Molecular structure
and properties calculations

> *In theory, there is no difference between theory and practice.*
> *In practice, there is.*
> Yogi Berra

Computational chemistry
Molecular modeling
Molecular quantum mechanics
Quantum chemistry

Computational chemistry (from wikipedia)

- a branch of chemistry that uses computers to assist in solving chemical problems.
- uses the results of theoretical chemistry, incorporated into efficient computer programs, to calculate the structures and properties of molecules and solids.
- it can (in some cases) predict hitherto unobserved chemical phenomena.
- widely used in the design of new drugs and materials.

Molecular modelling

- theoretical methods and computational techniques used to model or mimic the behaviour of molecules.

- used for studying molecular systems ranging from small chemical systems to large biological molecules and material assemblies.

Major goals of computational chemistry

- to create efficient **mathematical approximations** and **computer programs**
  - to calculate the **properties of molecules**
  - to apply these programs to concrete physico-chemical systems

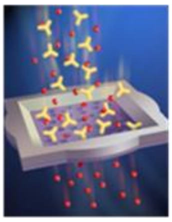
Properties:

- structure (conformers, tautomers)
- absolute and relative energies
- interaction energies
- electronic charge distributions
- dipoles and higher multipole moments
- vibrational spectra (frequencies and intensities)
- UV-Vis, ESR, and NMR spectra

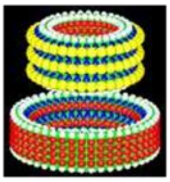
- MO energies and shapes (UPS and XPS) spectra
- reactivity
- etc.

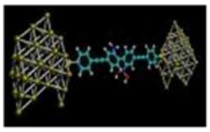
Molecular structures and molecular properties by quantum chemical methods

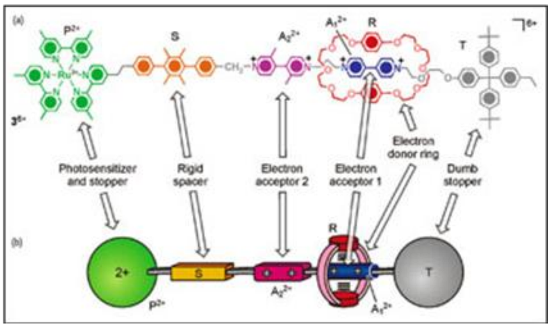
Bio



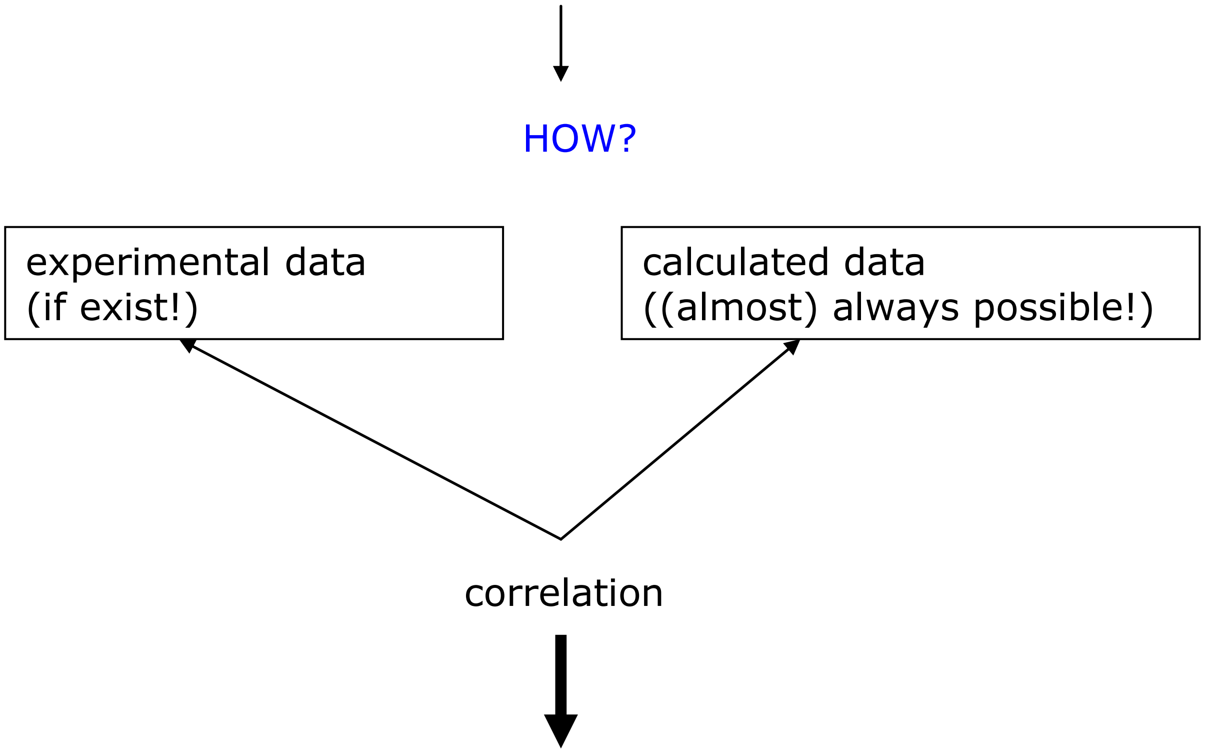
Nanotechnology







STRUCTURE ⇔ PROPERTIES



HOW?

experimental data
(if exist!)

calculated data
((almost) always possible!)

correlation

Implications:

Why computing molecules?
Electronic structure calculations - waht are they useful for?

- molecular modeling (drug design)
- predicting structure and properties (new materials)
- confirm and complement (explain) the experimental data
- structures for highly reactive species (free radicals)
- ...

Computational chemistry is one of the most rapidly advancing and exciting fields in the natural sciences today.
"Many experimental chemists use various kinds of spectroscopy in their research even though they are not spectroscopists. In a similar manner, more and more scientists are applying computational techniques as another weapon in their arsenal"

Delano P. Chong in *Recent Advances in Density Functional Methods, Part I*, World Scientific, 1995

METHODS

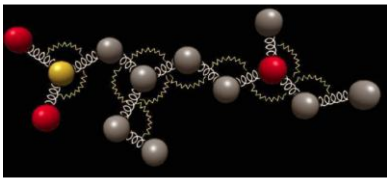
1. EMPIRICAL APROACHES

a) molecular mechanics (MM methods) (empirical force field calculations)

- based on classical-mechanical model of molecular structures

- potential energy of molecules is calculated based on a given force field

- electrons are not considered explicitely (they find their optimum distribution once the position of nuclei are known) = Born-Oppenheimer approximation



- atoms are considered spheres with net charges

- interactions
    - based on classical potentials
    - determine the spatial distribution of atoms and their energies

- the potential energy of the molecular system:

$$E = E_{covalent} + E_{non\text{-}covalent}$$

$$E_{covalent} = E_{bond} + E_{angle} + E_{dihedral}$$
$$E_{non\text{-}covalent} = E_{van\ der\ Waals} + E_{electrostatic}$$ (non-bonded energy)

- stretching energy: $\sum_{bonds} k_b (r - r_0)^2$
- bending energy: $\sum_{angles} k_\theta (\theta - \theta_0)^2$
- torsion energy: $\sum_{torsions} A(1 + \cos\Phi)^2$
- non-bonded energy: $\sum_i \sum_j -\frac{A_{ij}}{r_{ij}^6} + \frac{B_{ij}}{r_{ij}^{12}} + \sum_i \sum_j \frac{q_i q_j}{r_{ij}}$ +H Bond

Each molecular mechanics method is characterized by its particular **force field**

i) *a set of equations* defining how the potential energy of a molecule varies with the positions of its constituent atoms

ii) *a series of atom types*, definig the characteristics of an element within a specific chemical context. The *atom type* depend on hybridization, charge and types of the other atoms to which it is bonded.

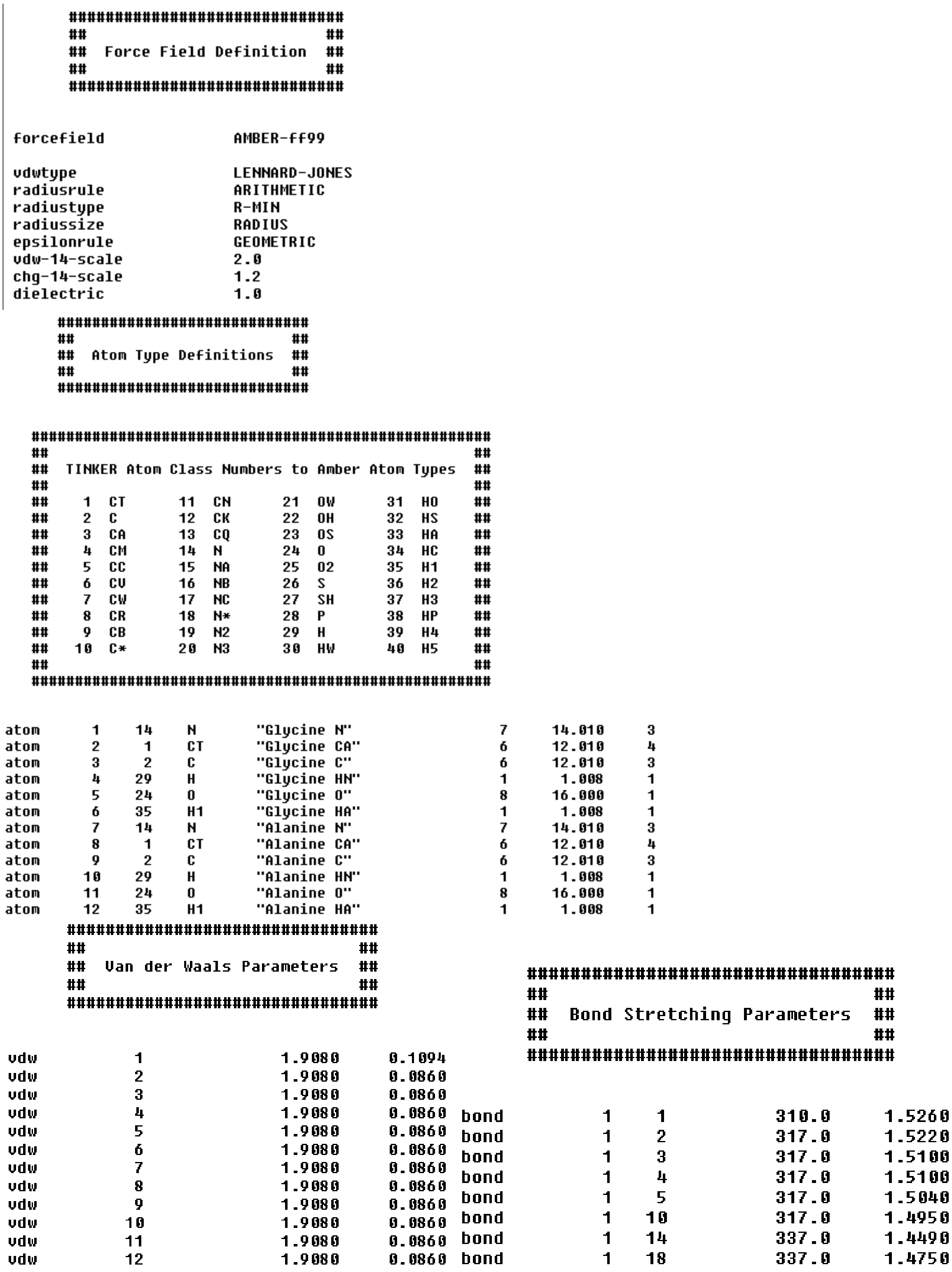
iii) one or more *parameter sets* that fit the equations and atom types to experimental data. A parameter set contains *force constants* (values used to relate atomic characteristics to energy components) and structural data (bond lengths, valence and dihedral angles)

Disadvantages of MM methods:

- each force field provides good results for a limited class of molecules (for wich it was parametrized). No force field is general!

- since empirical methods do not consider the electrons, they can not describe bond formation and breaking
- many molecular properties depending on subtle electronic effects are not reproducible by MM methods.

See AMBER parameter set!



```
      ##############################
      ##                          ##
      ##  Force Field Definition  ##
      ##                          ##
      ##############################


forcefield              AMBER-ff99

vdwtype                 LENNARD-JONES
radiusrule              ARITHMETIC
radiustype              R-MIN
radiussize              RADIUS
epsilonrule             GEOMETRIC
vdw-14-scale            2.0
chg-14-scale            1.2
dielectric              1.0
```

```
     ############################
     ##                        ##
     ##  Atom Type Definitions ##
     ##                        ##
     ############################
```

```
   ###################################################
   ##                                               ##
   ##  TINKER Atom Class Numbers to Amber Atom Types ##
   ##                                               ##
   ##    1  CT     11  CN     21  OW     31  HO     ##
   ##    2  C      12  CK     22  OH     32  HS     ##
   ##    3  CA     13  CQ     23  OS     33  HA     ##
   ##    4  CM     14  N      24  O      34  HC     ##
   ##    5  CC     15  NA     25  O2     35  H1     ##
   ##    6  CV     16  NB     26  S      36  H2     ##
   ##    7  CW     17  NC     27  SH     37  H3     ##
   ##    8  CR     18  N*     28  P      38  HP     ##
   ##    9  CB     19  N2     29  H      39  H4     ##
   ##   10  C*     20  N3     30  HW     40  H5     ##
   ##                                               ##
   ###################################################
```

```
atom      1    14    N     "Glycine N"          7     14.010    3
atom      2     1    CT    "Glycine CA"         6     12.010    4
atom      3     2    C     "Glycine C"          6     12.010    3
atom      4    29    H     "Glycine HN"         1      1.008    1
atom      5    24    O     "Glycine O"          8     16.000    1
atom      6    35    H1    "Glycine HA"         1      1.008    1
atom      7    14    N     "Alanine N"          7     14.010    3
atom      8     1    CT    "Alanine CA"         6     12.010    4
atom      9     2    C     "Alanine C"          6     12.010    3
atom     10    29    H     "Alanine HN"         1      1.008    1
atom     11    24    O     "Alanine O"          8     16.000    1
atom     12    35    H1    "Alanine HA"         1      1.008    1
```

```
      #################################
      ##                             ##
      ##  Van der Waals Parameters   ##
      ##                             ##
      #################################


vdw           1               1.9080      0.1094
vdw           2               1.9080      0.0860
vdw           3               1.9080      0.0860
vdw           4               1.9080      0.0860
vdw           5               1.9080      0.0860
vdw           6               1.9080      0.0860
vdw           7               1.9080      0.0860
vdw           8               1.9080      0.0860
vdw           9               1.9080      0.0860
vdw          10               1.9080      0.0860
vdw          11               1.9080      0.0860
vdw          12               1.9080      0.0860
```

```
##################################
##                              ##
##  Bond Stretching Parameters  ##
##                              ##
##################################


bond         1     1           310.0      1.5260
bond         1     2           317.0      1.5220
bond         1     3           317.0      1.5100
bond         1     4           317.0      1.5100
bond         1     5           317.0      1.5040
bond         1    10           317.0      1.4950
bond         1    14           337.0      1.4490
bond         1    18           337.0      1.4750
```

b) molecular dynamics (MD methods)

- consists in following the temporal evolution of a system of interacting atoms
  - by integrating their equations of motions: $F_i = m_i a_i$
  - being given a set of positions and velocities => evolution in time is completely determined
- forces acting on atoms are derived from the potential:

$$F_i = -\nabla_{r_i} V(r_1, \dots, r_N)$$

The simplest choice for *V* is to write it as a sum of pairwise interactions:

$$V(r_1, \dots, r_N) = \sum_i \sum_{j>i} \Phi\left(\left|r_i - r_j\right|\right)$$

The potential functions representing the non-bonded energy are formulated as a sum over interactions between the particles of the system. The simplest choice is the "pair potential", in which the total potential energy can be calculated from the sum of energy contributions between pairs of atoms.

Lennard-Jones potential:

$$\Phi_{LJ}(r) = 4\varepsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right]$$

The intramolecular bonding interactions must be also considered:

- stretching energy: $\sum_{bonds} k_b (r - r_0)^2$
- bending energy: $\sum_{angles} k_\theta (\theta - \theta_0)^2$
- torsion energy: $\sum_{torsions} A(1 + \cos \Phi)^2$

Force field are also nedeed in this case (CHARMM)

Advantages of MD methods

- can treat very large molecular systems (thousands of atoms)
- used in biophysics (proteins, viruses)

Disadvantages of MD methods

- not able to model bond forming/breaking since electronic structure methods does not enter these models

2. QUANTUM MECHANICAL APROACHES

a) semiempirical methods (AM1, PM3, PPP, INDO, MINDO, ...)
  - approximate methods in which some quantities (electron repulsion integrals) are taken from experiment, some small quantities are neglected and some quantities are estimated by fitting to experimental data; empirical parameters and functions are used to compensate for errors introduced by removing the integrals
  - model only the valence electrons
  - limited to hundred of atoms
  - can be used to study ground and excited molecular states
  - parametrized to reproduce experimental reference data or accurate high-level theoretical predictions

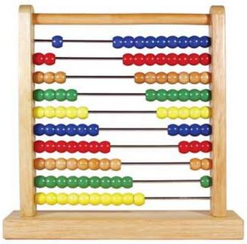
b) non empirical methods → Ab Initio ← wave function
→ DFT ← electronic density

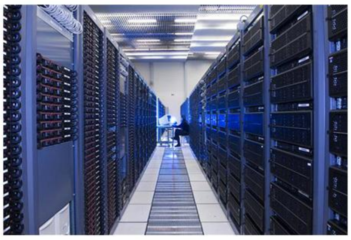
(quantum mechanics methods)
(electronic structure methods)
  - do not require empirical parameters
  - can be used for any molecular system
  - limited to tens of atoms
  - can be used to study ground and excited molecular states

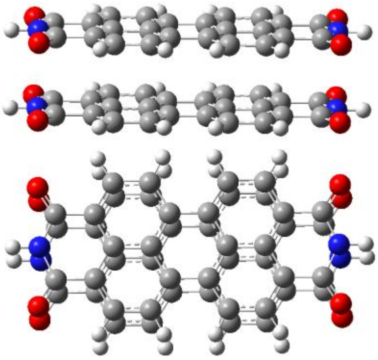
Quantum mechanics methods are based on the following principles:
- nuclei and electrons are distinguished from each other
- electron-electron and electron-nuclear interactions are considered explicitly.
- interactions are governed by nuclear and electron charges (potential energy) and electron motions
- interactions determine the spatial distribution of nuclei and electrons and their energies

What level of theory?

1. The choice of level of theory to use is usually determined by the **property** being calculated and the **cost** of the calculation in terms of computer usage.
2. In general **the higher the level of theory the greater the computer cost**. For example Hartree-Fock based ab-initio methods are more costly than semiempirical methods. Correlated methods, depending on the sophistication, can add additional cost to the Hartree-Fock method.
3. For the Hartree-Fock and Density Functional methods an additional factor is the **quality (size) of the basis sets** used. In general the larger the number of functions used in the basis set the more flexible (better) the basis set. Each additional function adds extra cost to the calculation.


MM methods


QM methods


Job CPU time
AM1: 6s
DFT: 55m 22s
MP2: 8h 14m 4s

PTCDI dimer

Type of calculations

1. Single point calculations

► *energy, wave-function and other requested properties at a single fixed geometry*

- usually done first at the beginning of a study on a new molecule to check out the nature of the wave-function
- it is also frequently carried out after a geometry optimization, with a larger basis set or more superior method than is possible with the basis set and method used to optimize geometry

These calculations are performed:

- to obtain basic information about a given molcule
- to check a molecular geometry to be used as a starting geometry in an optimization process
- to compute more accurate energies and other properties for an optimized geometry at a lower theoretical level.
- …

2. Geometry optimization calculations

►calculation of the wave-function and energy at a starting geometry and then searching for the geometry which will give the lowest energy (local or global energy minimum)

Potential energy surface (PES)

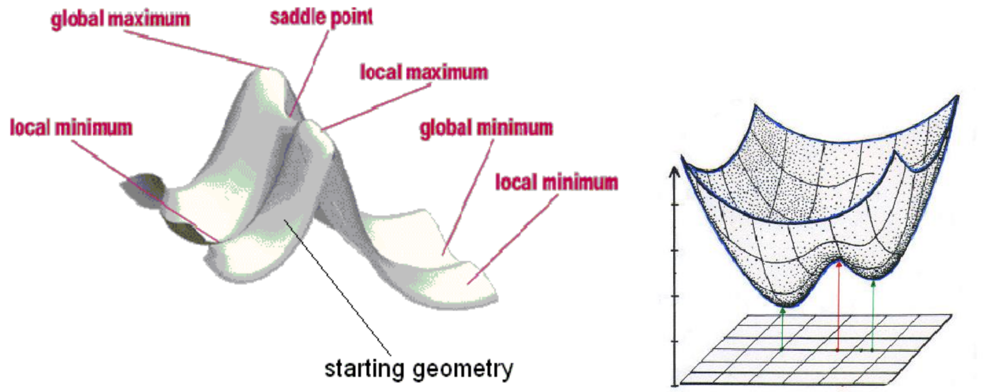
- specify the way in which the energy of a molecular system varies with small changes in its structure
  - mathematical relationship linking the molecular structure and the resultant energy

local minimum – the lowest energy point in a limited region of PES
global minimum – the lowest energy point anywhere on the PES

saddle points → maximum in one direction and minimum in all other directions
correspond to transition state structures

all minima + saddle points = stationary points ($\nabla E = 0 \quad (F = 0)$)

Geometry optimizations usually locate the stationary points closest to the starting geometry!!!



Potential energy surfaces

2-dim. PES
with two
minima
and one TS

Conformational search

- necessary to distinguish between a local minimum and the global minimum.

how?

- altering the initial geometry slightly (usually by dihedral angles) and then performing another optimization

3. Frequency calculations

- to predict the IR and Raman spectra of molecules (frequencies, intensities and normal modes)
- to compute polarizability and hyperpolarizability tensor
- to compute force constants for a geometry optimization
- to identify the nature of stationary points on the PES (check if an optimized geometry corresponds or not to an energy minimum)
- to compute zero-point vibrational energies, thermal energy corrections, enthalpy and entropy of the system

- should only be carried out at the geometry obtained from an optimization run and with the same basis set and method.

For a local or a global minimum all the calculated frequencies will be positive (real)

For any stationary point other than a minimum some of the frequencies will be negative (imaginary frequencies)

scaling factors see: CCCBDB (Computational Chemistry Comparison and Benchmark DataBase)
http://cccbdb.nist.gov/

4. Magnetic properties calculations

NMR spectra
    - chemical shifts, spin-spin couplings
ESR spectra
    - hyperfine coupling constants and hyperfine coupling tensors

Molecular properties calculated by Gaussian 2003

http://www.gaussian.com/g_ur/m_jobtypes.htm
Atomic charges: **Pop**
Dipole moment: **Pop**
Electron affinities via propagator methods: **OVGF**
Electron density: **cubegen**
Electronic circular dichroism: **TD**
Electrostatic potential: **cubegen**, **Prop**
Electrostatic-potential derived charges: **Pop=Chelp**, **ChelpG** or **MK**
Frequency-dependent polarizabilities/hyperpolarizabilities: **Polar CPHF=RdFreq**
High accuracy energies: **CBS-QB3**, **G2, G3**, **W1U**
Hyperfine coupling constants (anisotropic): **Prop**
Hyperfine spectra tensors (incl. *g* tensors): **NMR** and **Freq=(VibRot, Anharmonic)**
Hyperpolarizabilities: **Freq**, **Polar**
Ionization potentials via propagator methods: **OVGF**
IR and Raman spectra: **Freq**
Pre-resonance Raman spectra: **Freq CPHF=RdFreq**
Molecular orbitals: **Pop=Regular**
Multipole moments: **Pop**
NMR shielding and chemical shifts: **NMR**
NMR spin-spin coupling constants: **NMR=SpinSpin**
Optical rotations: **Polar=OptRot CPHF=RdFreq**
Polarizabilities: **Freq**, **Polar**
Thermochemical analysis: **Freq**
UV/Visible spectra: **CIS**, **Zindo**, **TD**
Vibration-rotation coupling: **Freq=VibRot**
Vibrational circular dichroism: **Freq=VCD**

Program packages in molecular electronic structure calculations

1. Gaussian
   http://www.gaussian.com/

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

3. DeFT
   http://lavoisier.dq.ufscar.br/download/chem/dft/

4. DALTON
   http://www.kjemi.uio.no/software/dalton/dalton.html

5. Mopac
   http://comp.chem.umn.edu/WWW/MOPAC/MOPAC.html

Molecular structure and properties visualization programs

1. GaussView
   http://www.gaussian.com

2. Molekel
   http://www.cscs.ch/molekel/

3. Raswin
   http://www.umass.edu/microbio/rasmol/getras.htm#raswin

4. Hyperchem
   http://www.hyper.com/

5. Molden
   http://www.ccl.net/cca/software/SOURCES/FORTRAN/molden/index.shtml

What shall we learn?

- the theory behind "molecular modeling"
- to use some molecular visualization packages
- to use program packages designed for molecular electronic structure theory
- to do calculations at different levels of theory and to interpret the results
- to make correlations between the experimental and theoretical data

Contents of the course

Hartree-Fock Theory
Basis sets
Electron Correlation Methods
Basis set superposition error
Density Functional Theory
Geometry optimizations
Calculation of vibrational spectra
Calculation of NMR and ESR spectra
Calculation of UV-VIS spectra

Can we do research?

pure theoretical studies
coupled experimental and theoretical investigation on the structure and properties of molecular systems

Where can we publish the results?

Journal of Molecular Structure
Journal of Molecular Structure (Theochem)
Journal of Molecular Spectroscopy
Chemical Physics
Chemical Physics Letters
Journal of Molecular Modelling
International Journal of Quantum Chemistry
Journal of Computational Chemistry
Journal of Chemical Physics A
The Journal of Chemical Physics
Molecular Physics
Chemical Reviews
Theoretical Chemistry Accounts

… and many others

Bibliography

1. A.R. Leach, *Molecular Modelling - Principles and Applications*, Prentice Hall, 2001
2. J.A. Pople, D.L.Beveridge*, Aproximate Molecular Orbitals Theory*, McGraw-Hill, New York, 1970
3. W.J. Hehre, L.Radom, P.v.R.Schleyer, J.A.Pople*, Ab Initio Molecular Orbital Theory*, John Willey & Sons, New York, 1986
4. F. Jensen, *Introduction to Computational Chemistry*, John Wiley and Sons, New York, 2001
5. D. C. Young, *Computational Chemistry*, John Wiley and Sons, 2001
6. A. Szabo, N.S. Ostlund, *Modern Quantum Chemistry; Introduction to Advanced Electronic Structure Theory*, McGraw-Hill Publishing Company, New York, 1989

7. R.G. Parr, W.Yang*, Density Functional Theory of Atoms and Molecules*, Oxford University Press, New York, 1989
8. C. J. Cramer, *Essentials of Computational Chemistry*, John Wiley & Sons (2002)
9. J.B. Foresman, A. Frisch Exploring Chemistry With Electronic Structure Methods: A Guide to Using Gaussian, Gaussian Inc.
10. P.M.W. Gill, DFT, *HF and the SCF*

Web resources

A mathematical and computational review of Hartree-Fock SCF methods in quantum chemistry by P. Echenique and J.L. Alonso

Quantum Chemistry-Computational Chemistry by D. Sherrill

Basic principles and Hartree-Fock theory by B.C. Hoffman

Orbital Functionals in DFT by E.K.U. Gross

Dichte-Funktional Theorie in der Chemie by M.Hoffman

Jan Labanowski's Basis Set Document

Grading

1. Midterm examination (end of november) (20%)
2. Final examination (40%)
3. Summary of a research paper (25%)
4. Research project related to your own interest (Optional) (15%)

Examples of research reports

1. Scaling the calculated vibrational frequencies
2. Calculation of NMR spectra: the influence of the method and basis set
3. Calculation of ESR spectra for paramagnetic compounds
4. Computational studies in molecular electronics
5. Modelling the intra and inter-molecular hydrogen bonds
6. Computational recipes for large molecules: the ONIOM method
7. Modelling the hydrogen bond interactions
8. Basis set superposition error – is it important?
9. Adsorbed molecules on metallic surfaces
10. Semiempirical methods: are they reliable?

…

Constructing Z-matrices

Z-matrix = a complete set of internal coordinates

- it is used to specify the geometry of a molecule

Internal coordinates

- bond lengths
- bond (valence) angles
- dihedral (torsional) angles

In a Z-matrix:

1-st atom is the origin
2-nd atom is defined by the distance to atom 1
3-rd atom is defined by a distance (to atom 1 or atom 2) and a bond angle
4-th, 5-th, ... atoms are defined by a distance, a bond angle and a dihedral angle with respect to already defined atoms

3N-6 variables are defined

The six missing variables correspond to the three translations and three rotations of the whole molecule (translations and rotations do not change de energy of the molecule)

The orientation of the molecule in space is not defined!

- Bond angles of 180 grades must be avoided in a definition path, as these make the dihedral angles undetermined
- numeric values in a Z-matrix are interpreted as constants; alpha-numeric symbols are used for variables

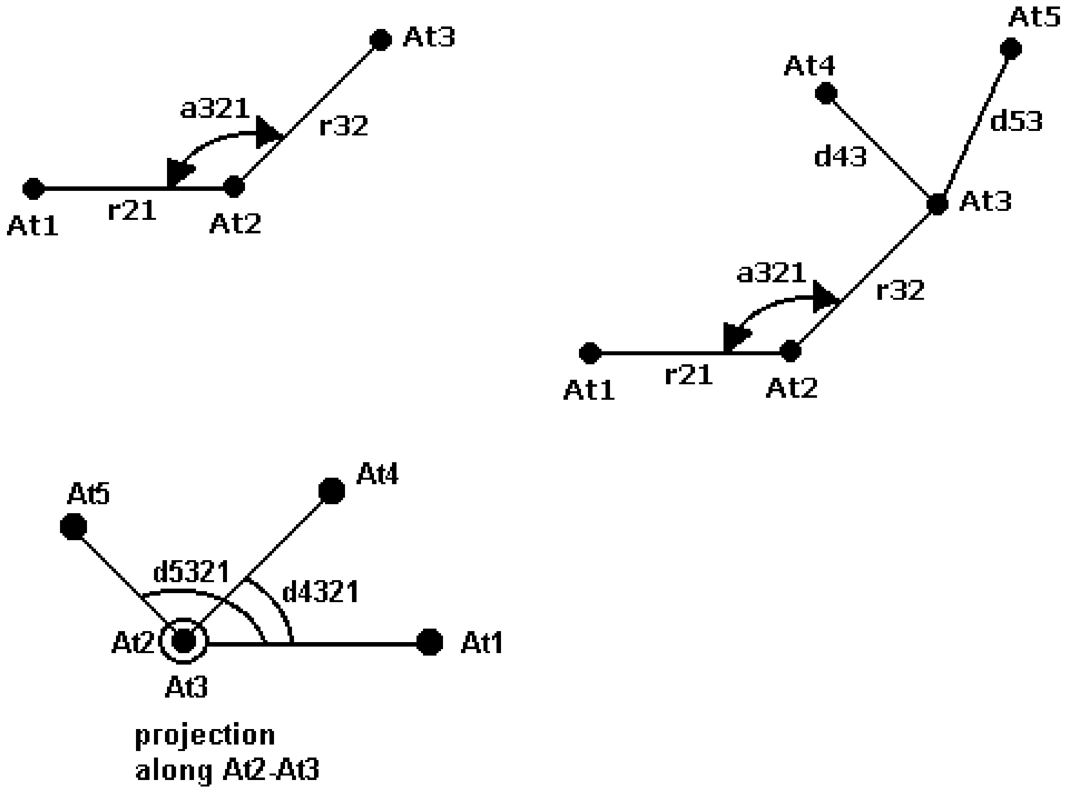
Dummy atoms

- can help in constructing Z-matrices and to impose a given molecular symmetry
- geometrical points that help to define atoms, but without chemical meaning

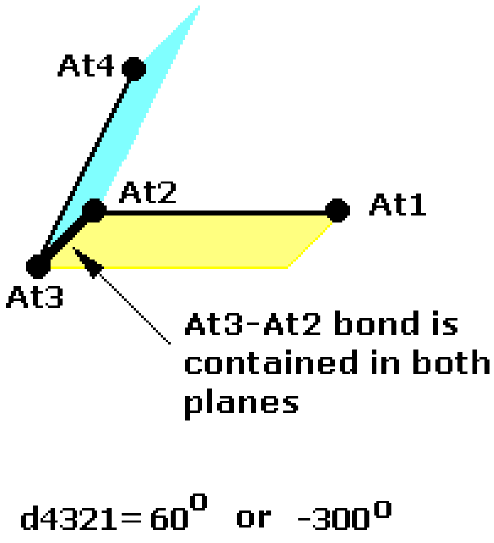
Convention

First bond (At1-At2) is parallel to z-axis in a Cartesian system

Dihedral angles - positive - clockwise rotations
- negative

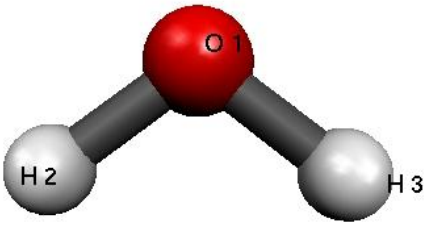


Bond lengths, bond angles and dihedral angles definitions

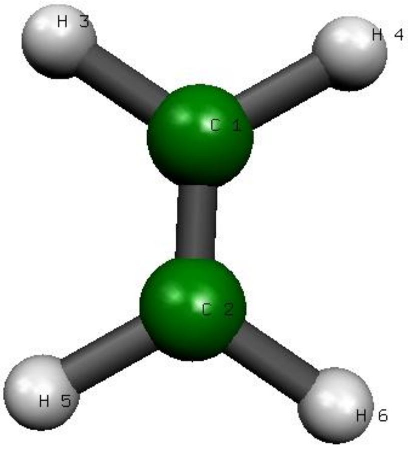


Dihedral angles definition

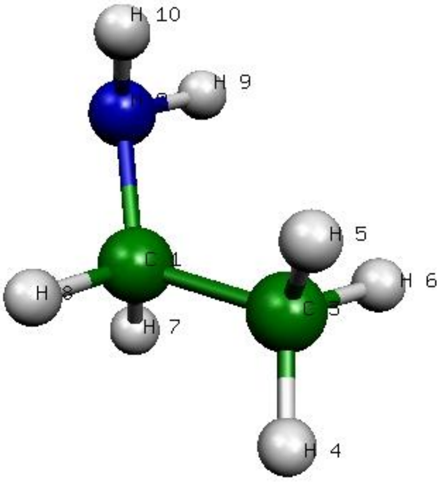
Examples

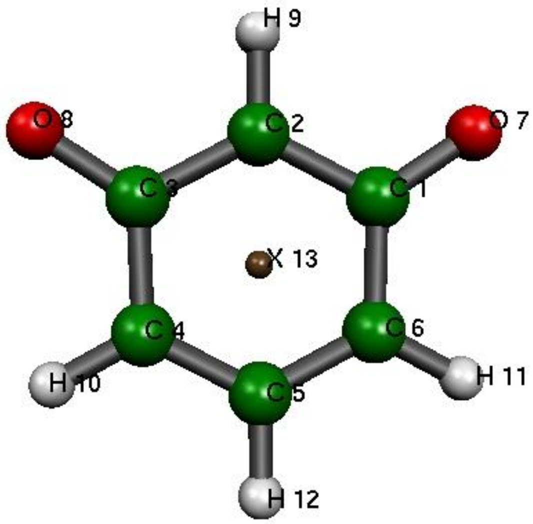
<table>
<tr><th colspan="3">Water (C_{2v})</th></tr>
<tr><td>Molecular structure</td><td>Atom label</td><td>Z-matrix and variables</td></tr>
<tr><td> </td><td>1
2
3</td><td>O1
H2 O1 r21
H3 O1 r21 H2 a

r21=0.97
a=104.5</td></tr>
</table>

<table>
<tr><th colspan="3">Ethylene (C_{2h})</th></tr>
<tr><td>Molecular structure</td><td>Atom label</td><td>Z-matrix and variables</td></tr>
<tr><td> </td><td>1
2
3
4
5
6</td><td>C
C 1 rcc
H 1 rch 2 a
H 1 rch 2 a 3 d1
H 2 rch 1 a 4 d1
H 2 rch 1 a 3 d1

rcc 1.09
a 122.0
rch 1.09
d1=180.</td></tr>
</table>

Ethylamine (Cs)		
Molecular structure	**Atom label**	**Z-matrix and variables**
	**1**	**C1**
	2	**N2 C1 r21**
	3	**C3 C1 r31 N2 a321**
	4	**H4 C3 r43 C1 a431 N2 d1**
	5	**H5 C3 r53 C1 a531 N2 d2**
	6	**H6 C3 r53 C1 a531 N2 -d2**
	7	**H7 C1 r71 N2 a712 C3 -d3**
	8	**H8 C1 r71 N2 a712 C3 d3**
	9	**H9 N2 r92 C1 a921 C3 d2**
	10	**H10 N2 r92 C1 a921 C3 -d2**
		Variables:
		r21 1.45
		r31 1.5
		a321 117.
		r43 1.1
		a431 110.0
		d1 180.
		r53 1.1
		a531 110.0
		d2 60.
		r71 1.1
		a712 109.
		d3 120.0
		r92 1.0
		a921 110.0

o-Benzosemiquinone (C_{2v})		
Molecular structure	**Atom label**	**Z-matrix and variables**
	**1**	**X**
	2	**C1 X r1x**
	3	**C2 X r2x C1 a2x1**
	4	**C3 X r1x C2 a2x1 C1 d1**
	5	**C4 X r4x C3 a4x3 C2 d1**
	6	**C5 X r5x C3 a5x3 C2 d1**
	7	**C6 X r4x C1 a4x3 C2 d1**
	8	**O1 C1 r11 C2 a112 C3 d1**
	9	**O3 C3 r11 C2 a112 C1 d1**
	10	**H2 C2 r22 X a22x C4 d1**
	11	**H4 C4 r44 C3 a443 C2 d1**
	12	**H6 C6 r44 C1 a443 C2 d1**
	13	**H5 C5 r55 X a55x C2 d1**
		Variables:
		r1x=1.462
		r2x=1.382
		a2x1=61.0
		r4x=1.419
		r5x=1.358
		a4x3=59.4
		a5x3=119.0
		r11=1.27
		a112=118.8
		r22=1.078
		a22x=180.0
		r44=1.073
		a443=116.3
		r55=1.072
		a55x=180.0
		Constants:
		d1=180.0

Charge and multiplicity

The multiplicity of a molecule is determined by the number of the unpaired electrons that it contains.

usually: ground states = singlets (no unpaired electron or closed shell molecules)
free radicals = open shell molecules: dublets, triplets, etc.
S – total spin of a molecule
= ½ *total number of unpaired electrons
2S+1 = multiplicity
$<S^2> = S(S+1)$ is the expectation value of the total spin
Spin contamination: calculated $<S^2> \neq S(S+1)$

Number of unpaired electrons	S	Multiplicity S(S+1)		$<S^2>$
0	0	1	singlet	0
1	0.5	2	doublet	0.75
2	1	3	triplet	2
3	1.5	4	quartet	3.75