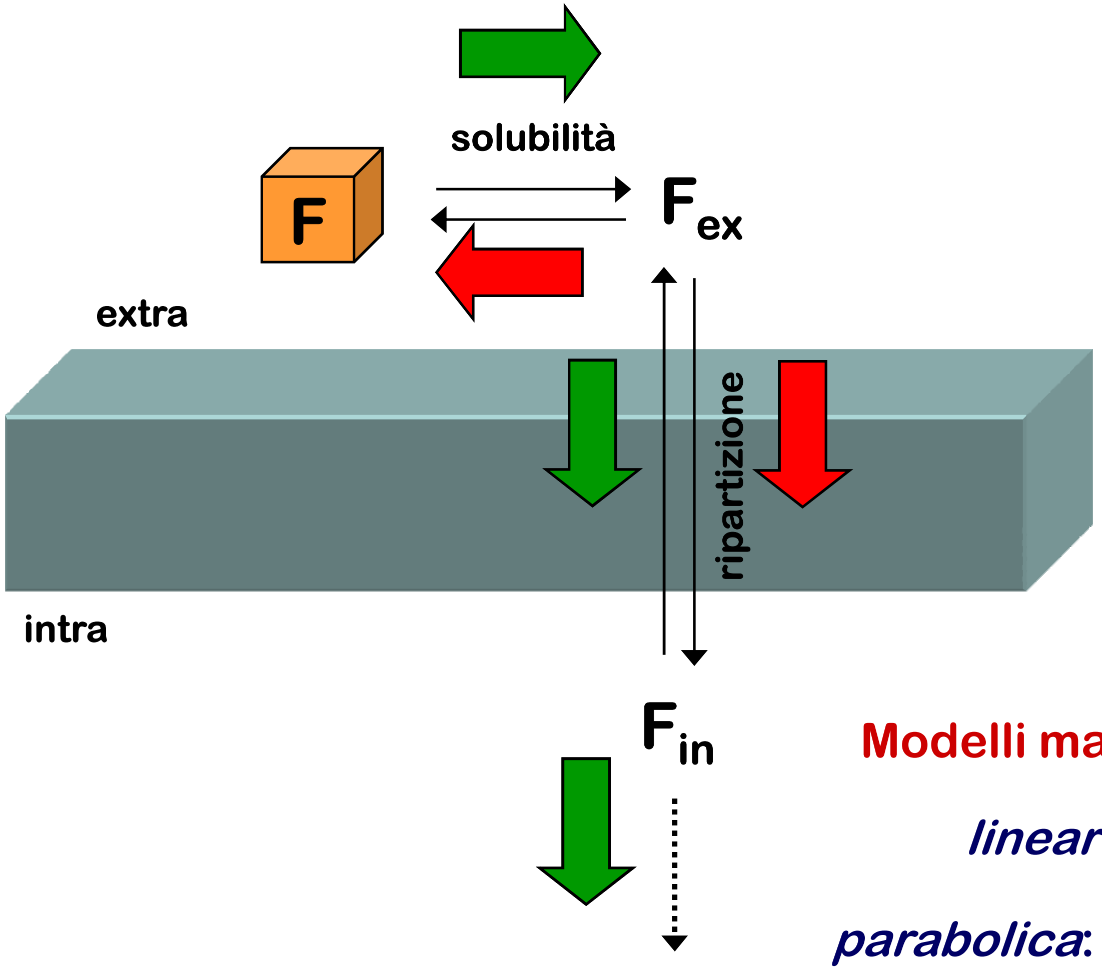
$pK_a = 8.24$

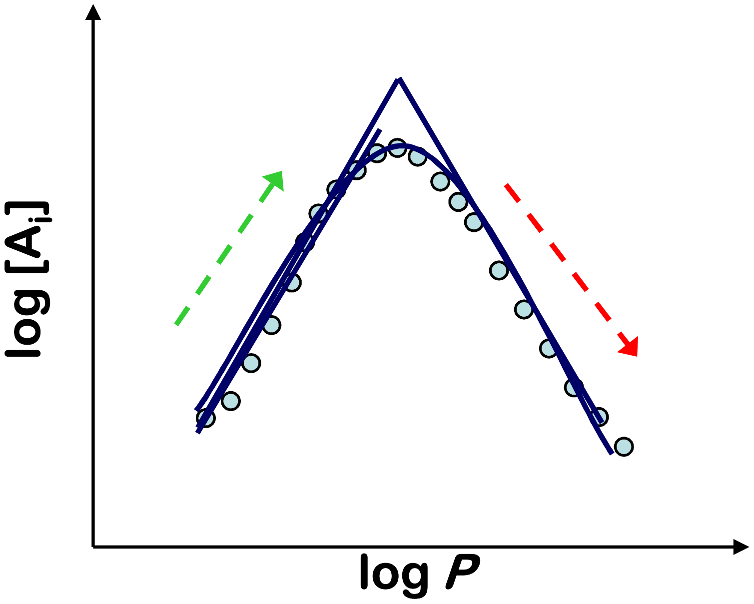




∠ 28.5°

Farmacocinetica e logP (o ClogP!!)





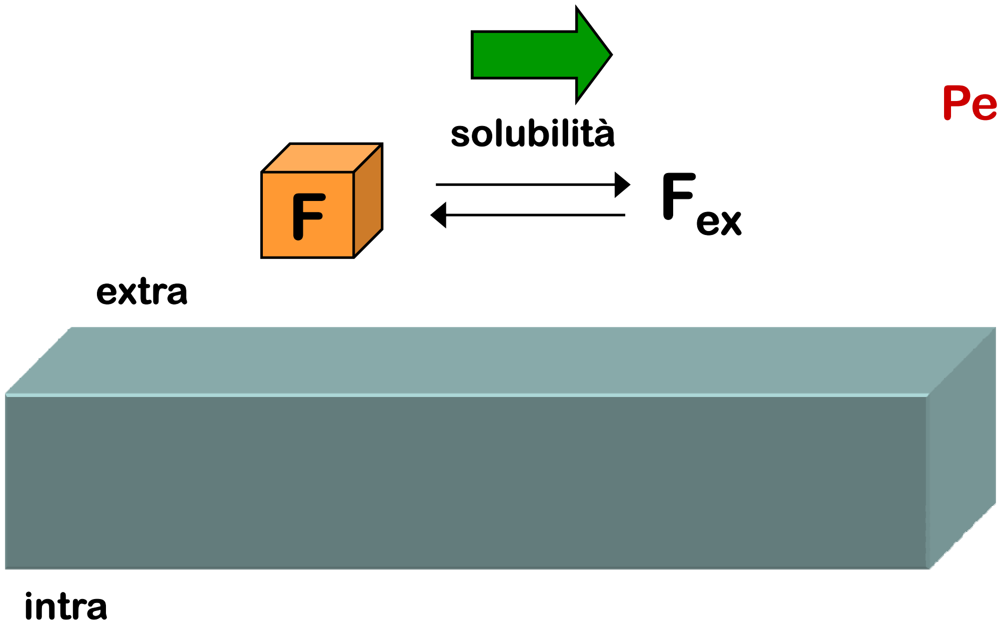
Modelli matematici:

lineare: $\log [A_i] = a \log P + c$

parabolica: $\log [A_i] = -a (\log P)^2 + b \log P + c$

bilineare: $\log [A_i] = a \log P - b \log (\beta P + 1) + c$

logP (o ClogP!!) e… solubilità!



Per composti liquidi immiscibili in acqua:

la solubilità in acqua è direttamente correlata al log P.

Per composti solidi :

la solubilità in acqua dipende anche dall'energia richiesta per disgregare la struttura cristallina o amorfa.

$$\log P = -0.89 (\log S) -0.015\, MP + 6.5$$

dove "S" è la solubilità in acqua espressa in μmoli/lt ed "MP" è il punto di fusione del composto solido. Di conseguenza, ci possono essere composti con più alti valori di logP che sono comunque più solubili in virtù del loro basso punto di fusione e viceversa. Così come, il medisimo composto (con il medesimo log P) può avere solubilità diverse in funzione di un diverso stato di aggregazione in fase solida (amorfo vs cristallino, polimorfismi….).

Bannerjee, Yalkowsky and Valvoni (1980) *Envir.Sci.Tech*,14,1227