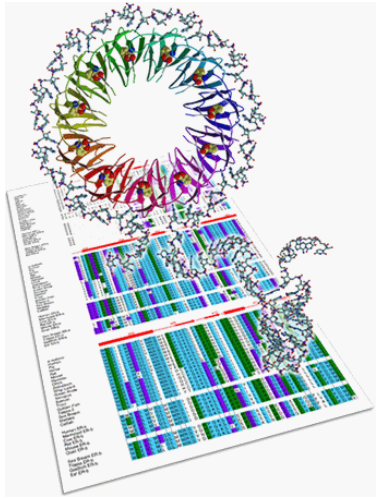
# Escuela Complutense de Verano
# Especialista en Bioinformática



## Interacciones entre proteínas y pequeños ligandos (I)

Federico Gago
Departamento de Farmacología
Universidad de Alcalá, Madrid

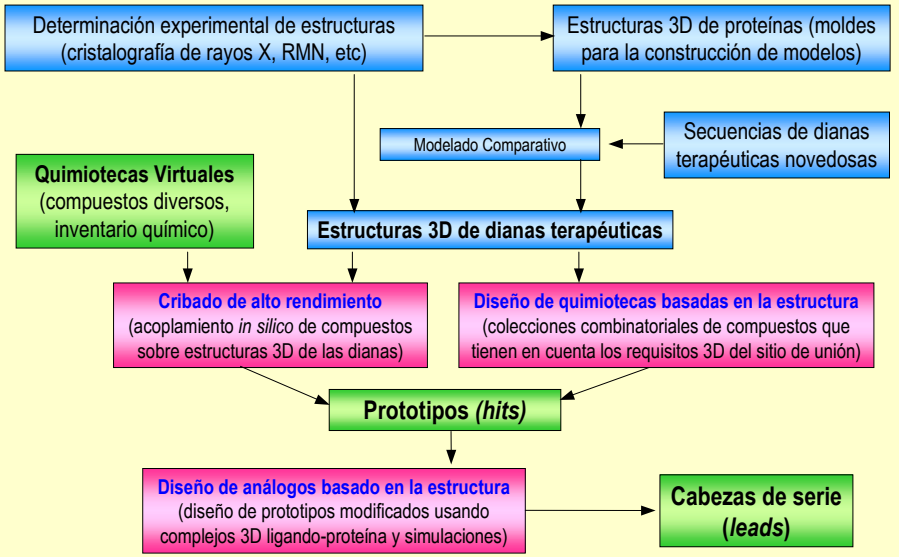
## Interacciones entre proteínas y moléculas pequeñas

1. Concepto de ligando y sitio de unión. Ejemplos.
2. Bases de datos estructurales y programas asociados.
3. Caracterización estructural de moléculas pequeñas y sus complejos con proteínas.
4. Acoplamiento ligando-receptor ("docking"): algoritmos y programas.
5. Cribado virtual.

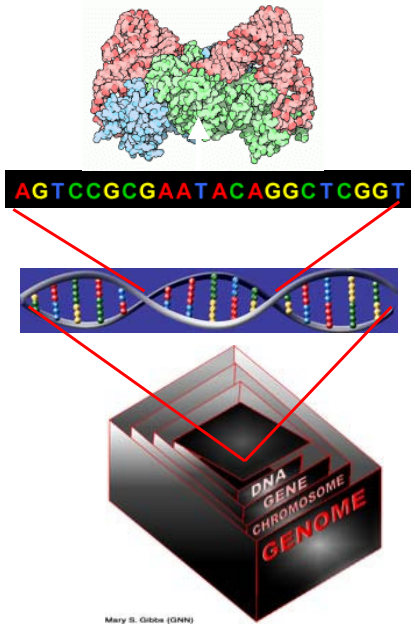
---

6. Relaciones estructura-actividad: QSAR y 3D-QSAR.
7. Diseño de nuevos ligandos.

## Integración de Metodologías Informáticas y Basadas en la Estructura para el Descubrimiento de Fármacos




## Gene expression = Protein production



# How Proteins Work

Proteins recognize and reversibly bind to other molecules: *cofactors, substrates, inhibitors*… Also *ions* and other *proteins*.

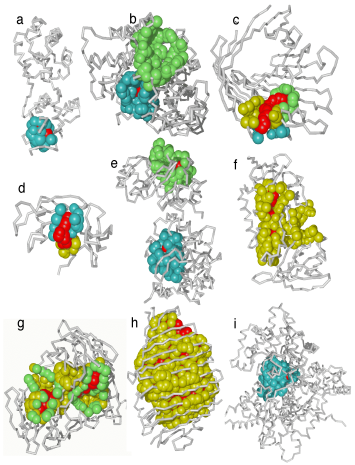
The bound small molecule is called a **ligand**.

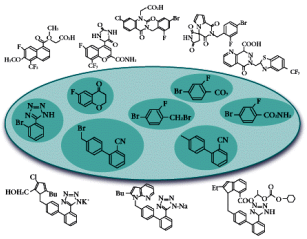
In the case of enzymes, the region that associates with substrates and products is called the **active site**.

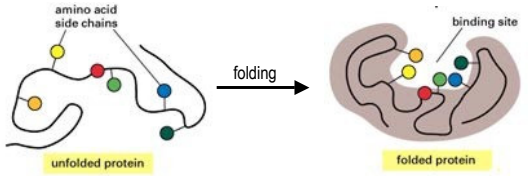
The region of a protein that associates with activator or inhibitor molecules is called an **allosteric site**.

Proteins can have > 1 binding site for different ligands

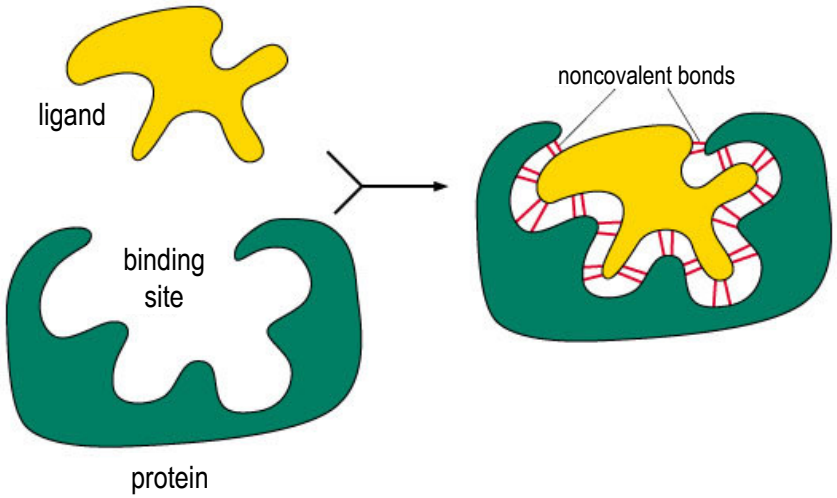
Proteins fold in such a way that they create specific sites that are the right *size*, *shape*, and *polarity* for their ligands.



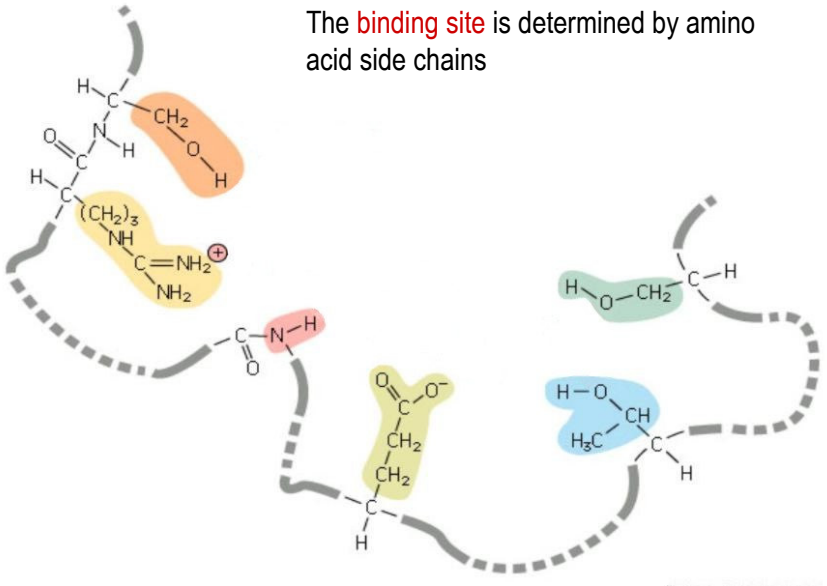





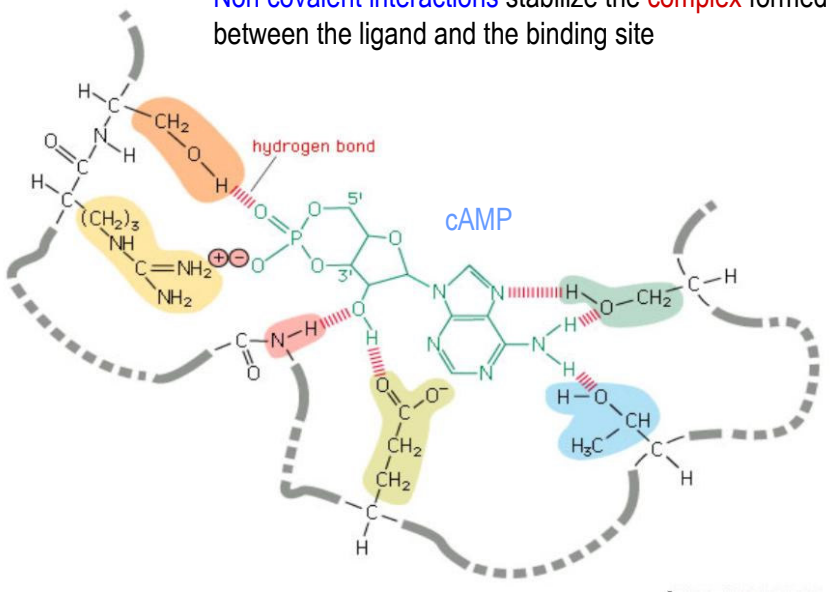

## Ligand binding is highly selective




The binding site is determined by amino acid side chains



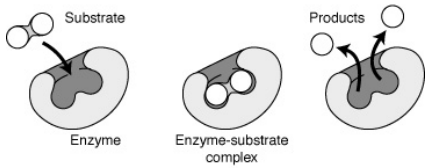
Non covalent interactions stabilize the complex formed between the ligand and the binding site




Sitios de unión: complementariedad de forma

... y complementariedad electrostática

## Enzymes: A special case of protein function

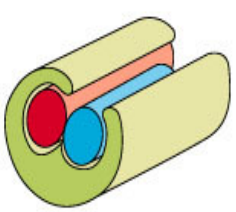



Enzymes bind and assist in the chemical transformation of other molecules

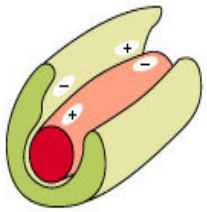
**Substrate**: molecule acted upon by an enzyme (analogous to ligand)

**Catalytic site**: substrate-binding site (analogous to ligand-binding site)

## How do enzymes catalyze reactions?



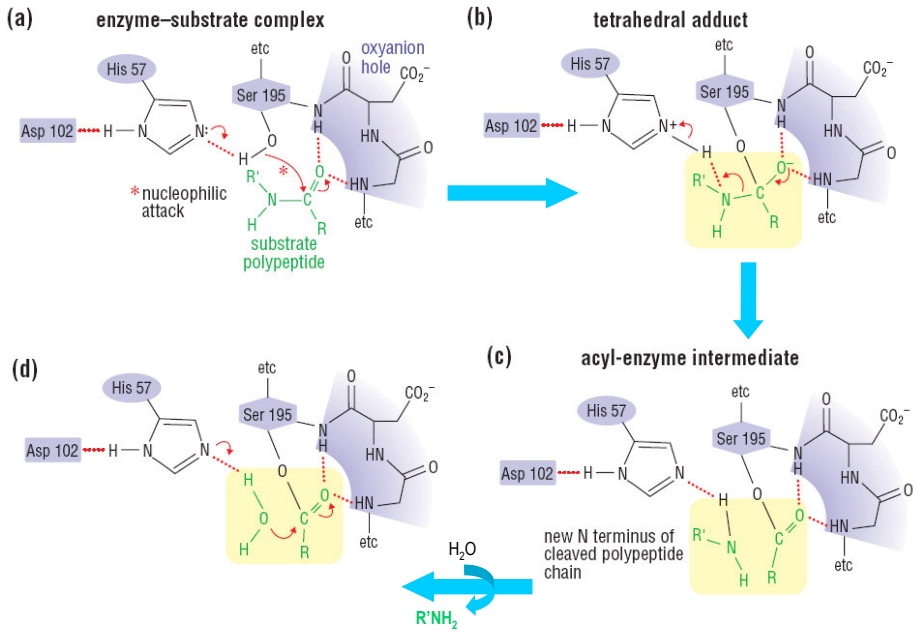
(A) enzyme binds to two substrate molecules and orients them precisely to encourage a reaction to occur between them

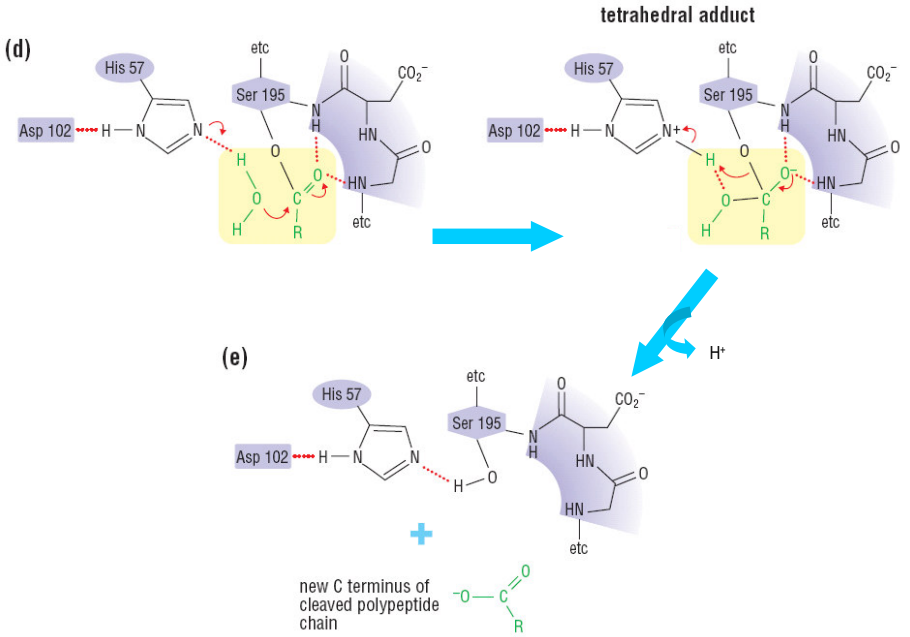


(B) binding of substrate to enzyme rearranges electrons in the substrate, creating partial negative and positive charges that favor a reaction

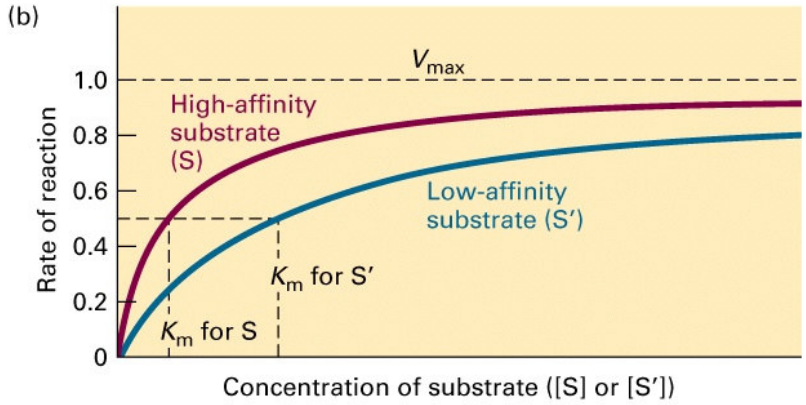
(C) enzyme strains the bound substrate molecule, forcing it toward a transition state to favor a reaction

The chemical steps in peptide hydrolysis catalyzed by the serine protease chymotrypsin

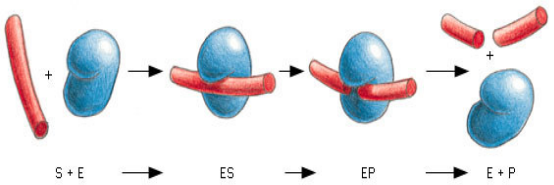


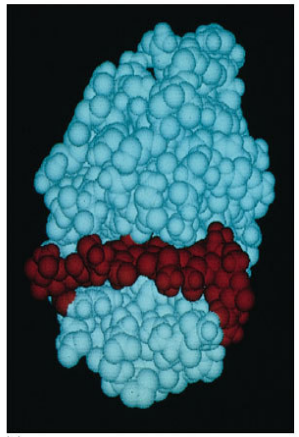




Kinetics of an enzymatic reaction are described by $V_{max}$ and $K_m$

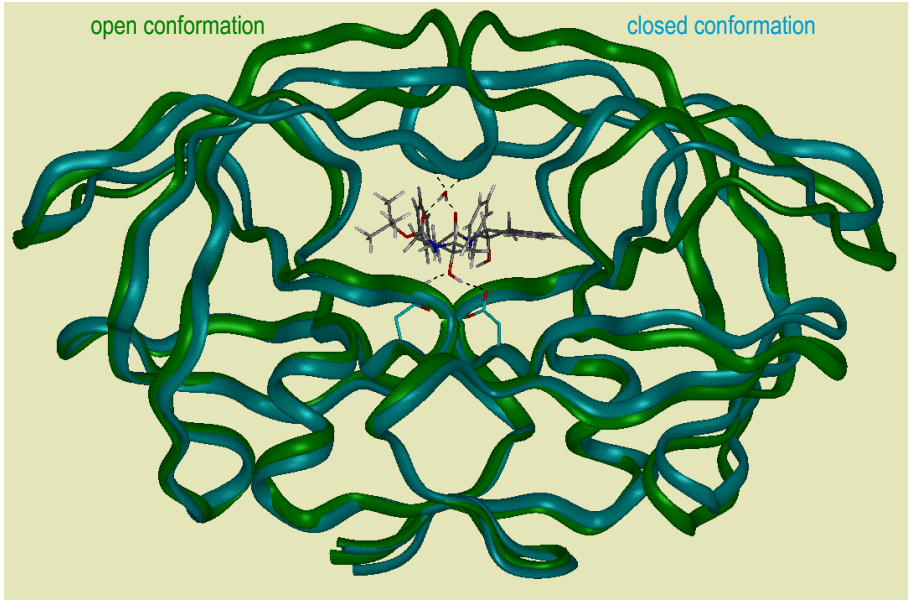



Lysozyme catalyzes the cutting of a polysaccharide chain

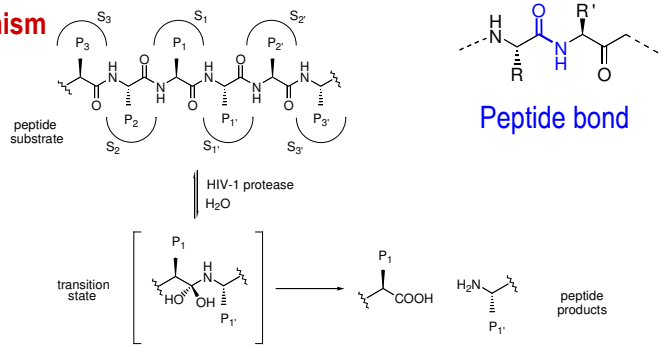


# HIV-1 protease

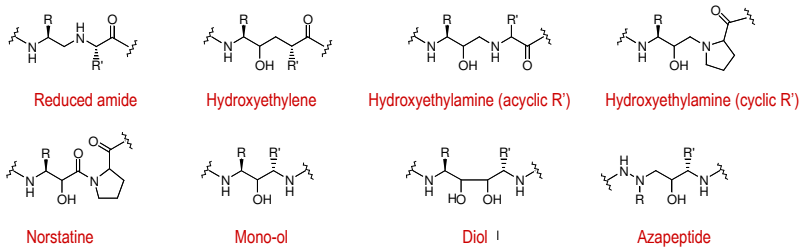



## Hydrolytic mechanism in HIV-1 protease




Peptide bond

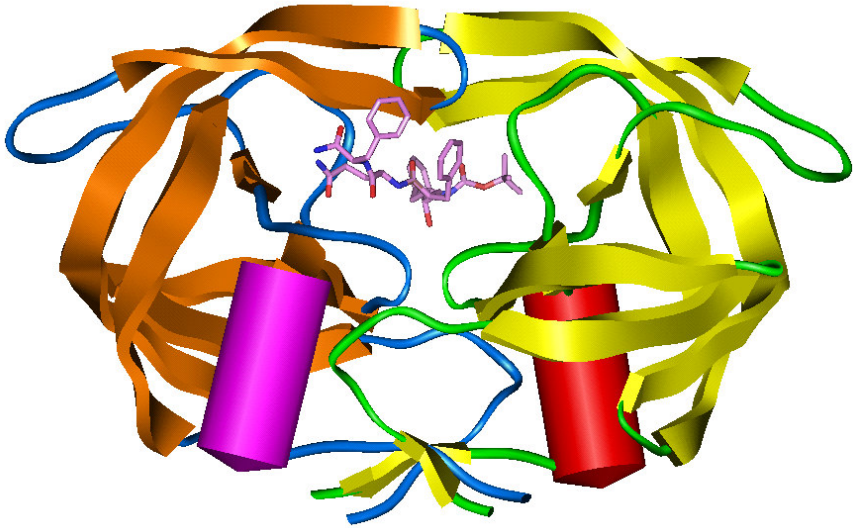
Examples of non-hydrolyzable isosteres of the peptide bond cleaved by HIV-1 protease



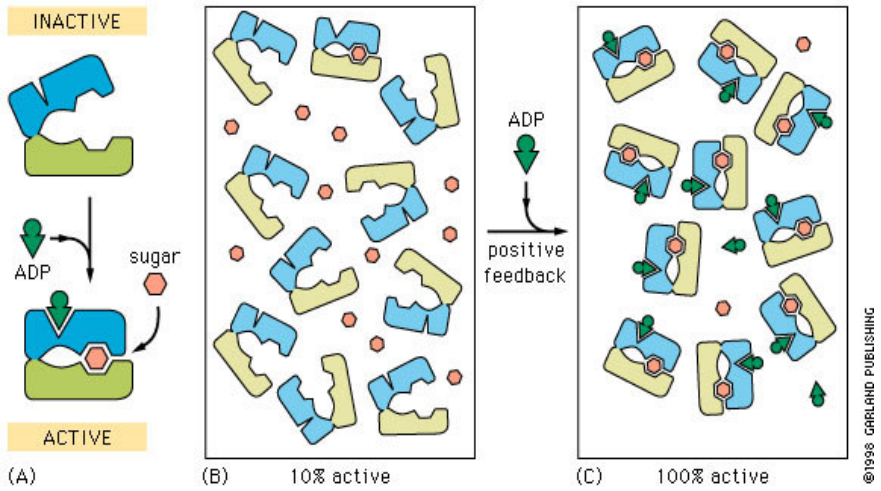
Reduced amide | Hydroxyethylene | Hydroxyethylamine (acyclic R') | Hydroxyethylamine (cyclic R')

Norstatine | Mono-ol | Diol | Azapeptide

# HIV-1 protease in complex with inhibitor QF-34

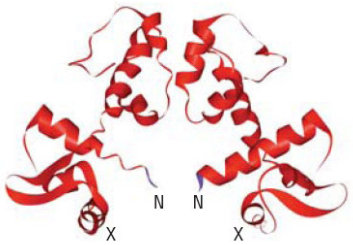


# Enzyme activation caused by an allosteric change

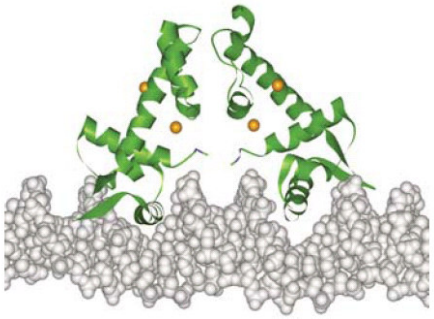


## Binding of gene regulatory proteins to DNA is often controlled by ligand-induced conformational changes

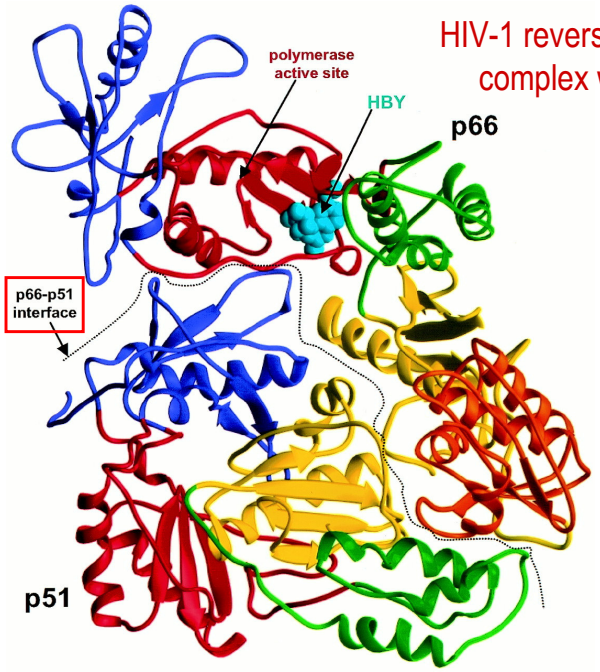
Iron binding regulates the repressor of the diphtheria toxin gene:




aporepressor DtxR (PDB 1dpr)



ternary complex of **repressor** (green), metal ion ($Fe^{2+}$) and DNA (PDB 1fst).

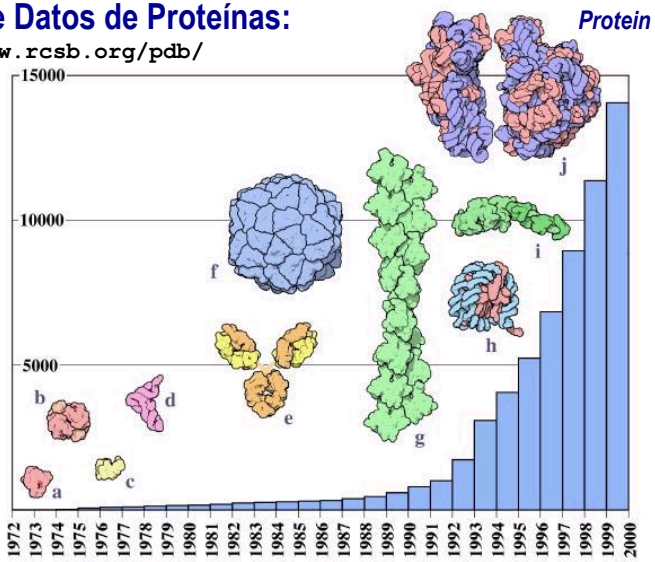



HIV-1 reverse transcriptase in complex with inhibitor HBY

## Banco de Datos de Proteínas:

*Protein Data Bank*

http://www.rcsb.org/pdb/




(a) mioglobina (b) hemoglobina (c) lisozima (d) ARN de transferencia
(e) anticuerpos (f) virus (g) actina (h) el nucleosoma
(i) miosina (j) el ribosoma

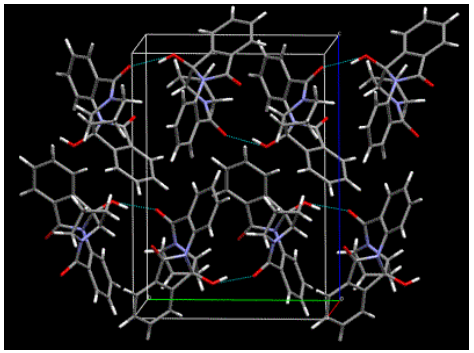
Cortesía de David Goodsell, TSRI

## Cambridge Structural Database

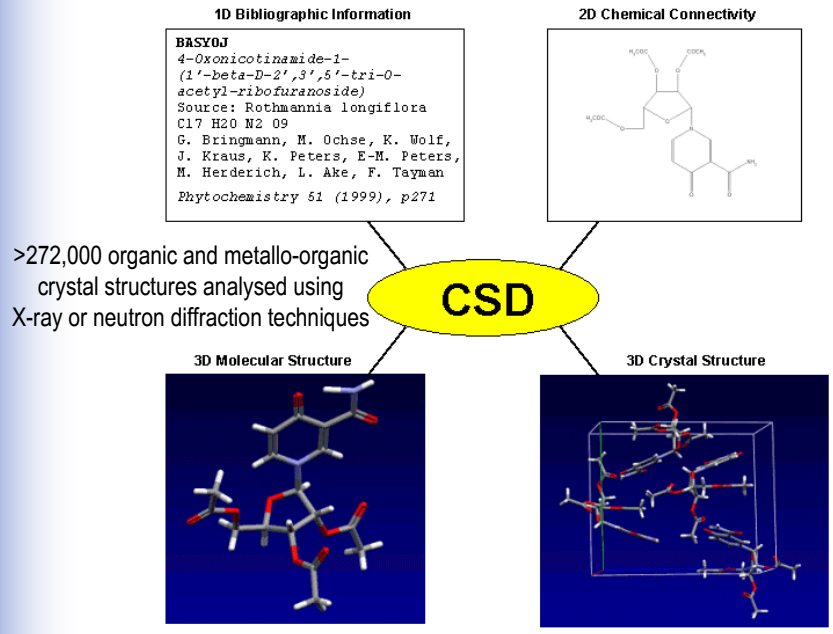
The Cambridge Crystallographic Data Centre (CCDC) builds, maintains and distributes the Cambridge Structural Database (CSD), a searchable database of organic and metallo-organic crystal structures.

The CCDC also produce and distribute software products which make use of the data contained in the CSD.

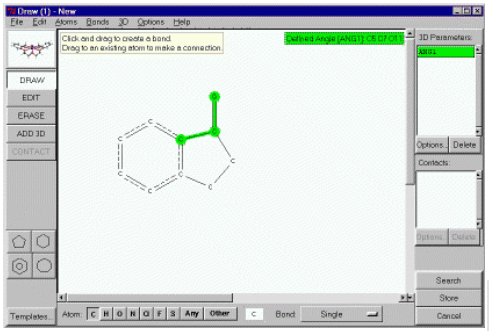
# Increasing the Value of Crystallographic Databases



- Derived knowledge bases
- Knowledge-based applications programs
- Data mining tools for protein-ligand complexes

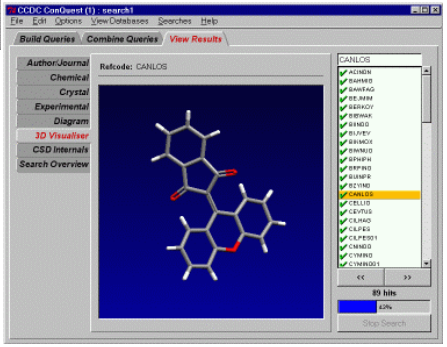



## ConQuest



**ConQuest provides a full range of text/numeric database search options, in addition to more complex search functionality, including:**

- **Chemical substructure searching**
- **Geometrical searching**
- **Intermolecular non-bonded contact searching**

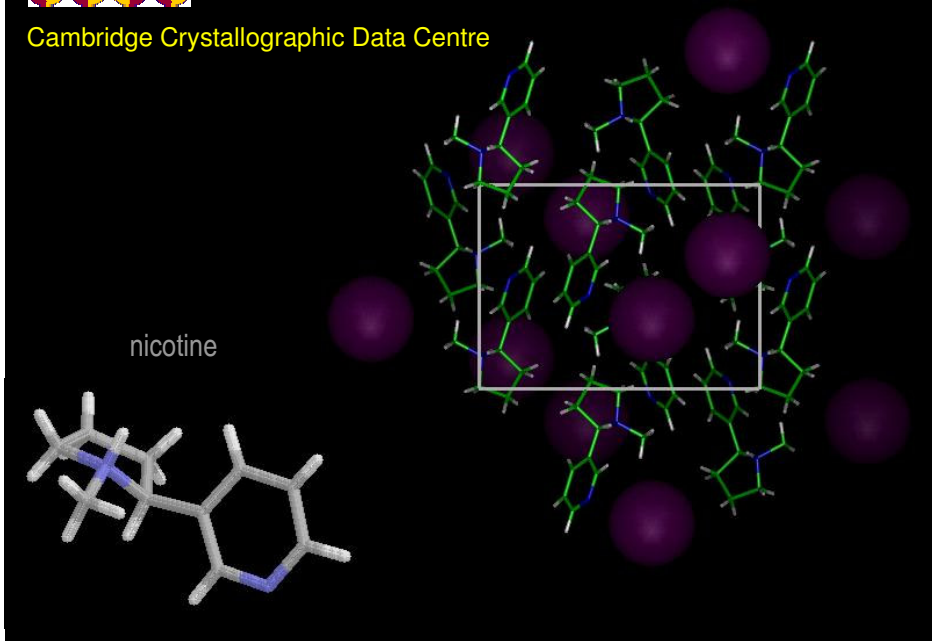


## Cambridge Crystallographic Data Centre

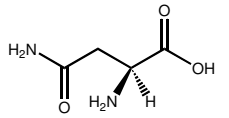
**http://www.ccdc.cam.ac.uk/**

```
#DOXSIS  33870428        16  9  0  0  0  4  4 28  0  0 3013220000001000000000086
137C
R=0.
211
C   6
I1
C2
C5
C8
C10
H3
H6
H9
H12
H15
 0  3
HEADER    CSD ENTRY DOXSIS
COMPND    NICOTINE MONOHYDROGEN IODIDE
CRYST    13.705    9.974    8.436  90.00  90.00  90.00 P212121
ATOM      1  I1  NICO    1        1.583   1.561   0.355  1.00  0.00
ATOM      2  N1  NICO    1        5.355  -5.377  10.252  1.00  0.00
ATOM      3  C1  NICO    1        5.744  -4.111  10.057  1.00  0.00
ATOM      4  C2  NICO    1        5.375  -3.343   8.953  1.00  0.00
ATOM      5  C3  NICO    1        4.573  -3.927   7.982  1.00  0.00
ATOM      6  C4  NICO    1        4.180  -5.249   8.151  1.00  0.00
ATOM      7  C5  NICO    1        4.584  -5.922   9.302  1.00  0.00
ATOM      8  C6  NICO    1        5.840  -1.905   8.899  1.00  0.00
ATOM      9  C7  NICO    1        4.752  -0.841   8.748  1.00  0.00
ATOM     10  C8  NICO    1        5.461   0.374   8.127  1.00  0.00
ATOM     11  C9  NICO    1        6.826  -0.125   7.655  1.00  0.00
ATOM     12  N2  NICO    1        6.722  -1.627   7.699  1.00  0.00
....
CONECT    1    0
CONECT    2    3    7
CONECT    3    2    4   14
....
MASTER        0    0    0    0    0    0    0    0   28    0   28    0
END
9220
5620
3700
1260
4940
4730
3960
7280
2420
```

Cambridge Crystallographic Data Centre

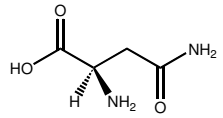



# The Importance of Chirality

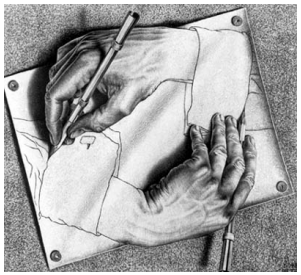


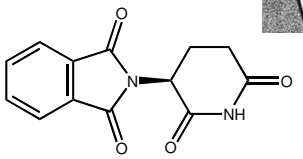

Bitter

***Asparagine***



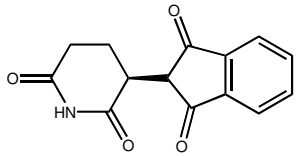

Sweet






Extreme teratogen

***Thalidomide***

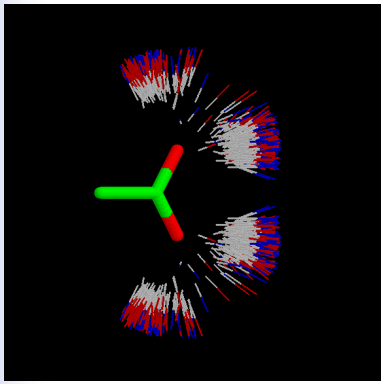


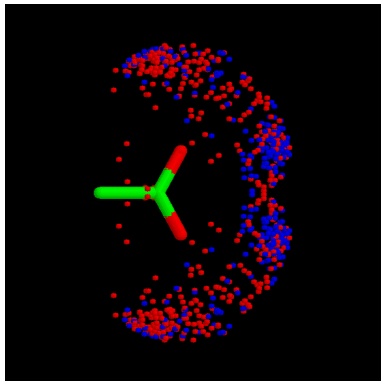

Anti-morning sickness

IsoStar and SUPERSTAR

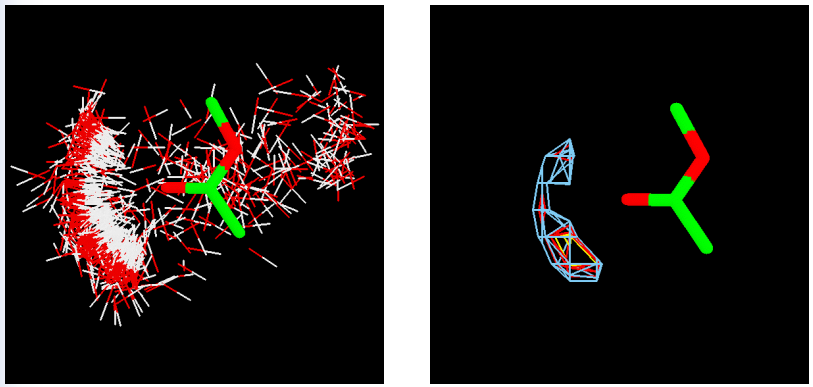
- IsoStar - knowledge base of information about intermolecular interactions
- SuperStar - program for predicting binding points in an enzyme active site
- SuperStar predictions based solely on IsoStar data

# IsoStar Scatterplots





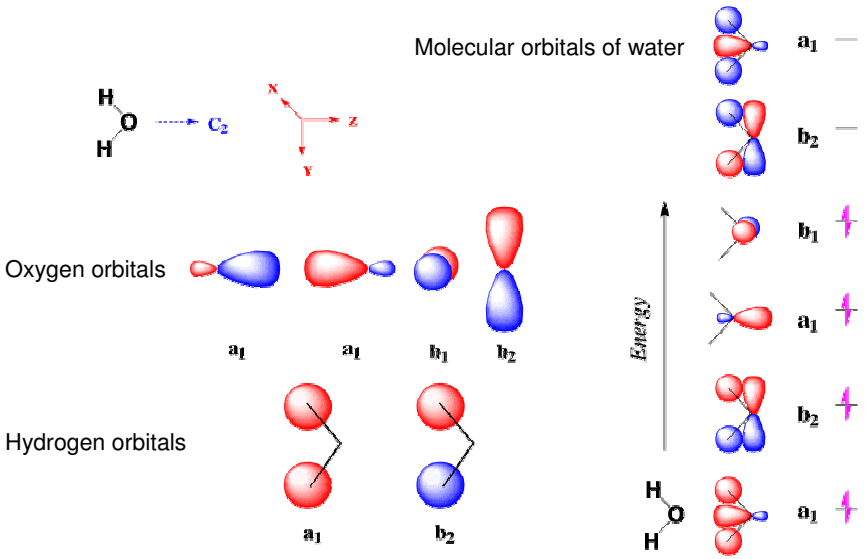
# IsoStar Density Surfaces



A probability surface derived from the observed positions of hydrogen-bonding hydrogen atoms around aliphatic esters.

## Quantum Chemistry

Atomic orbitals can be combined to give molecular orbitals




## *Ab initio* METHODS

* Hartree-Fock method
* Electron correlation methods
  - variational methods

    Configuration Interaction with double excitations (CID)

    Configuration Interaction with single and double excitations (CISD)
  - perturbation methods

    Møller and Plesset (MP2, MP3, MP4)

    Quadratic Convergence CI method (QCISD)
  - density functional methods (DFT)

    BP86 - developed by Becke and Perdew in 1986

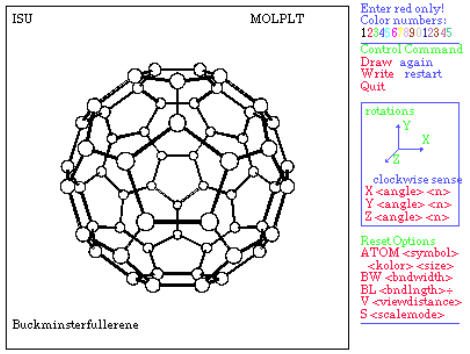
    BLYP - developed by Becke, Lee, Yang and Parr

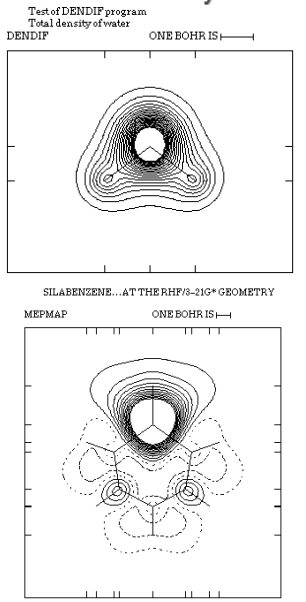
    B3LYP - a modification of BLYP in which a 3-parameter functional developed by Axel Becke is used.

## GAMESS

General Atomic and Molecular Electronic Structure System

http://www.msg.ameslab.gov/GAMESS/GAMESS.html



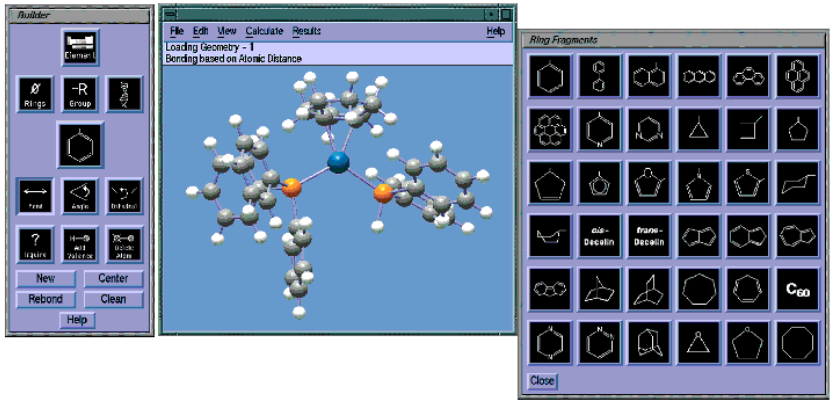
Gaussian

http://www.gaussian.com/

wavefun.com
chemistry software tools
for education and industry

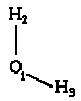
Spartan

http://www.wavefun.com/




# Some sample Gaussian z-matrices

## Water ($C_{2v}$)




With **variables**:

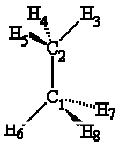
```
o
h 1 11
h 1 11 2 a1

11 0.96
a1 104.0
```

With **values**:

```
o
h 1 0.96
h 1 0.96 2 104.0
```

## Ethane ($D_{3d}$)




```
c
c 1 11
h 2 12 1 a1
h 2 12 1 a1 3  120.0
h 2 12 1 a1 3 -120.0
h 1 12 2 a1 3  180.0
h 1 12 2 a1 6  120.0
h 1 12 2 a1 6 -120.0

11 1.54
12 1.09
a1 110.0
```

# **SEMI-EMPIRICAL METHODS**: levels of approximation

**CNDO** Complete Neglect of Differential Overlap (Developed by John Pople - assumes atomic orbitals to be spherical when evaluating the two-electron integrals)

**INDO** Intermediate Neglect of Differential Overlap

**NDDO** Neglect of Diatomic Differential Overlap

**MINDO/3** Modified INDO (Developed by Michael Dewar - uses a set of parameters to approximate the two-electron repulsion integrals)

**ZINDO** Includes parameters for transition metals

**MNDO** Modified NDO (Developed by Michael Dewar and Walter Thiel in 1977)

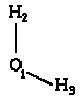
**AM1** Austin Model 1 (Developed by Michael Dewar and Andrew Holder in 1986)

**PM3** Parametric Model 3 (Developed by Jimmy Stewart in 1988)

# Sample input for MOPAC

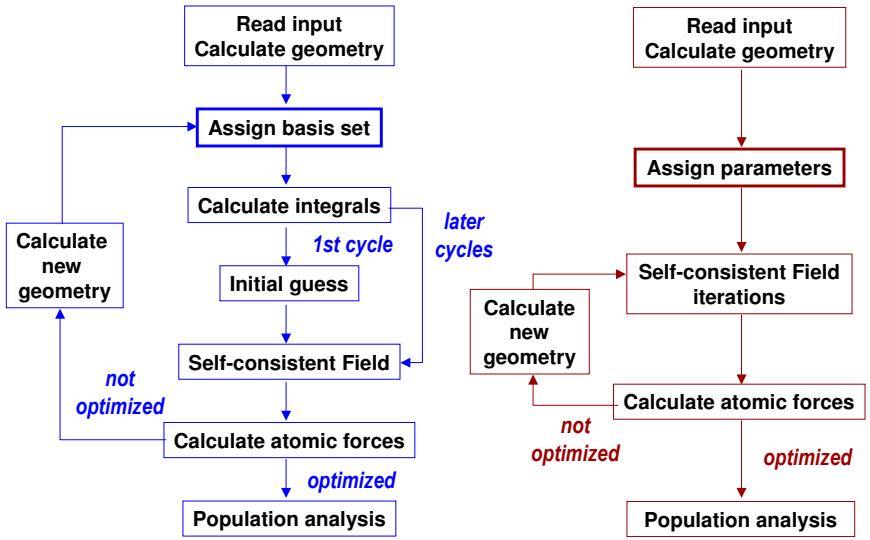
```
PM3 EF PRECISE
H2O (water)
MOPAC input as a Z-matrix
O
H          0.96000  1                                    1
H          0.96000  1      104.00000  1                 1 2
```

## Water ($C_{2v}$)




```
AM1 EF PRECISE
H2O (water)
MOPAC input in Cartesian coordinates
O         0.0000  0  0.0000  0  0.0000  0
H         0.9600  1  0.0000  0  0.0000  0
H        -0.2322  1  0.9315  1  0.0000  0
```

## Typical flow charts for **an *ab initio* optimization** and a corresponding **semi-empirical calculation**



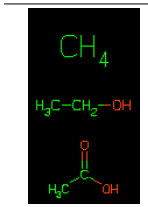
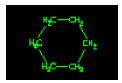

# SMILES

## Simplified Molecular Input Line Entry Specification

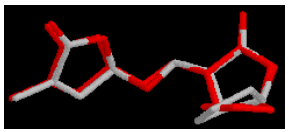
### Rules

1. Atoms are represented by atomic symbols: B, C, N, O, F, P, S, Cl, Br, and I.
2. Double bonds are `=', triple bonds are `#'.
3. Branching is indicated by parentheses.
4. Ring closures are indicated by pairs of matching digits.

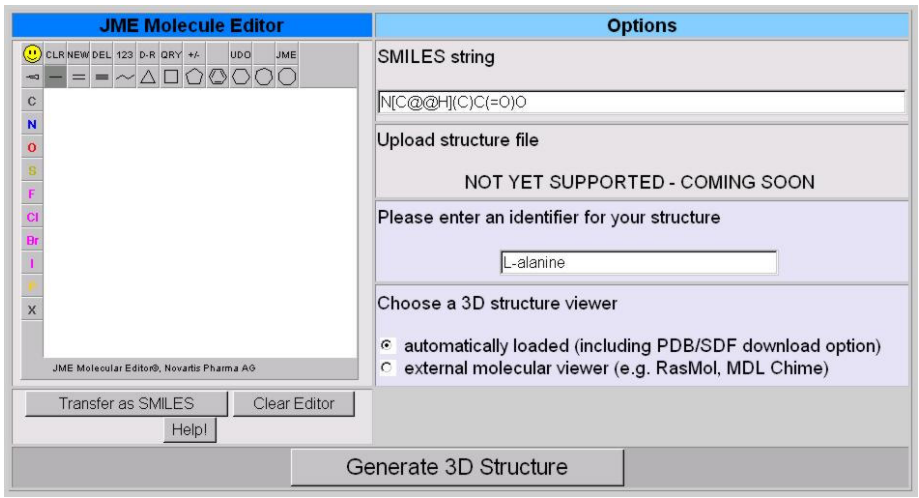
### Examples

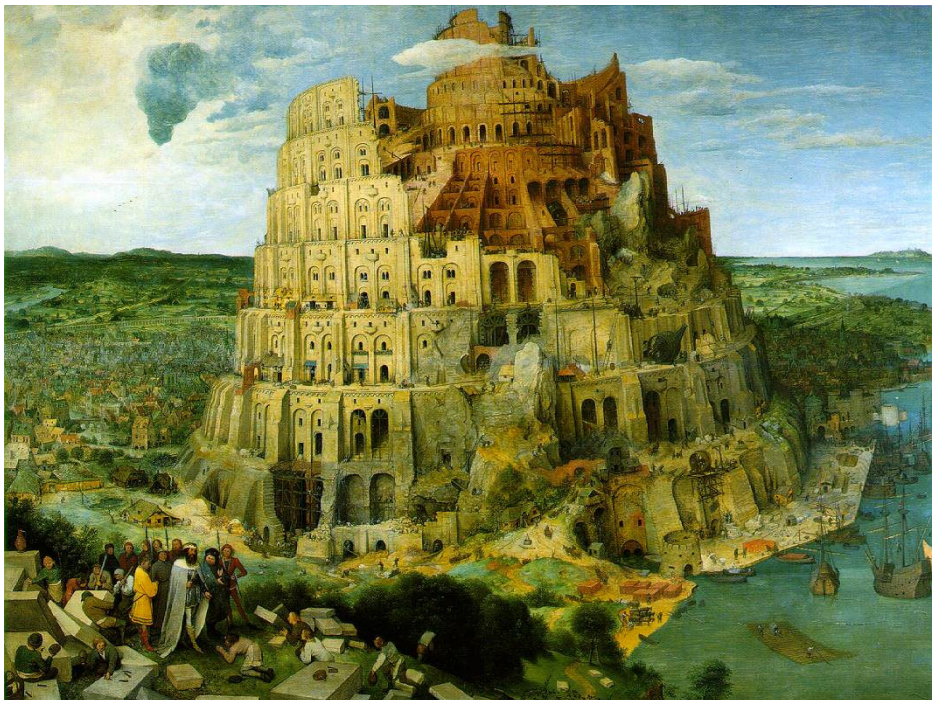
| Depiction | SSMILES | Name | Remark |
|---|---|---|---|
|  | C | methane | hydrogens fill normal valence |
| | CCO | ethanol | a single bond is assumed to join adjacent atoms |
| | CC(=O)O | acetic acid | parentheses are used to indicate branching |
|  | C1CCCCC1 | cyclohexane | bonds can also be represented by pairs of matching digits |

## Automatic generation of three-dimensional atomic COoRdINAtes



http://www2.chemie.uni-erlangen.de/software/corina/free_struct.html



## BABEL A program designed to interconvert a number of file formats currently used in molecular modelling

**Input type codes:**

```
alc -- Alchemy file
prep -- AMBER PREP file
bs -- Ball and Stick file
bgf -- MSI BGF file
car -- Biosym .CAR file
boog -- Boogie file
caccrt -- Cacao Cartesian file
cadpac -- Cambridge CADPAC file
charmm -- CHARMm file
c3d1 -- Chem3D Cartesian 1 file
c3d2 -- Chem3D Cartesian 2 file
cssr -- CSD CSSR file
fdat -- CSD FDAT file
gstat -- CSD GSTAT file
dock -- Dock Database file
dpdb -- Dock PDB file
feat -- Feature file
fract -- Free Form Fractional file
gamout -- GAMESS Output file
gzmat -- Gaussian Z-Matrix file
gauout -- Gaussian 92 Output file
g94 -- Gaussian 94 Output file
gr96A -- GROMOS96 (A) file
gr96N -- GROMOS96 (nm) file
hin -- Hyperchem HIN file
sdf -- MDL Isis SDF file
m3d -- M3D file
macmol -- Mac Molecule file
macmod -- Macromodel file
micro -- Micro World file
mm2in -- MM2 Input file
mm2out -- MM2 Output file
mm3 -- MM3 file
mmads -- MMADS file
mdl -- MDL MOLfile file
molen -- MOLIN file
mopcrt -- Mopac Cartesian file
mopint -- Mopac Internal file
mopout -- Mopac Output file
pcmod -- PC Model file
pdb -- PDB file
psin -- PS-GVB Input file
psout -- PS-GVB Output file
msf -- Quanta MSF file
schakal -- Schakal file
shelx -- ShelX file
smiles -- SMILES file
spar -- Spartan file
semi -- Spartan Semi-Empirical file
spmm -- Spartan Mol. Mechanics file
mol -- Sybyl Mol file
mol2 -- Sybyl Mol2 file
wiz -- Conjure file
unixyz -- UniChem XYZ file
xyz -- XYZ file
xed -- XED file
```

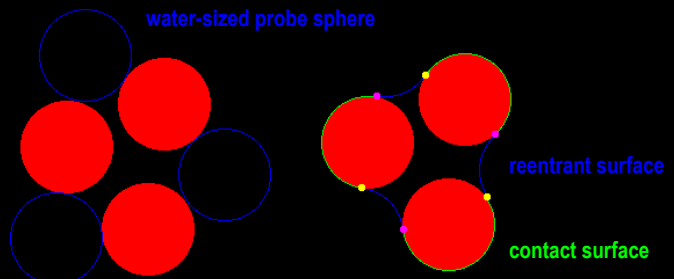
## BABEL A program designed to interconvert a number of file formats currently used in molecular modelling

**Output type codes:**

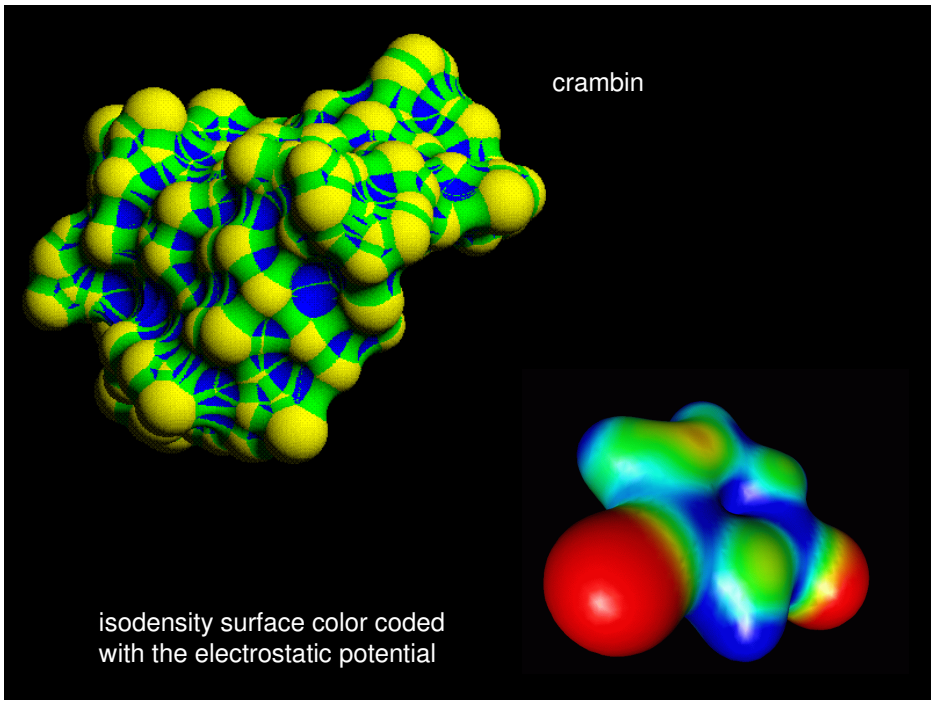
```
diag -- DIAGNOTICS file
t -- Alchemy file
bs -- Ball and Stick file
bmin -- Batchmin Command file
caccrt -- Cacao Cartesian file
cacint -- Cacao Internal file
cache -- CAChe MolStruct file
c3d1 -- Chem3D Cartesian 1 file
c3d2 -- Chem3D Cartesian 2 file
d -- ChemDraw Conn. Table file
con -- Conjure file
contmp -- Conjure Template file
cssr -- CSD CSSR file
feat -- Feature file
fhz -- Fenske-Hall ZMatrix file
gamin -- Gamess Input file
gcart -- Gaussian Cartesian file
g -- Gaussian Z-matrix file
gotmp -- Gaussian Z-matrix tmplt file
hin -- Hyperchem HIN file
icon -- Icon 8 file
i -- IDATM file
macmol -- Mac Molecule file
k -- Macromodel file
micro -- Micro World file
mi -- MM2 Input file
mo -- MM2 Ouput file
mm3 -- MM3 file
mmads -- MMADS file
mdl -- MDL Molfile file
ac -- Mopac Cartesian file
ai -- Mopac Internal file
pc -- PC Model file
p -- PDB file
report -- Report file
spar -- Spartan file
mol -- Sybyl Mol file
mol2 -- Sybyl Mol2 file
maccs -- MDL Maccs file file
xed -- XED file
unixyz -- UniChem XYZ file
x -- XYZ file
```

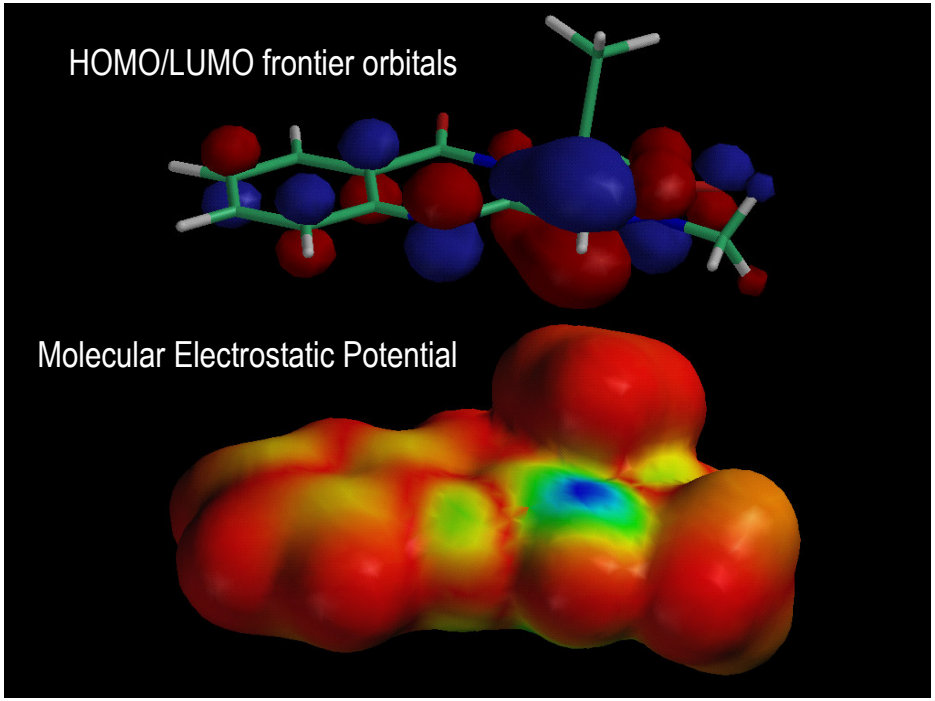
**ftp://ccl.osc.edu/pub/chemistry/software/UNIX/babel/**

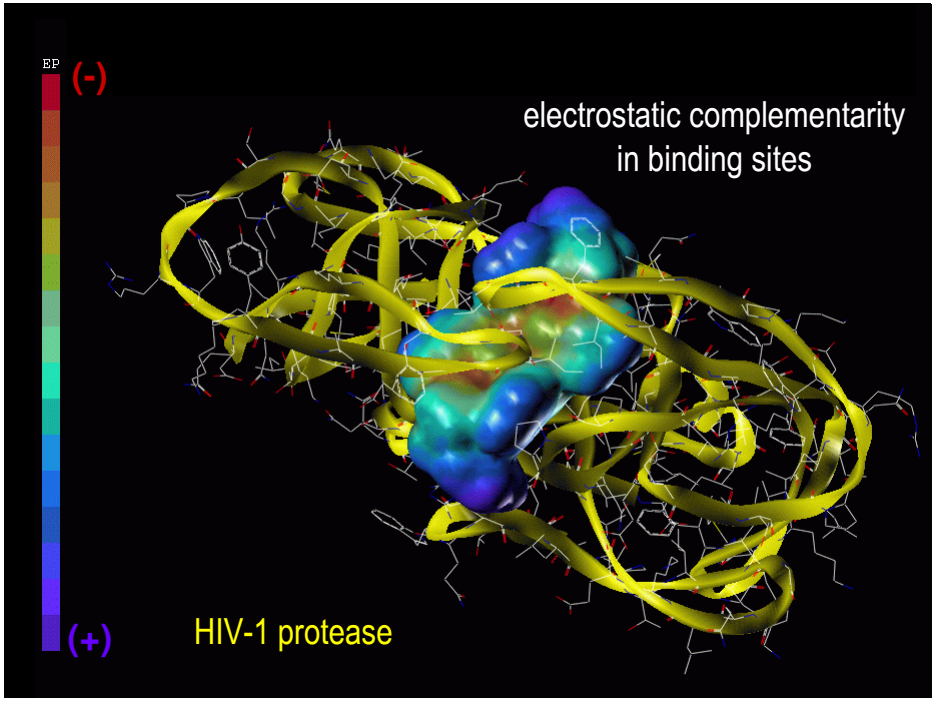
## Molecular Surfaces




The solvent-accessible surface (Lee and Richards, 1971) is traced out by the probe sphere center as it rolls over the molecule.






## MOLECULAR MECHANICS (MM)

✓ A computational technique used to model the conformational behaviour and energetic properties of molecules.

✓ The molecule is treated at the atomic level, i.e. the electrons are not treated explicitly.

✓ MM uses an Energy Function, defined so that given a particular conformation, (i.e. given a set of spatial coordinates for all the atoms) the energy of the molecule can be calculated.

✓ The energy function is empirical, i.e. it is not entirely derived from rigorous theories.

✓ The energy function makes a distinction between 'bonded' and 'non-bonded' interactions.

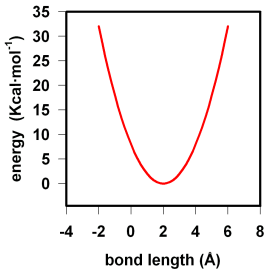
## Mecánica Molecular

$$E_{potencial} = E_{enlazada} + E_{no\text{-}enlazada}$$

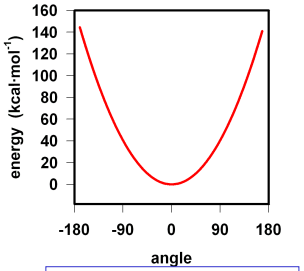
$$E_{enlazada} = \Sigma_i\ E_{enlaces} + \Sigma_i\ E_{ángulos} + \Sigma_i\ E_{diedros}$$

$$E_{no\text{-}enlazada} = \Sigma_i\ E_{electrostática} + \Sigma_i\ E_{van\ der\ Waals}$$

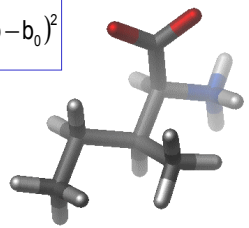
## TÉRMINOS ENLAZADOS

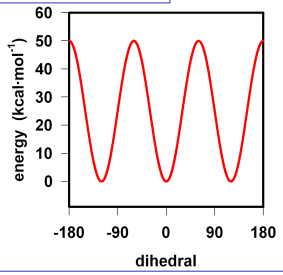



$$E_{enlaces} = \sum_{enlaces} \frac{1}{2} k_b (b - b_0)^2$$




$$E_{ángulos} = \sum_{ángulos} \frac{1}{2} k_\theta (\theta - \theta_0)^2$$



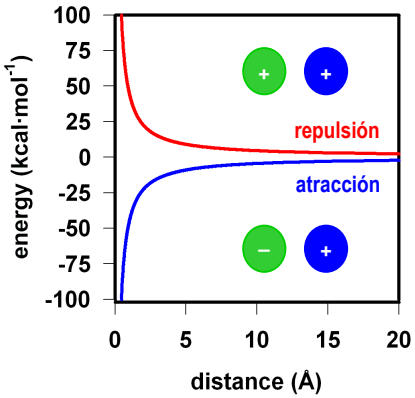


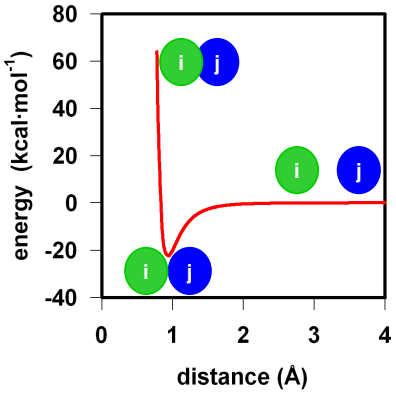

$$E_{dihedros} = \sum_{dihedros} \frac{1}{2} k_d [1 + \cos(\varphi - \varphi_0)]$$

## TÉRMINOS NO-ENLAZADOS

$$E_{electrostática} = \frac{1}{4\pi\varepsilon_0\varepsilon} \sum_{ij} \frac{q_i q_j}{r_{ij}}$$

$$E_{Lennard-Jones} = \sum_{ij} \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}}$$



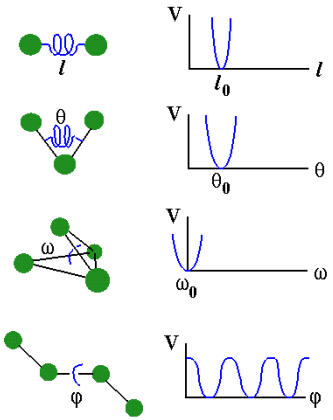

## Empirical Potential Energy Function

Bonds

Angles

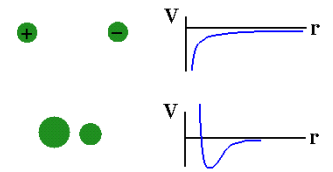
Improper Dihedrals

Torsions



Electrostatics

van der Waals

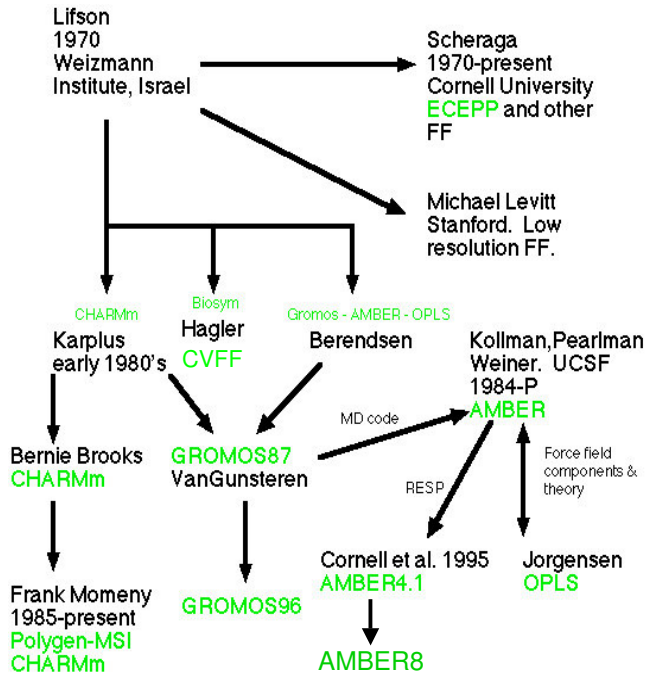


Resumen de interacciones incluidas en un **campo de fuerzas** representativo de mecánica molecular

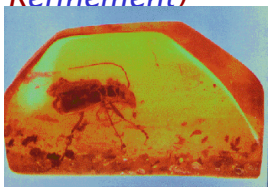
$$V = \sum_{enlaces} k_b (b - b_0)^2 + \sum_{ángulos} k_\theta (\theta - \theta_0)^2 +$$
$$+ \sum_{diedros} \sum_{n=1}^{N} K_\varphi^{(n)} [1 + \cos(n\phi - \delta)] + \sum_{impropios} K_\omega (\omega - \omega_0)^2$$
$$+ \sum_{i,j} 4\varepsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] + \sum_{i,j} \left( \frac{q_i q_j}{D r_{ij}} \right)$$

✓ La función empírica de energía potencial es diferenciable con respecto a las coordenadas atómicas.

✓ Esto proporciona el valor y la dirección de la fuerza que actúa sobre cada átomo y puede así utilizarse en una simulación de dinámica molecular.



*AMBER (Assisted Model Building with Energy Refinement)*



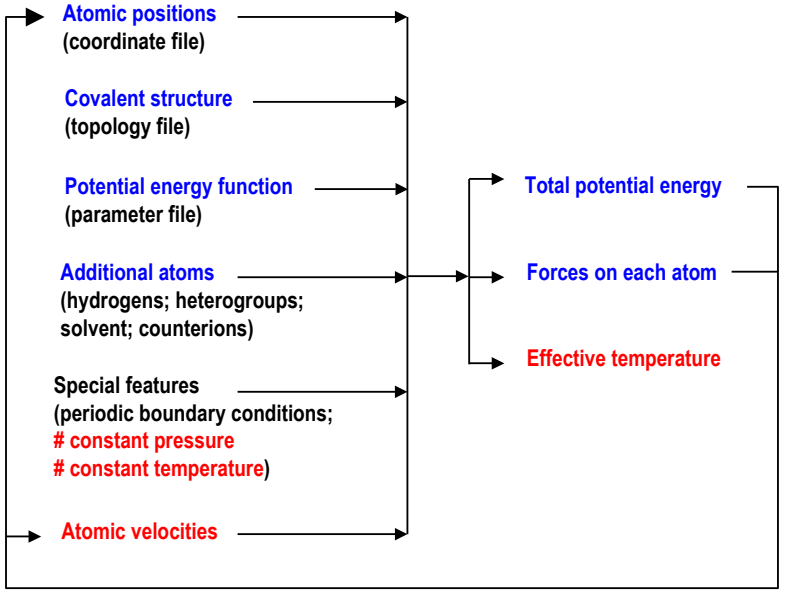
*CHARMm® (Chemistry at HARvard Macromolecular Mechanics)*

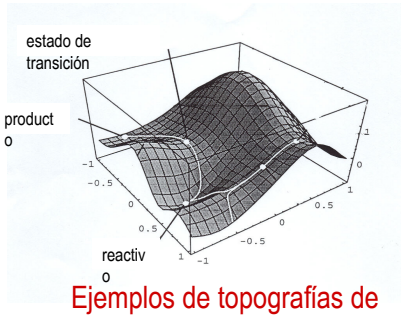
*CVFF (Consistent-Valence Force Field)*

*GROMOS (GROningen MOlecular Simulation package)*

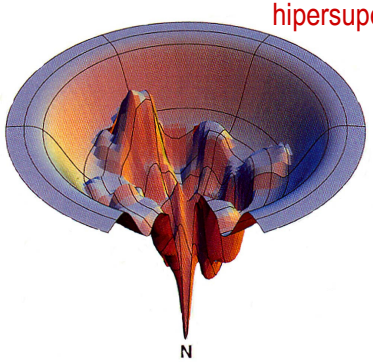
*OPLS (Optimized Potentials for Liquid Simulations)*

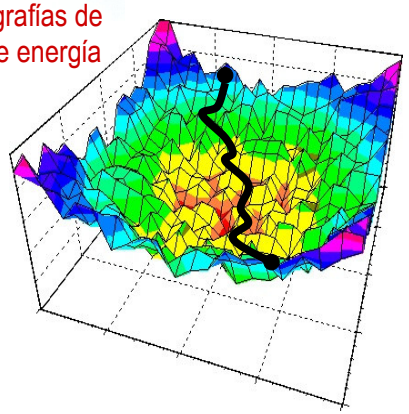
ALGORITHMS FOR **ENERGY MINIMIZATION** AND **MOLECULAR DYNAMICS**




Ejemplos de topografías de hipersuperficies de energía



# Dinámica Molecular


Segunda ley de Newton:

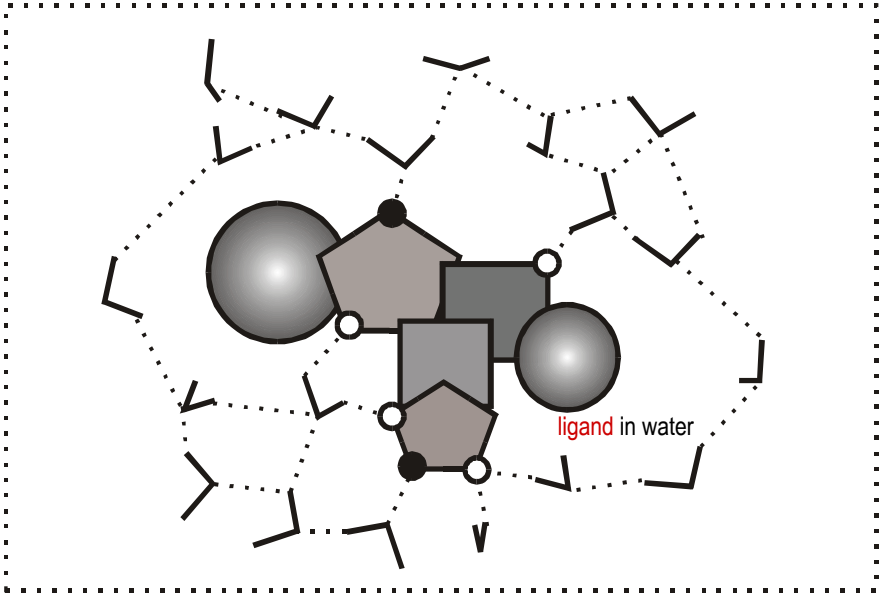
$$-\frac{dE}{dx} = F = m \cdot a = m \cdot \frac{v}{t} = m \cdot \frac{d^2x}{dt^2}$$

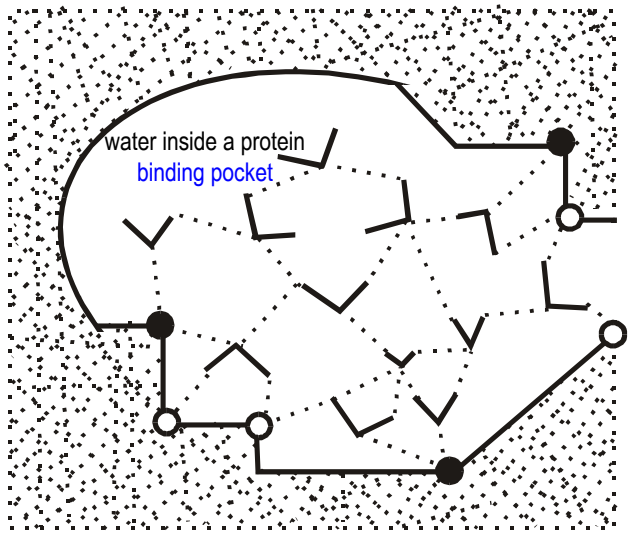
Las simulaciones de Dinámica Molecular permiten el estudio de procesos dinámicos complejos que ocurren en los sistemas biológicos, por ejemplo:

- ✓ Estabilidad de proteínas
- ✓ Cambios conformacionales
- ✓ Plegamiento de proteínas
- ✓ Reconocimiento molecular: ligandos, proteínas, ADN...
- ✓ Transporte de iones en sistemas biológicos

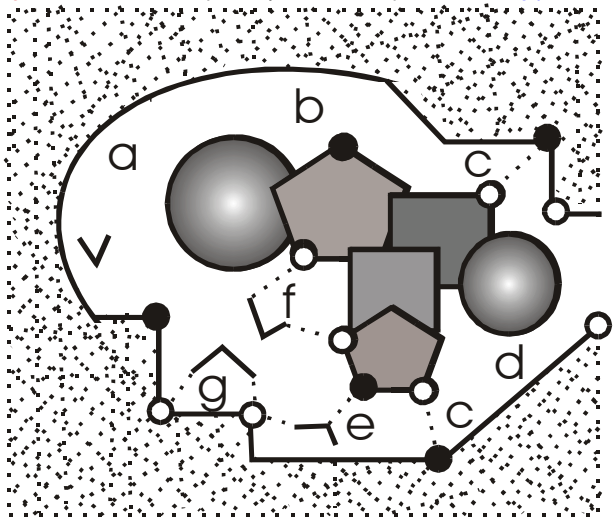
y proporcionan un medio valioso para llevar a cabo estudios de:

- determinación de estructuras por difracción de rayos X y espectroscopía de RMN
- diseño de nuevos fármacos



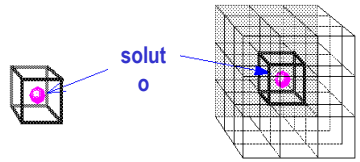



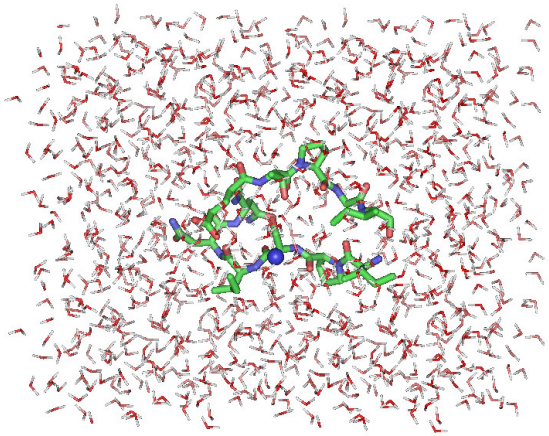

ligand buried inside a partially desolvated protein binding pocket



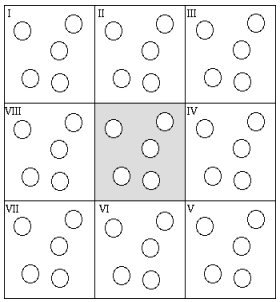

**Conclusion:** Any comprehensive method that attempts to model ligand binding must also consider the energy of solvation and entropic contributions to the binding process.

## TRATAMIENTO EXPLÍCITO DEL DISOLVENTE

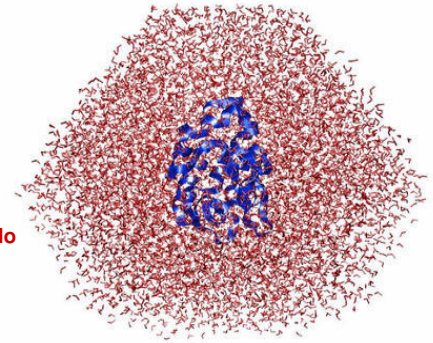


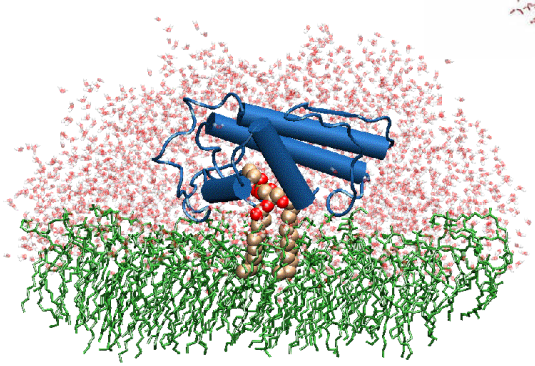
CONDICIONES DE LÍMITE PERIÓDICO

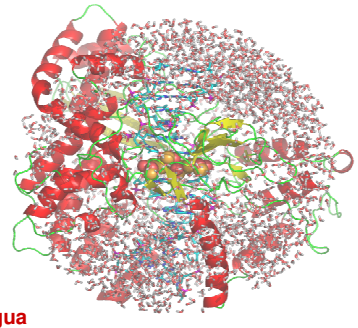



## TRATAMIENTO EXPLÍCITO DEL DISOLVENTE

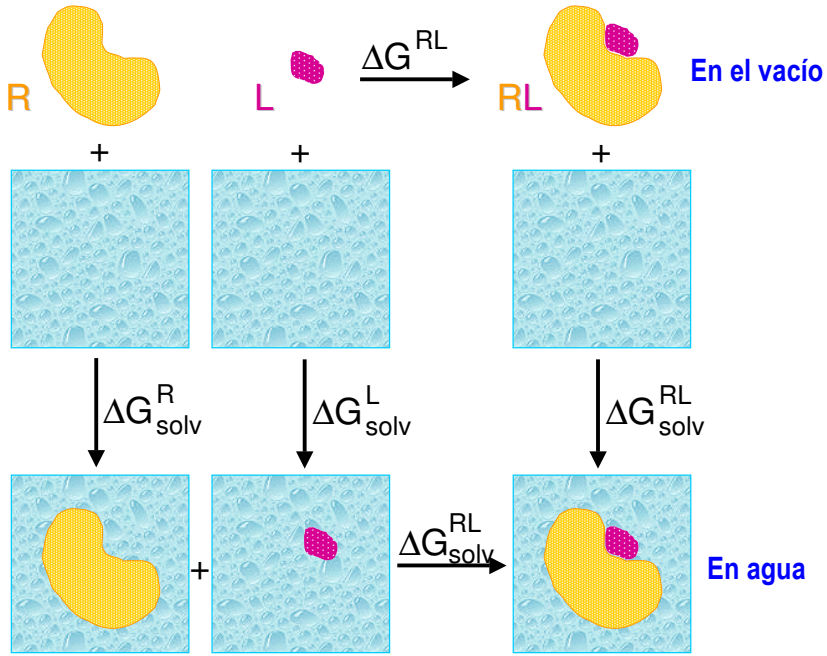


octaedro truncado





"gotas" o "conchas" de agua




## "DelPhi - A Macromolecular Electrostatics Modelling Package":

**Kim A. Sharp, Anthony Nicholls & Barry Honig**
Department of Biochemistry and Molecular Biophysics, Columbia University, New York

**- Klapper, I.; Hagstrom, R.; Fine, R.; Sharp, K.; Honig, B. "Focusing of Electric Fields in the Active Site of Cu-Zn Superoxide Dismutase: Effects of Ionic Strength and Amino-acid Modification."** ***Proteins (1986) 1, 47-59.***

**- Gilson, M. K.; Sharp, K. A.; Honig, B. H. "Calculating the Electrostatic Potential of Molecules in Solution: Method and Error Assessment"** ***J. Comput. Chem. (1987) 9, 327-335.***

**- Gilson, M. K.; Honig, B. "Calculation of the Total Electrostatic Energy of a Macromolecular System: Solvation Energies, Binding Energies, and Conformational Analysis."** ***Proteins (1988) 4, 7-18.***

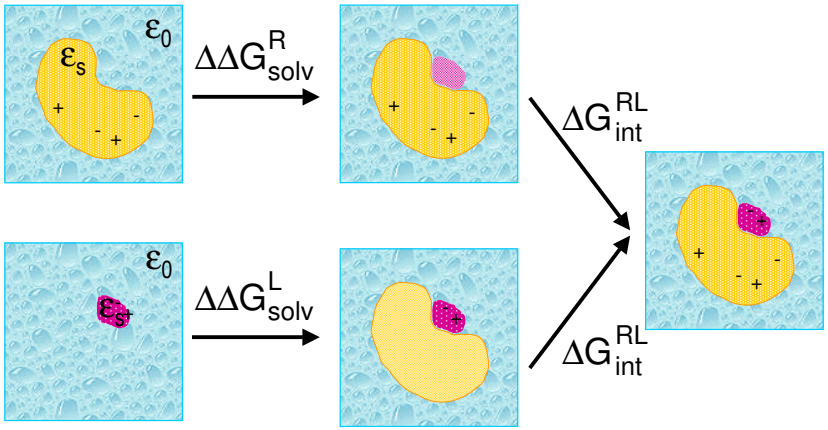
**- K. Sharp, K.; Honig, B. "Electrostatic Interactions in Macromolecules: Theory and Applications."** ***Ann. Rev. Biophys. Biopys. Chem. (1990) 19, 301-332.***

**- Nicholls, A.; Honig, B. "A Rapid Finite Difference Algorithm, Utilizing Successive Over-Relaxation to Solve the Poisson-Boltzmann Equation."** ***J. Comput. Chem. (1991) 12, 435-445.***

The original reference to the use of the finite difference method for macromolecular electrostatics is: J. Warwicker and H. C. Watson, *J. Mol. Biol. (1982) 157, 671.*

Ecuación de Poisson: $\nabla^2\phi(r) = -\frac{4\pi\rho(r)}{\varepsilon}$

**Ecuación de Poisson-Boltzmann:** $\nabla\cdot[\varepsilon(r)\nabla\phi(r)] - k'\phi(r) = -4\pi\rho(r)$




# Affinity *vs.* Specificity

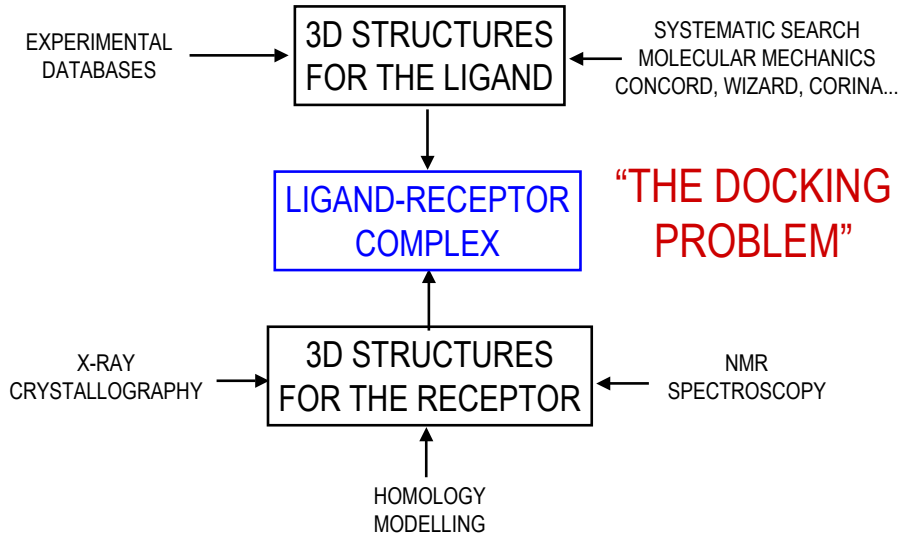
$$\text{Ligand} + \text{Receptor} \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} \text{Ligand-Receptor}$$

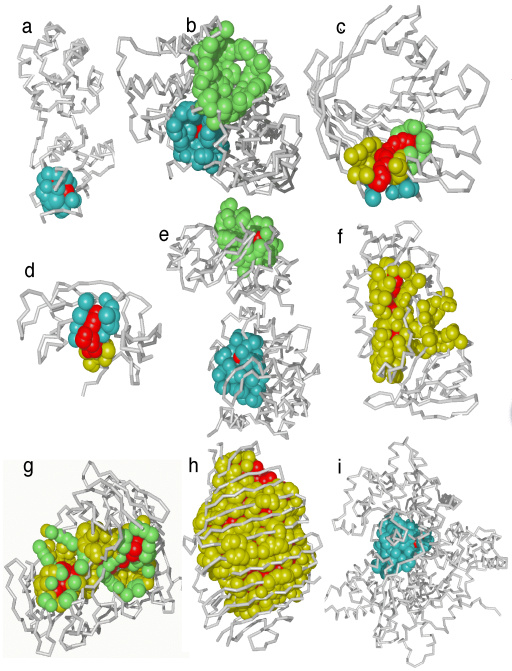
$$K_d = \frac{k_{off}}{k_{on}} = \frac{[\text{Ligand}]\,[\text{Receptor}]}{[\text{Ligand-Receptor}]}$$

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

| **Binding constant** | **Binding energy** |
|---|---|
| $\Delta K_d$ | $\Delta G$ (kcal/mol) |
| 2x | 0.5 |
| 5x | 1.0 |
| 13x | 1.5 |
| 29x | 2.0 |
| 68x | 2.5 |
| 158x | 3.0 |

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



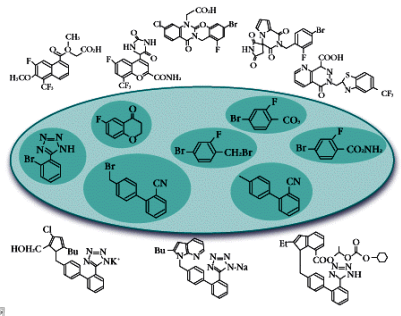



← Dianas validadas

\+

Quimioteca Combinatoria

↓



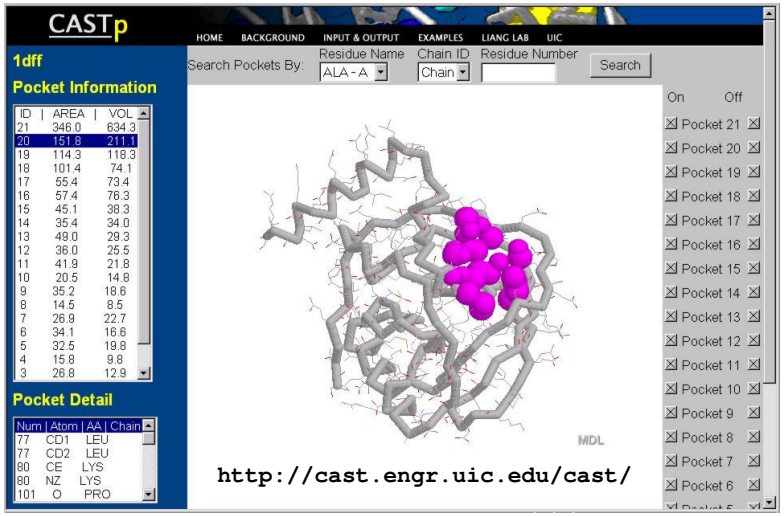
II **EL PROBLEMA DEL *"DOCKING"***

Complejos Ligando-Receptor

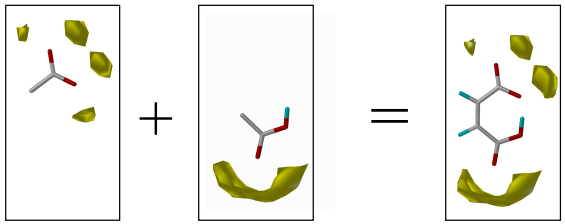
# CASTp

*A Server for Identification of Protein Pockets & Cavities*

- Identifies all pockets and cavities.
- Measures the volume and area analytically.



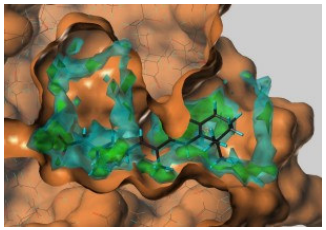

# SuperStar



- Calculate binding positions for specific probe atoms in protein active sites
- Identify functional groups in binding-site
- Look up relevant IsoStar scatterplots and overlay on functional groups
- Contour - combining by taking products

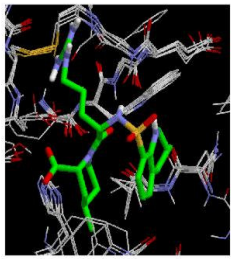
# SuperStar Features

map for aromatic CH carbon probe generated at the binding site of the protein-ligand complex 1CPS.



- Cavity detection
- Surface or pharmacophore point display
- Metal coordination
- Hyperlinking to IsoStar scatterplots
- Choice of CSD- or PDB-based maps
- Gaussian fits

# Relibase



**Relibase**: A program for searching protein-ligand databases.

**Version 4.0, October 2000**



Install 3D visualization software

Comments/bugs/queries to:

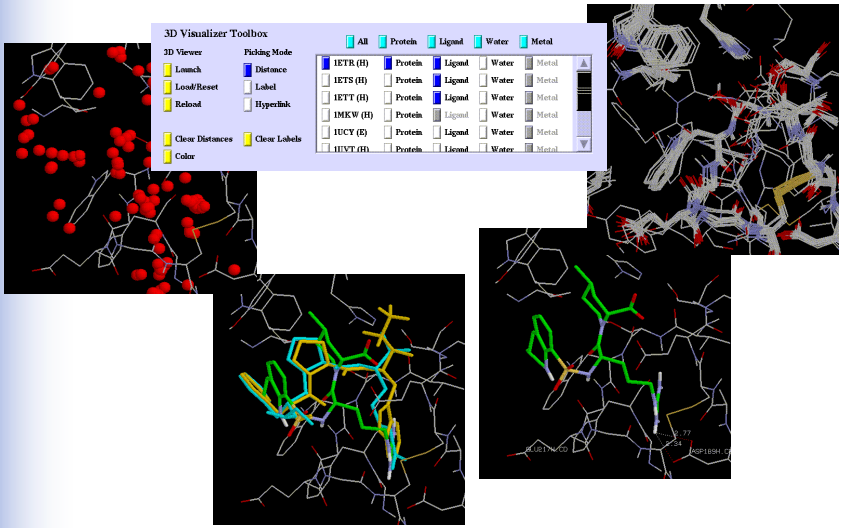
*Cambridge Crystallographic Data Centre*
*12 Union Road*
*Cambridge CB2 1EZ*
*United Kingdom*
*tel: (44) (1223) 336022*
*e-mail: support@ccdc.cam.ac.uk*

**http://relibase.ccdc.cam.ac.uk/**

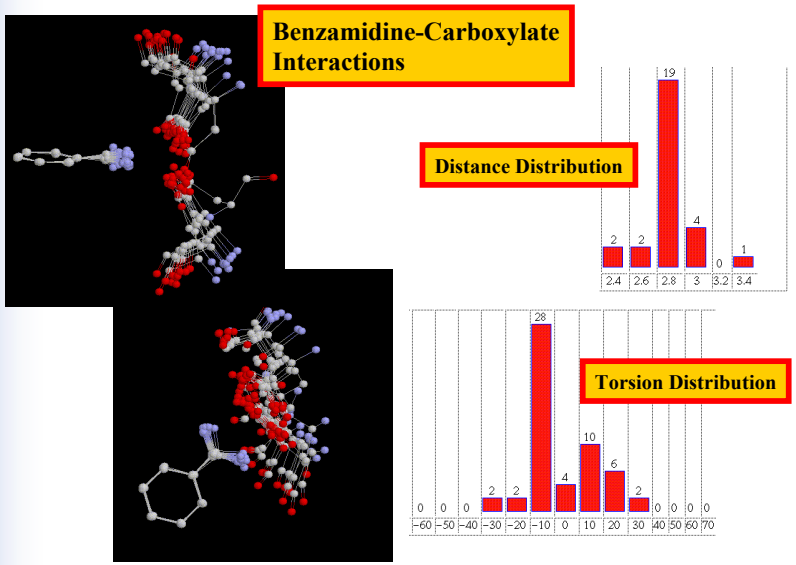
**http://relibase.ebi.ac.uk/**

**http://relibase.rutgers.edu/**

# Binding Site Superposition




# Analysis of 3D Queries




# Relibase+

- Protein-ligand database system
- Based on original software developed by Manfred Hendlich and colleagues at Merck and Marburg University
- Enables searching of PDB and of in-house proprietary databases

Relibase: **http://relibase.ccdc.cam.ac.uk/**

# Some Relibase+ Options

- Text searching
- Sequence searching
- 2D substructure and similarity searching
- 3D substructure searching
- Logical combination of hit lists
- Searching for intermolecular interactions
- Auto-superposition of similar binding sites
- Scripting facility based on Python

# LIGPLOT

http://www.biochem.ucl.ac.uk/bsm/ligplot/ligplot.html

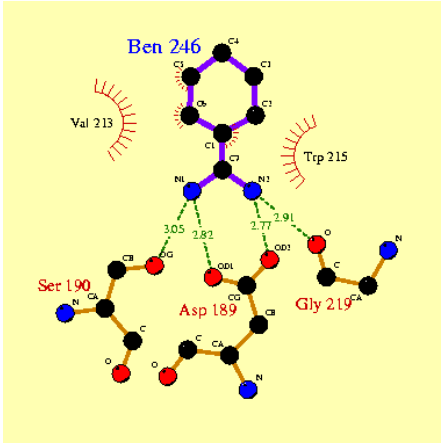
Program for automatically plotting protein-ligand interactions (by A. Wallace & R. Laskowski)

Automatically generates schematic diagrams of protein-ligand interactions for a given PDB file.

hydrogen bonds: dashed lines between the atoms involved.

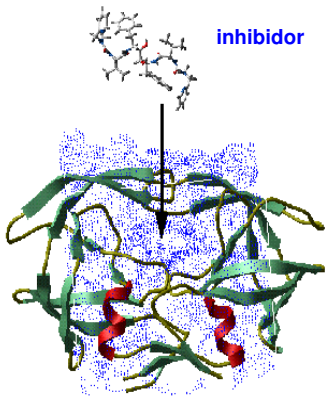
hydrophobic contact: an arc with spokes radiating towards the ligand atoms they contact. The contacted atoms are shown with spokes radiating back.

Atom accessibilities can also be depicted; the ligand atoms can be colour-coded to indicate their accessibility to solvent.

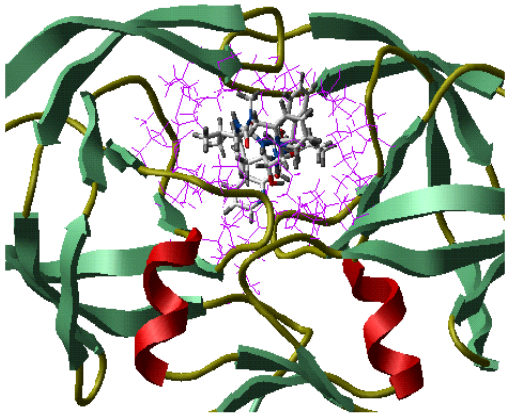



benzamidine (PDB code 2TBS)

## ACOPLAMIENTO LIGANDO-RECEPTOR (*DOCKING*




Proteasa de VIH-1



Complejo enzima-inhibidor

# Virtual ("in silico") screening

## Docking/scoring programs

Docking engines: search the conformational space in the binding site

Scoring functions: discrimination of correctly docked from misdocked conformations

# MOLECULAR DOCKING

- SYSTEMATIC SEARCH (*brute force algorithm*):

All binding orientations of all conformers of the ligand and the receptor (impractical for most situations).

- AUTOMATED SEARCH:

GEOMETRIC METHODS: Matching of ligand and receptor site descriptors (descriptors, grids, fragments...).

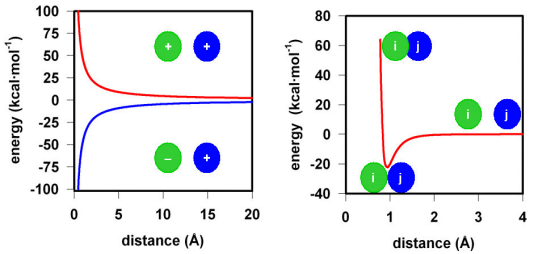
FORCE FIELD METHODS: Minimizing the ligand-receptor interaction energy - Molecular dynamics and Monte Carlo simulations.

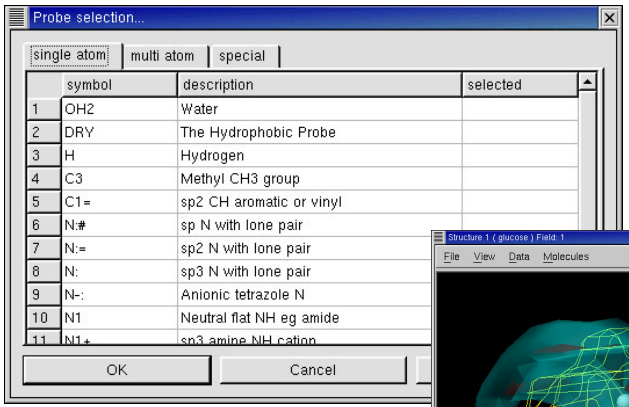
## "GRID: A Computational Procedure for Determining Energetically Favorable Binding Sites on Biologically Important Macromolecules"

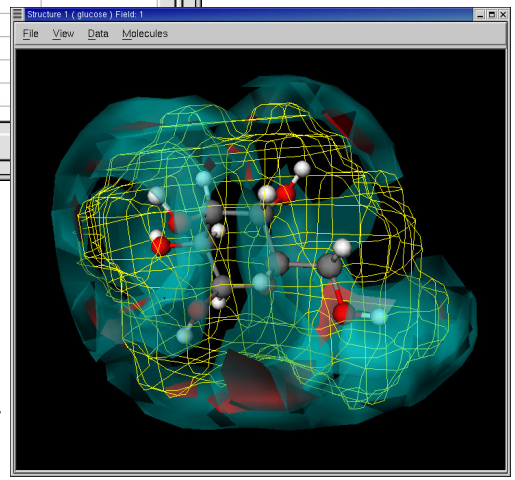
Peter Goodford, Oxford University

***J. Med. Chem.*** *28*, 849-857 (1985)
***ibid.*** *32*, 1083-1094 (1989); *36*, 140-147 (1993); *36*, 148-156 (1993)

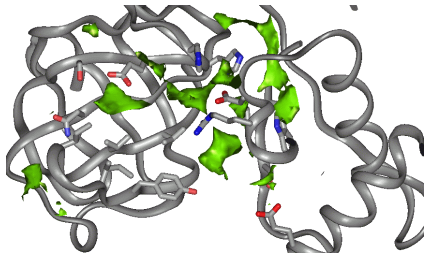
http://www.moldiscovery.com/



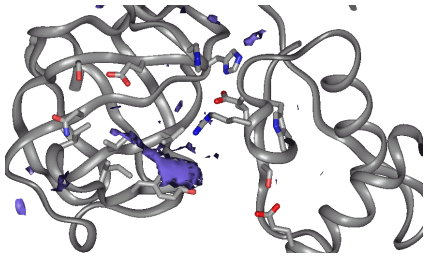



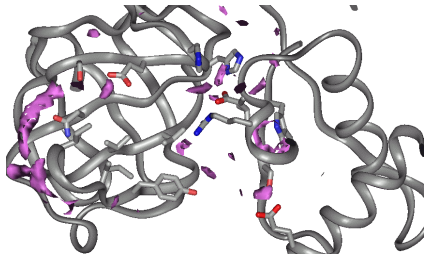

http://www.moldiscovery.com/



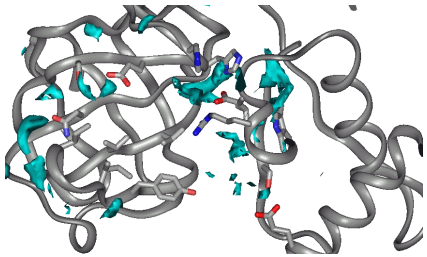
**Aromatic carbon probe**
Grid point value range: -5.45 to 5.0 kcal/mol
Contour level: -2.5 kcal/mol



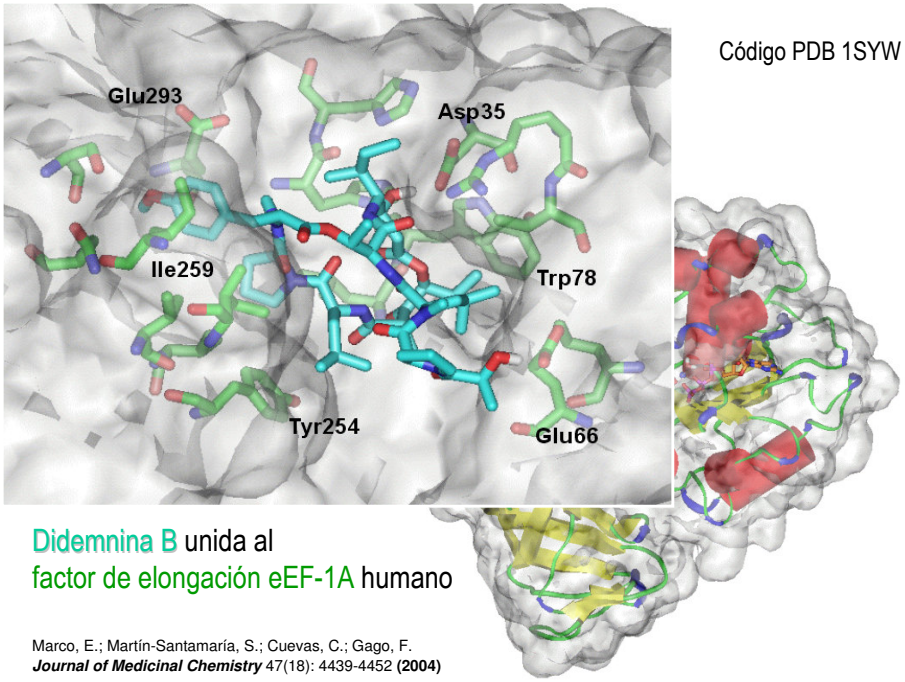
**Hydrophobic probe**
Grid point value range: -2.86 to 0.0 kcal/mol
Contour level: -1.0 kcal/mol



**Carbonyl oxygen probe**
Grid point value range: -8.03 to 5.0 kcal/mol
Contour level: -5.0 kcal/mol



**Hydroxyl oxygen probe**
Grid point value range: -12.30 to 5.0 kcal/mol
Contour level: -7.0 kcal/mol




Código PDB 1SYW

Didemnina B unida al
factor de elongación eEF-1A humano

Marco, E.; Martín-Santamaría, S.; Cuevas, C.; Gago, F.
***Journal of Medicinal Chemistry*** 47(18): 4439-4452 **(2004)**

# Why Use Molecular Docking?

- Most detailed representation of binding site
  - overcomes simplifications of pharmacophores
  - identifies both conservative and novel solutions
  - provides impetus for *de novo* design/optimisation
- Broad range of analyses applicable
  - diverse scoring/selection criteria
- Quality/throughput of available methods
  - good enough, despite technical limitations

# "THE DOCKING PROBLEM"

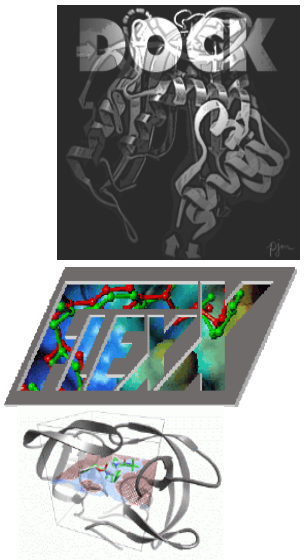
➢ SITE/LIGAND REPRESENTATION
(treatment of H atoms?)

➢ JUXTAPOSITION OF THE LIGAND AND SITE FRAMES OF REFERENCE (docking engine)

➢ EVALUATION OF COMPLEMENTARITY
(scoring functions)

AIM: To obtain the lowest free energy structure(s) for the receptor-ligand complex

## Examples of docking algorithms




Rigid ligand:

Fast shape matching (DOCK)

Flexible ligand:

Fast shape matching (DOCK 4.0)

Incremental construction (FlexX)

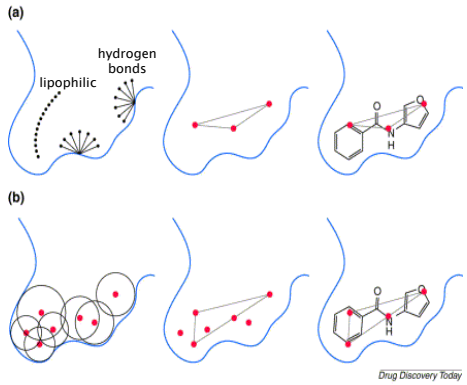
Simulated annealing (AutoDock 2.4)

Monte Carlo simulations (MCDOCK)

Genetic algorithm (AutoDock 3.0, GOLD, GAMBLER)

FlexX algorithm

DOCK algorithm




surface of the receptor pocket

FlexX matches triangles of interaction sites onto complementary ligand atoms.

DOCK fills the binding site with spheres, and sphere centers are then matched to the ligand atoms to determine plausible ligand–receptor complexes.

# PROGRAM **DOCK**

"A Geometric Approach to Macromolecule-Ligand Interactions"
I. D. Kuntz, J. M. Blaney, S. J. Oatley, R. Langridge, T. E. Ferrin
*J. Mol. Biol. 161*, 269-288 (1982)

"Using Shape Complementarity as an Initial Screen in Designing Ligands for a Receptor Binding Site of Known Three-Dimensional Structure"
R. L. DesJarlais, R. P. Sheridan, G. L. Seibel, J. S. Dixon, I. D. Kuntz, R. Venkataraghavan
*J. Med. Chem. 31*, 722-729 (1988)

"Automated Docking with Grid-Based Energy Evaluation"
E. C. Meng, B. K. Soichet, I. D. Kuntz
*J. Comp. Chem. 13*, 505-524 (1991)

**RECEPTOR COORDINATES**

**SITE CHARACTERIZATION**

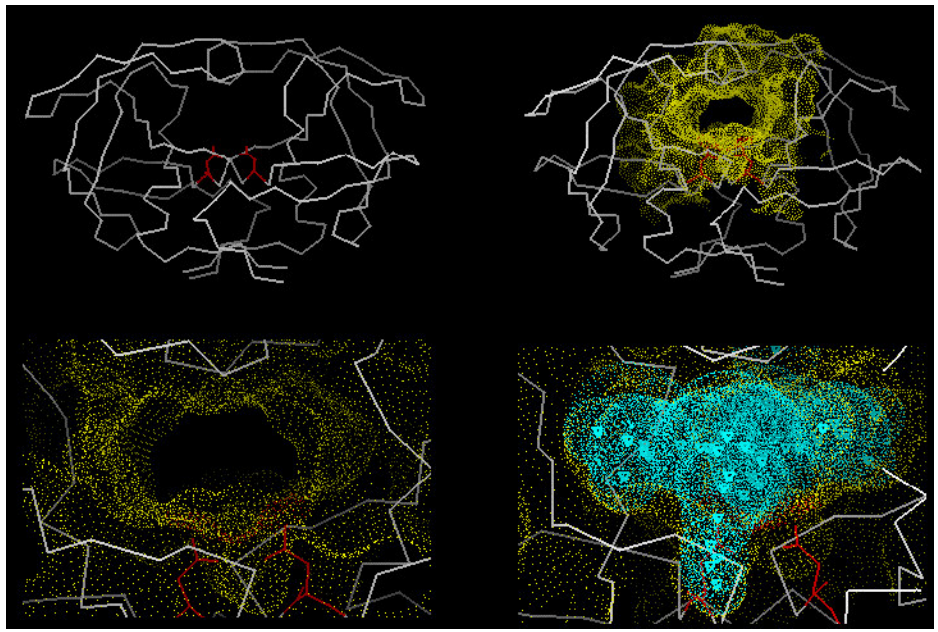
| | |
|---|---|
| MS | molecular "dot" surface |
| SPHGEN | negative image of site |

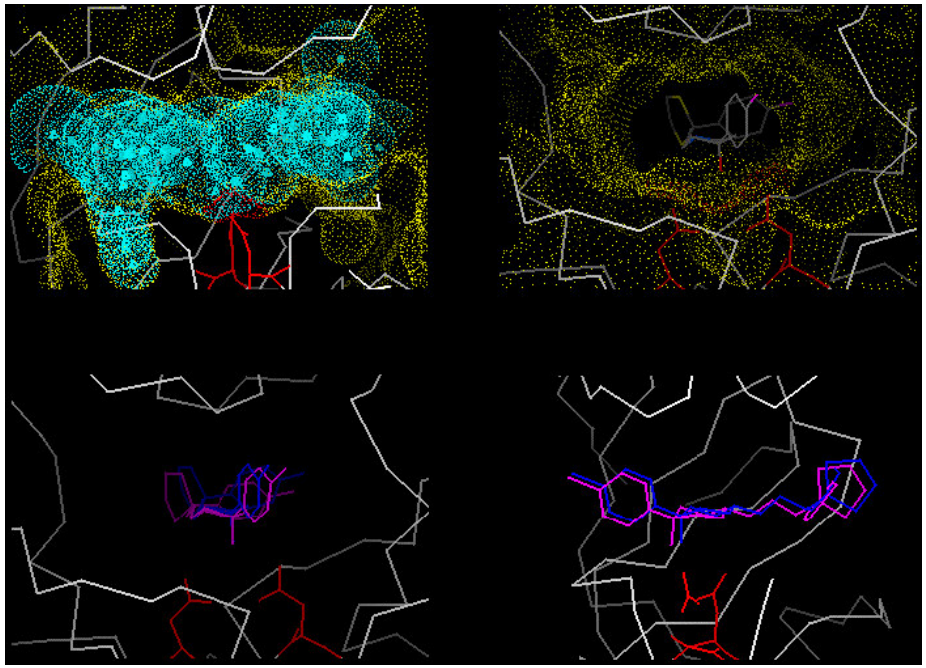
**GRID CALCULATION**

| | |
|---|---|
| DISTMAP | contact scoring |
| CHEMGRID | force-field scoring |

**DOCKING AND SCORING**

LIGAND COORDINATES → DOCK: Matching, Orientation, Scoring





Docking accuracy
[Rms deviations (non hydrogen atoms, in Å) from the X-ray pose]
(top solution of each docking tool)

| ligand | Docking method | | |
|---|---|---|---|
| | DOCK | FlexX | GOLD |
| deoxythymidine | 0.82 | 0.78 | 0.72 |
| 5-iododeoxyuridine | 9.33 | 1.03 | 0.77 |
| 5-iodouracil-anhydrohexitol | 1.16 | 0.88 | 0.63 |
| dhbt (not publicly available) | 2.02 | 3.65 | 0.93 |
| 6-(3-hydroxy-propyl-thymine) | 1.02 | 4.18 | 0.49 |
| 6-[6-hydroxymethy-5-methyl-2,4-dioxo-hexahydro-pyrimidin-5-yl-methyl]-5-methyl-*1*H-pyrimidin-2,4-dione | 9.62 | 13.30 | 2.33 |
| (North)-methanocarbathymidine | 7.56 | 1.11 | 1.19 |
| aciclovir | 3.08 | 2.71 | 2.74 |
| ganciclovir | 3.01 | 6.07 | 3.11 |
| penciclovir | 4.10 | 5.96 | 3.01 |

Only one set of protein (TK) coordinates used: pdb code 1kim

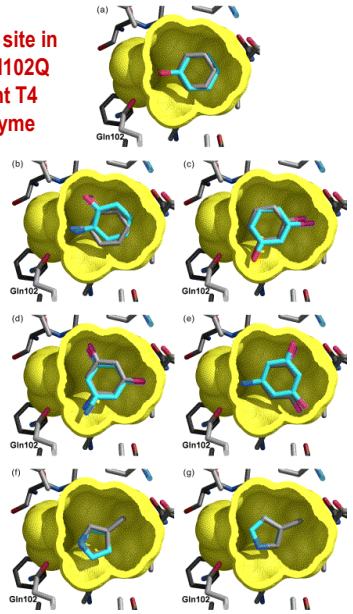
## Scoring functions

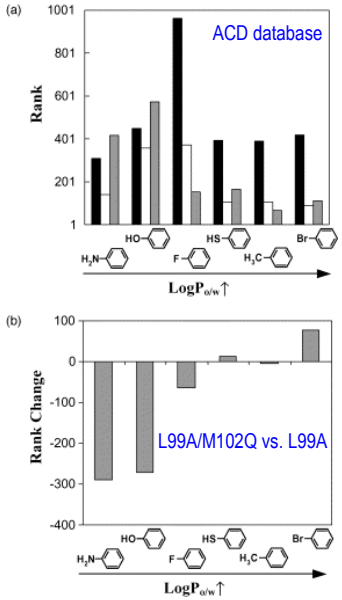
Knowledge-based: statistical analysis of 3D complex structures to derive a sum of *potentials of mean force* between receptor and ligand atoms

Force field-based: calculation of van der Waals and electrostatic interaction energies between the receptor and the ligand atoms

Empirical: the binding free energy is broken down into a number of different *weighted* contributions (supposed to be additive: number of hydrogen bonds, ionic interactions, apolar contacts, entropy penalties...)

**Binding site in L99A/M102Q mutant T4 lysozyme**

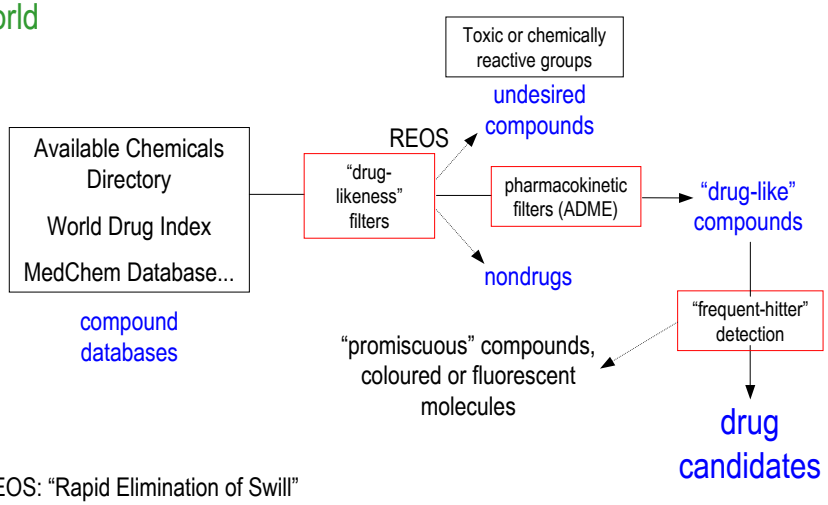



**Wei BQ, Baase WA, Weaver LH, Matthews BW, Shoichet BK.**
**A model binding site for testing scoring functions in molecular docking.**
***J. Mol. Biol.* (2002) 322:339-355**

# *In silico* **VIRTUAL SCREENING** and FOCUSED LIBRARY DESIGN

Near-perfect structures in an imperfect world

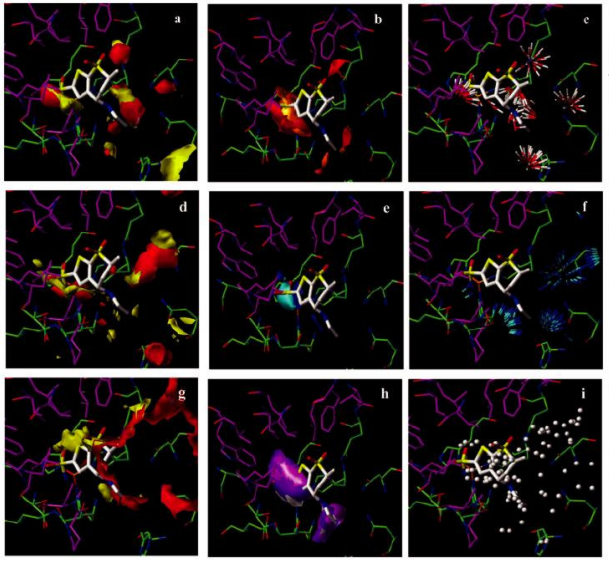


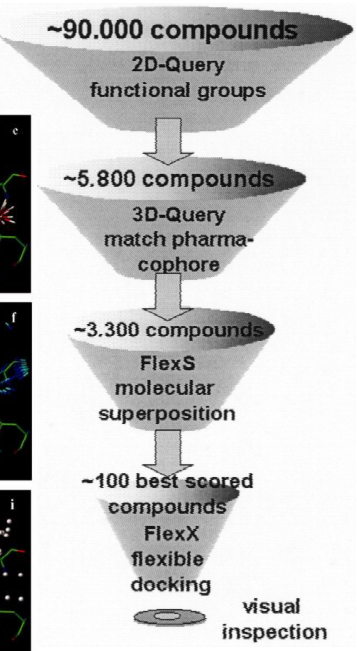

REOS: "Rapid Elimination of Swill"

3588 *J. Med. Chem.* **2002**, *45*, 3588–3602

**Successful Virtual Screening for Novel Inhibitors of Human Carbonic Anhydrase: Strategy and Experimental Confirmation**

Sven Grüneberg,† Milton T. Stubbs, and Gerhard Klebe*
*Institute of Pharmaceutical Chemistry, University of Marburg, Marbacher Weg 6, D-35032 Marburg, Germany*



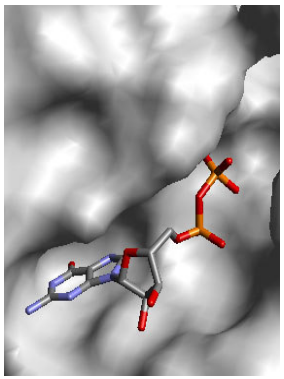



University of Oxford

## Screensaver Lifesaver

searching for anti-cancer drugs by distributed computational chemistry

http://www.chem.ox.ac.uk/ccdd/ccdd.html



NATIONAL FOUNDATION FOR CANCER RESEARCH
research for a cure

| | |
|---|---|
| Superoxide dismutase | Vascular Endothelial Growth Factor |
| RAS proteins | Insulin Tyrosine Kinase |
| Cyclooxygenase (COX-2) | c-ABL Tyrosine Kinase |
| Fibroblast Growth Factor Receptor | CDK-2 |
| RAF | Farnesyltransferase |
| Protein-Tyrosine-Phosphatase 1B | VEGFr1 |

## fightAIDS@home

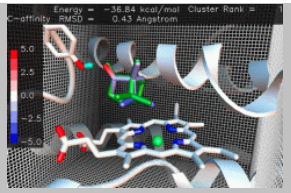
the Olson laboratory
computing toward a cure
powered by ENTROPIA

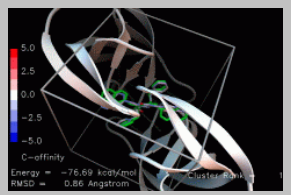
THE SCRIPPS RESEARCH INSTITUTE

http://FightAidsathome.scripps.edu

Dr. Garrett Morris

fight AIDS @home
computing toward a better inhibitor
powered by AUTODOCK





# Applications of Ligand-Protein Docking in Drug Design

**Existing methods**
Given a protein, find potential binding ligands from a chemical database

Compound Database
Compound 1
...
Compound n

Protein

Successfully docked compounds as putative ligands

***Science* (1992) *257*: 1078**

**New method**
Given a ligand, find potential potein targets from a protein database

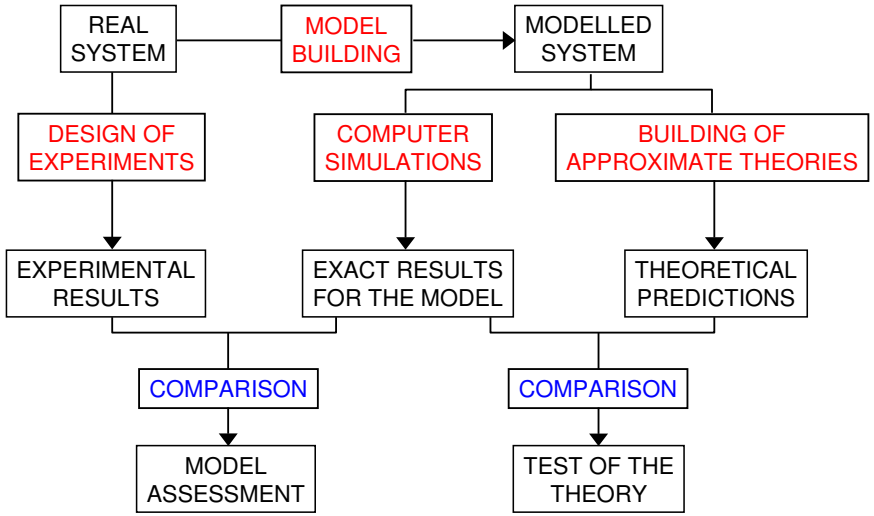
Protein Database
Protein 1
...
Protein n

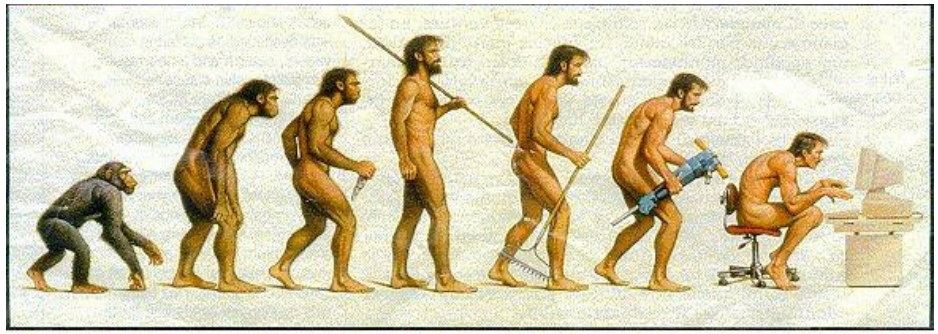
Ligand

Successfully docked proteins as putative targets

***Proteins* (2001) *43*: 217**

# CONNECTION BETWEEN EXPERIMENT, THEORY AND COMPUTER SIMULATION



**Somewhere, something went terribly wrong**

## QUESTIONS WELCOME

PREGUNTAS, POR FAVOR

E-mail: federico.gago@uah.es