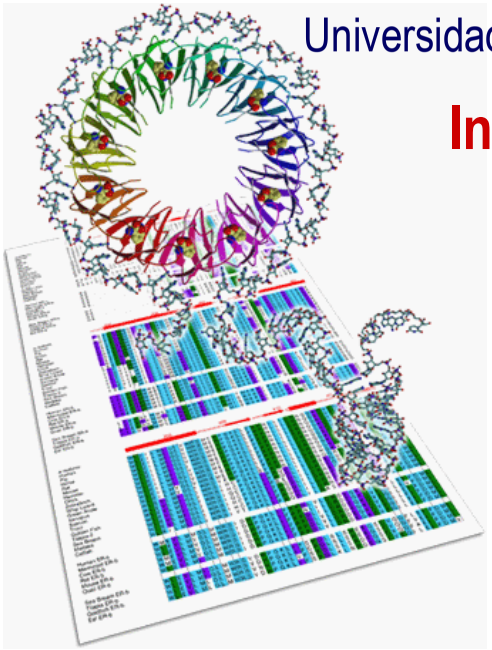
BioInfo 2004

Curso de Doctorado: BIOINFORMÁTICA
Universidad Autónoma de Madrid. Marzo-Abril 2004

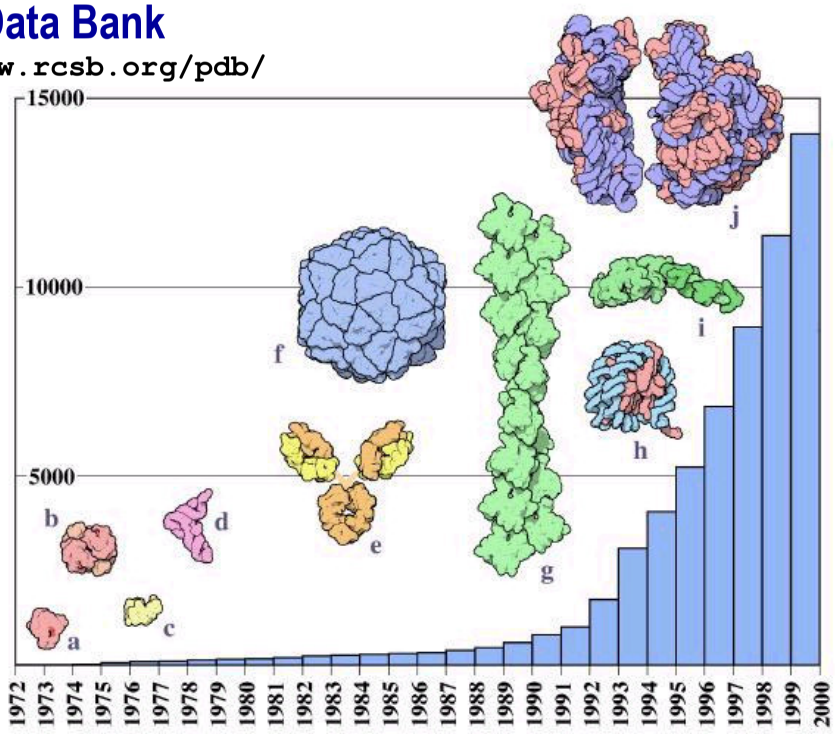


Interacciones entre proteínas y moléculas pequeñas (II)

Federico Gago
Departamento de Farmacología
Universidad de Alcalá, Madrid

Protein Data Bank

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



(a) myoglobin (b) hemoglobin (c) lysozyme (d) transfer RNA
(e) antibodies (f) viruses (g) actin (h) the nucleosome
(i) myosin (j) the ribosome

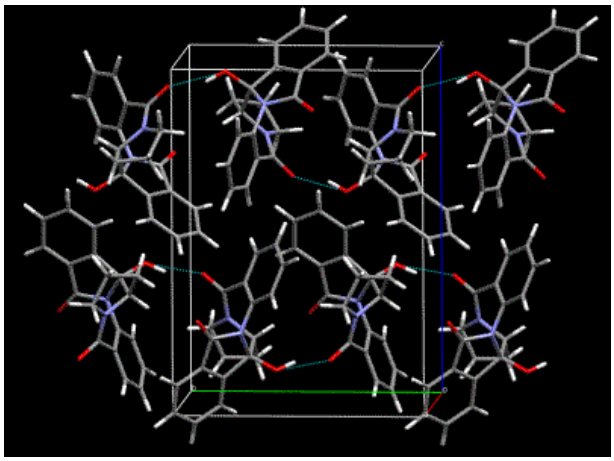
Courtesy of David Goodsell, TSRI

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

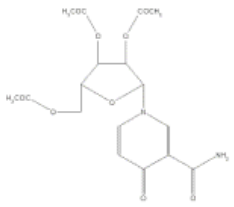


- Derived knowledge bases
- Knowledge-based applications programs
- Data mining tools for protein-ligand complexes

1D Bibliographic Information

BASYOJ
4-Oxonicotinamide-1-(1'-beta-D-2',3',5'-tri-O-acetyl-ribofuranoside)
Source: Rothmannia longiflora
C17 H20 N2 09
G. Bringmann, M. Ochse, K. Wolf, J. Kraus, K. Peters, E-M. Peters, M. Herderich, L. Ake, F. Tayman
Phytochemistry 51 (1999), p271

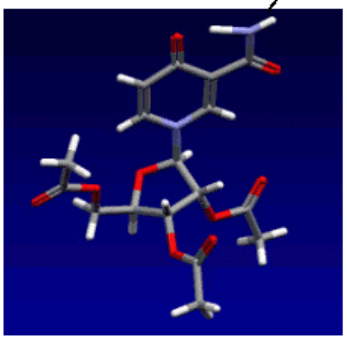
2D Chemical Connectivity



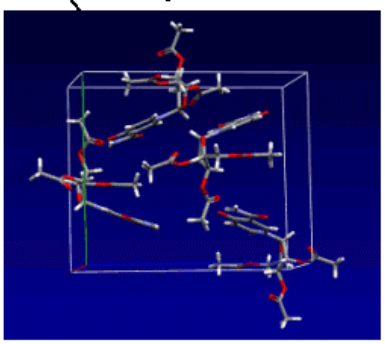
>272,000 organic and metallo-organic crystal structures analysed using X-ray or neutron diffraction techniques

CSD

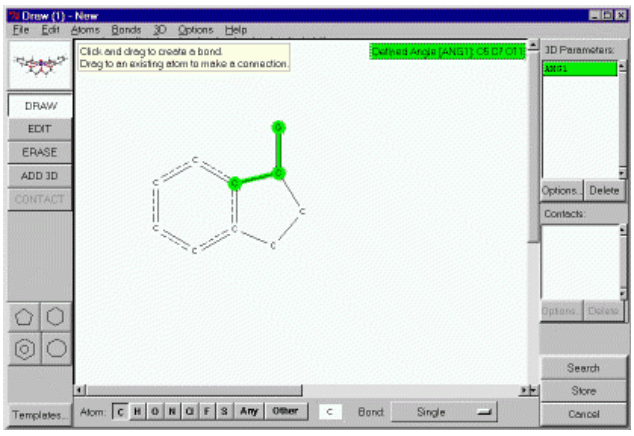
3D Molecular Structure



3D Crystal Structure

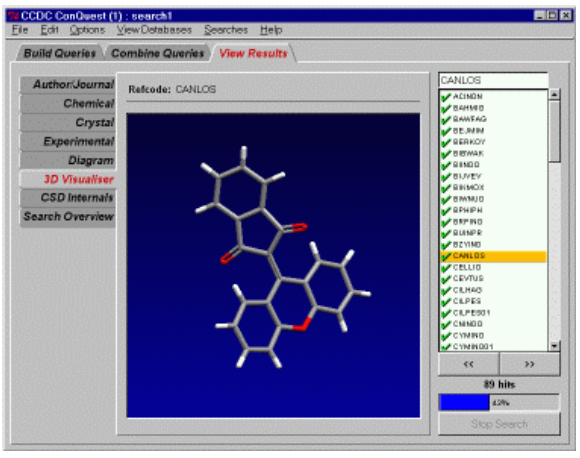


ConQuest



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

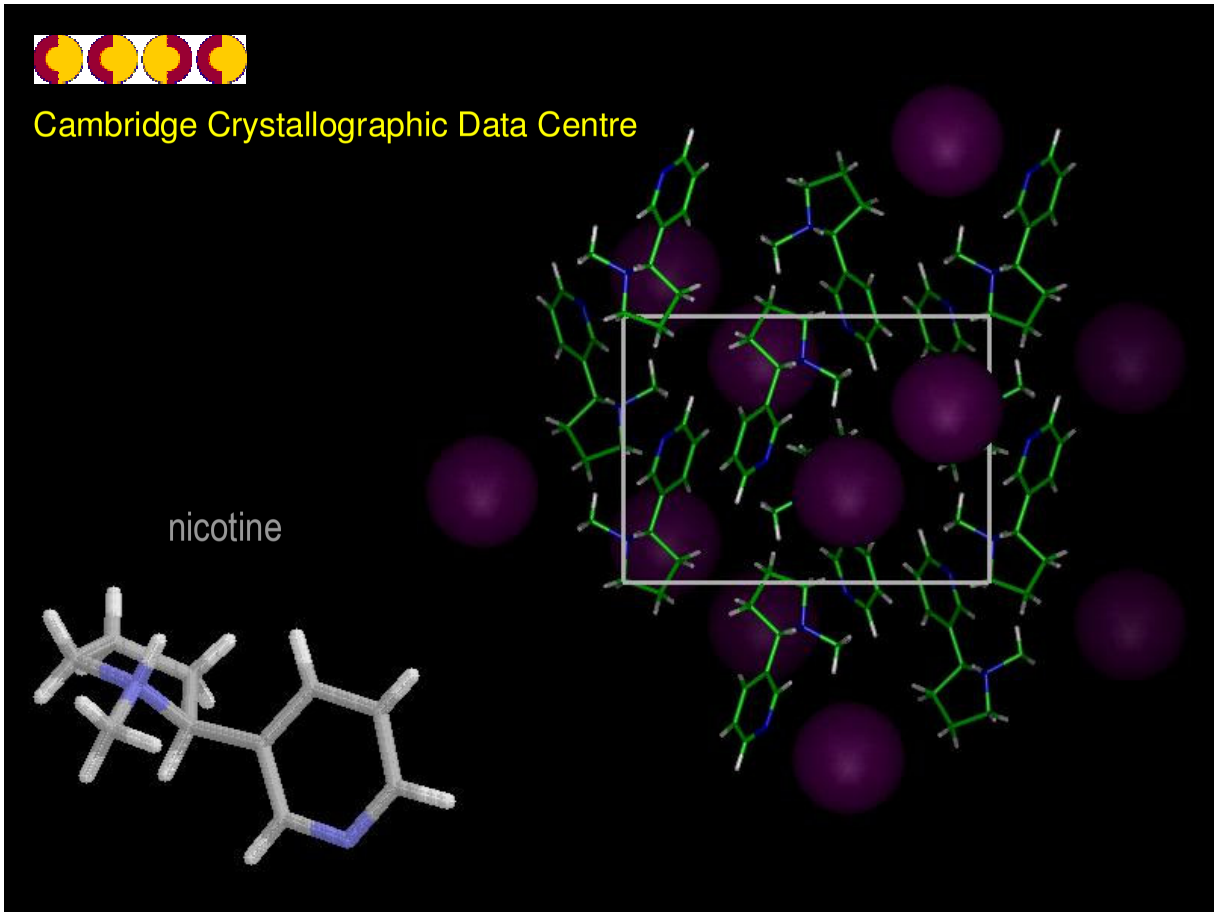


Cambridge Crystallographic Data Centre

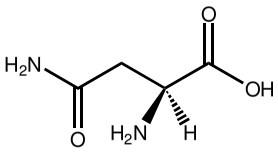
http://www.ccdc.cam.ac.uk/

```
#DOXSIS  33870428        16  9  0  0  0  4  4 28  0  0 30132200000010000000000086
137(
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
```

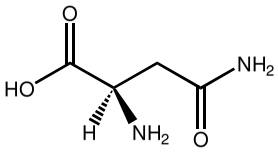


The Importance of Chirality

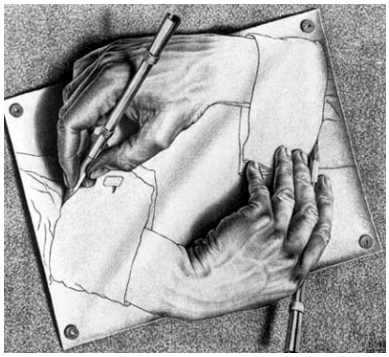


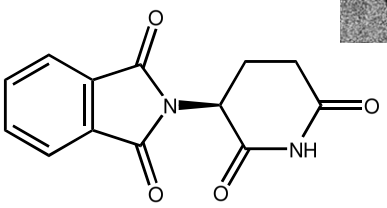
Bitter

Asparagine



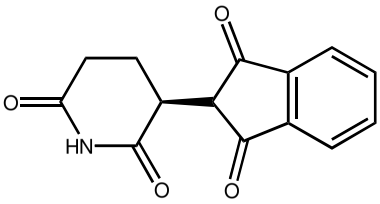
Sweet





Extreme teratogen

Thalidomide

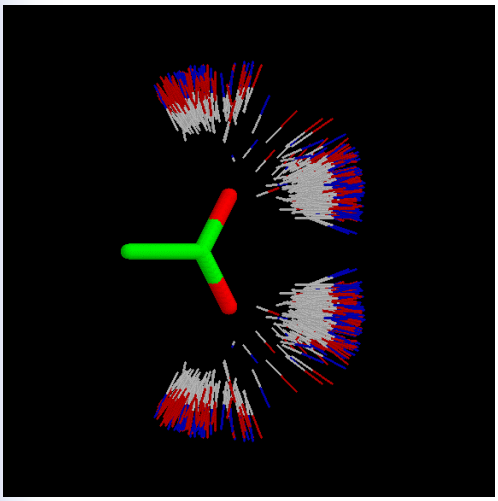


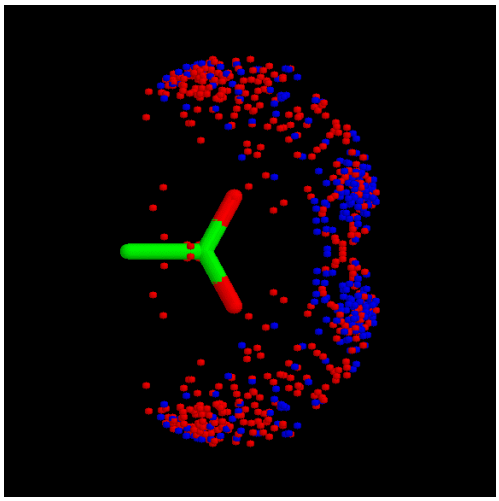
Anti-morning sickness

and SUPERSTAR

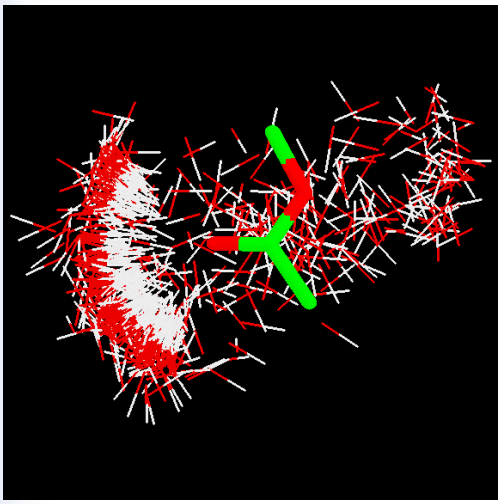
- IsoStar - knowledge base of information about intermolecular interactions
- SuperStar - program for predicting binding points in an enzyme active site
- SuperStar predictions based solely on IsoStar data

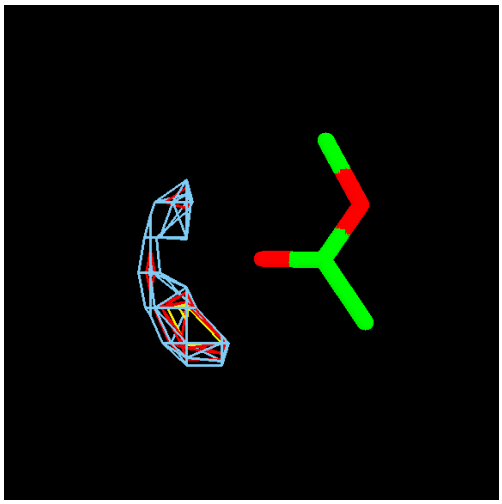
IsoStar Scatterplots





IsoStar Density Surfaces

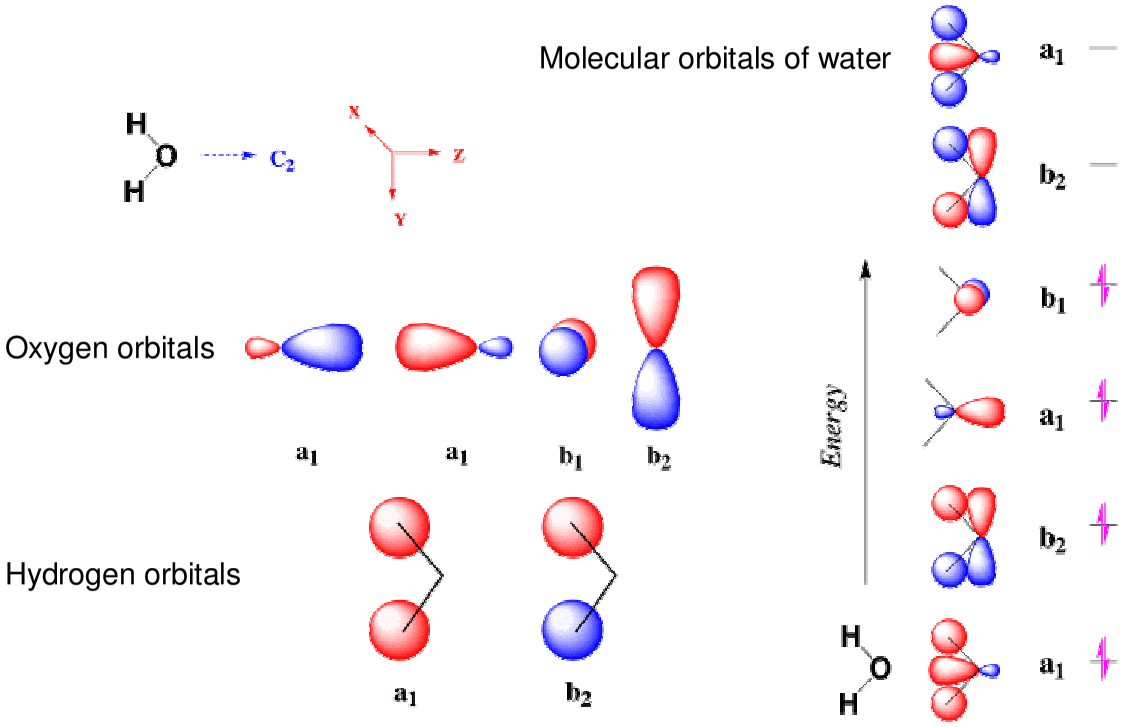




A probability surface derived from the observed positions of hydrogen-bonding hydrogen atoms around aliphatic esters.

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

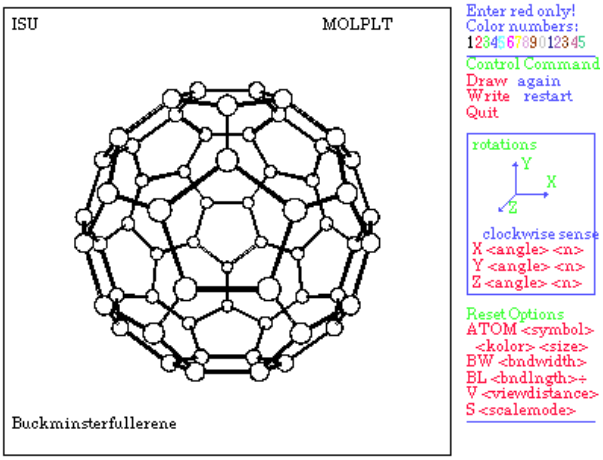
  BLYP - developed by Becke, Lee, Yang and Parr

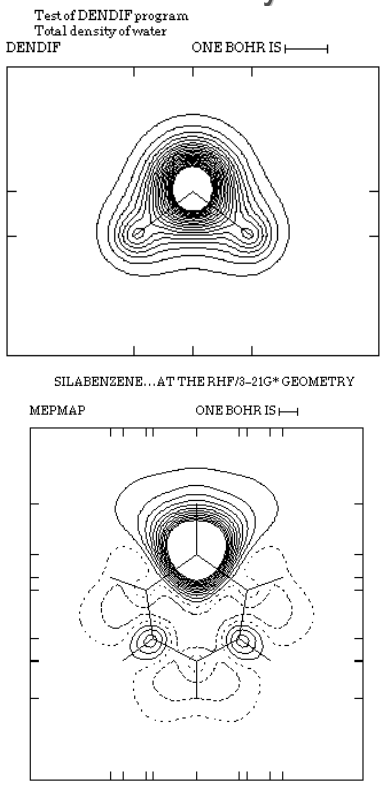
  B3LYP - a modification of BLYP in which a 3-parameter functional developed by Axel Becke is used.

GAMESS

General Atomic and Molecular Electronic Structure System

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



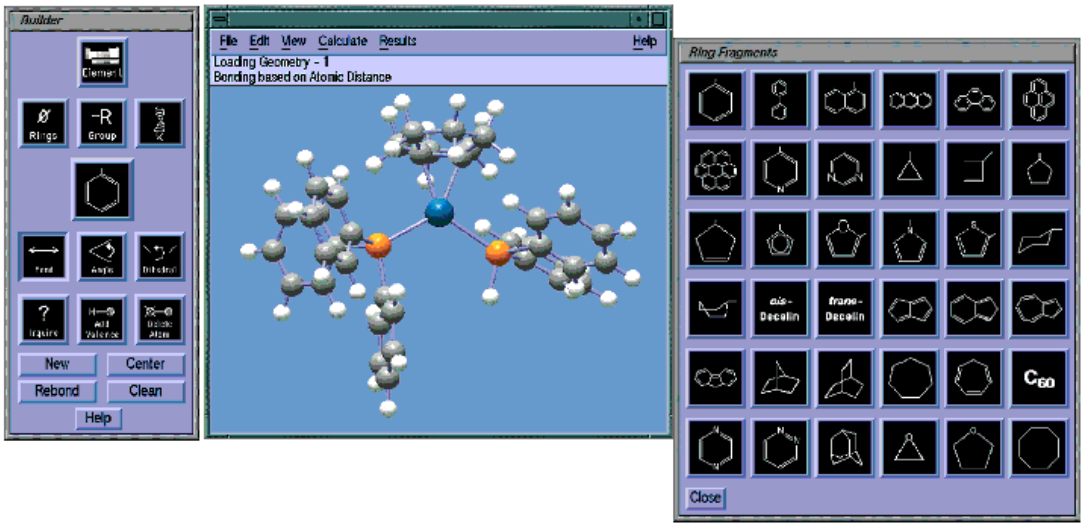


Gaussian

http://www.gaussian.com/

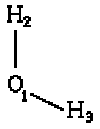
Spartan

http://www.wavefun.com/



Some sample Gaussian z-matrices

Water (C_{2v})



With variables:

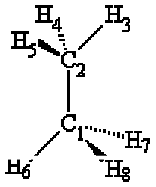
```
o
h 1 l1
h 1 l1 2 a1

l1 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 l1
h 2 l2 1 a1
h 2 l2 1 a1 3  120.0
h 2 l2 1 a1 3 -120.0
h 1 l2 2 a1 3  180.0
h 1 l2 2 a1 6  120.0
h 1 l2 2 a1 6 -120.0

l1 1.54
l2 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)

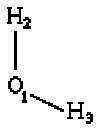
AM1 Austin Model 1 (Developed by Michael Dewar and Andrew Holder in 1986)

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

Sample input for MOPAC

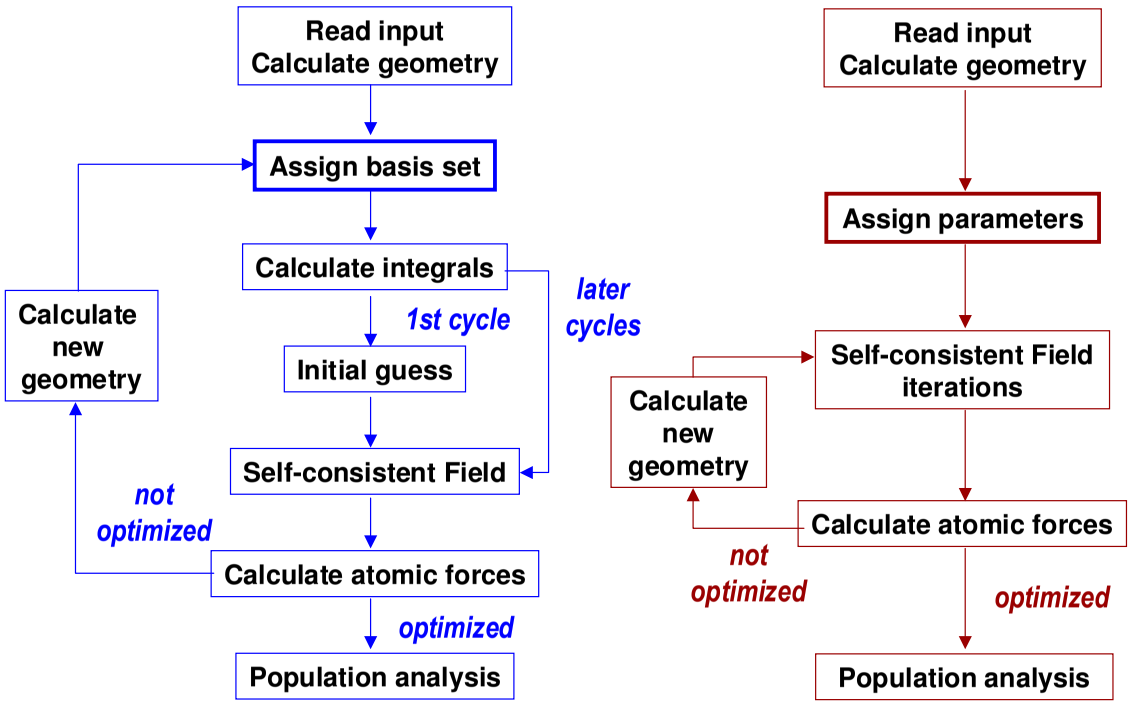
```
PM3 EF PRECISE
H2O (water)
MOPAC input as a Z-matrix
O
H          0.96000  1                                    1
H          0.96000  1     104.00000  1                   1 2
```

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**



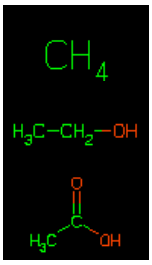
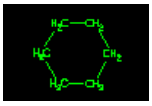
SMILES

Simplified Molecular Input Line Entry Specification

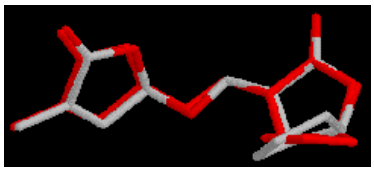
Rules

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.

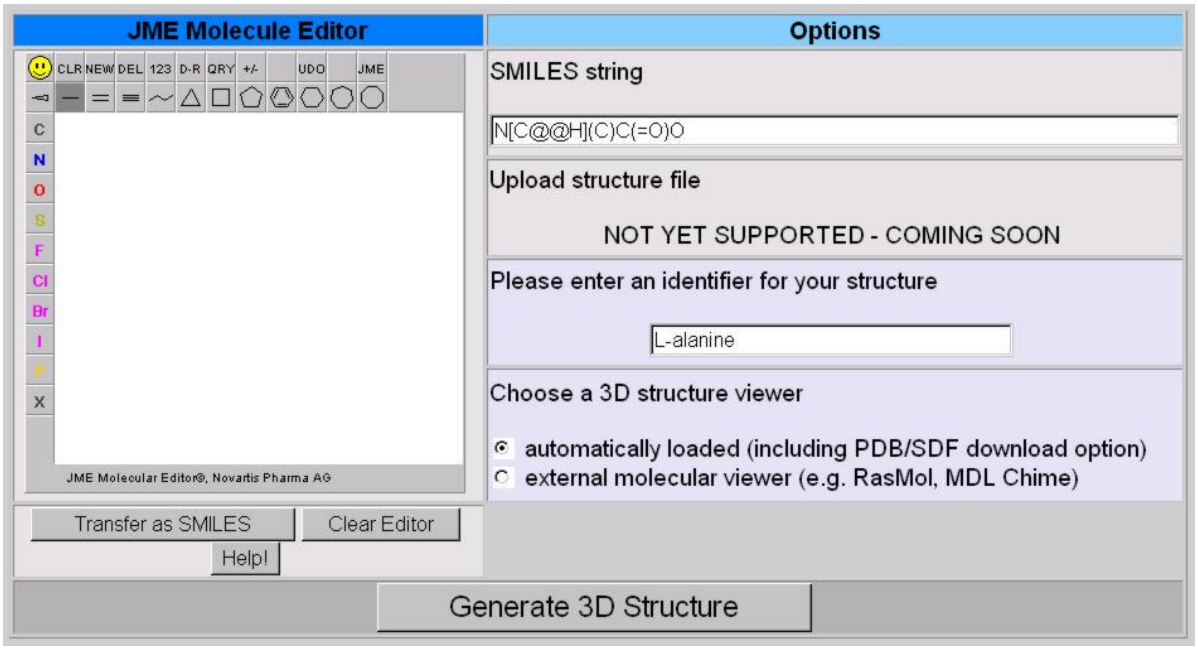
Examples

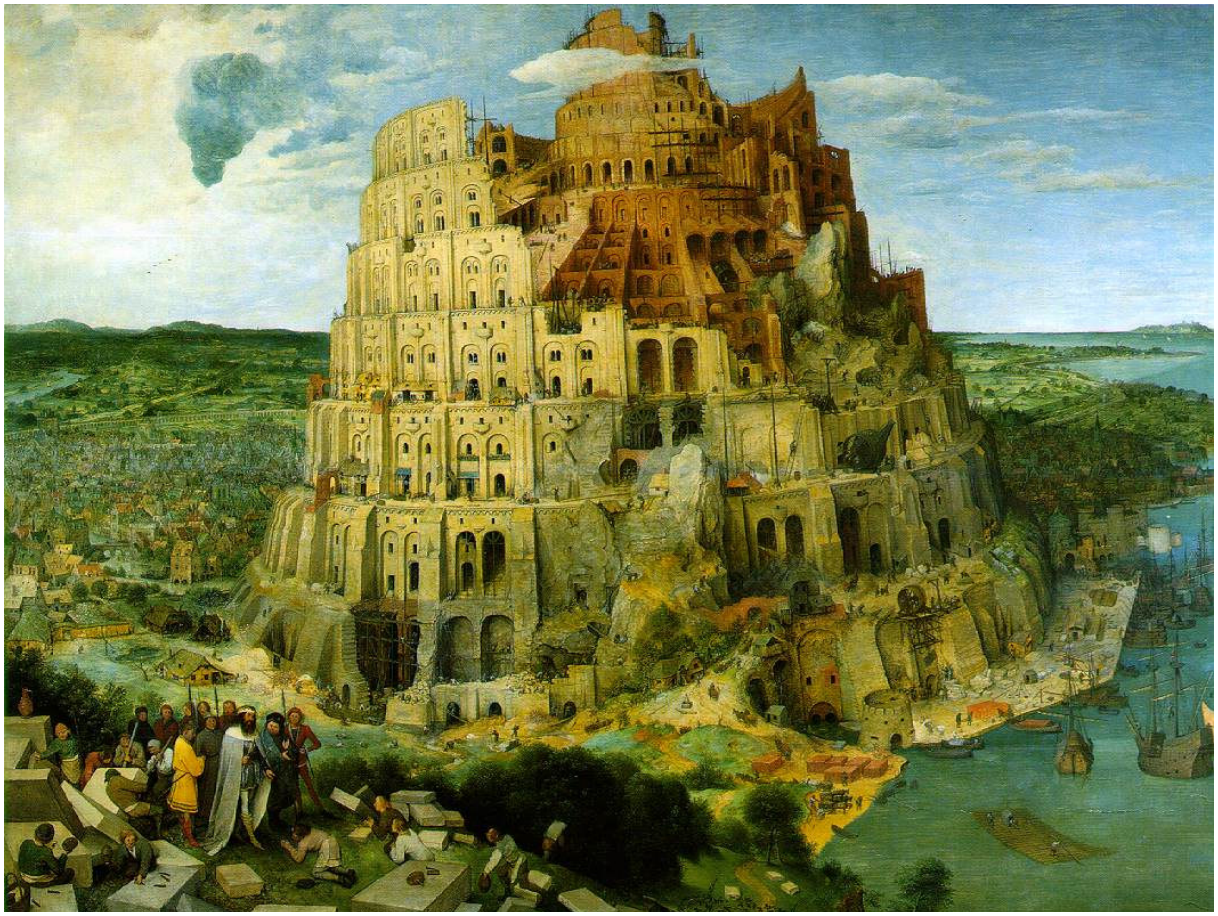
Depiction	SSMILES	Name	Remark
	C	methane	hydrogens fill normal valence
	CCO	ethanol	a single bond is assumed to join adjacent atoms
	CC(=O)O	acetic acid	parentheses are used to indicate branching
	C1CCCCC1	cyclohexane	bonds can also be represented by pairs of matching digits

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
```

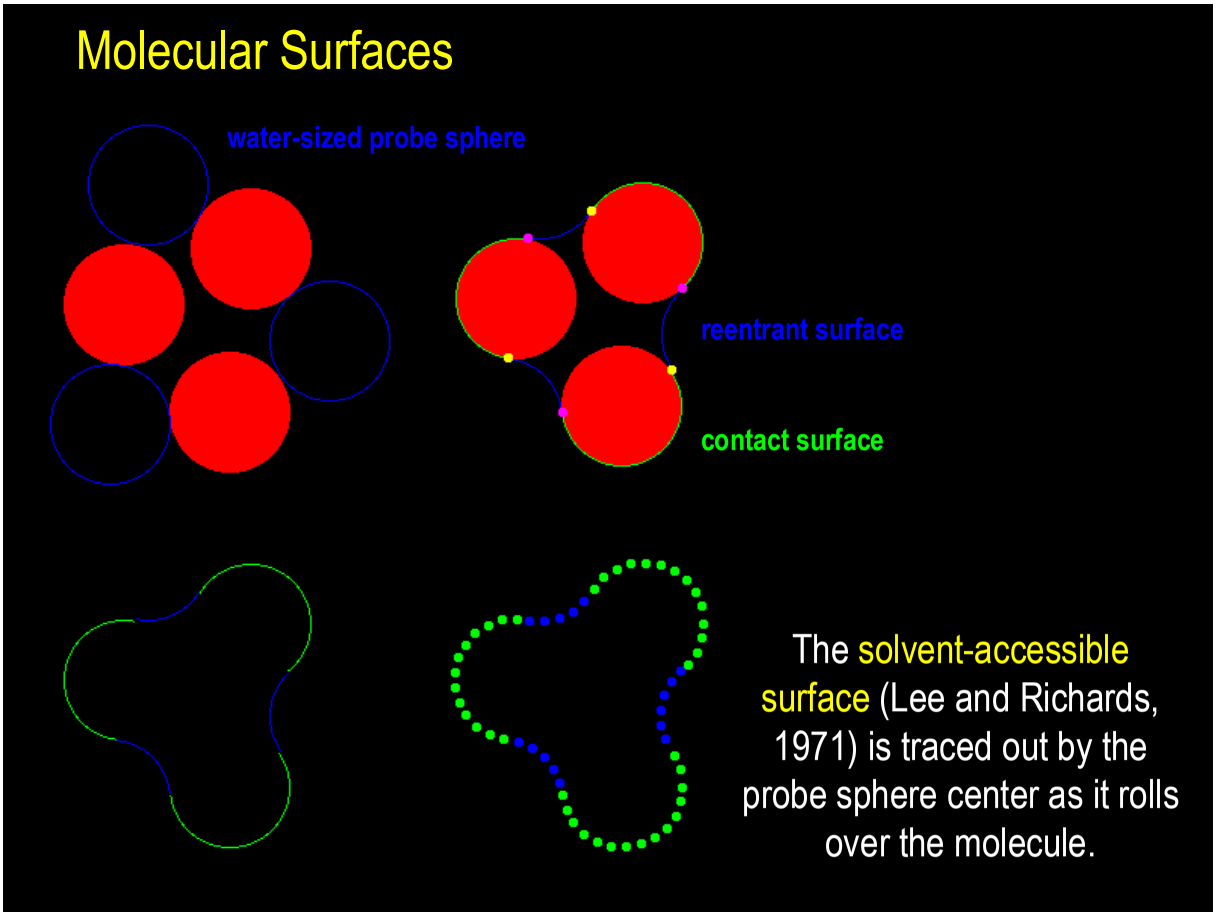
BABEL A program designed to interconvert a number of file formats currently used in molecular modelling

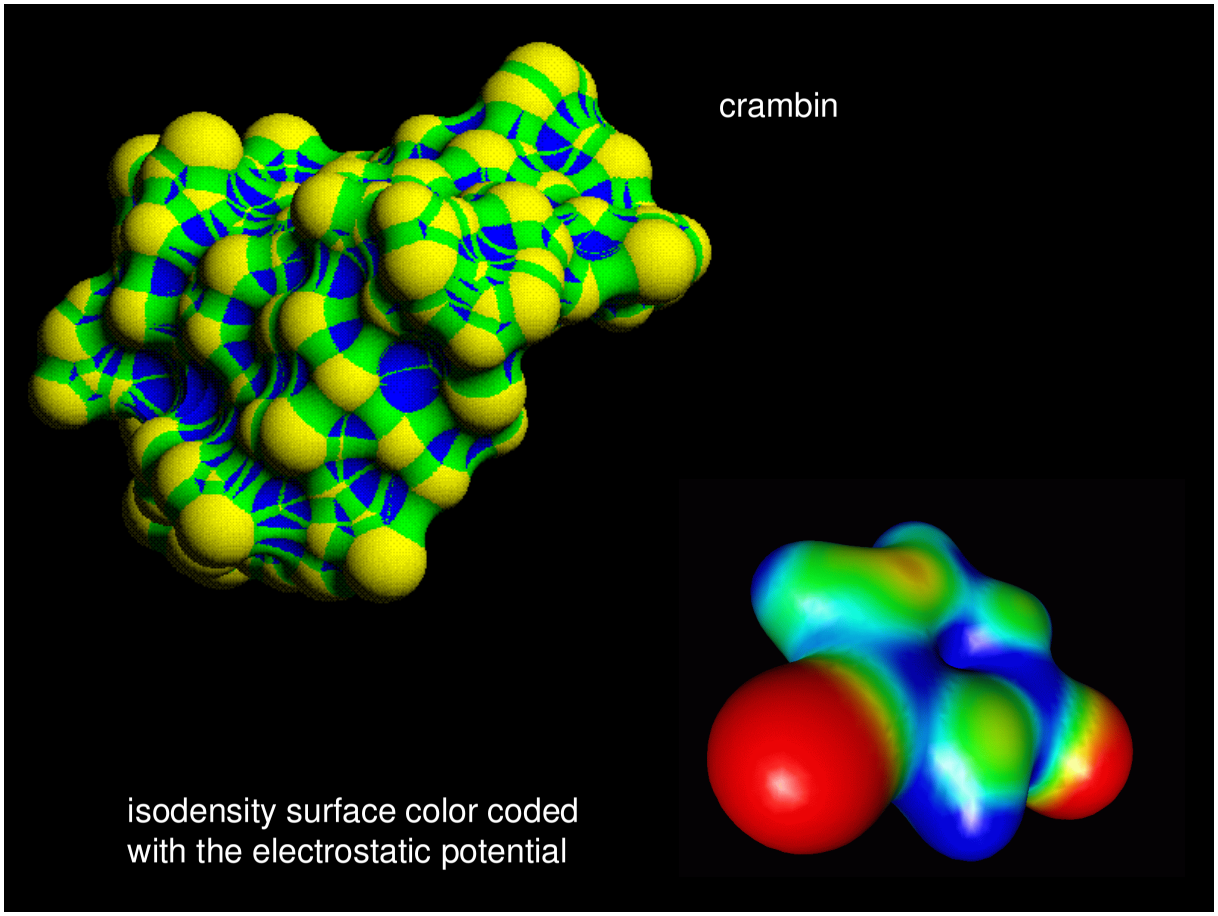
Output type codes:

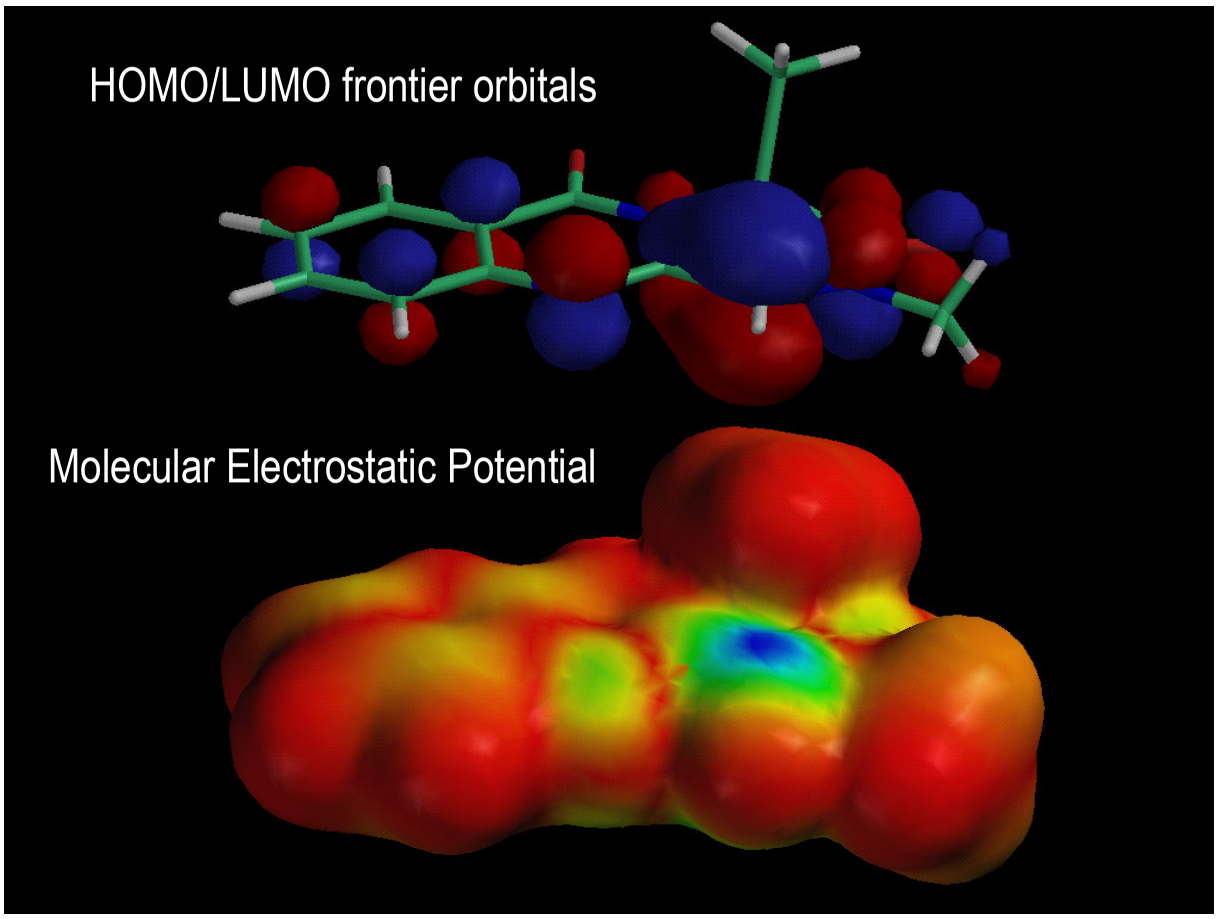
```
diag -- DIAGNOTICS file
t -- Alchemy file
bs -- Ball and Stick file
bmin -- Batchmin Command file
caccrt -- Cacao Cartesian file
cacint -- Cacao Internal file
cache -- CAChe MolStruct file
c3d1 -- Chem3D Cartesian 1 file
c3d2 -- Chem3D Cartesian 2 file
d -- ChemDraw Conn. Table file
con -- Conjure 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
```

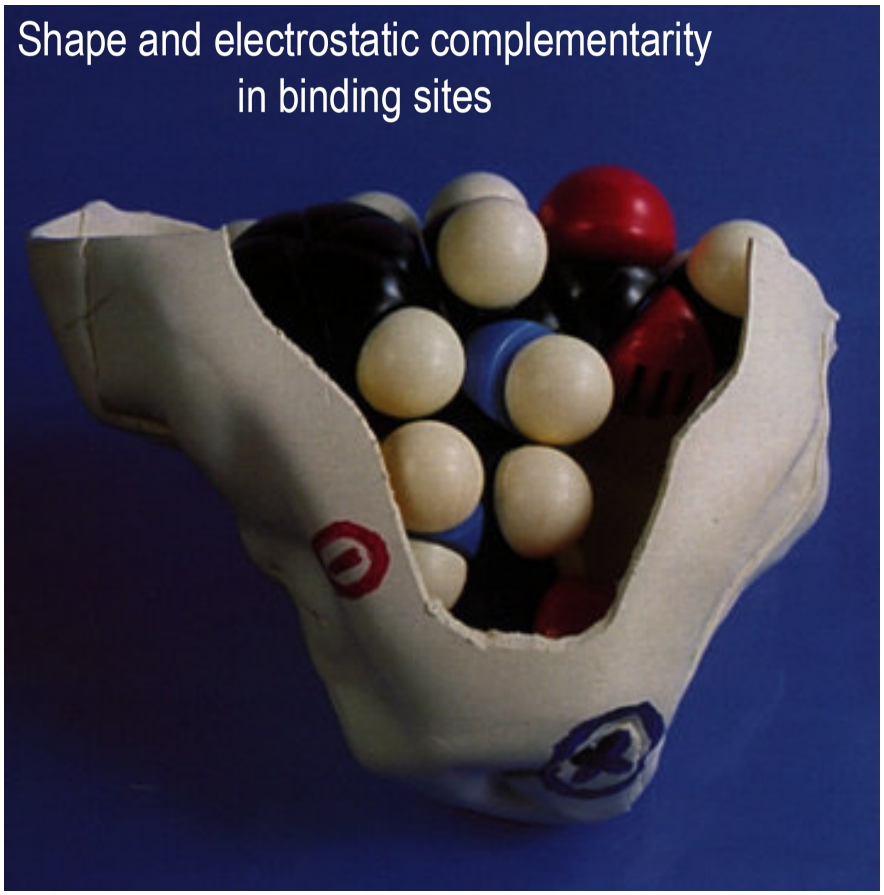
ftp://ccl.osc.edu/pub/chemistry/software/UNIX/babel/

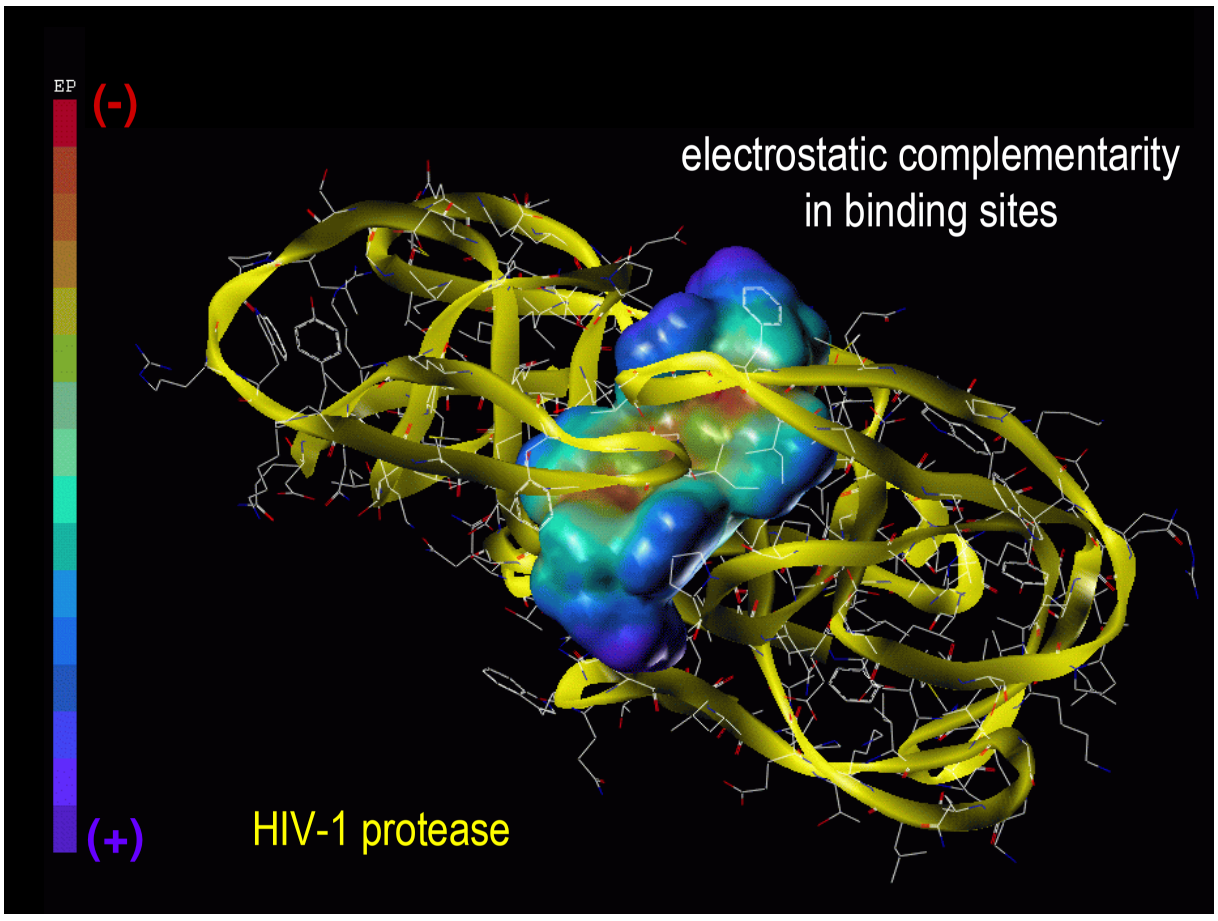
Molecular Surfaces




crambin
isodensity surface color coded
with the electrostatic potential


HOMO/LUMO frontier orbitals
Molecular Electrostatic Potential


Shape and electrostatic complementarity
in binding sites


EP
(-)
(+)
electrostatic complementarity
in binding sites
HIV-1 protease

MOLECULAR MECHANICS (MM)

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

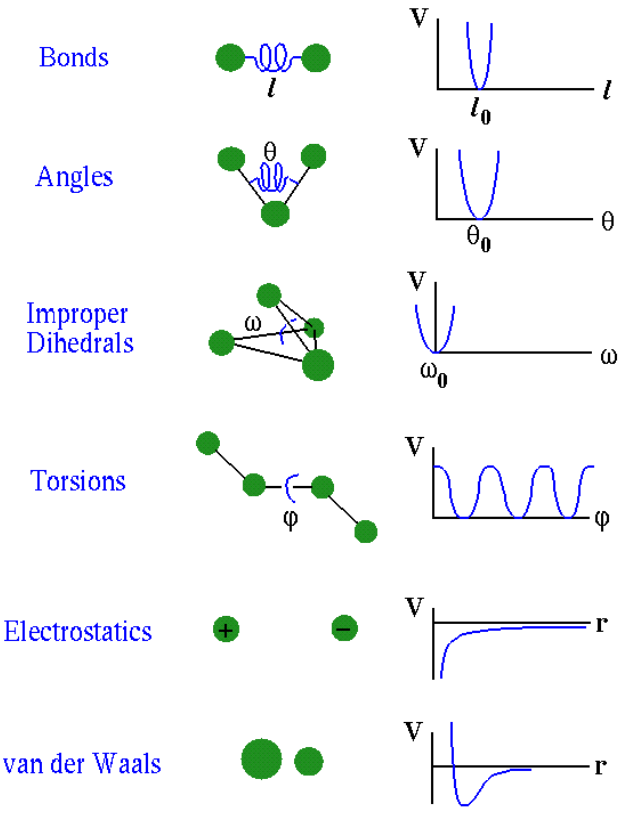
✓ The molecule is treated at the atomic level, i.e. the electrons are not treated explicitly.

✓ 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.

Empirical Potential Energy Function



Summary of interactions included in a representative molecular mechanics

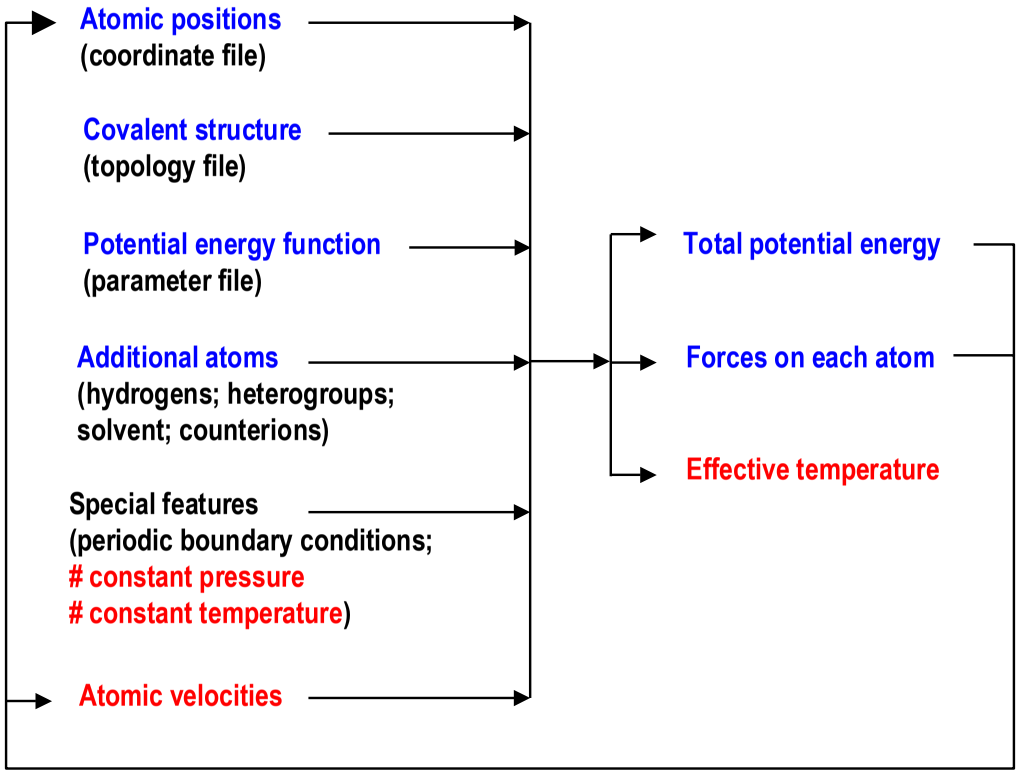
force field

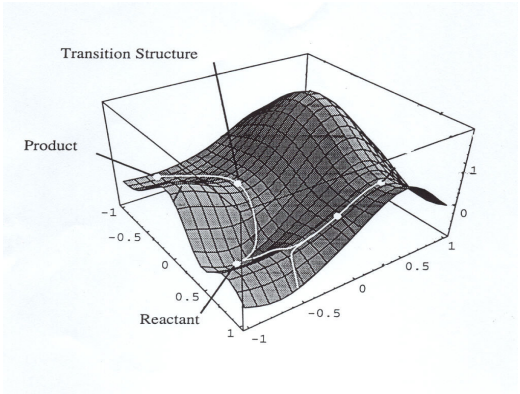
$$E_{\text{total}} = \sum_{\text{bonds}} K_r(r - r_{eq})^2 + \sum_{\text{angles}} K_\theta(\theta - \theta_{eq})^2 + \sum_{\text{dihedrals}} \frac{V_n}{2}[1 + \cos(n\phi - \gamma)] + \sum_{i<j} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\epsilon R_{ij}}\right] + \sum_{\text{H-bonds}} \left[\frac{C_{ij}}{R_{ij}^{12}} - \frac{D_{ij}}{R_{ij}^{10}}\right]$$

✓ The empirical potential energy function is differentiable with respect to the atomic coordinates.

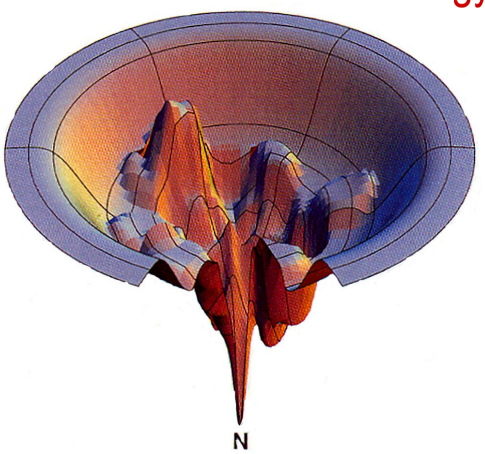
✓ This gives the value and the direction of the force acting on each atom and can thus be used in a molecular dynamics simulation.

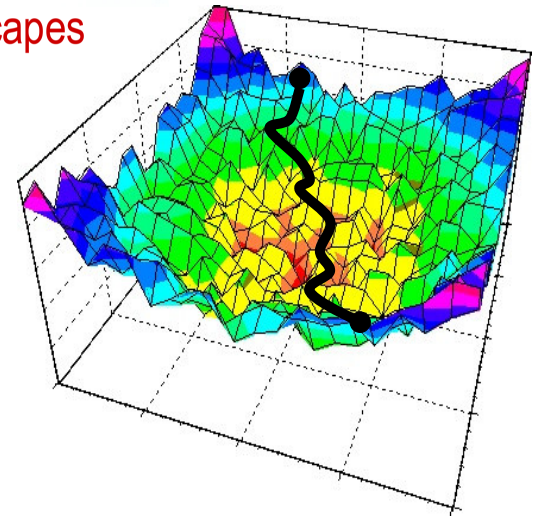
ALGORITHMS FOR ENERGY MINIMIZATION AND MOLECULAR DYNAMICS

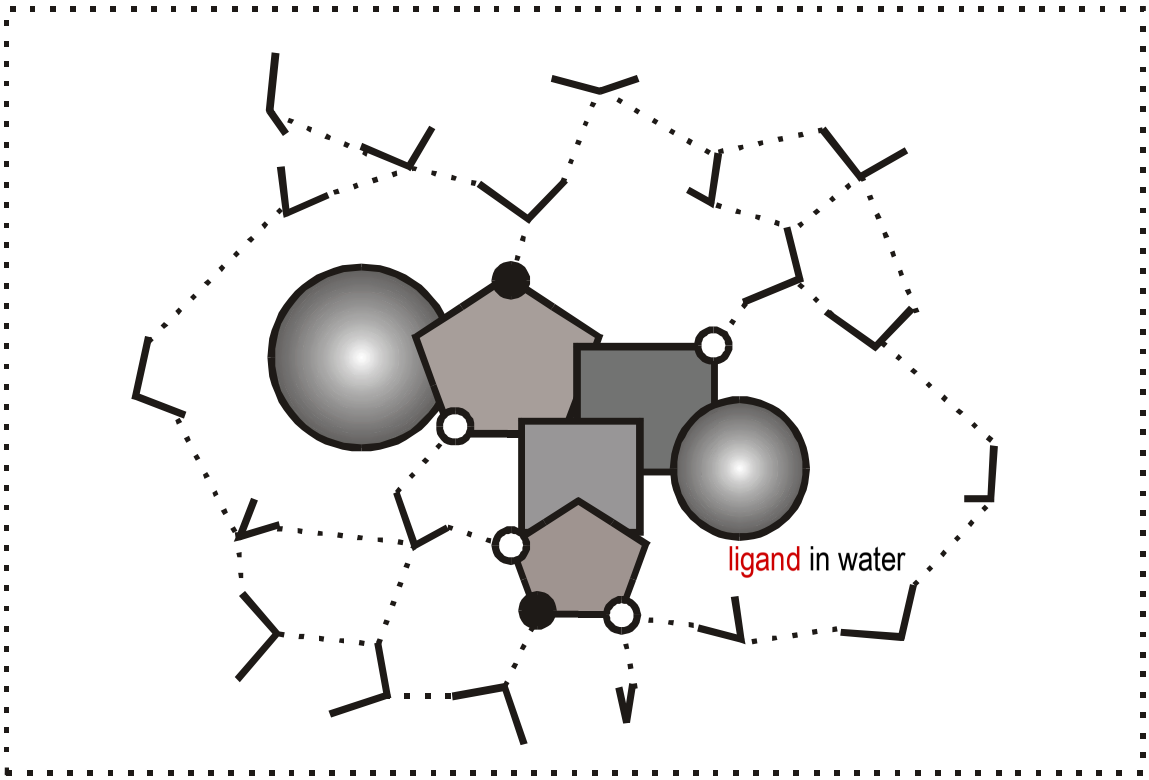


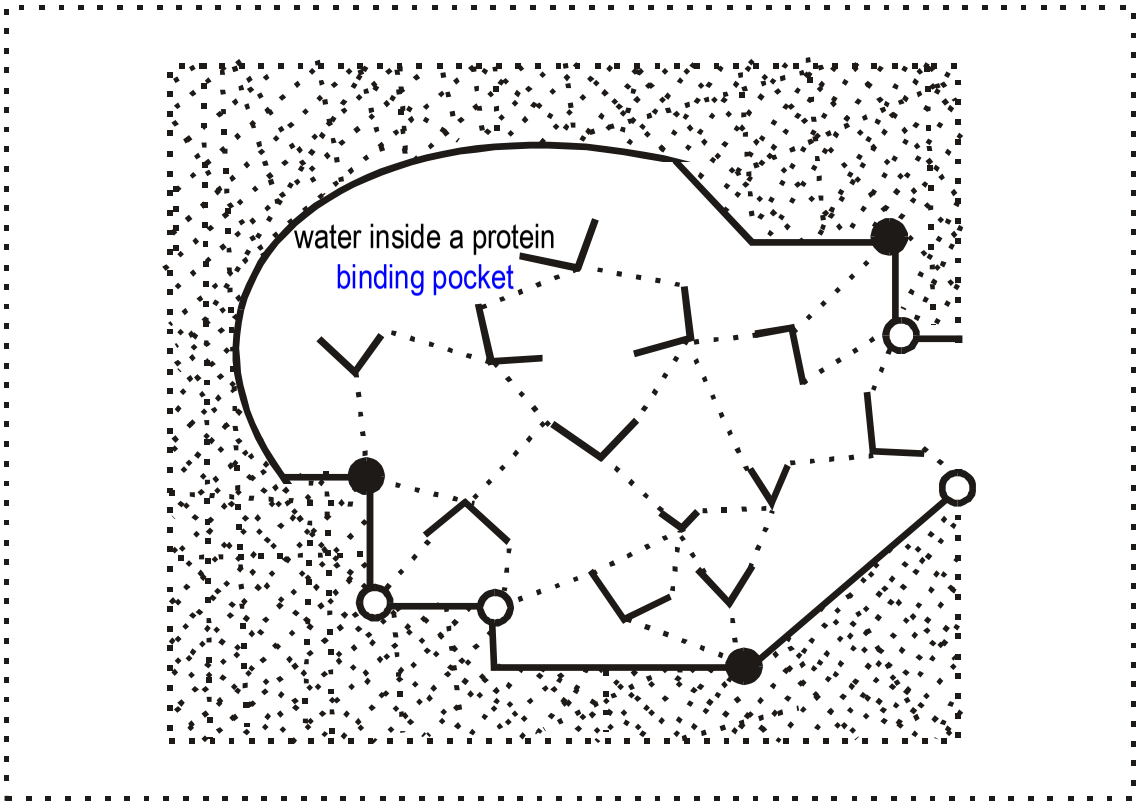


Energy landscapes

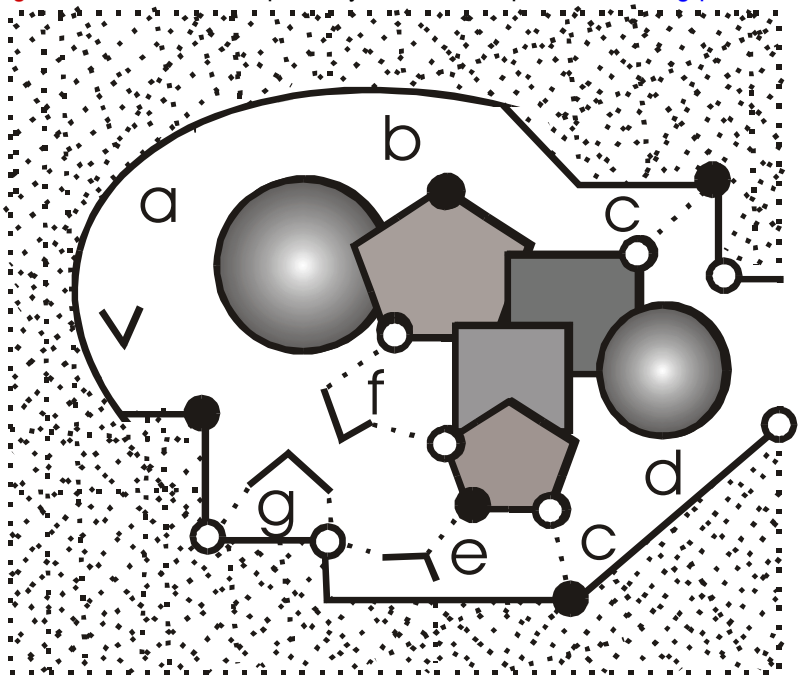




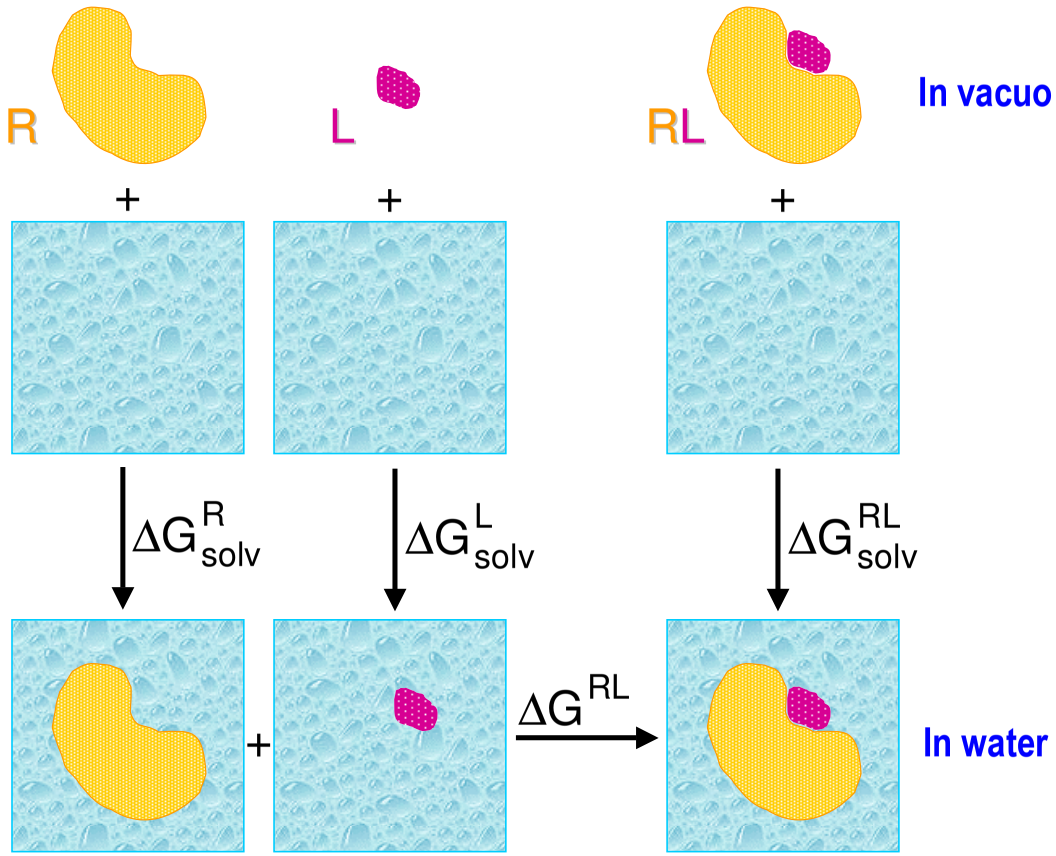

ligand in water


water inside a protein
binding pocket

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.



"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.***

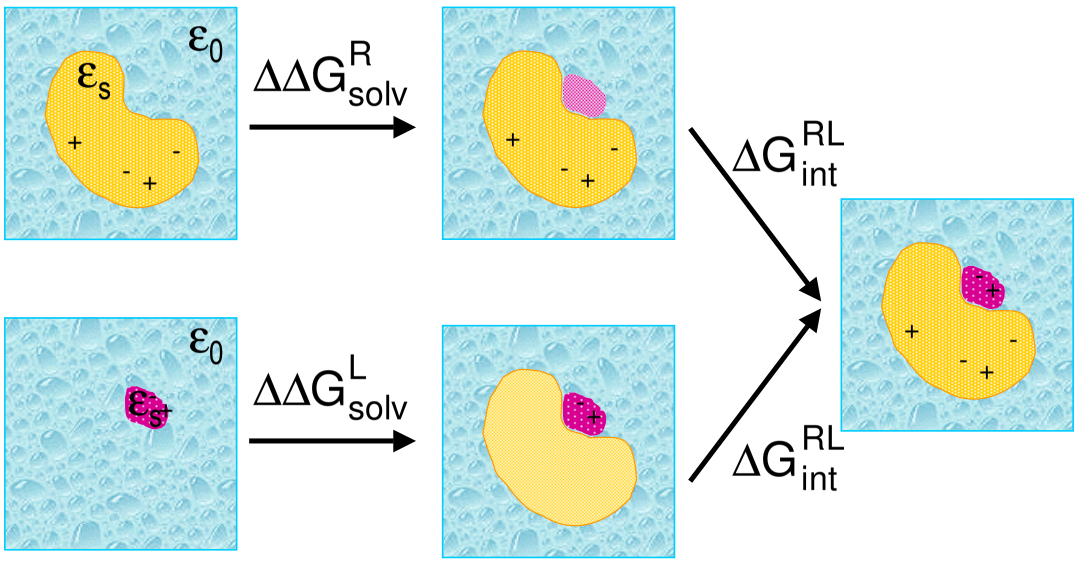
- Gilson, M. K.; Sharp, K. A.; Honig, B. H. "Calculating the Electrostatic Potential of Molecules in Solution: Method and Error Assessment" ***J. Comput. Chem. (1987) 9, 327-335.***

- 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.*



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 ΔK_d	**Binding energy** ΔG (kcal/mol)
2x	0.5
5x	1.0
13x	1.5
29x	2.0
68x	2.5
158x	3.0

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

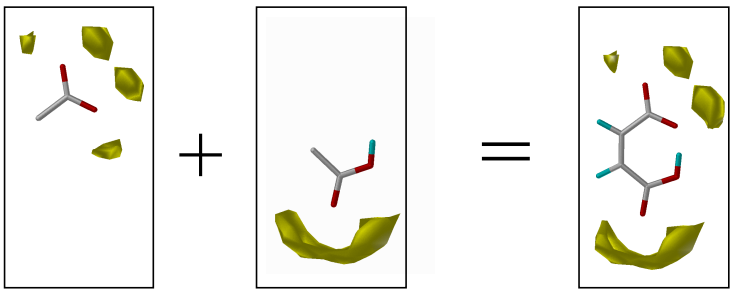
"GRID: A Computational Procedure for Determining Energetically Favorable Binding Sites on Biologically Important Macromolecules"

Peter Goodford, Oxford University

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

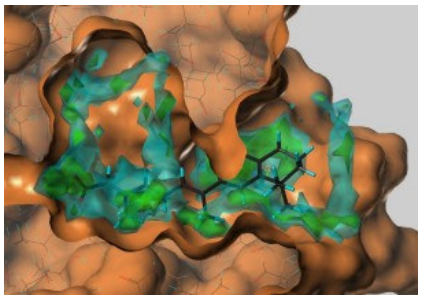
http://www.moldiscovery.com/

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+

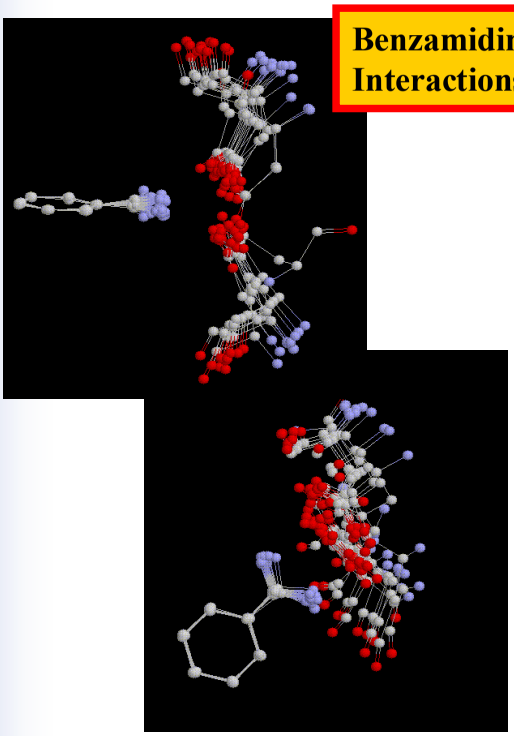
- Protein-ligand database system
- Based on original software developed by Manfred Hendlich and colleagues at Merck and Marburg University
- Enables searching of PDB and of in-house proprietary databases

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

Some Relibase+ Options

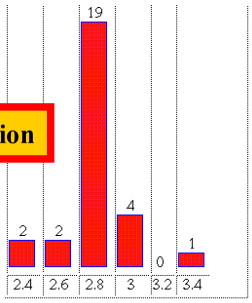
- Text searching
- Sequence searching
- 2D substructure and similarity searching
- 3D substructure searching
- Logical combination of hit lists
- Searching for intermolecular interactions
- Auto-superposition of similar binding sites
- Scripting facility based on Python

Analysis of 3D Queries

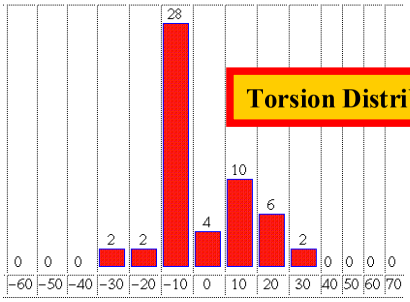


Benzamidine-Carboxylate Interactions

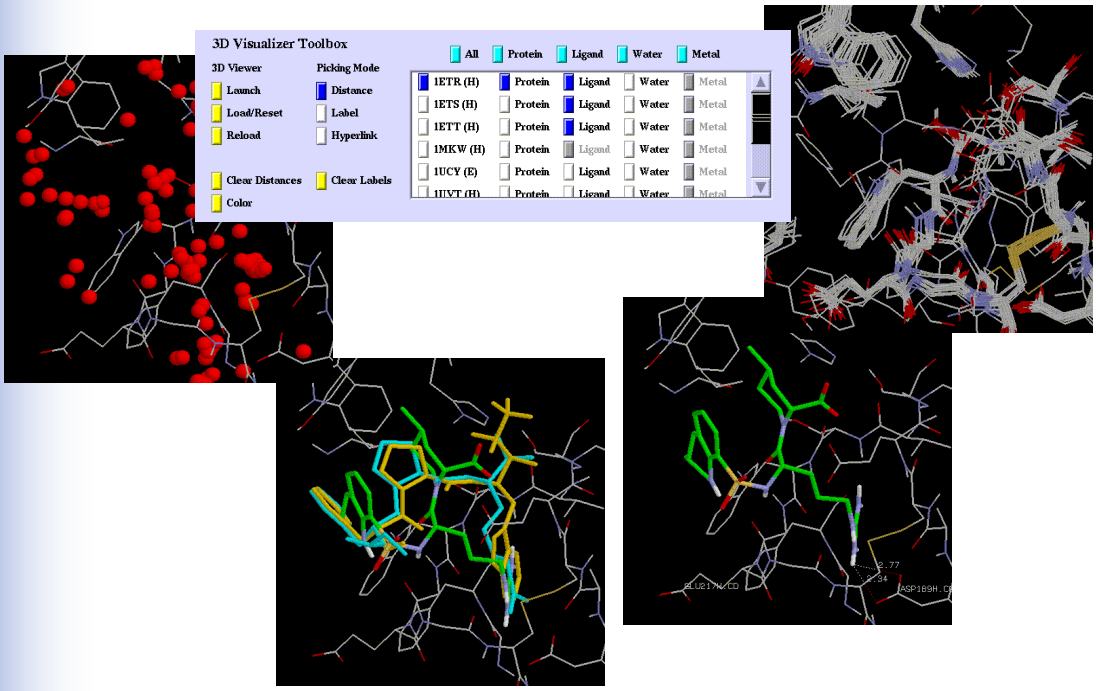
Distance Distribution



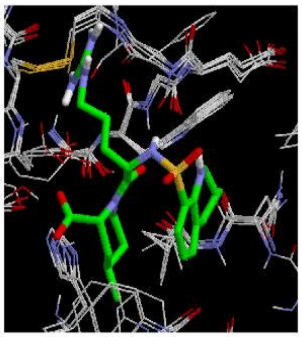
Torsion Distribution



Binding Site Superposition



Relibase



Relibase: A program for searching protein-ligand databases.

Version 4.0, October 2000

Relibase is copyright Manfred Hendlich 1994-1999 and Cambridge Crystallographic Data Centre 1999, 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/

LIGPLOT

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

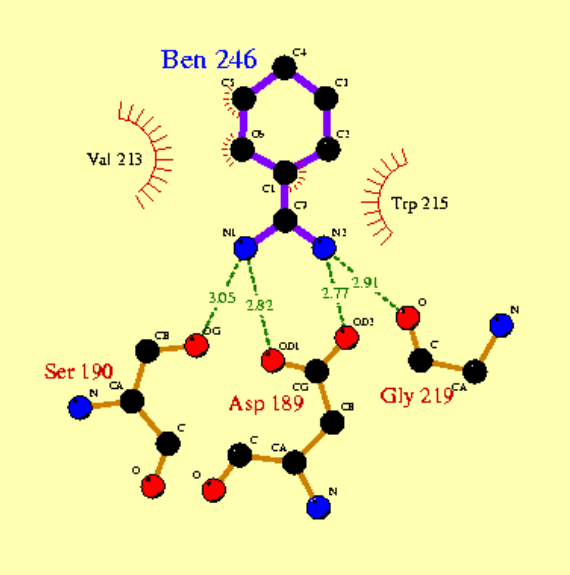
Program for automatically plotting protein-ligand interactions (by A. Wallace & R. Laskowski)

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.

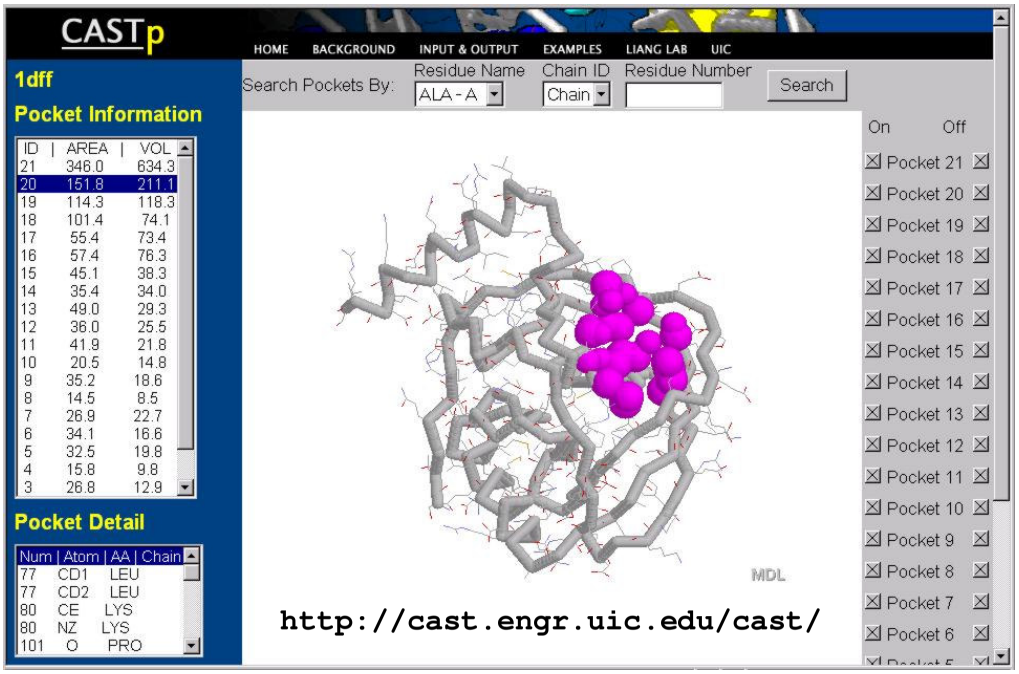


benzamidine (PDB code 2TBS)

CASTp

A Server for Identification of Protein Pockets & Cavities

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



Ligand Protein Database

http://lpdb.scripps.edu

variables	VIP*
number of hydrogen and ionic bonds	1.50
interaction surface area (Å^2)	1.31
calculated molecular refractivity (ligand)	1.31
molecular weight (ligand)	1.29
number of atoms (ligand)	1.28
number of donors (ligand)	1.20
number of rotatable bonds (ligand)	1.16
number of acceptors (ligand)	1.14
ClogP (ligand)	1.01

* variable influence on projection parameter

O. Roche, R. Kiyama & C. L. Brooks
J. Med. Chem. **44**, 3592-3598 (2001)

Break - Descanso

Top 10 ways to tell you drink too much coffee

10 Juan Valdéz names his donkey after you
9 You get a speeding ticket even when you're parked
8 You grind your coffee beans in your mouth
7 You sleep with your eyes open
6 You watch videos in fast-forward
5 You lick your coffeepot clean
4 Your eyes stay open when you sneeze
3 The nurse needs a scientific calculator to take your pulse
2 You can type sixty words a minute with your feet
1 You can jump-start your car without jumper cables.