Introduction to Computational Chemistry and the working environment

Computational Chemistry
Elective Course
Chemistry Degree
4th year

Contents

1. Computational Chemistry
    1. The science
    2. The course

2. The working environment
3. Molecular potential energy
    1. Molecular geometry
    2. Gaussian input
    3. Gaussian output

4. Graphical building of molecular structures

1. Computational Chemistry

1.1. The science

Computational Chemistry (CC) is a branch of chemistry that is characterized not by the object under study, but by the tools used.

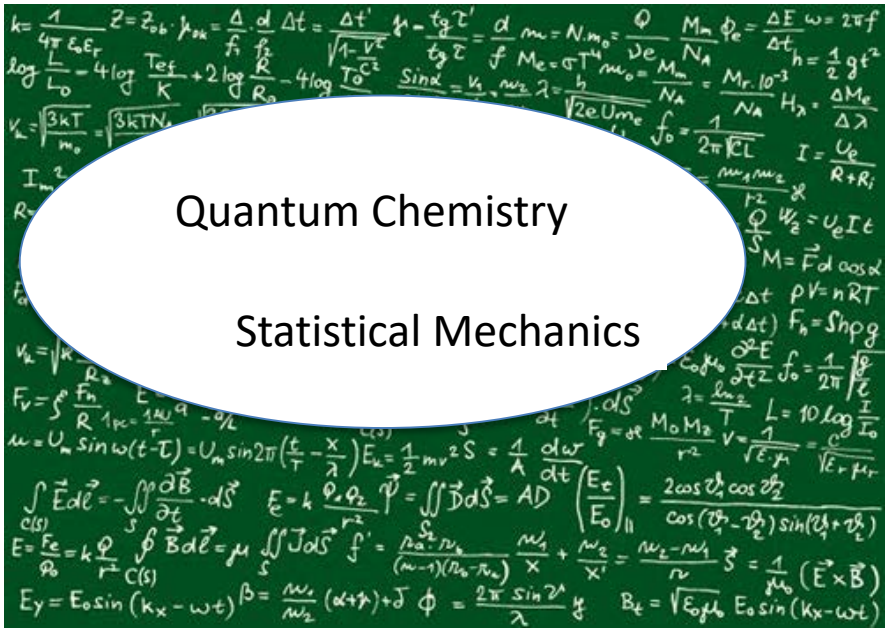
Computational Chemistry studies chemical problems at a microscopic level (atomic-molecular level) by means of equations obtained from **Quantum Mechanics** (that allows the characterization of electron structure of atoms and molecules) and **Statistical Mechanics** (that allows us to obtain macroscopic properties from microscopic constituents)

Theoretical Chemistry is the mathematical description of chemistry. **Computational Chemistry** also involves a computational implementation to solve the corresponding equations. The complexity in the resolution of the proposed equations involves the introduction of approximations and the use of computers.

1. Computational Chemistry

1.1. The science

Theoretical Chemistry



Computational Chemistry

Solutions

(based on approximations and models)

1. Computational Chemistry

Quantum Chemistry provides an accurate description of the behavior of electrons, and hence the structure of atoms and molecules. However, the resulting equations have only been solved exactly for the hydrogen atom. Its application to other chemical problems is performed based on two strategies:

• The use of **approximations**: very few chemical problems can be solved exactly. But that does not mean that an approximate solution can not be useful:

- Sometimes an approximate solution can be more accurate than the best of experiments

- Often only a certain degree of accuracy is required in the solution. For example, for energies it is generally accepted that the necessary accuracy for chemical applications is about 1 kcal·mol^{-1}

- An approximate or qualitative solution can sometimes provide a more thorough understanding of the problem than can be obtained by experiment. For example, a calculation of an approximate dissociation energy can tell us more about the bond being studied than an experiment that accurately measures the dissociation energy

1. Computational Chemistry

• The use of **models**: a model is a simplified representation of the real system that can be solved.

For example, we can study a diatomic molecule modelling it as two masses connected by a harmonic spring

The model may be realistic if quantum mechanics or experimental measurements are used to find the value of the force constant of the spring, so that it represents the real system as closely as possible. This is the basis for the representation of molecules using molecular mechanics, that we will study in this course

1. Computational Chemistry

Thus, we have theories, approximations and models to solve chemical problems. These elements may not be exact, but it is important to know the degree of applicability of each theory, model or approximation to the problem under study. Not knowing the exact solution does not imply that Computational Chemistry can not provide valid knowledge about the problem under analysis.

In general, there are two ways to use Computational Chemistry:

- In an **interpretive** way: reproducing an experiment and obtaining detailed knowledge that accounts for the observed results, and from there developing a working hypothesis

- **Predictively**: studying changes in composition that allow us to obtain the desired properties

1. Computational Chemistry

1.2. The course

The course of Computational Chemistry is an optional subject of 6 credits

The focus of the course is highly practical, aiming to show the use of Computational Chemistry tools for solving problems. The course is divided into 20 sessions of 3h (except the exam, that is divided into 2 sessions of 1.5 h) with three types of activities:

- 9 Practical sessions (about 45 h)
- 8 Theoretical Seminars (about 12 h)
- Exam (3 h)

The **practical sessions** are carried out on the computer and are focused on learning methods and programs needed to solve chemical problems at the atomic-molecular level. The classes usually begin with a theoretical introduction by the teacher to place the ideas that are to be worked in that session within a broader framework. Most of the time, the session will be devoted to student practical work.

In the **seminars**, the theoretical elements that are used in the practical sessions will be presented in greater detail and rigor, with special emphasis on the Hartree-Fock method.

1. Computational Chemistry

Sessions

All sessions are 3h long, except when indicated

#	Date	Practical Content	Theoretical Content
1	11/9	Working environment (I)	
2	18/9	Working environment (II)	Seminar HF (I)
3	23/9	Electronic structure (I)	Seminar HF (II)
4	25/9	Electronic structure (II)	Seminar HF (III)
5	30/9	Electronic structure (III)	Seminar HF (IV)
6	2/10	Optimization (I)	Optimization methods
7	14/10	Optimization (II)	DFT
8	16/10	Optimization (III)	
9	21/10	Reactivity (I)	Fundamental of reactivity
10	23/10	Reactivity (II)	
11	28/10	Semiempirical calculations	Semiempirical methods
12	30/10	Spectroscopy (I)	Post HF Methods (I)
13	4/11	Spectroscopy (II)	Post HF Methods (II)
14	6/11	Solvent effects (I)	Molecular mechanics and Continuum models
15	11/11	Solvent effects (II)	
16	13/11	Molecular dynamics simulations (I)	Molecular dynamics
17	18/11	Molecular dynamics simulations (II)	
18	25/11	Applications (I)	
19	27/11	Applications (II)	
20	2/12 (1,5h)	Exam (I)	
21	4/12 (1,5h)	Exam (II)	

1. Computational Chemistry

Assessment

The assessment of the Computational Chemistry course will include:

- **Final exam**: test based on the completion of a project with a written report that must be defended orally (60%)

- Assessment of the **participation in oral presentations** (10%)

- Assessment of the **reports** corresponding to the practical sessions (20%)

- **Continuous** evaluation of each student, based on regular attendance to classroom activities, participation and degree of involvement in the teaching-learning process (10%)

1. Computational Chemistry

Assessment

Specifically, the following items will be assessed:

- A report to be handed in for each practical session (with a maximum of 2 sides) in which one of the exercises is presented, collecting: objectives, methods, results, and conclusions.

- A full report of the final exercise

- The oral presentation, to be held on the date of the examination (02-04/12/2019)

- Continuous evaluation (regular class attendance and classroom activities, participation and degree of implication)

1. Computational Chemistry

Assessment

1 report to be handed in for each practical session (1 exercise chosen from those discussed in the session)

What have I done?

How did I do it?

What are the results?

What can I conclude?

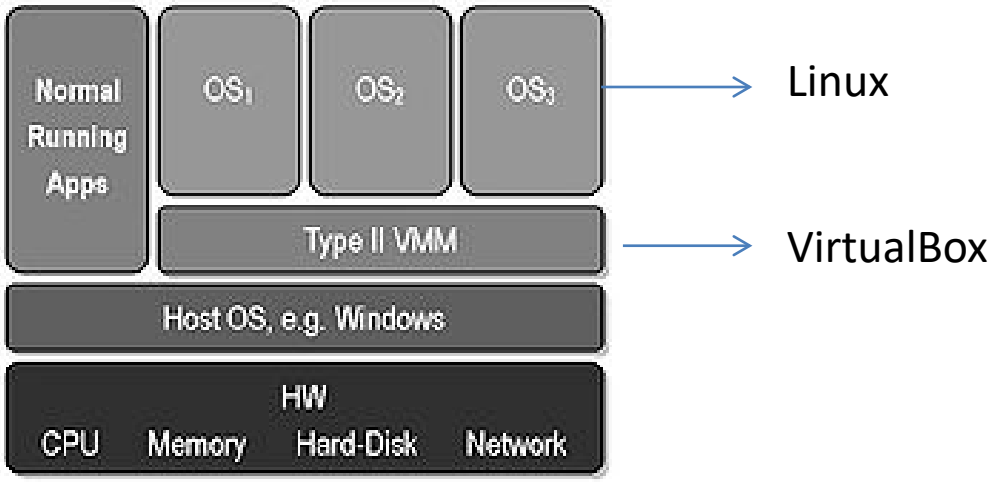
Attention:

- Numbering of figures and tables
- Use correct number of significant figures (we will discuss this)
- English is optional (+1)

2. Practical Sessions: Working environment

To ensure that we all have a homogeneous computing environment, with the same programs, the same versions and identical operating system, we will work on what is called a virtual machine.

A virtual machine is a program that can simulate a computer within another one. For example, in our case we will have a computer with the Linux operating system and with all the necessary programs within another running a Windows/Mac/Linux operating system.



What do you need if you want to use your own computer?

- **Oracle VM VirtualBox (www.virtualbox.org)**
- **QTC.ova (in software.uv.es)**

Software.uv.es → download QTC.ova

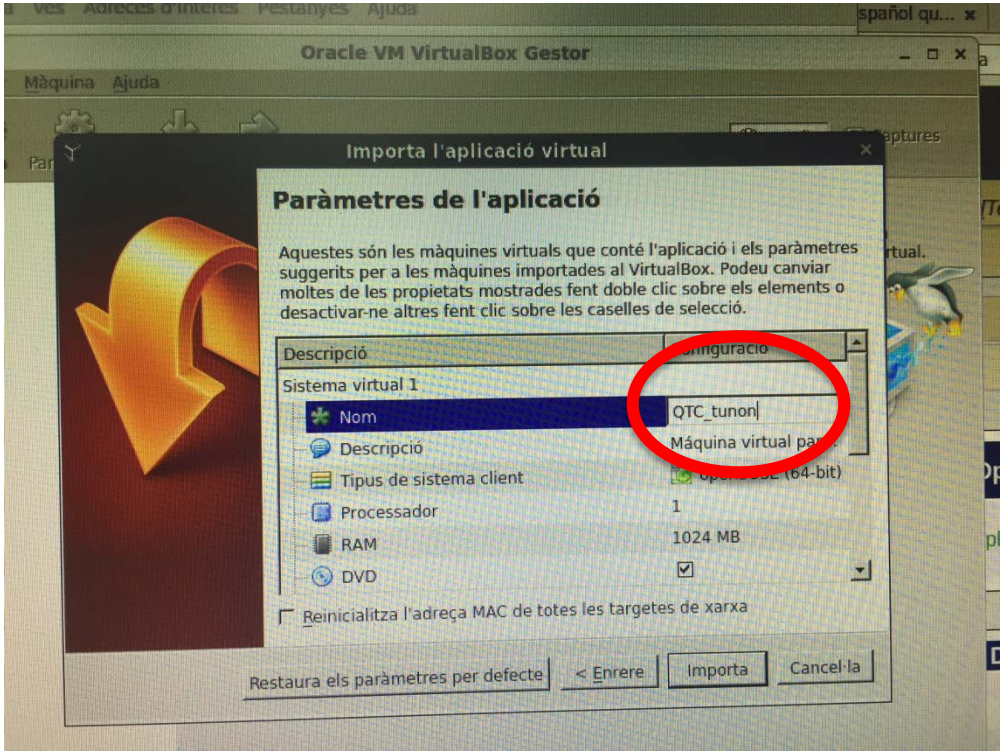
Aplicacions > Eines del sistema > Oracle Virtual Box

File > Import

Select Descargas/QTC.ova and press next

Change name of the Virtual Machine to QTC_xxxx (where xxxx is your surname or username)

Continue



2. Practical Sessions: Working environment

Linux is an operating system distributed freely, initially developed in 1991 by a Finnish student, Linus Torvalds. Although Linux is, strictly speaking, the operating system, the core part of the interaction between the core and the user (or application programs) is usually handled with GNU tools, developed by Richard Stallman

A Linux distribution is a distribution of Linux-based software that includes software packages to meet the needs of a specific group of users. They are usually composed, wholly or largely, of free software, but often incorporate proprietary applications or drivers. The most popular are: Red Hat, Ubuntu, Fedora, ...

In addition to the Linux kernel, distributions usually include libraries and tools from the GNU project and the X Window System. Depending on the type of user at which the distribution is directed, it also includes other software such as word processors, spreadsheets, media players, administrative tools, etc.. If GNU project tools are included, the term distribution GNU / Linux is also used .

2. Practical Sessions: Working environment

Linux basic commands:

ls
This shows the contents of the folder you indicate; it also accepts certain arguments that may be interesting.

To display the files and folders along with the rights you have, sizes, etc:

$ ls -l

Furthermore, the arguments may overlap. If we want to display the files in the same way as before, but also showing hidden files:

$ ls -la

$ represents the computer prompt. Usually, it includes the working directory and an arrow, like this:
$ /home/student/direc->

2. Practical Sessions: Working environment

cd

Change directory. We can use it with absolute or relative paths. In the absolute path we indicate all the way from the root directory (/). For example, wherever we are, if we write in the console ...

$ cd /etc/X11

... we will be taken to this folder directly. Similarly if we write ...

$ cd /

... we will be sent to the root file system.

$ cd ..

... We will be sent to the parent directory (.. [two points] always refer to the higher directory, and . [one point] to the current directory)

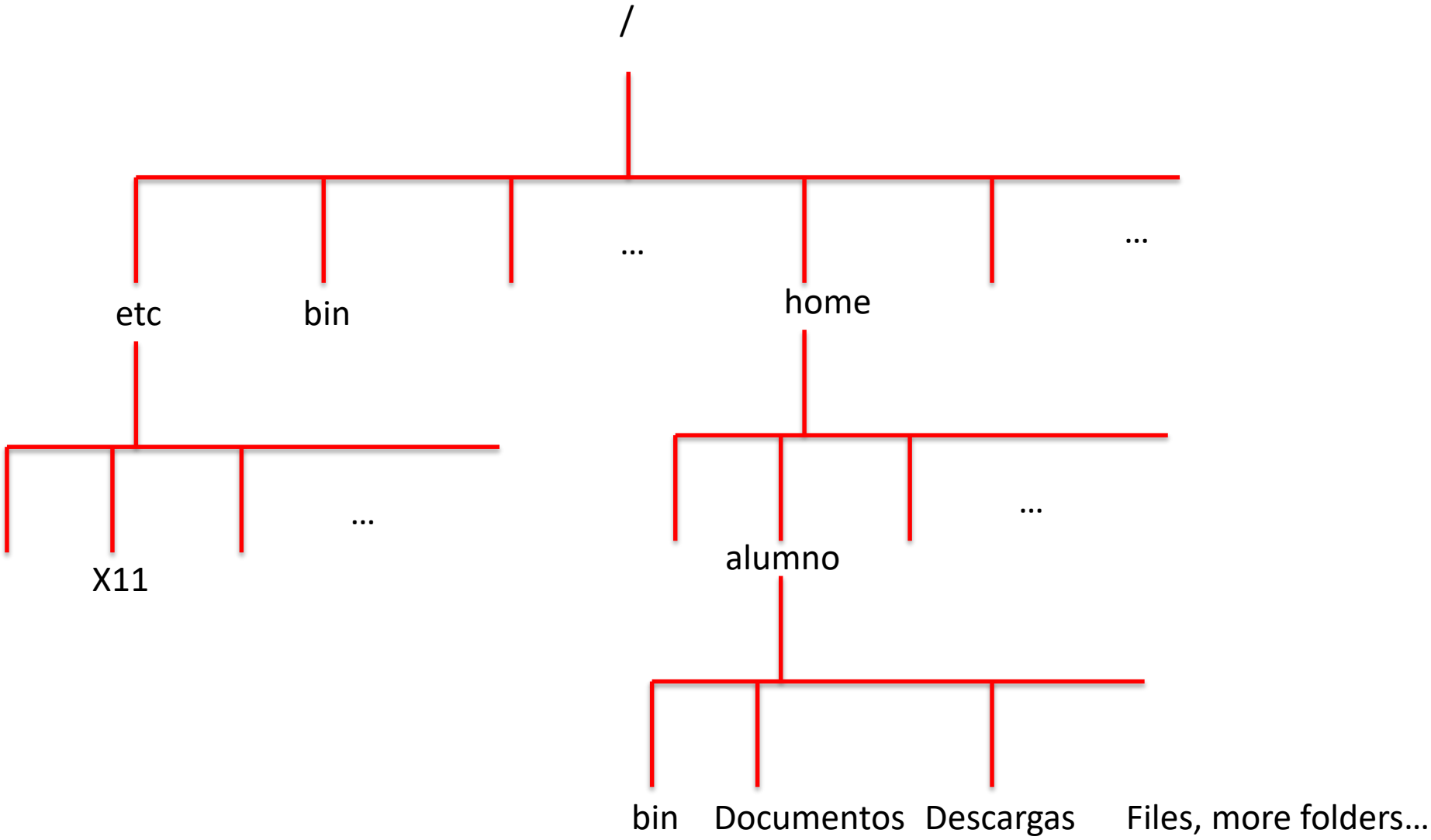
Relative paths refer to the folder where we are. Imagine you are in your home directory and want to go to a folder called temporary within your personal folder.
We should write ...
$ cd temporary

$ cd

This puts us straight into the starting directory

2. Practical Sessions: Working environment

Directory tree



2. Practical Sessions: Working environment

mkdir

Make directory. Create a folder with the name that you indicate. Again we can use absolute paths or relative paths. We can write the whole path preceding the directory we want to create or, if we are in the folder that will contain it, it is enough to put the name:

$ mkdir /home/alumno/session1

If you are in /home/alumno you can just type ...

$ mkdir session1

pwd

This command tells you which directory you are in

2. Practical Sessions: Working environment

rm

Remove: Delete the file or folder that you indicate. As before, you can enter either the full path or the file name. This will be ignored from now, since it has become clear with the two previous commands.

To delete a file:
$ rm filename

To delete an empty folder:
rm foldername

To delete a folder containing files and/or other folders that may contain even more files or folders:
rm -r foldername

Other options: "-f" do not ask for a confirmation to delete
"-v" will show what is erased

2. Practical Sessions: Working environment

cp

copy. Copy the file indicated to where you specify. Here we can also play with the routes, both for the source file, and for the destination. You can also put the name you want to use for the copy. For example, if we were in /etc/X11 and would like to make a backup of xorg.conf in our personal folder:

$ cp xorg.conf /home/alumno/xorg.conf.backup

mv

move. Same as above, only instead of making a copy, directly move the file with the name that you indicate, which can be different from the original:

$ mv /etc/page.html /home/alumno/new_page.html

Another very practical command is to rename a file. Simply enter the new name in the second argument with the same path of the first. In this example we assume that we are in the folder that contains the file page.html:

$ mv page.html new_page.html

2. Practical Sessions: Working environment

find

Find the file or folder that you specify:

$ find / -name page

The above command would look everywhere folders and files called page from root directory /. If we were sure that it is located in /var for example, we should use:

$ find /var -name page

If we are not sure of the name we can indicate it with wildcards. Suppose we seek a name containing the string "pep" in the same folder as before:

$ find /var -name *pep*

You have other options. For example, we can choose to find files or folders over 1500 KB:

$ find / -size +1500

Or the files or folders containing the name "pepi" and having less than 1000 KB:

$ find / -name *pepi* -size -1000

2. Practical Sessions: Working environment

clear
Clear the screen or console leaving it as if we just opened it.

$ clear

ps
process status: status of the processes. It shows us what we want to know about the processes running on your system. Each process is identified by a number called PID. If we do...

$ ps -A

…will show a list of all processes, their PID to the left and the name to the right. If you want even more information:

$ ps aux

2. Practical Sessions: Working environment

kill
Kill the process we indicate with PID:

$ kill PID

Where PID is the PID number of the process to be killed. Sometimes the process does not "die" at all, but you can force the system to safely kill it as follows
$ kill -9 PID

man
Manual. This is another powerful command in Linux. Each program or command normally comes with a complete help file on their use and their arguments. If you don't know how a command or application is used and what arguments the program has you can type in the console:
$ man XXX

Where XXX is the command or application. Sometimes the information offered by man can become excessive. Almost all commands and applications accept the argument "-help" to display more summarized some help. For example kill:

$ kill -help

2. Practical Sessions: Working environment

Exercise 1: Practice move and copy

Open a web browser and download the compressed file Ejemplos.tar from aulavirtual. The file will be saved in /home/alumno/Descargas

Move Ejemplos.tar from /home/alumno/Descargas to /home/alumno
If you are in /home/alumno you can type:

$ mv Descargas/Ejemplos.tar . (the dot indicates the current directory)
(it will disappear from Descargas … take a look)

Uncompress the file
$ tar –xvf Ejemplos.tar
Now you have a new folder
/home/alumno/Ejemplos with several files (including h2oc.com and h2oi.com)

Create a directory **/home/alumno/session1**
(Do not use special symbols such as ´* $ % & ….. In any file or folder name)
Copy the files h2oc.com and h2oi.com from
/home/alumno/Ejemplos to /home/alumno/session1
Go to /home/alumno/session1 and list the content (cd and ls)

2. Practical Sessions: Working environment

Visualization of files

more

more: display the contents of the file whose name is indicated page after page, i.e. showing the maximum content that occupies one screen at a time

$ more filename

The command terminates automatically at the end of the file. If the file content is longer than one screen, press the spacebar to display the next page. **q** to quit the listing

less

less: display the contents of a file. The difference between this and the more command is that it does not end automatically when reaching the end of the file

$ less filename

cat

cat: dumps the file content on screen directly, unpaginated until the end

$ cat filename

2. Practical Sessions: Working environment

The vi file editor

vi is a text EDITOR, not a WORD PROCESSOR. It is made to write and edit simple text files, programs, etc.. It does not justify paragraphs, nor use different fonts, nor write in multiple columns, insert graphics, etc.

$ vi filename
Edit a file called 'filename'. If it does not exist, then it is created

There are three modes or states for **vi**:

• Command Mode: This is the mode where the program is each time it starts.
The keys perform actions (commands) that allow you to move the cursor, execute commands for text editing, exit vi, saving changes, etc..

• Insertion or text mode: This mode is used to insert the text. Several
Commands can be used to enter in this mode.

• Line mode: commands are written in the last line at the bottom of the screen.

2. Practical Sessions: Working environment

Moving the cursor:

Command (Keys)	Action
Arrows	Moving in the direction of the arrow
h	Move Left
l	Move right
k	Move up
j	Move down
1G	Move the cursor to the beginning of the file
G	Move the cursor to the end of file

Change text command mode:

Command	Action
i	Inserts text to the left of the cursor
a	Inserts text to the right of the cursor
A	Insert text at the end of the line where the cursor is
I	Insert text at the beginning of the line where the cursor is
o	Open a line below the current
O	Open a line above the current

2. Practical Sessions: Working environment

Clear text:

Command	Action
x	Delete the character under the cursor
dd	Deletes the line where the cursor is
ndd	Deletes the next n lines
D	Deletes from where the cursor is to the end of the line
dw	Deletes from where the cursor is to the end of word

Cut and paste:

- Cut text to be moved using any of the commands used to clear text.
- Move the cursor to where you want to paste the text.
- Paste text with the **p** command.

Copy and paste:

- Use the command **yy**, whose funtion is to copy the line where the cursor is located.
- Move the cursor (with any of the commands used to move the cursor to the text) to where you want to paste the text.
- Paste text with the **p** command.

Undo changes:

- You can undo the last change made, using the command **u**

2. Practical Sessions: Working environment

LINE MODE:
To enter line mode from command mode, use one of the following keys: **/ ? :**

To return to command mode from the last line mode, you must press the **ENTER** key (to end the command) or **ESC** (which interrupts the command).

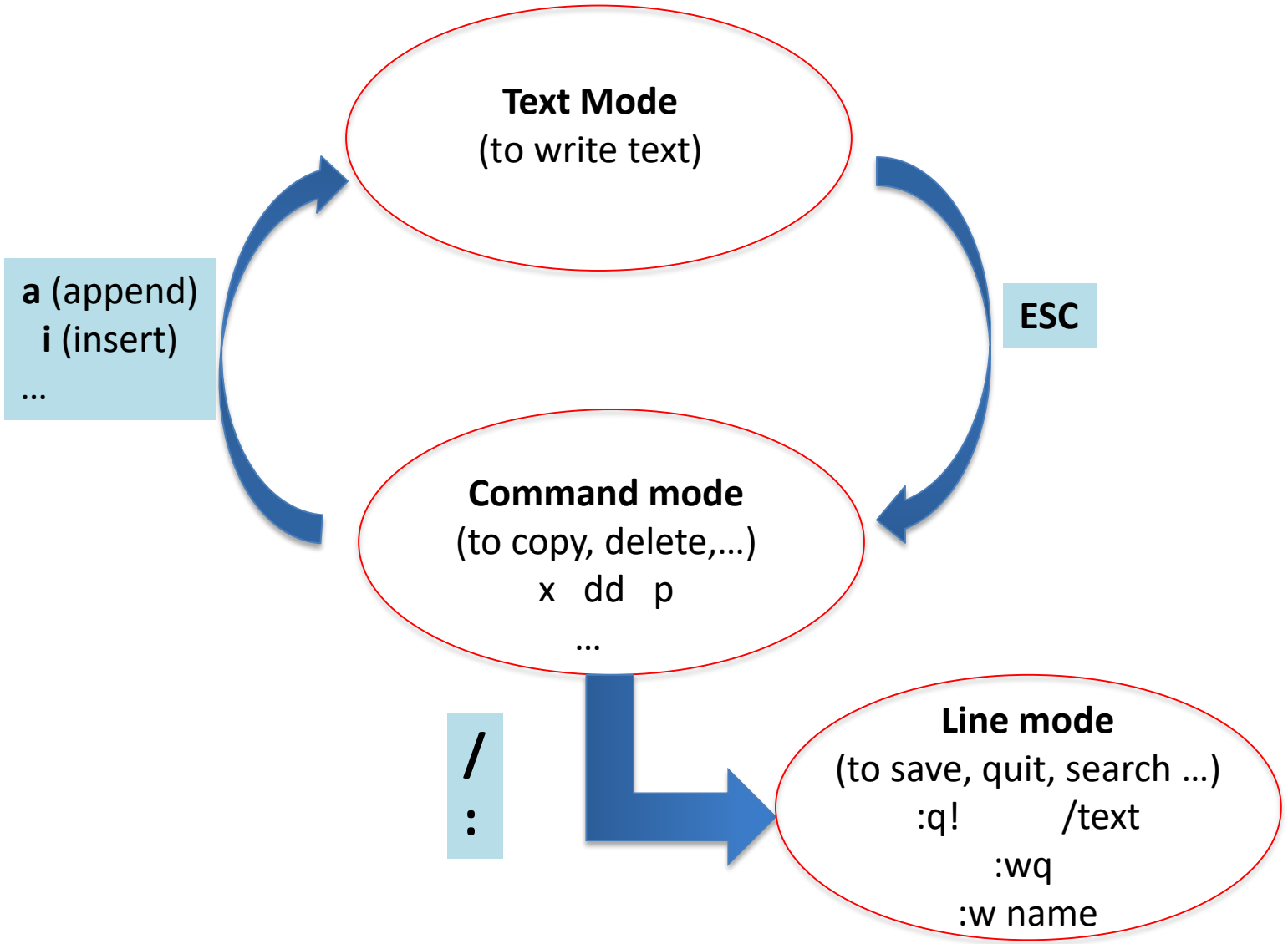
Search text:

Command	Action
/text	Search forward the string "text"
?text	Search backward for the string "text"

Exit vi, save, not save changes, etc..:

Command	Action
:q	Exit if no change was made
:q!	Exit without saving changes
:w	Save Changes
:w file1	Save changes on file1
:wq	Save changes and exit

2. Practical Sessions: Working environment



2. Practical Sessions: Working environment

Exercise 2: Create in the directory session1 a file named perfil1, where you must write:
Your name (line 1)
E-mail address (line 2)
Phone number (line 3)

Next edit the file and modify it by adding the postal address in the second line and save the file with the name perfil2

3. Molecular Potential Energy

To study chemical systems we need to know their energy and how it depends on several factors: geometry, environment, composition, ...

To calculate the energy of a system constituted by nuclei and electrons the best tool we have is Quantum Mechanics, which establishes that the energy can be calculated using the Schrödinger equation:

$$\hat{H}(R_N, r_e)\Psi(R_N, r_e) = E(R_N, r_e)\Psi(R_N, r_e)$$

Where R_N and r_e refer the coordinates of nuclei and electrons, respectively. The usual approximation is to consider the nuclei as point particles of charge Z_N. Thus the Hamiltonian, as the sum of the kinetic and potential energies of nuclei and electrons, can be written (in atomic units) as:

$$\hat{H}(R_N, r_e) = \hat{T}_N(R_N) + \hat{T}_e(r_e) + \hat{V}_{NN}(R_N) + \hat{V}_{ee}(r_e) + \hat{V}_{Ne}(R_N, r_e)$$

Where T refers to the kinetic energy of nuclei (N) and electrons (e) and V to potential energies due to the electrostatic interactions between the point charges.

3. Molecular Potential Energy

$$\hat{H}(R_N, r_e) = \hat{T}_N(R_N) + \hat{T}_e(r_e) + \hat{V}_{NN}(R_N) + \hat{V}_{ee}(r_e) + \hat{V}_{Ne}(R_N, r_e)$$

$$\hat{T}_N(R_N) = -\sum_I \frac{\hbar^2}{2M_I} \nabla^2_{R_I}$$

$$\hat{T}_e(r_e) = -\sum_i \frac{\hbar^2}{2m_i} \nabla^2_{r_i}$$

$$\hat{V}_{NN}(R_N) = \sum_I \sum_{J>I} \frac{Z_I Z_J}{R_{IJ}} \qquad R_{IJ} = \left| \vec{R}_I - \vec{R}_J \right|$$

$$\hat{V}_{ee}(r_e) = \sum_I \sum_{J>I} \frac{1}{R_{ij}} \qquad R_{ij} = \left| \vec{r}_i - \vec{r}_j \right|$$

$$\hat{V}_{Ne}(R_N, r_e) = -\sum_I \sum_i \frac{Z_I}{R_{Ii}} \qquad R_{Ii} = \left| \vec{R}_I - \vec{r}_i \right|$$

3. Molecular Potential Energy

The usual way of solving the problem is to use the Born-Oppenheimer approximation: Because the mass of the nuclei is much greater than that of the electrons ($M_N >> m_e$) one can obtain the electronic structure by decoupling the movement of the nuclei.

Put simply, one can say that we shall solve the molecular Hamiltonian assuming that the nuclei are in a fixed position, i.e. ignoring their kinetic energy and assuming that the term V_{NN} is a constant.

Thus, the remaining term is what we call the electronic Hamiltonian:

$$\hat{H}_{elec}(r_e; R_N) = \hat{T}_e(r_e) + \hat{V}_{ee}(r_e) + \hat{V}_{Ne}(r_e; R_N)$$

And the resulting wave equation allows us to obtain the electronic structure:

$$\hat{H}_{elec}(r_e; R_N)\Psi_{elec}(r_e; R_N) = E_{elec}(R_N)\Psi_{elec}(r_e; R_N)$$

Most of the course will be devoted to the resolution of this equation by using approximate methods.

3. Molecular Potential Energy

Molecular Potential Energy refers to the energy that corresponds to the molecule for a fixed configuration of the nuclei. For a molecule with N nuclei:

$$V(R_N) = E_{ele}(R_N) + V_{NN}(R_N)$$

Where V_{nn} is the internuclear repulsion energy and E_{ele} the electron energy:

$$\hat{H}_{ele}\Psi_{ele} = E_{ele}\Psi_{ele}$$

• The potential energy is not the total energy of molecules, and we should also take into account the kinetic energy of the nuclei.

• It depends on 3N-6 coordinates (since normally neither the position nor the orientation relative to the axis of the laboratory system influence the total energy).

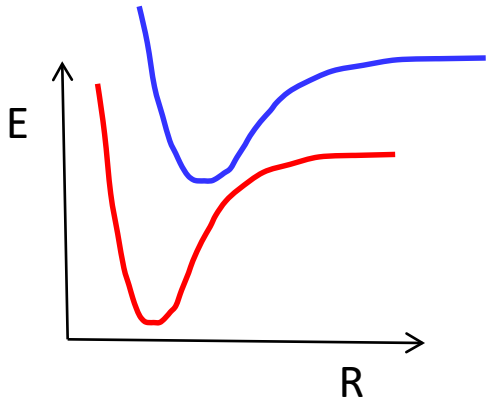
• The potential energy provides the forces acting on the nuclei in the field created by the electrons, and therefore tells us which configurations are stable (those for which the forces are cancelled) as well as the system's response to an external perturbation (such as a collision with another molecule).

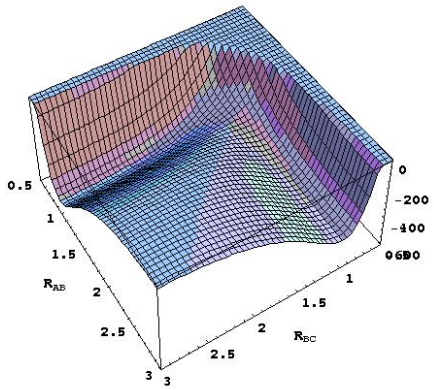
3. Molecular Potential Energy

A Potential Energy Surface is the representation of the set of values for the Molecular Potential Energy corresponding to each of the possible arrangements of the nuclei as a function of the variables on which it depends.

Chemistry can be considered as an exercise on the potential energy surface (PES):

- A vibrational spectrum is obtained from the motion of a molecule on a potential energy well.
- A reaction is the transition of molecules from one well to another on the PES.
- An electron transition is the passage of a PES to another.





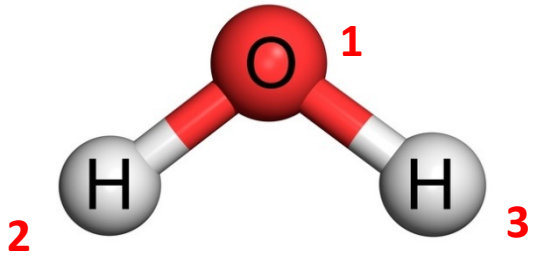
3. Molecular Potential Energy

3.1. Specifying Molecular Geometry

Consider the water molecule. The geometry can be specified, for example giving the cartesian coordinates of three atoms with respect to any reference system of the laboratory.

If we choose the Angstrom as the unit of distance (1 Å = 10^{-10} m)

O 0.000, 0.000, 0.000
H 0.757, 0.586, 0.000
H -0.757, 0.586, 0.000



Then, 9 values have to be specified. However, the potential energy depends only on 3*3-6=3 variables. The position of the center of mass and the orientation of the axes of inertia do not normally affect the energy.
To reduce the number of variables one can specify only the internal coordinates: bond distances, bond angles and torsion angles (z-matrix)

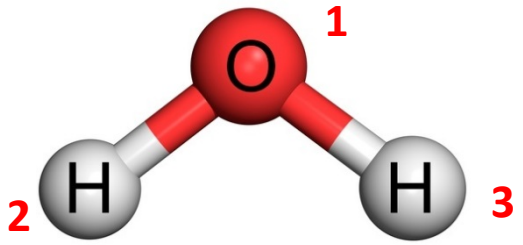
O → Is taken as the origin
H 1 0.958 → To specify the second atom you only need to give its distance from the first
H 1 0.958 2 104.5 → To specify the third atom the distance from the first atom and the angle with the second are given

3. Molecular Potential Energy

To clarify what values are to be treated as variables they can be indicated by a name. The values for the variables are given at the end of the z-matrix.

O
H 1 r1
H 1 r2 2 a2

r1 0.958
r2 0.958
a2 104.5



We can reduce the number of variables using molecular symmetry

O
H 1 r1
H 1 r1 2 a2

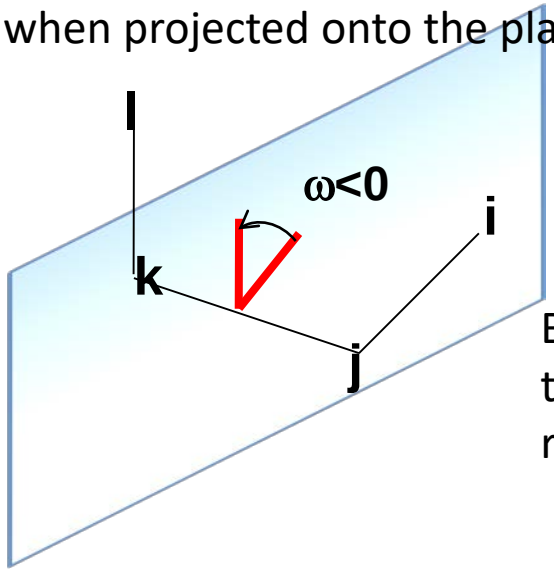
r1 0.958
a2 104.5

3. Molecular Potential Energy

Exercise 3: Create in the directory Session1 a file called h2o where you write the z-matrix corresponding to a molecule of water.

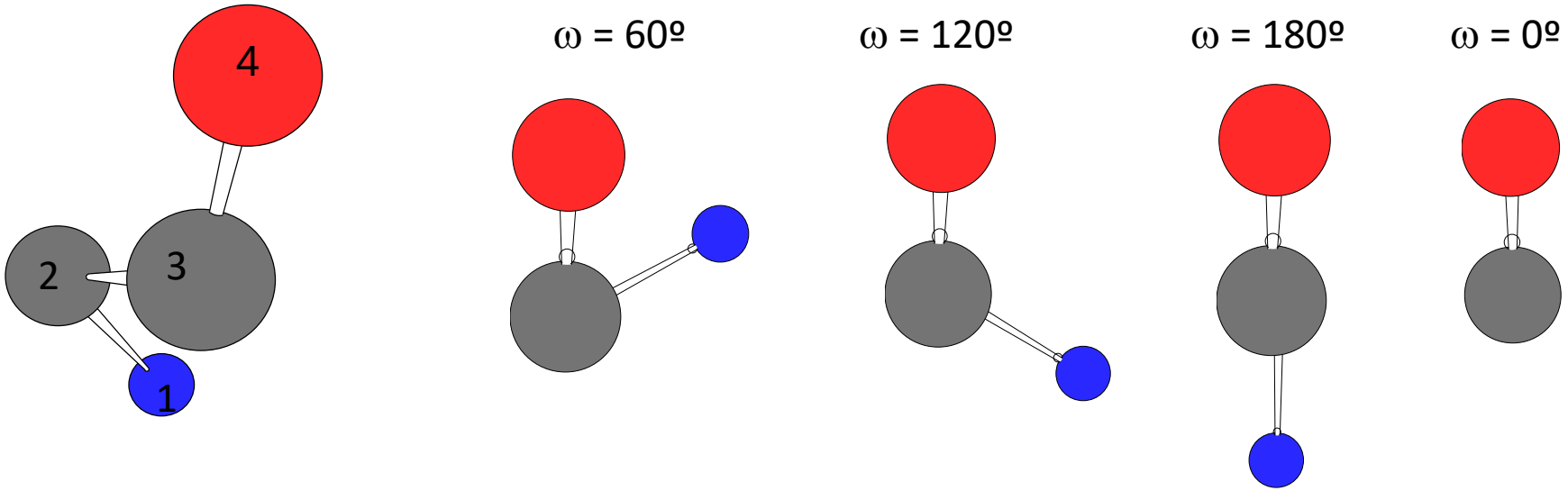
Next display the molecule with the program molden
$ molden h2o

Let us consider four atoms linked in sequence, ijkl. The dihedral angle or torsion associated with this sequence is defined as the angle formed between the bonds ij and kl when projected onto the plane which bisects the bond jk.



By convention, the dihedral angle is taken as positive if the angle of the bond in front of the plane has to be rotated clockwise and negative if counterclockwise.

The twist angle is periodic. The existence interval can be defined between $0 < \omega < 2\pi$ or between $-\pi < \omega < +\pi$

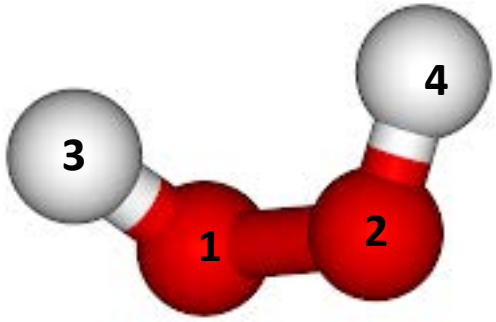


3. Molecular Potential Energy

Exercise 4: Create in the directory sesion1 two files named h2o2c and h2o2t where you should write the z-matrixes corresponding to the hydrogen peroxide with the hydrogen atoms are in *cis* and *trans* conformations, respectively.

h2o2c

```
O
O  1 r1
H  1 r2  2 a2
H  2 r2  1 a2  3 z2

r1 1.48
r2 0.96
a2 104.5
z2 0.0
```



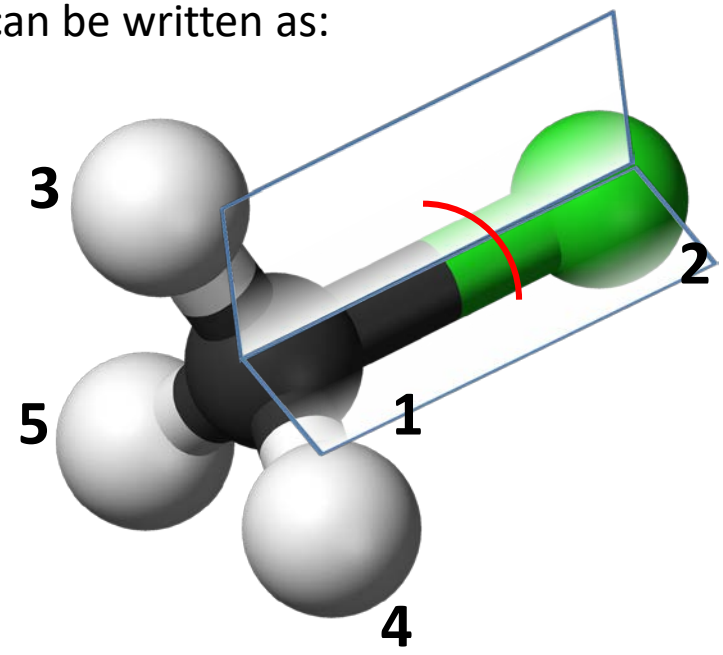
h2o2t

```
O
O  1 r1
H  1 r2  2 a2
H  2 r2  1 a2  3 z2

r1 1.48
r2 0.96
a2 104.5
z2 180.0
```

3. Molecular Potential Energy

The dihedral angle may also be defined for 4 atoms that are not bound consecutively, observing the angle defined by each group of three atoms.
For example, the z-matrix of chloroethane can be written as:

```
C
Cl  1 r1
H   1 r2  2 a2
H   1 r2  2 a2  3 z2
H   1 r2  2 a2  3 –z2

r1 1.77
r2 1.09
a2 109.47
z2 120.
```

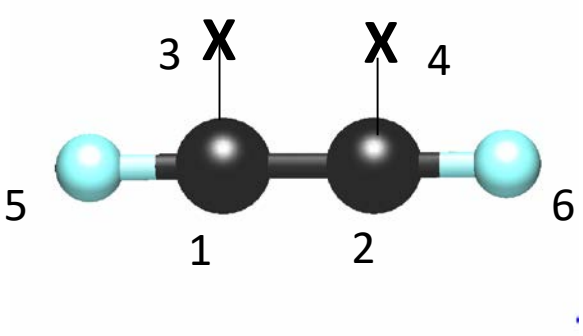


3. Molecular Potential Energy

Difficulties may arise when there are three consecutive atom in line because they do not define a single plane. In that case, 'ghost atoms' (dummy atoms) can help to define the molecular geometry. They are simply reference points that avoid uncertainties as described.

Example: Z-Matrix of acetylene

```
C
C  1  r1
X  1  1.0   2  90.
X  2  1.0   1  90.  3    0.
H  1  r2    3  90.  2  180.
H  2  r2    4  90.  1  180.

r1 1.20
r2 1.09
```



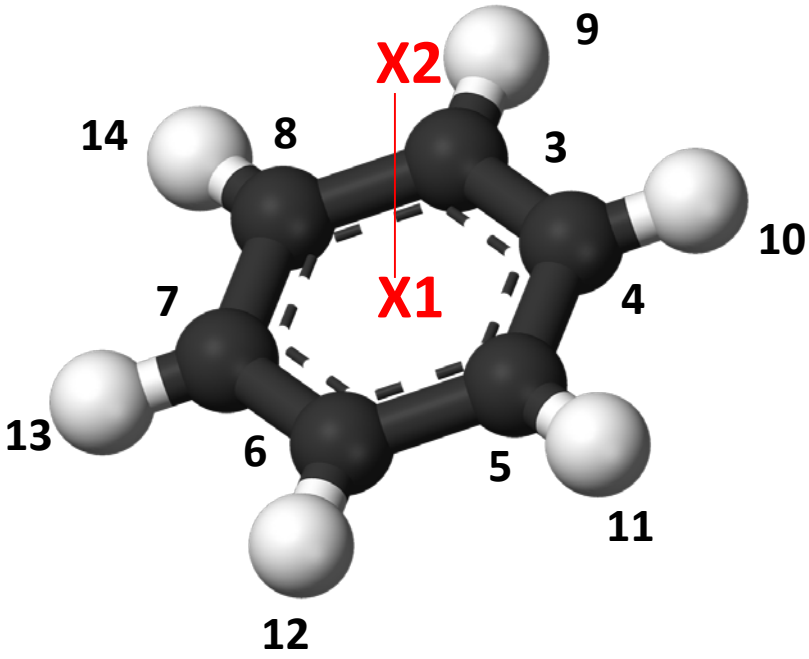
The position of the dummy does not affect the energy. On the other hand, if we know that the molecule is linear, the angles between atoms must be constant

3. Molecular Potential Energy

Dummy atoms can be also useful to define rings preserving the symmetry

Example: benzene z-matrix

```
X
X  1  1.0
C  1  rcx  2  90.
C  1  rcx  2  90.  3  60.
C  1  rcx  2  90.  4  60.
C  1  rcx  2  90.  5  60.
C  1  rcx  2  90.  6  60.
C  1  rcx  2  90.  7  60.
H  1  rhx  2  90.  3   0.
H  1  rhx  2  90.  4   0.
H  1  rhx  2  90.  5   0.
H  1  rhx  2  90.  6   0.
H  1  rhx  2  90.  7   0.
H  1  rhx  2  90.  8   0.

rcx 1.3
rhx 1.39
```



Bond lengths in Å			
Bond	**Length**	**Bond**	**Length**
H--H	0.74	H--C	1.09
C--C	1.54	H--N	1.01
N--N	1.45	H--O	0.96
O--O	1.48	H--F	0.92
F--F	1.42	H--Cl	1.27
Cl-Cl	1.99	H--Br	1.41
Br-Br	2.28	H--I	1.61
I--I	2.67		
C--C	1.54	C=C	1.34
C--N	1.47	C≡C	1.20
C--O	1.43	C=O	1.23
C--S	1.82	O--O	1.48
C--F	1.35	O=O	1.21
C--Cl	1.77		
C--Br	1.94	N--N	1.45
C--I	2.14	N≡N	1.10

3. Molecular Potential Energy

3.2. Gaussian input

The program that we will use to solve the electron wave equation is Gaussian, in both the G03 and G09 version. The online manual can be found at:
http://www.gaussian.com/g_tech/g_ur/g09help.htm

The input for a calculation with the program Gaussian ('name'.com) has the following structure:

```
%chk=h2oi.chk
#hf/3-21G

Water in internal coordinates

0 1
O
H  1 r1
H  1 r1  2 a2

r1 0.958
a2 104.5
```

File management. The .chk file in binary format contains information that may be relevant for subsequent calculations

Command line. It specifies the type of calculation to perform

Blank line

Title line

Blank line

Charge and multiplicity (2S+1)

Molecular Geometry

Blank line

Blank line

3. Molecular Potential Energy

3.2. Input of Gaussian program

In the directory Ejemplos there must be an input for water molecule in internal coordinates (h2oi.com) or create it from h2o file

Download Ejemplos.tar from the aula virtual
If we are in /home/alumno
$ mv Descargas/Ejemplos.tar .
$ tar –xvf Ejemplos.tar
$ cd Ejemplos

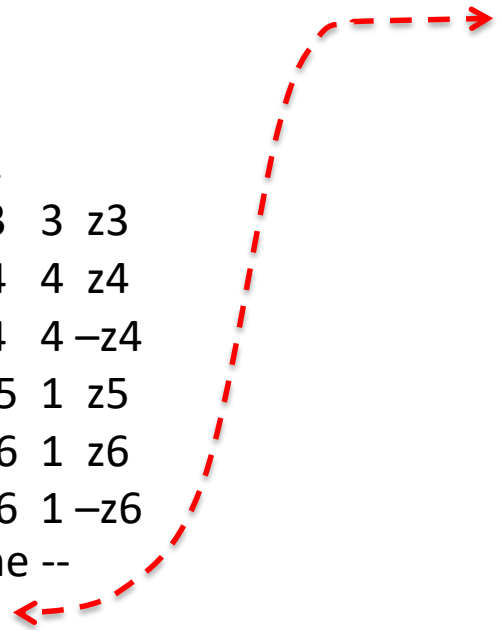
Input files can be visualized with the program molden or the program Gaussview (command gv)
$ molden h2oi.com
$ gv h2oi.com

Exercise 5: Write a Gaussian input for ethanol (**etoh.com**) using internal coordinates and keeping the same options for the calculation (hf/3-21G)
Visualize it with molden to verify that the geometry is correct

3. Molecular Potential Energy

3.2. Gaussian input

```
#hf/3-21G
  -- blank line --
ethanol
  -- blank line --
0 1
O
C  1 r1
C  2 r2  1 a2
H  1 r3  2 a3   3  z3
H  2 r4  1 a4   4  z4
H  2 r4  1 a4   4 –z4
H  3 r5  2  a5  1  z5
H  3 r6  2  a6  1  z6
H  3 r6  2  a6  1 –z6
  -- blank line --
```



```
r1 1.43
r2 1.54
r3 0.96
r4 1.09
r5 1.09
r6 1.09
a2 109.5
a3 105.0
a4 109.5
a5 109.5
a6 109.5
z3 180.0
z4 60.0
z5 180.
z6 60.0
-- blank line --
```

Caution:

-leave blank spaces to separate different variables or numbers
-numbers with a decimal point
-blank lines are important!

Visualize the molecule as:

$ molden etoh.com

3. Molecular Potential Energy

3.3. Gaussian output

We can perform a calculation with the Gaussian program with the following command:

$ g09 < etoh.com > etoh.log &

Indicates the input

Indicates the output

Indicates that the process is executed in background, leaving the console free

When the calculation finishes we can find the result in the folder where we have executed the command, and where the input is found.

We can display the result as:

$ more etoh.log

One can also display the molecular structure with the programs molden and Gaussview

$ molden etoh.log
$ gv etoh.log

3. Molecular Potential Energy

3.3. Gaussian output

Input orientation:

Center Number	Atomic Number	Atomic Type	Coordinates (Angstroms) X	and	Z
1	8	0	0.000000	0.000000	0.000000
2	6	0	0.000000	0.000000	1.430000
3	6	0	1.451668	0.000000	1.944063
4	1	0	-0.927289	0.000000	-0.248466
5	1	0	-0.513740	-0.889823	1.793849
6	1	0	-0.513740	0.889823	1.793849
7	1	0	1.451668	0.000000	3.034063
8	1	0	1.966137	0.889823	1.581246
9	1	0	1.966137	-0.889823	1.581246

3. Molecular Potential Energy

3.3. Gaussian output

Distance matrix (angstroms):

		1	2	3	4	5
1	O	0.000000				
2	C	1.430000	0.000000			
3	C	2.426256	1.540000	0.000000		
4	H	0.960000	1.917580	3.235215	0.000000	
5	H	2.067271	1.090000	2.162678	2.265803	0.000000
6	H	2.067271	1.090000	2.162678	2.265803	1.779646
7	H	3.363462	2.163413	1.090000	4.053940	2.488522
8	H	2.675410	2.163413	1.090000	3.537166	3.059760
9	H	2.675410	2.163413	1.090000	3.537166	2.488974

		6	7	8	9
6	H	0.000000			
7	H	2.488522	0.000000		
8	H	2.488974	1.779646	0.000000	
9	H	3.059760	1.779646	1.779646	0.000000

Stoichiometry C2H6O
Framework group CS[SG(C2H2O),X(H4)]
Deg. of freedom 13
Full point group CS NOp 2
Largest Abelian subgroup CS NOp 2
Largest concise Abelian subgroup CS NOp 2

3. Molecular Potential Energy

3.3. Gaussian output

Standard orientation:

Center Number	Atomic Number	Atomic Type	Coordinates (Angstroms) X	and	Z
1	8	0	-1.221023	-0.197515	0.000000
2	6	0	0.000000	0.546800	0.000000
3	6	0	1.194531	-0.425155	0.000000
4	1	0	-1.915832	0.464935	0.000000
5	1	0	0.043276	1.174846	0.889823
6	1	0	0.043276	1.174846	-0.889823
7	1	0	2.125241	0.142190	0.000000
8	1	0	1.152517	-1.053286	-0.889823
9	1	0	1.152517	-1.053286	0.889823

Rotational constants (GHZ): 35.7263392 9.0866050 7.9737914

3. Molecular Potential Energy

3.3. Gaussian output

Standard basis: 3-21G (6D, 7F)
There are 29 symmetry adapted basis functions of A' symmetry.
There are 10 symmetry adapted basis functions of A" symmetry.
Integral buffers will be 262144 words long.
Raffenetti 1 integral format.
Two-electron integral symmetry is turned on.
39 basis functions, 63 primitive gaussians, 39 cartesian basis functions
13 alpha electrons 13 beta electrons
nuclear repulsion energy 81.3542011755 Hartrees.

SCF Done: E(RHF) = -153.220477909 A.U. after 5 cycles
Convg = 0.6729D-04 -V/T = 2.0023
S**2 = 0.0000

3. Molecular Potential Energy

3.3. Gaussian output

```
 Alpha  occ. eigenvalues --  -20.43526 -11.21836 -11.15646  -1.34678  -1.00895
 Alpha  occ. eigenvalues --   -0.83394  -0.67878  -0.64634  -0.55745  -0.52613
 Alpha  occ. eigenvalues --   -0.52430  -0.46903  -0.42788
 Alpha virt. eigenvalues --    0.25980   0.30160   0.33351   0.34507   0.35881
 Alpha virt. eigenvalues --    0.37800   0.41721   0.44682   0.92582   0.95982
 Alpha virt. eigenvalues --    0.98722   1.03244   1.07189   1.13566   1.25660
 Alpha virt. eigenvalues --    1.30995   1.33168   1.36410   1.37263   1.42097
 Alpha virt. eigenvalues --    1.88034   1.93109   2.00937   2.15428   2.30582
 Alpha virt. eigenvalues --    3.40932
```

3. Molecular Potential Energy

3.3. Gaussian output

Mulliken atomic charges:

		1
1	O	-0.673672
2	C	-0.088065
3	C	-0.602659
4	H	0.371734
5	H	0.181513
6	H	0.181513
7	H	0.192667
8	H	0.218485
9	H	0.218485

Sum of Mulliken charges= 0.00000

3. Molecular Potential Energy

3.3. Gaussian output

Atomic charges with hydrogens summed into heavy atoms:

		1
1	O	-0.301938
2	C	0.274960
3	C	0.026978
4	H	0.000000
5	H	0.000000
6	H	0.000000
7	H	0.000000
8	H	0.000000
9	H	0.000000

Sum of Mulliken charges= 0.00000
Electronic spatial extent (au): <R**2>= 196.2866
Charge= 0.0000 electrons
Dipole moment (field-independent basis, Debye):
X= 0.2851 Y= 2.0128 Z= 0.0000 Tot= 2.0329

3. Molecular Potential Energy

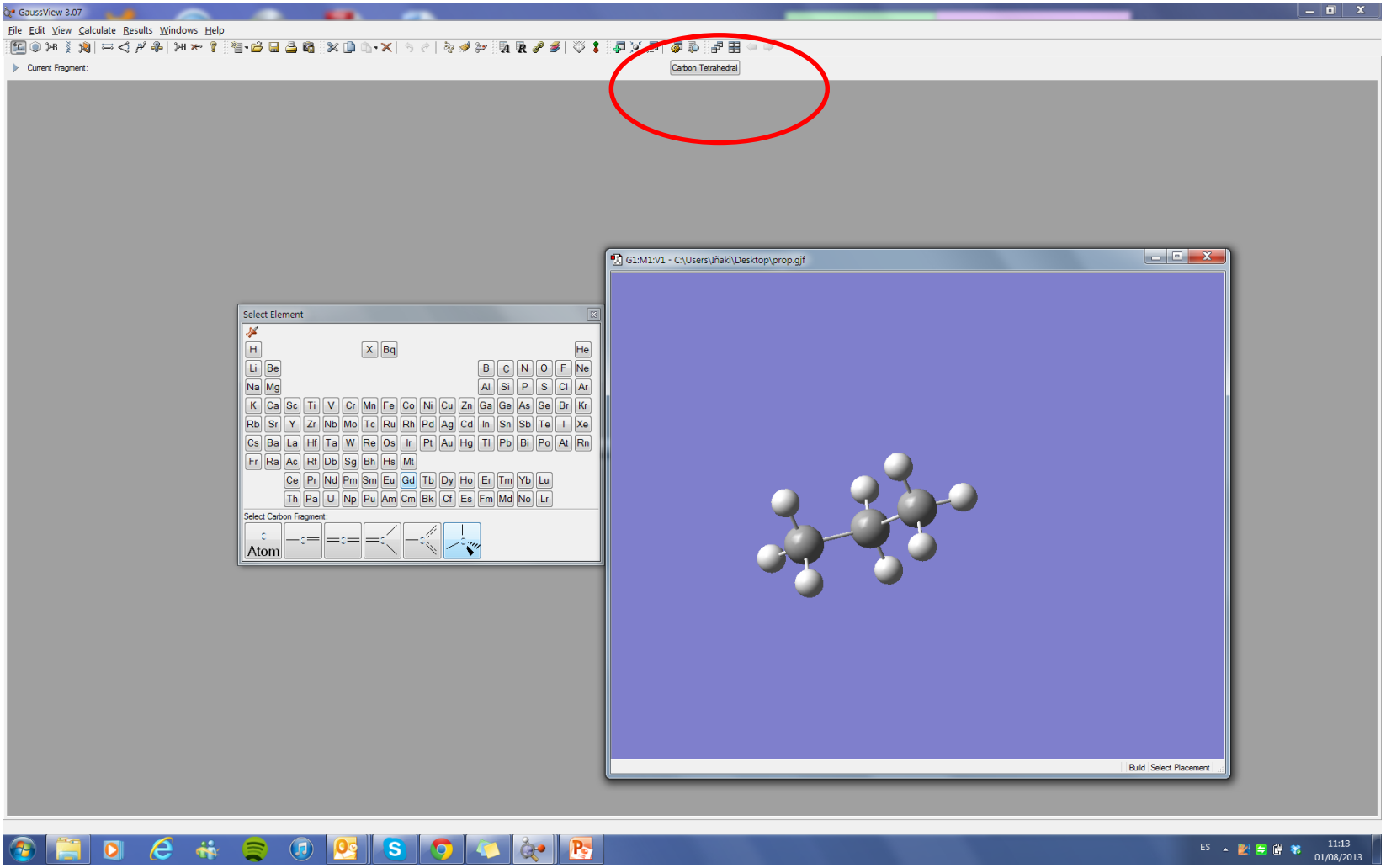
3.3. Gaussian output

1|1|UNPC-UNK|SP|RHF|3-21G|C2H6O1|PCUSER|09-Apr-2013|0||#HF/3-21G||etan
ol||0,1|O|C,1,1.43|C,2,1.54,1,109.5|H,1,0.96,2,105.,3,180.,0|H,2,1.09,
1,109.5,4,60.,0|H,2,1.09,1,109.5,4,-60.,0|H,3,1.09,2,109.5,1,180.,0|H,
3,1.09,2,109.5,1,60.,0|H,3,1.09,2,109.5,1,-60.,0||Version=x86-Win32-G0
3RevB.01|State=1-A'|HF=-153.2204779|RMSD=6.729e-005|Dipole=-0.6177796,
0.,0.507946|PG=CS [SG(C2H2O1),X(H4)]||@

HONESTY IN A LAWYER IS LIKE A HEN'S HIND LEGS.
-- MAGNUS OLESON, LAKE WOBEGON PATRIARCH, c.1875
Job cpu time: 0 days 0 hours 0 minutes 3.0 seconds.
File lengths (MBytes): RWF= 11 Int= 0 D2E= 0 Chk= 7 Scr= 1
Normal termination of Gaussian 03 at Tue Apr 09 10:44:53 2013.

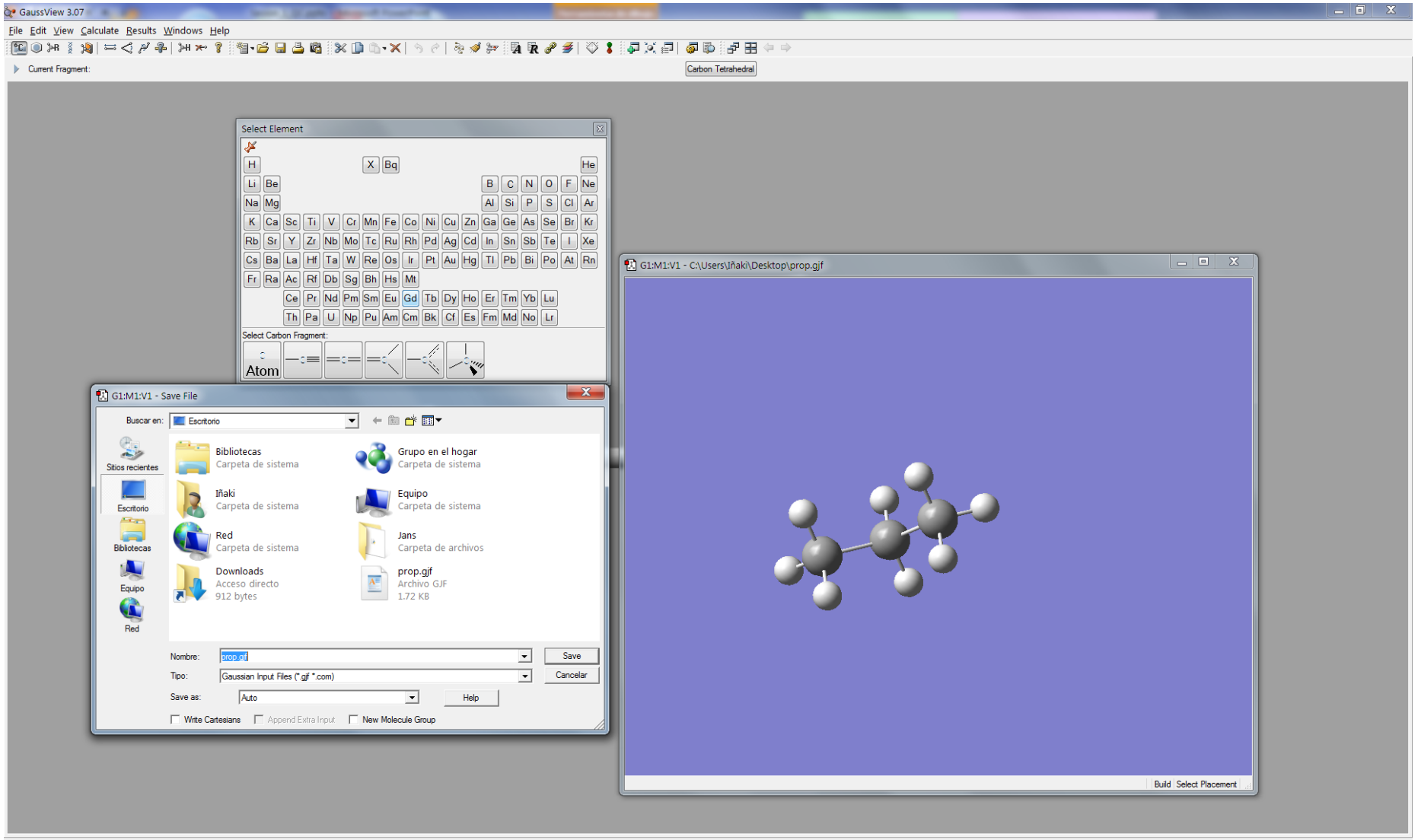
4. Graphical building of molecular structures

The z-matrixes, cartesian coordinates or even complete Gaussian inputs can be obtained by using the facilities of the GaussView program.



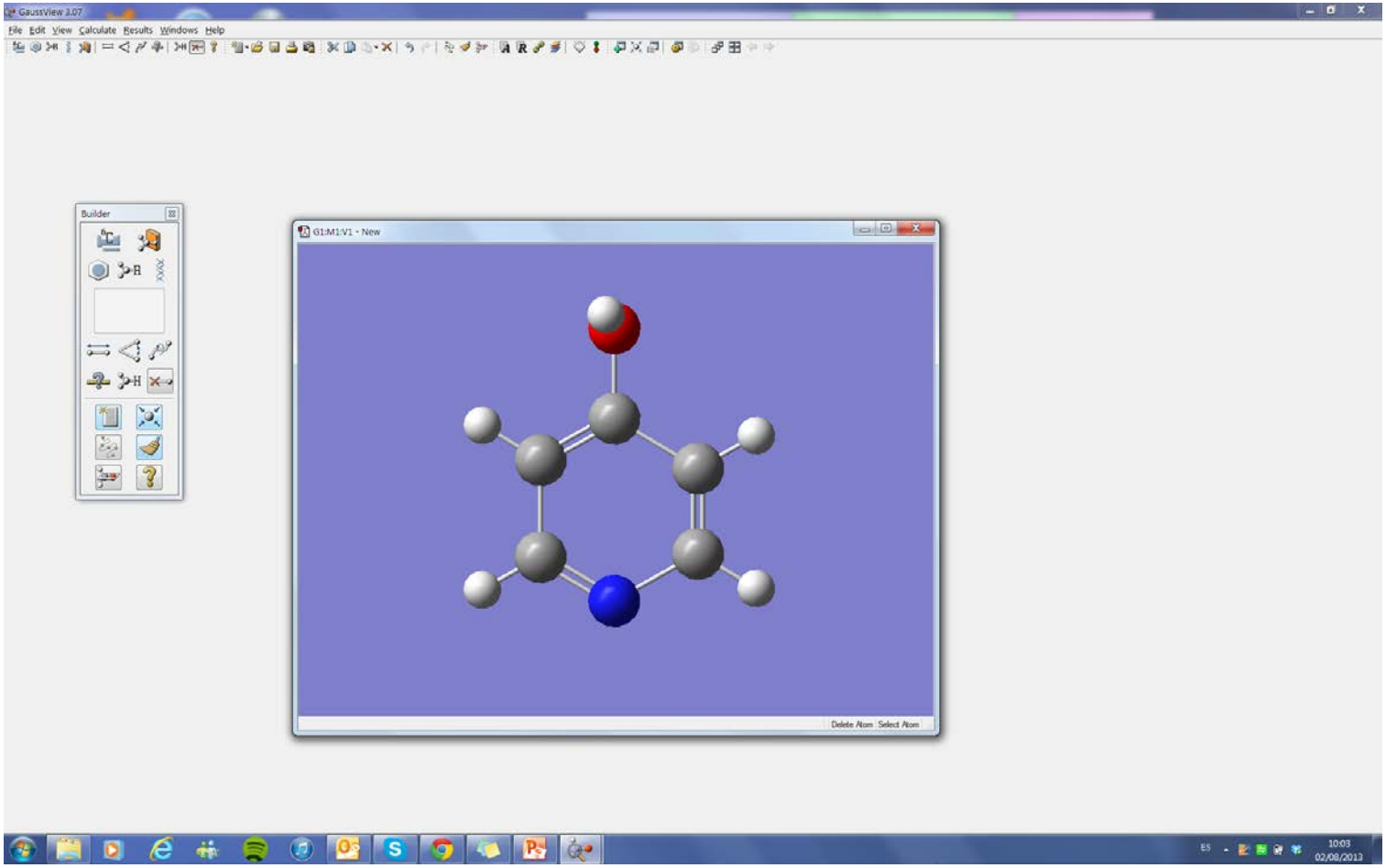
4. Graphical building of molecular structures

When the molecule is built, one can obtain the corresponding input with the ‘Save’ option from the ‘File’ menu



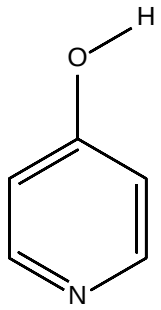
4. Graphical building of molecular structures

Available tools are shown using the option ‘Builder’ in the ‘View’ menu



4. Graphical building of molecular structures

Exercise 6: By means of the GaussView program build the Gaussian input for the 4-hydroxypyridine molecule and save it with the name 4hpir.com



Hartree-Fock calculations

Computational Chemistry

Elective Course

Chemistry Degree

4th year

Contents

- Energy calculations and electronic structure:
  - Basis sets
- Dissociation curves: LiH and HH
- **Concepts**: HF calculation and basis functions

Visualization

Exercise 1.

- Use Gaussview to visualize the following properties
  - Atomic charges
  - Molecular orbitals
- Systems:
  - **Formaldehyde** (HF/3-21G)
  - ethene, butadiene and hexatriene (HF/3-21G)
  - benzene & phenol (HF/STO-3G)
  - Use **guess=save** in the command line

Visualization

1- prepare the input (with Gaussview)

%chk=**file_name**.chk
#HF/3-21G guess=save

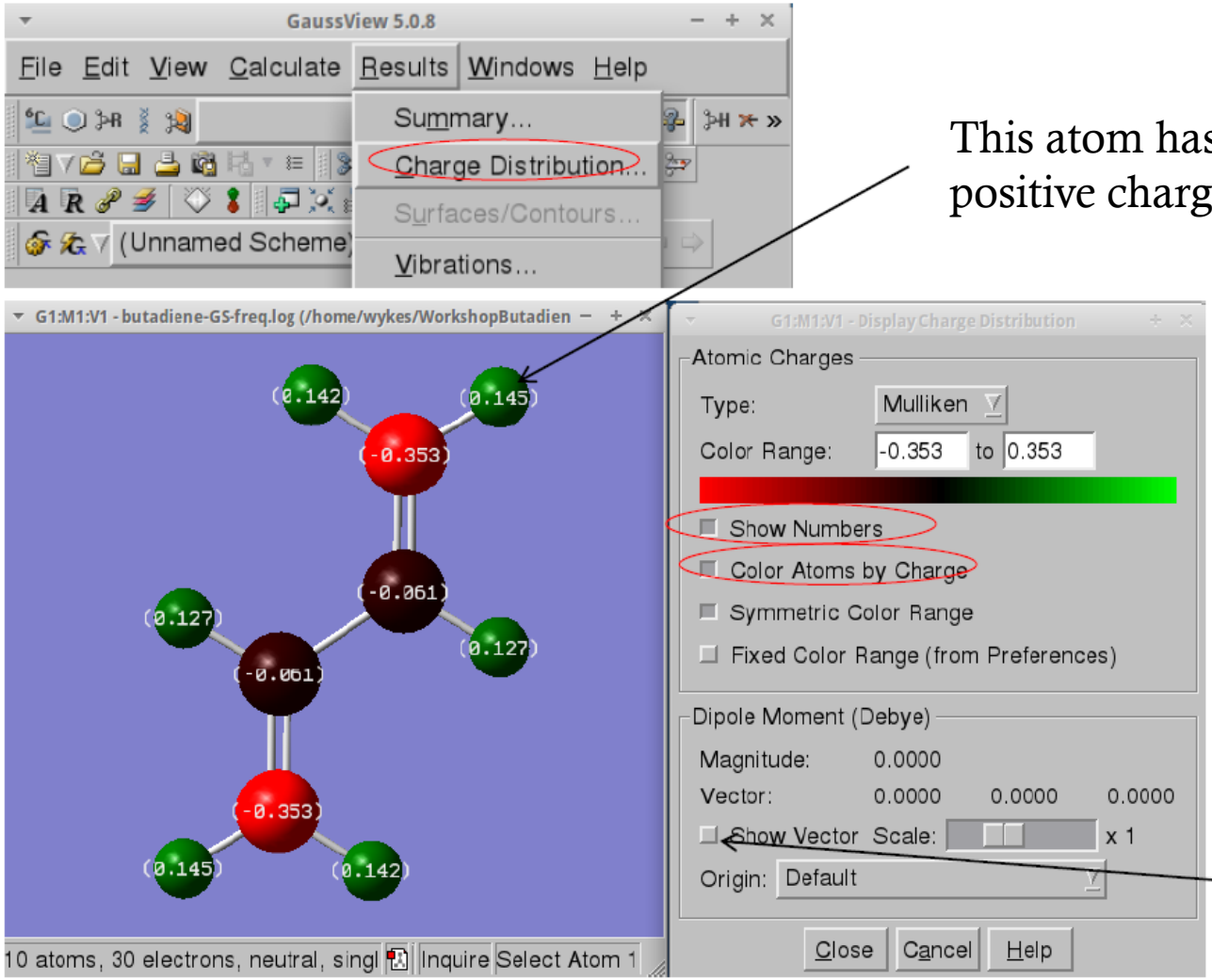
formaldehyde

0 1
Z-matrix

2- run the calculation **$ g09 < file_name.com > file_name.log &**

3- Open the chk file with Gaussview **$gv file_name.chk**

Atomic charges



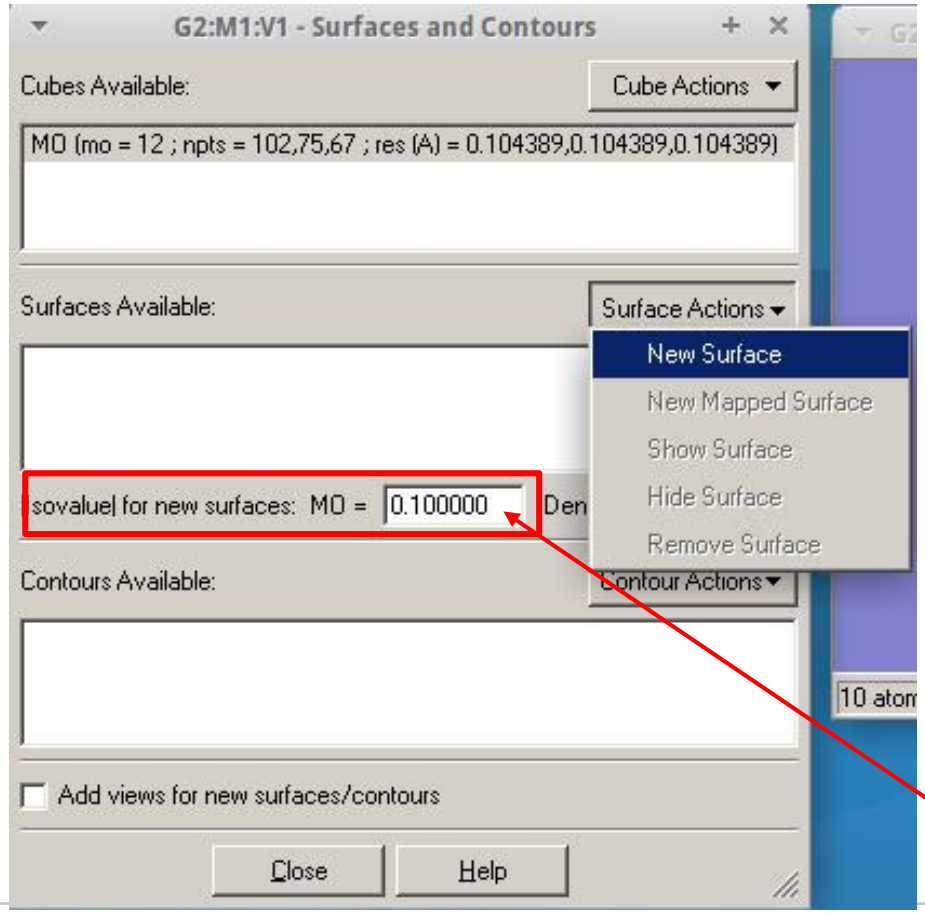
This atom has a positive charge

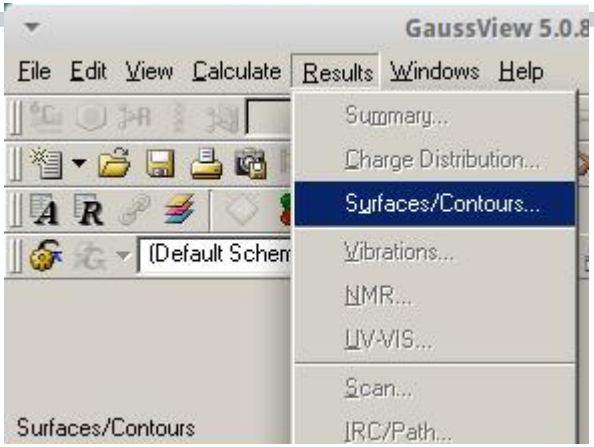
If the molecule has a non-zero dipole moment, an arrow indicating its direction can be shown be **clicking** this box

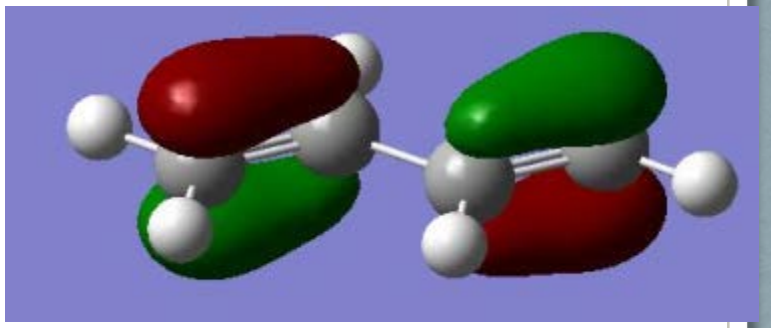
Molecular Orbitals HOMO and LUMO

From Gaussview select Surface/Contours from“Results” and open files *cube*

Represent HOMO, LUMO and HOMO-1 for formaldehyde

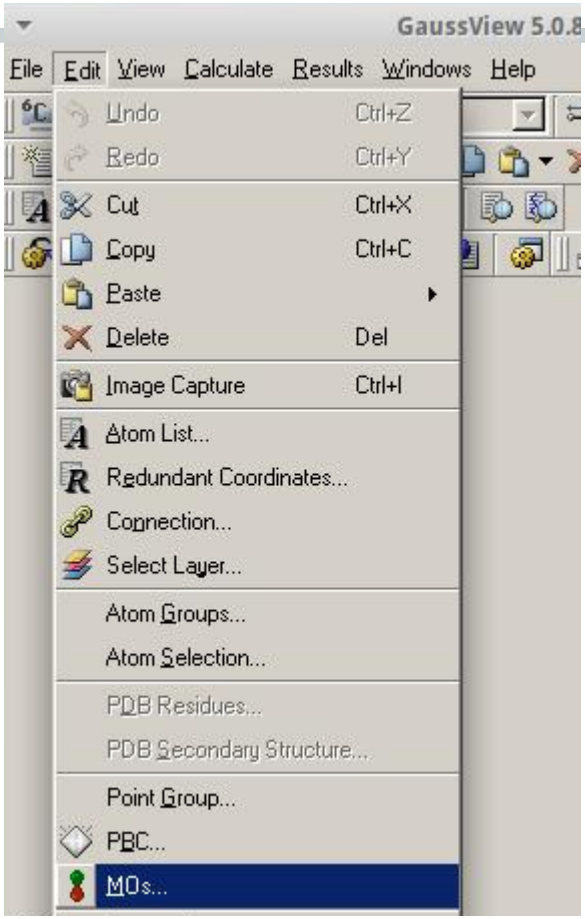


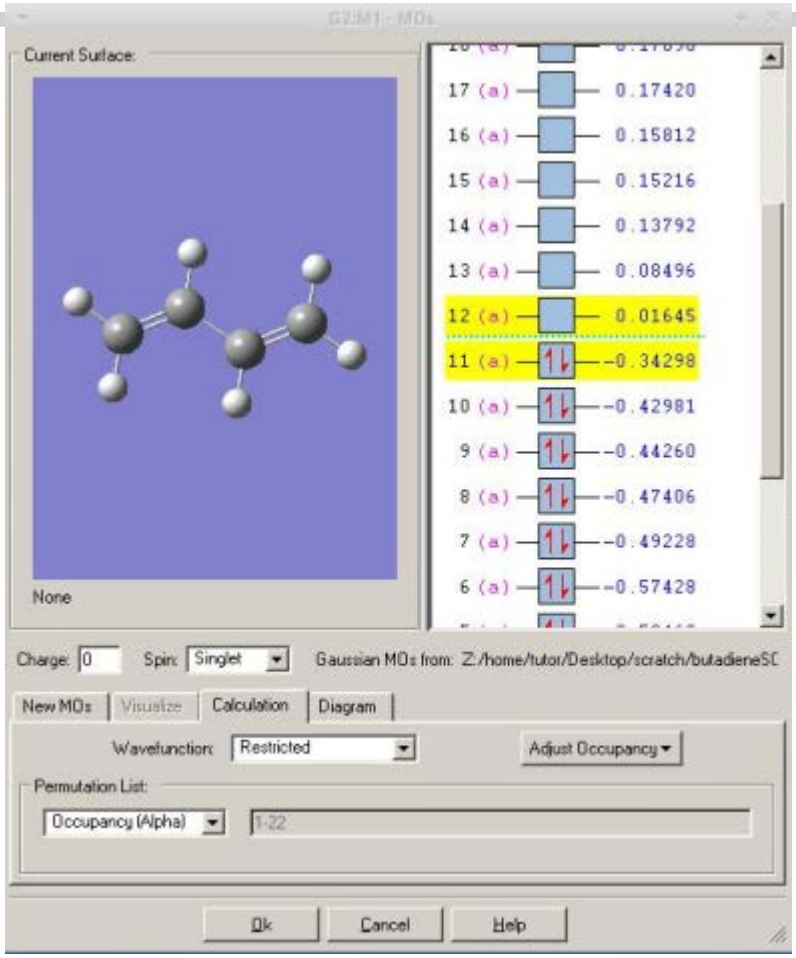




Choose the “isovalue”. The larger the value the smaller the volume specified by the MO surface

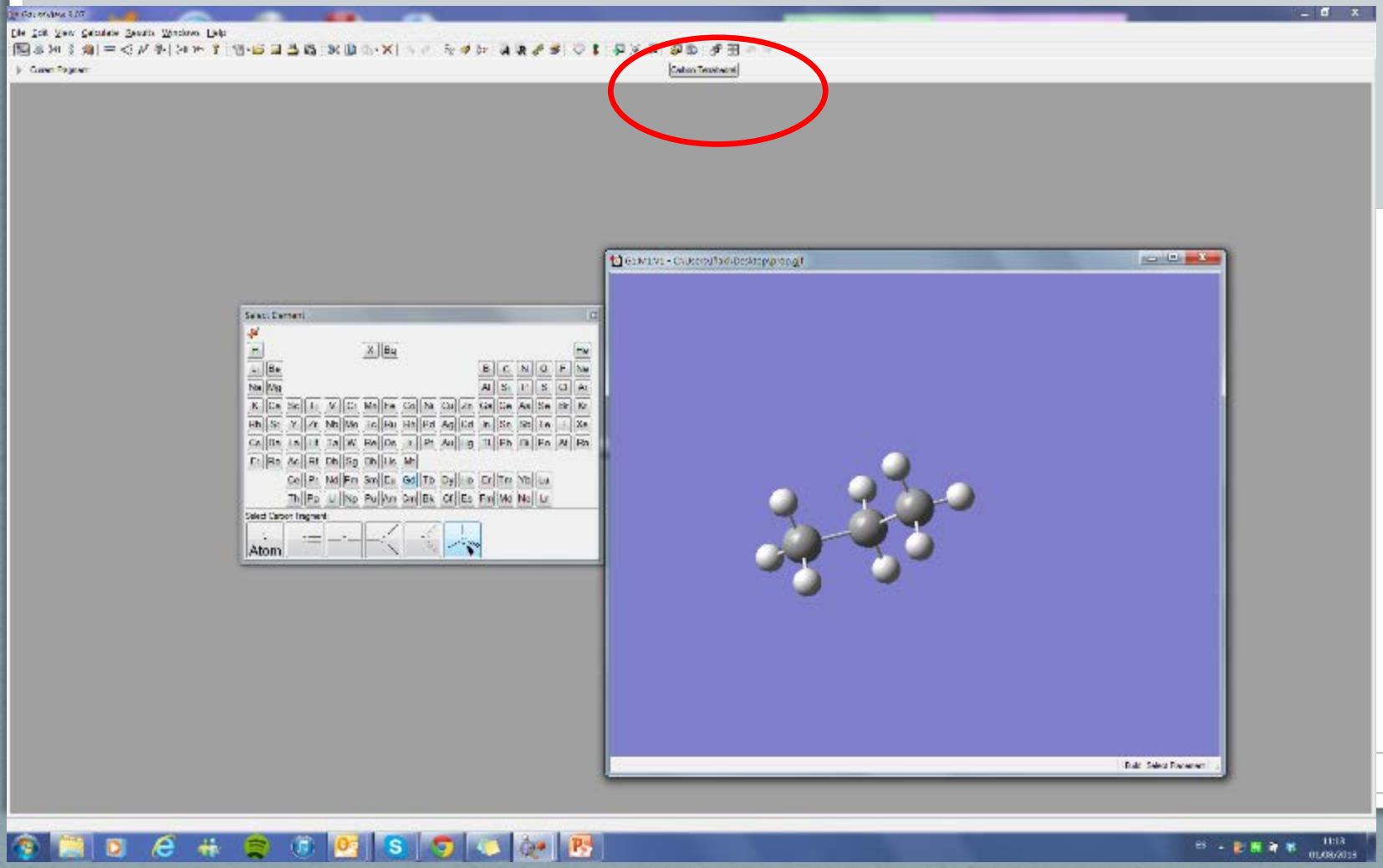
Molecular Orbitals





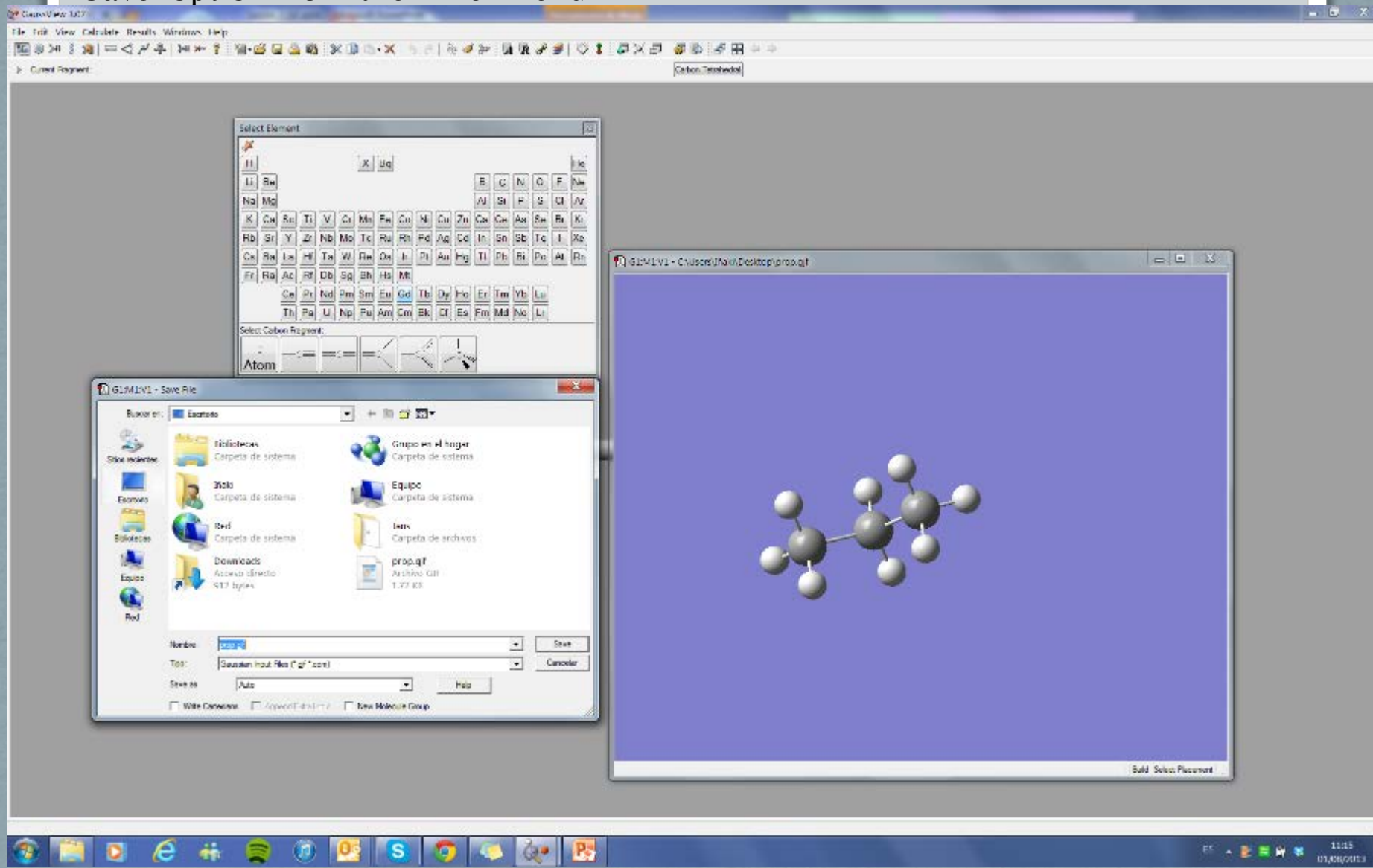
Graphical building of molecular structures

The z-matrixes, Cartesian coordinates or even complete Gaussian inputs can be obtained by using the GaussView program.



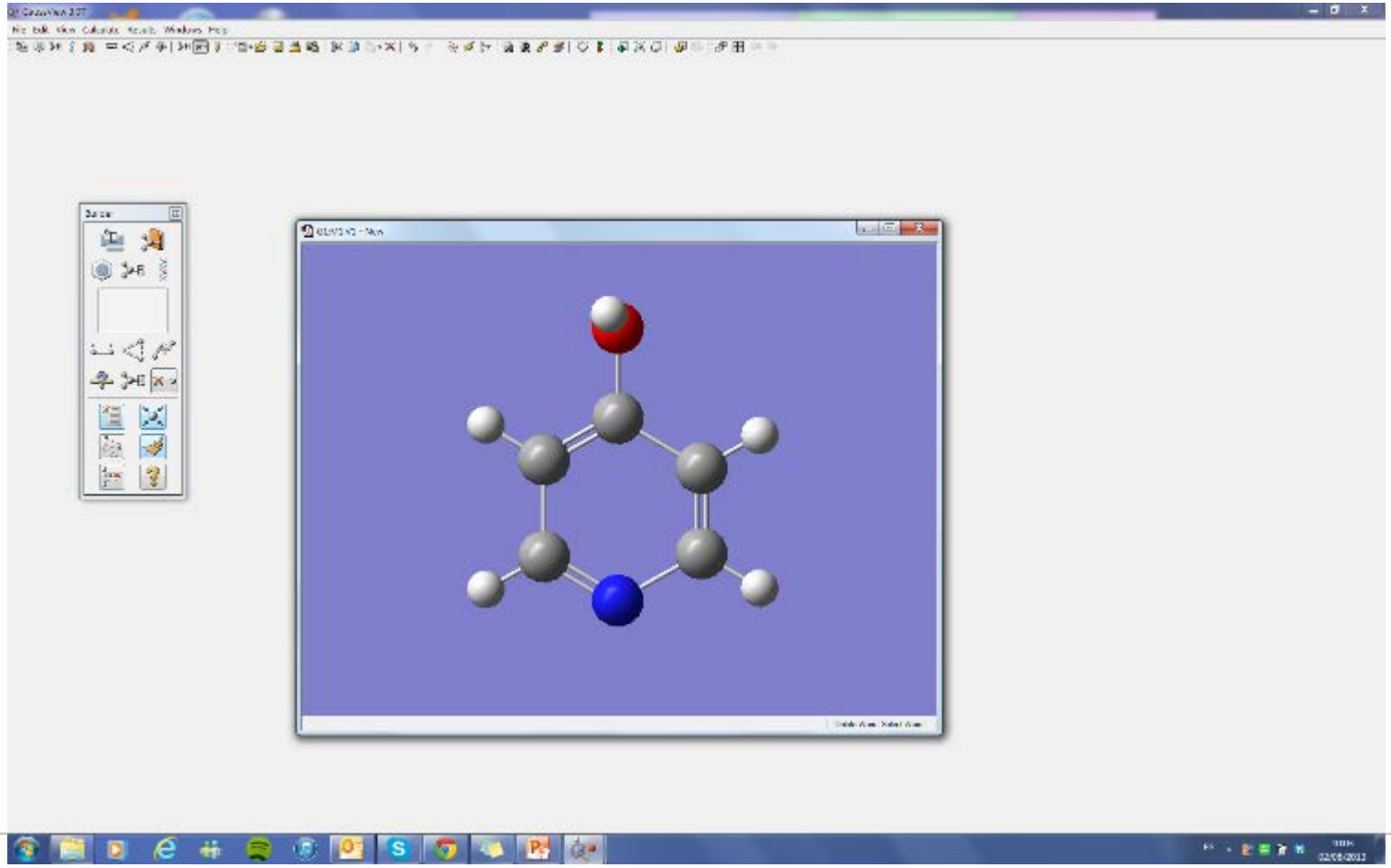
Graphical building of molecular structures

When the molecule is built one can obtain the corresponding input with the ‘Save’ option from the ‘File’ menu



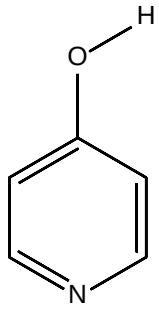
Graphical building of molecular structures

Available tools are shown using the option ‘Builder’ in the ‘View’ menu

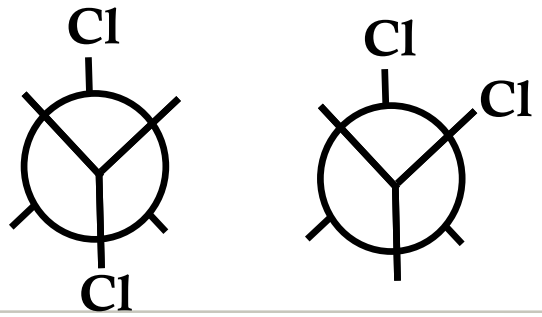


Graphical building of molecular structures

Exercise: By means of the GaussView program, build the Gaussian input for the 4-hydroxypyridine molecule and save it with the name 4hpir.com



Exercise: By means of the GaussView program, build the geometry of dichloroethane and then modify the geometry to obtain different conformations.



How are these calculations done on a computer?

$\{\phi_\mu \mid \mu = 1,2, \ldots K\}$ $\qquad \psi_i = \sum_{\mu=1}^{K} C_{\mu i}\, \phi_\mu \qquad i = 1,2, \ldots K$

$$\Psi_0 = \left| \psi_1 \bar{\psi}_1 \cdots \psi_a \bar{\psi}_a \cdots \psi_{N/2} \bar{\psi}_{N/2} \right\rangle \qquad f(\mathbf{r_1})\, \psi_i(\mathbf{r_1}) = \varepsilon_i\, \psi_i(\mathbf{r_1})$$

Eq. Roothaan: $\mathbf{F\,C = S\,C\,\varepsilon}$

$F_{\nu\mu} = \langle \phi_\nu | f | \phi_\mu \rangle$

$S_{\nu\mu} = \langle \phi_\nu | \phi_\mu \rangle$

$$\mathbf{C} = \begin{pmatrix} C_{11} & C_{12} & \cdots & C_{1K} \\ C_{21} & C_{22} & \cdots & C_{2K} \\ \vdots & \vdots & & \vdots \\ C_{K1} & C_{K2} & \cdots & C_{KK} \end{pmatrix} \qquad \varepsilon = \begin{pmatrix} \varepsilon_1 & 0 & 0 & 0 \\ 0 & \varepsilon_2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \varepsilon_K \end{pmatrix}$$

$\mathbf{F = F(C)}$

$\mathbf{F(C)\,C = S\,C\,\varepsilon}$

SCF Process (*Self-Consistent Field*)

How are these calculations done on a computer?

$$f\chi_i(\mathbf{x}_i) = \varepsilon_a \chi_a(\mathbf{x}_i)$$

$$\chi_i(\mathbf{x}_i) = \psi_j(\mathbf{r}_i)\eta(\omega_i) \rightarrow \eta(\omega_i)\begin{cases}\alpha(\omega_i)\\ \beta(\omega_i)\end{cases}$$

Spin-orbital

orbitals

$$f(\mathbf{r}_1)\,\psi_i(\mathbf{r}_1) = \varepsilon_i\,\psi_i(\mathbf{r}_1)$$

$$\psi_i = \sum_{\mu=1}^{K} C_{\mu i}\,\phi_\mu \qquad i = 1,2,\ldots K$$

Molecular orbitals

Atomic orbitals

Molecular orbitals are obtained as linear combinations of atomic orbitals (LCAO). The set of atomic orbitals used in the calculation is known as the **basis set**

Basis Sets: CGTO

- Expansion of the spatial orbitals for a set of basis functions $\{\phi_\mu\}$:

$$\psi_i = \sum_{\mu=1}^{K} C_{\mu i}\phi_\mu$$

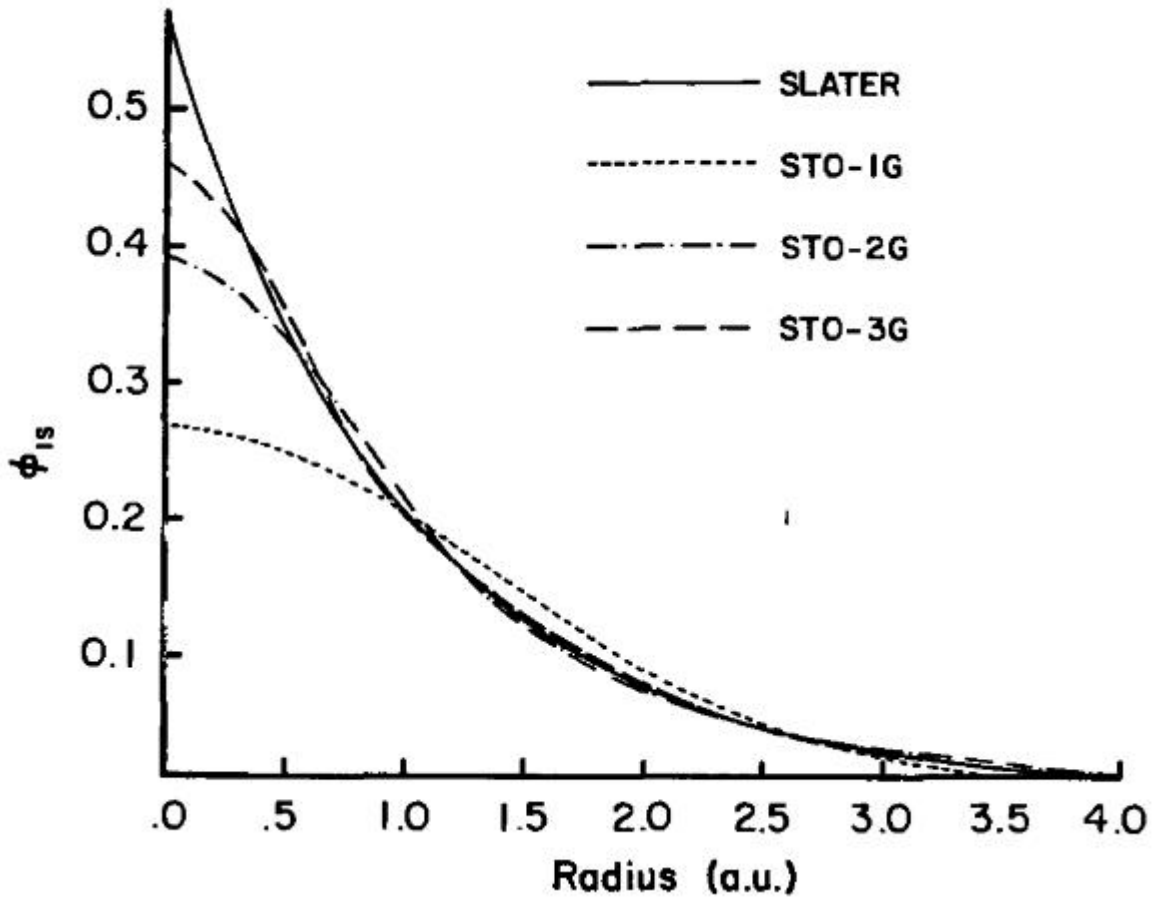
- **PROBLEM**: Find the basis set that best reproduces the orbitals with the minimal cost. Two types:
  - Slater functions (STO): function of $e^{-\zeta r}$: better at reproducing the orbitals, but the calculation is very costly
  - Gaussian functions (GTO): function of $e^{-a\cdot r^2}$: worse at reproducing the orbitals, but the calculation is very efficient

- **SOLUTION**: Adjust the STO as a fixed linear combination (contraction) of GTO (primitive):

CGTO: CONTRACTED GAUSSIAN TYPE ORBITALS

$$\phi_\mu^{CGF}(\mathbf{r}-\mathbf{R}_A) = \sum_{p=1}^{L} d_{p\mu} g_p(\alpha_{p\mu}, \mathbf{r}-\mathbf{R}_P)$$

Basis sets:
STO y GTO



CGTO

$$\phi_{\mu}^{CGF}(\mathbf{r}-\mathbf{R}_A)=\sum_{p=1}^{L} d_{p\mu} g_p(\alpha_{p\mu}, \mathbf{r}-\mathbf{R}_P)$$

- **L**: contraction length
- $\alpha_{\pi\mu}$, $\mathbf{d}_{\pi\mu}$: GTO exponents and contraction coefficients
- Primitive normalized functions GTO, 1s, 2p, 3d…:

$$g_{1s}(\alpha,\mathbf{r})=\left(8\alpha^3/\pi^3\right)^{1/4} e^{-\alpha r^2}$$

$$g_{2p_x}(\alpha,\mathbf{r})=\left(128\alpha^5/\pi^3\right)^{1/4} xe^{-\alpha r^2}$$

$$g_{3d_{xy}}(\alpha,\mathbf{r})=\left(2048\alpha^7/\pi^3\right)^{1/4} xye^{-\alpha r^2}$$

Types of basis function sets

- Minimal basis sets: Each AO $\rightarrow$ a CGTO function $\Rightarrow$ STO-3G
- Double ζ: Each AO $\rightarrow$ two CGTO functions $\Rightarrow$ cc-PVDZ
- Triple ζ: … $\Rightarrow$ cc-PVTZ
- Split-Valence: $\Rightarrow$ 3-21G, 4-31G, 6-31G, 6-311G…
  - Each *core* AO $\rightarrow$ one CGTO function
  - Each valence AO $\rightarrow$ two (or more) CGTO functions
- Extended functions: $\Rightarrow$ 6-31G*, 6-31G**, 6-31+G*, aug-cc-PVTZ
  - Polarization functions (*): Adding functions of upper □
  - Diffuse functions (+): Adding functions of very small α

All AO until complete valence are included

AS MORE functions are added MORE FLEXIBILITY and QUALITY in the Basis Set but the calculation is MORE COSTLY

Minimal basis sets: STO-3G

- 1 CGTO per AO until the valence shell:
  - H and He $\rightarrow$ 1 function: 1s
  - Li to Ne $\rightarrow$ 5 functions: 1s, 2sp (4 functions)
  - Na to Ar $\rightarrow$ 9 functions: 1s, 2sp, 3sp
  - K and Ca $\rightarrow$ 13 functions: 1s, 2sp, 3sp, 4sp
  - Sc to Kr $\rightarrow$ 18 functions: 1s, 2sp, 3sp, 3d, 4sp…
- Each CGTO is a contraction of 3 GTO
- The s & p AOs form a unique sp shell (4 functions)
- GTOs are adjusted to the corresponding STO form by minimizing the integral of the square of the difference between STO-CGTO

Split valence: N-X1(1)G

- *core* AO $\rightarrow$ one CGTO with N primitive: minimal basis
  - *core*: {1s} for 2º period, {1s,2sp} for 3th…
- valence AO $\rightarrow$ 2 (or 3) CGTO: doble-ζ basis set (or triple-ζ)
  - The most internal CGTO with X Gaussians
  - External CGTO with 1 Gaussian
- 3-21G :
  - *core* AO $\rightarrow$ 1 CGTO with 3 Gaussians
  - valence AO $\rightarrow$ 2 CGTO, one with 2 GTO and another with 1 GTO
- 4-31G: *core* with 4 GTO and valence with 3+1 GTO
- 6-31G: 1 *core* orbital with 6 GTO and 2 valence orbitals with 3+1 GTO
- 6-311G: *core* with 6 GTO and valence with 3+1+1 GTO

Split valence: 3-21G

- For Li to F atoms the contraction scheme is:

$$\phi_{1s}\left(\mathbf{r}\right) = \sum_{i=1}^{3} d_{i,1s} g_{1s}\left(\alpha_{i,1s}, \mathbf{r}\right)$$

$$\phi'_{2s}\left(\mathbf{r}\right) = \sum_{i=1}^{2} d'_{i,2s} g_{1s}\left(\alpha'_{i,2sp}, \mathbf{r}\right); \quad \phi'_{2p}\left(\mathbf{r}\right) = \sum_{i=1}^{2} d'_{i,2p} g_{2p}\left(\alpha'_{i,2sp}, \mathbf{r}\right)$$

$$\phi''_{2s}\left(\mathbf{r}\right) = g_{1s}\left(\alpha''_{2sp}, \mathbf{r}\right); \quad \phi''_{2p}\left(\mathbf{r}\right) = g_{2p}\left(\alpha''_{2sp}, \mathbf{r}\right)$$

Polarization functions

- Functions with upper ℓ values are added
  - First row atoms $\rightarrow$ ***p*** functions (or even upper ℓ)
  - Second row atoms $\rightarrow$ one or more sets of 6 ***d*** functions (xx, yy, zz, xy, yz, zx) or even upper ℓ values. These functions incorporate the AO "deformation" due to the inclusion of the atom in the molecule (polarization)
- 6-31G* or 6-31G(d) adds a set of polarization functions to all atoms EXCEPT for H
- 6-31G** or 6-31G (d,p) adds a set of polarization functions to all atoms INCLUDING H
- 6-31G(3df,2p): adds 2 GTO functions of ***p*** type to H, and 3 GTO of ***d*** type and 1 of ***f*** type to all other atoms

Diffuse functions: augmented basis sets

- To the previous basis sets, diffuse GTO functions of s type can be added (α very small)
- 6-31+G* or 6-31++G(3df,2p), for instance (first '+' is for all atoms except H, second '+' includes diffuse functions on H's)
- They allow a good description of situations where the electron density extends far from the nuclei (anions, calculation of polarizabilities…)

Basis set exercises

Exercise 2. Perform the following HF calculations with different basis sets using standard geometries (see Table in Session 1).

- Systems:

  H_2, HF, CO, N_2, CH_4, H_2O

- Basis sets:

  STO-3G, 3-21G, 4-31G, 6-31G* and 6-31G**

- Record the following properties in tables:
  - Total energy: all molecules
  - IP (Koopmans Theorem)
    - First IP of H_2, CH_4, H_2O, HF
  - Mulliken population analysis of CH_4, H_2O, HF
  - Dipolar moments of CO, H_2O, HF

Basis set exercises

Exercise 2. Perform the following HF calculations with different basis sets using standard geometries (see Table in Session 1).

- Geometries:

H_2, HF, CO, N_2, CH_4, H_2O

d(H-H)=0.74 Å

d(H-F)=0.92 Å

d(C-H)=1.09 Å HCH= 109.47 degrees

d(O-H)=0.96 Å HOH= 104.5 degrees

Basis set exercises

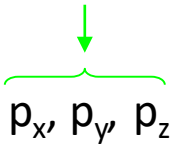
Exercise 2. Perform the following HF calculations with different basis sets using standard geometries (see Table in Session 1).

	STO-3G	3-21G	6-31G**
E(hartrees)			
IP (hartrees)			
μ (Debyes)			
Q_X (u.a.)			
Q_H (u.a.)			

Analysis of a simple system: The H_2 molecule

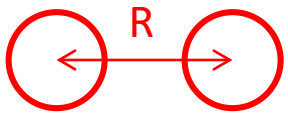
- Basis set:

Minimal Basis	H s
DZ	H s s´
DZ+P	H s s´ p



Basis set	Total number of functions
MB (STO-3G)	2
DZ (3-21G)	4
DZ+P (6-31G**)	10

- Geometry:



R (H–H) = 1.4 au = 0.74 Å

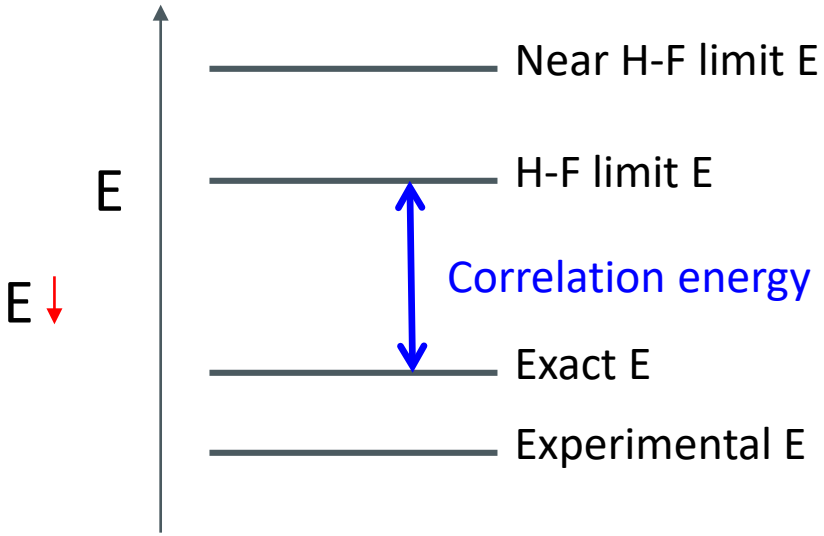
(1 au = 0.529177 Å)

The hydrogen molecule

- Total energies:

Basis set	energy (au)
MB	–1.117
DZ	–1.127
DZ+P	–1.131
H-F Limit	–1.134

Qualitative scale of total energies:



- Dissociation energy of H_2 in minimal basis set:

$E_{MB}(H) = -0.467$ au $\qquad E_{MB}(H_2) = -1.117$ au

$\Delta E_{MB} = 2\,E_{MB}(H) - E_{MB}(H_2) = 0.183$ au $\equiv 4.98$ eV $\qquad \Delta E_{exp} = 4.75$ eV

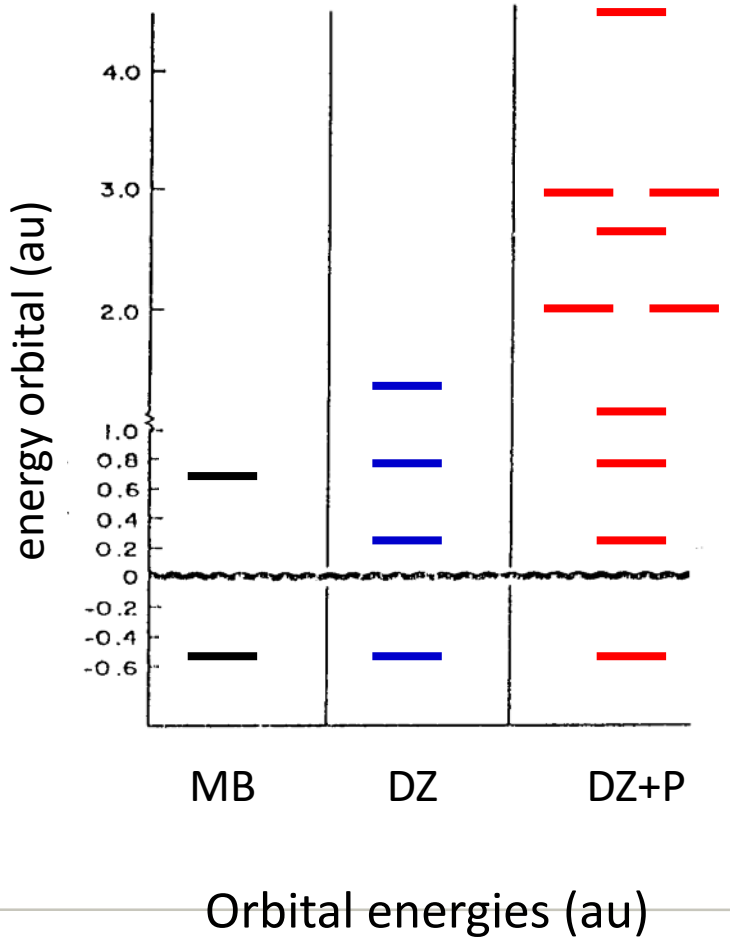
error≈5 %

au of energy (or **Hartree**)
1 au = 627.51 kcal·mol^{-1}
1 au = 27.2114 eV

The hydrogen molecule

- Ionization potentials (Koopmans theorem):

Basis Set	IP (au)	
MB	0.578	
DZ	0.596	
DZ+P	0.595	
Near H-F limit	0.595	2 %
Experimental	0.584	



Basis sets and bond length

Exercise 3. Determine the Hartree-Fock minimum energy distances (equilibrium geometry) for H_2 and HF with STO-3G, 3-21G and 6-31G(d,p) basis sets.

```
#HF/STO-3G Scan

Scan of H2

0 1
H
H 1 r1

r1 0.6 10 0.03
```

→ Perform 10 calculations with the energy changing r1 in steps of 0.03 Å, starting at 0.6 Å

```
#HF/STO-3G Scan

Scan of HF

0 1
H
F 1 r1

r1 0.8 10 0.03
```

Represent the energy versus the interatomic distance and estimate the position of the minimum.

Basis sets and bond length

Results of the HF/STO-3G scan for hydrogen molecule

Summary of the potential surface scan:

N	r1	SCF
1	0.6000	-1.10113
2	0.6300	-1.10928
3	0.6600	-1.11440
4	0.6900	-1.11698
5	0.7200	-1.11745
6	0.7500	-1.11615
7	0.7800	-1.11339
8	0.8100	-1.10940
9	0.8400	-1.10438
10	0.8700	-1.09851
11	0.9000	-1.09191

Represent the energy (SCF column) versus the interatomic distance (r1) and estimate the position of the minimum. This is the equilibrium bond length for this basis set. Repeat the calculations for other basis sets

The hydrogen molecule

Exercise 3. Determine the Hartree-Fock minimum energy distances (equilibrium geometry) for H_2 with STO-3G, 3-21G and 6-31G(d,p) basis sets

- Optimal SCF Bond Distances (R_{opt}):

Basis Set	R_{opt} (au) / (Å)
MB (STO-3G)	1.346 / 0.712
DZ (3-21G)	1.380 / 0.730
DZ+P(6-31G**)	1.385 / 0.733
Experimental	1.401 / 0.741

0.016 au = (≈1 %)

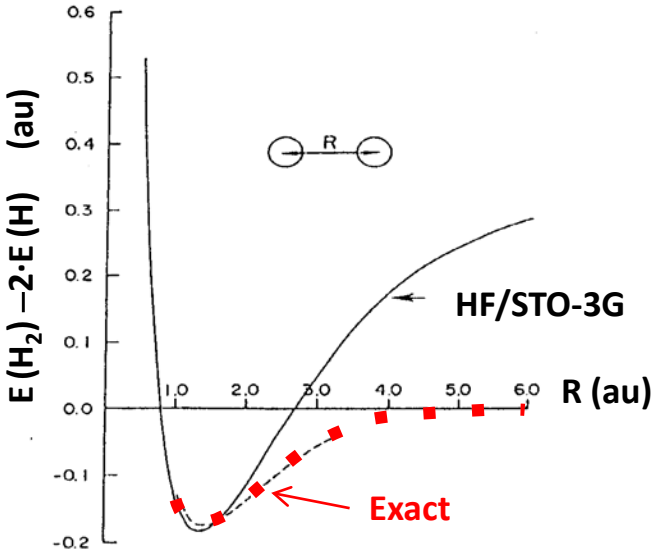
The hydrogen molecule

Exercise 4. Determine the RHF dissociation curve of H_2. Compare the energy of the molecule ($E(H_2)$) with that of the separated atoms ($2 \cdot E(H)$) obtained with the same basis set.

Represent $E(H_2)-2 \cdot E(H)$ vs the intermolecular distance

```
#HF/STO-3G Scan

Scan of H2

0 1
H
H 1 r1

R1 10.0 38 -0.25
```

```
#HF/STO-3G

Hydrogen atom

0 2
H
```



The hydrogen molecule

- Potential energy curves:

  MOs H_2 RHF in **M**inimal **B**asis **S**et $1s_A$ $1s_B$ H_A–H_B

(σ_u^*) ——— $\psi_2(i) = \frac{1}{\sqrt{2(1-s)}}(\phi_A(i) - \phi_B(i))$

$(\sigma_g)^2$ ↑↓ $\psi_1(i) = \frac{1}{\sqrt{2(1+s)}}(\phi_A(i) + \phi_B(i))$ $s = \langle \phi_A | \phi_B \rangle$

$\Psi_0({}^1\Sigma_g^+, R = R_{opt}) = |\psi_1 \ \bar{\psi}_1\rangle$ Spin-orbitals with spatial part ψ_1 and spin α and β

$$\Psi_0({}^1\Sigma_g^+, R = R_{opt}) = \frac{1}{2(1+s)}\left(|\phi_A\bar{\phi}_A\rangle + |\phi_B\bar{\phi}_B\rangle + |\phi_A\bar{\phi}_B\rangle + |\phi_B\bar{\phi}_A\rangle\right)$$

$H_A^-\ H_B^+$ $H_B^-\ H_A^+$ $H_A\bullet\ H_B\bullet$ $H_B\bullet\ H_A\bullet$

When $R \to \infty$, $s \to 0$:

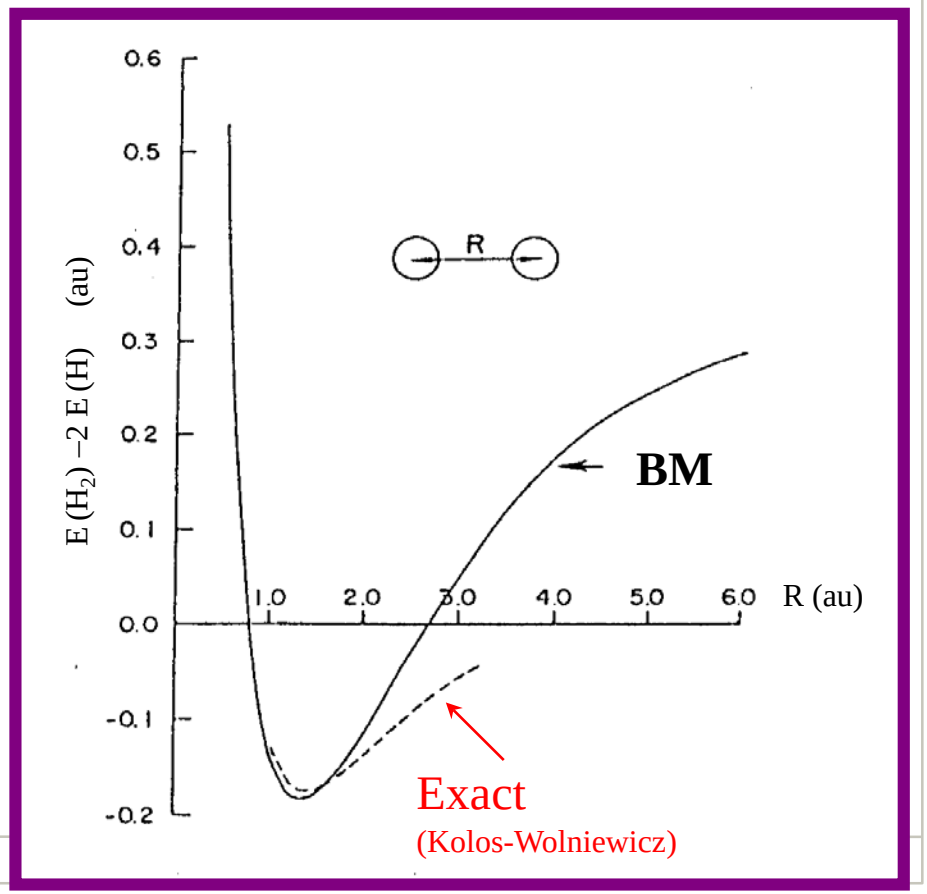
$$E_0 = \frac{\langle \Psi_0 | H | \Psi_0 \rangle}{\langle \Psi_0 | \Psi_0 \rangle} = \frac{1}{4}\left(4E(H) + 2E(H^-)\right) = E(H) + \frac{1}{2}E(H^-)$$

The hydrogen molecule

When $R \to \infty$, $s \to 0$:

$$E_0 = \underbrace{E(H) + \frac{1}{2}E(H^-)}_{\neq 2\,E(H)}$$

!!!!



The hydrogen molecule

- Potential energy curves:

True ground state of H_2 at $R \to \infty$: two $H(^2S)$ atoms

$$\Psi_0\left({}^1\Sigma_g^+, R=\infty\right) = \frac{1}{\sqrt{2}}\left(\left|\phi_A\bar{\phi}_B\right\rangle - \left|\bar{\phi}_A\phi_B\right\rangle\right)$$

Equivalent to

$$\Psi_0\left({}^1\Sigma_g^+, R=\infty\right) = \frac{1}{\sqrt{2}}\left(\left|\psi_1\bar{\psi}_1\right\rangle - \left|\psi_2\bar{\psi}_2\right\rangle\right)$$

Suggesting at $R=R_{opt}$

$$\Psi_{CI} = \underset{(\sigma_g)^2}{\left|\psi_1\bar{\psi}_1\right\rangle} + \lambda \underset{(\sigma_u^*)^2}{\left|\psi_2\bar{\psi}_2\right\rangle} \qquad \lambda < 0 \quad \text{If } R\to\infty, \quad \lambda \to -1$$

Left-right correlation:

- allows the correct dissociation, since
- covalent terms ↑ and ionic terms ↓ ⇒ Optimal bond distances → longer

The hydrogen molecule

- Potential energy curves:

Alternative: unrestricted solution? If the two spatial parts can be different then one can tend towards to ϕ_A and the other to ϕ_B

$$\Psi_{UHF} = \left|\psi_1^{\alpha}(1)\ \overline{\psi}_1^{\beta}(2)\right\rangle$$

- If R small $\quad \psi_1^{\alpha} = \psi_1^{\beta} = \psi_1 = c_A\phi_A + c_B\phi_B \ \Big\} \ c_A = c_B$ Restricted solution

- If R large

$$\left.\begin{aligned}\psi_1^{\alpha} &= c_A\phi_A + c_B\phi_B \\ \psi_1^{\beta} &= c_B\phi_A + c_A\phi_B\end{aligned}\right\} \ c_A \neq c_B$$

Unrestricted solution

- If $R\rightarrow\infty$

$$\Psi_{UHF}\ (R\rightarrow\infty) = \left|\phi_A(1)\ \overline{\phi}_B(2)\right\rangle$$

Dissociation Energy: OK!

Wavefunction: ~~OK~~!

H_2: The dissociation problem

- Closed-shell system, equilibrium geometry: RHF solution is OK
- At large internuclear distances the description tends towards two H atoms, at infinite distance 2 degenerated orbitals with 1 electron each ⇒ UHF
- At intermediate R_{AB} there can be more than one solution → it depends on the trial orbitals

UHF Exercises

Exercise 5. Determine the STO-3G UHF dissociation curves of H_2. Compare the energy to that of the separated atoms with the same basis set (E_{H2}-$2E_H$).

- The dissociation problem: H_2
  - H_2: UHF STO-3G calculation with R_{AB} from 0.7 to 10 Å, by using different trial functions $c_A=c_B$, $c_A \neq c_B$

```
#UHF/STO-3G Scan guess=mix

Scan of H2

0 1
H
H 1 r1

r1 10.0 38 -0.25
```

```
#UHF/STO-3G Scan

Scan of H2

0 1
H
H 1 r1

R1 10.0 38 -0.25
```

UHF Exercises

Exercise 6. Determine the UHF and RHF dissociation curves of H_2 and LiH using the 3-21G basis set. Compare the energy to that of the separated atoms with the same basis set.

- The dissociation problem: H_2, LiH
  - H_2: UHF 3-21G calculation with R_{AB} from 0.7 to 10 Å, by using different trial functions $c_A=c_B$, $c_A \neq c_B$
  - LiH: the same but from 1.2 to 7.5 Å

```
#UHF/3-21G Scan guess=mix

Scan of H2

0 1
H
H 1 r1

r1 10.0 38 -0.25
```

```
#RHF/3-21G Scan

Scan of H2

0 1
H
H 1 r1

r1 10.0 38 -0.25
```

Optimization of Molecular Structures

Computational Chemistry
Elective Course
Chemistry Degree
4th year

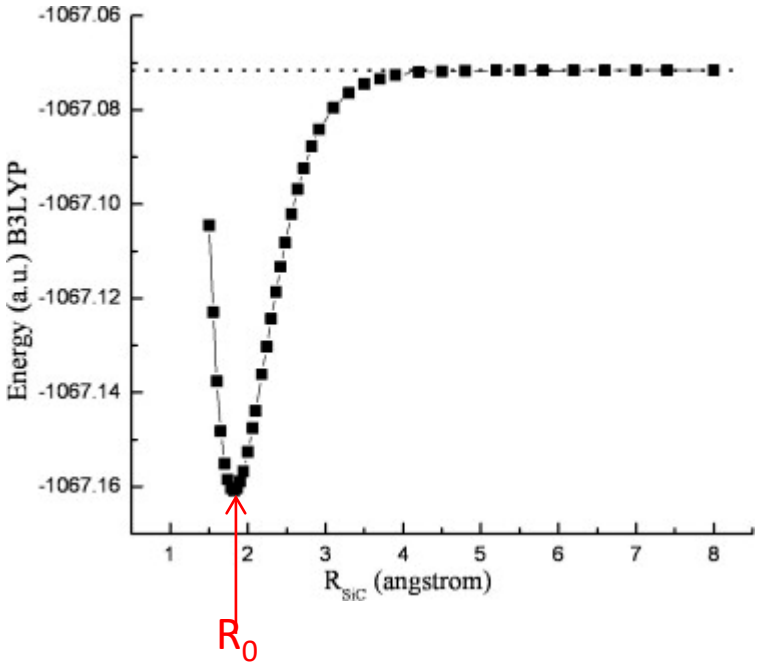
Content

1. Introduction to optimization
2. HF molecular optimization: basis set dependence
3. Frequency calculations and thermochemistry
4. DFT applied to molecular optimization
5. Conformational analysis

1. Optimization

1.1. Minimum energy structures

By studying the dissociation curves of diatomic molecules, we have obtained the molecular potential energy as a function of the distance:

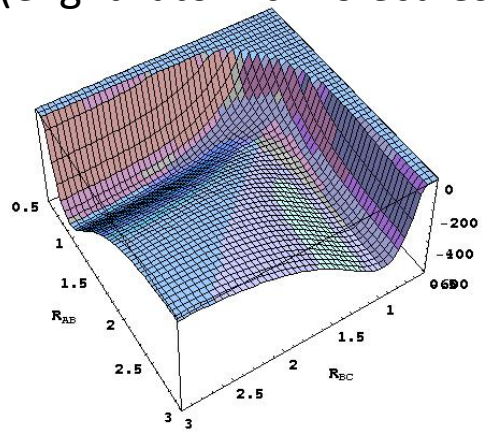


The minimum of the curve corresponds to an important structure. The derivative of the energy is zero, which means that the forces cancel. It is a stationary structure. But still, any small movement originates a force which tends to recover the structure with distance $R = R_0$. The minimum corresponds to a stable structure, observable in the laboratory.

1.1. Minimum energy structures

There are two ways to locate the minimum:

- By drawing the complete potential energy curve (potential energy surface). This procedure can be long and tedious, especially when the potential energy depends on more than one variable (e. g. triatomic molecules or larger)



- By using a method of optimization (direct search of the minimum of the potential energy function)

1.2. Optimizing a function

Suppose a function with just one variable, x. Using a Taylor series expansion we can calculate the value at a point from the value of the function and its derivatives in another:

$$f(x_{n+1}) = f(x_n) + \frac{1}{1!} f'(x_n)(x_{n+1} - x_n) + \frac{1}{2!} f''(x_n)(x_{n+1} - x_n)^2 + \ldots \quad (1)$$

Where the first derivative is the gradient (g) and the second the Hessian (h):

$$f'(x_n) = \left(\frac{df(x)}{dx} \right)_{x=x_n} = g(x_n) \qquad f''(x_n) = \left(\frac{d^2 f(x)}{dx^2} \right)_{x=x_n} = h(x_n) \quad (2)$$

1.2. Optimizing a function

It must be kept in mind that for a molecule of N atoms, with 3N Cartesian coordinates, the gradient is a vector 3N and the Hessian a 3N x 3N matrix. They are normally calculated in Cartesian and then transformed into the required coordinate system (e.g.: internal). In general for coordinates q

$$\mathbf{g} = \begin{bmatrix} \dfrac{\partial E}{\partial q_1} \\ \dfrac{\partial E}{\partial q_2} \\ \vdots \\ \vdots \\ \dfrac{\partial E}{\partial q_{3N}} \end{bmatrix} \qquad \mathbf{h} = \begin{bmatrix} \dfrac{\partial^2 E}{\partial q_1^2} & \dfrac{\partial^2 E}{\partial q_1 \partial q_2} & \cdots & \dfrac{\partial^2 E}{\partial q_1 \partial q_{3N}} \\ \dfrac{\partial^2 E}{\partial q_2 \partial q_1} & \dfrac{\partial^2 E}{\partial q_2^2} & \cdots & \dfrac{\partial^2 E}{\partial q_2 \partial q_{3N}} \\ \vdots & \vdots & \cdots & \\ \dfrac{\partial^2 E}{\partial q_{3N} \partial q_1} & \cdots & \cdots & \dfrac{\partial^2 E}{\partial q_{3N}^2} \end{bmatrix} \qquad (3)$$

Optimization methods can be classified depending on whether they make use only of the function, the function and its gradient, or function, gradient and Hessian.

Gradient methods

Steepest slope method: This method looks for a new point in the direction in which the value of the function decreases, i.e. in the direction of the gradient

$$x_{n+1} = x_n - g(x_n)\cdot\lambda$$

Where λ is the size of the 'step size' of minimization that can be set by the user or by the algorithm at each step of the optimization. A step size value sufficiently small (and $\lambda > 0$) ensures that

$$f(x_{n+1}) \leq f(x_n)$$

$$f(x_{n+1}) = f(x_n) + \frac{1}{1!} f'(x_n)(x_{n+1} - x_n) + \ldots = f(x_n) + g(x_n)(x_{n+1} - x_n) + \ldots =$$

$$= f(x_n) - \boldsymbol{\lambda}\cdot g(x_n) g(x_n) + \ldots = f(x_n) - \boldsymbol{\lambda}\cdot g^2(x_n) + \ldots$$

Quasi-Newton methods

We want the new point x_{n+1} to be a minimum, so its first derivative should be zero. Taking a Taylor expansion of the derivative at that point :

$$g(x_{n+1}) = g(x_n) + h(x_n)(x_{n+1} - x_n) + \ldots = 0$$

Then the desired point is:

$$x_{n+1} = x_n - \frac{g(x_n)}{h(x_n)} \tag{4}$$

Due to the approximate nature of the truncated series, the point is usually obtained iteratively, modifying the above equation to introduce a 'step size' in the optimization. Considering that, in general, the gradient is a vector and the Hessian a matrix:

$$\boldsymbol{x}_{n+1} = \boldsymbol{x}_n - \lambda \cdot \boldsymbol{h}^{-1} \cdot \boldsymbol{g} \tag{5}$$

Quasi-Newton methods

This method requires at least the knowledge of the function and its gradient. The Hessian can be estimated as:

$$h(x_n) = \frac{g(x_{n+1}) - g(x_n)}{(x_{n+1} - x_n)} \tag{5}$$

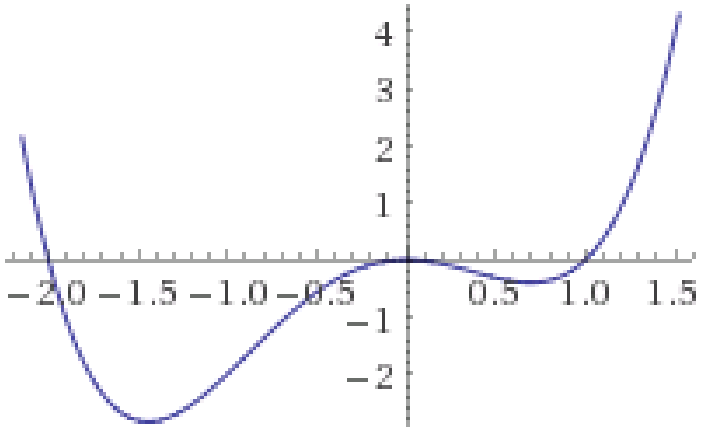
The estimation of the Hessian (equation (5)) becomes an undetermined problem when we have more than one variable. The different methods (Berny, BFGS,...) are different in the way they obtain the Hessian

Limitations:

* Works well in environments where the function is approximately quadratic
* if the Hessian is analytical the methods allow us to find any type of stationary point (first derivative zero, i.e. minima and maxima)
* if the initial estimate of the Hessian is not suitable, the method may not converge

Example using gradient methods

$$f(x) = x^4 + x^3 - 2x^2$$



(x from -2 to 1.5)

We want to find the minimum of this function starting from x = 2

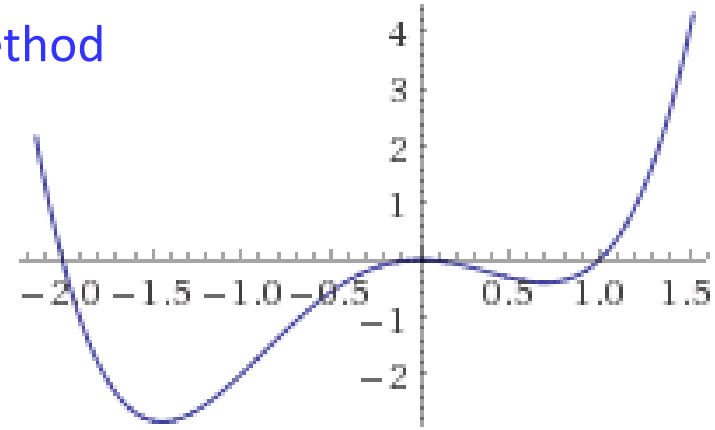
$$g(x) = 4x^3 + 3x^2 - 4x$$

$$x_{n+1} = x_n - g(x_n)\cdot\lambda \quad \lambda = 0.01$$

n	x_n	$f(x_n)$	$g(x_n)$	x_{n+1}
1	2	16.000	36.000	1.640
2	1.640	6.266	19.153	1.448
5	0.243	-0.101	-0.738	0.317
8	0.488	-0.303	-0.773	0.565
12	0.678	-0.396	-0.087	0.687
15	0.692	-0.397	-0.006	0.693
16	0.693	-0.397	-0.003	0.693

Example using the Quasi-Newton method

$$f(x) = x^4 + x^3 - 2x^2$$



(x from -2 to 1.5)

We want to find the minimum of this function starting from x = 2

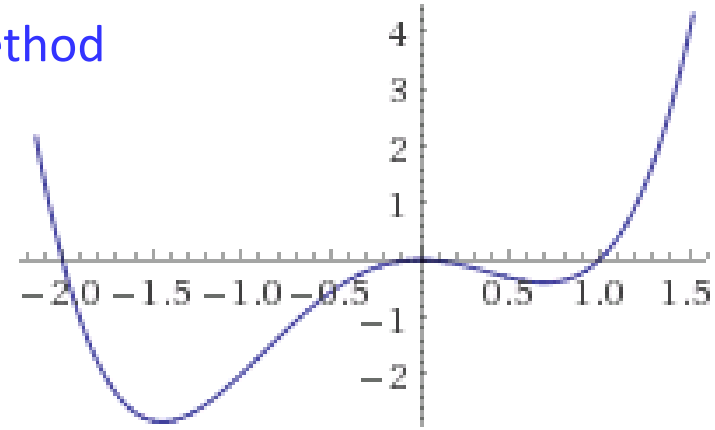
$$g(x) = 4x^3 + 3x^2 - 4x \qquad h(x) = 12x^2 + 6x - 4$$

$$x_{n+1} = x_n - \frac{g(x_n)}{h(x_n)}$$

n	x_n	$f(x_n)$	$g(x_n)$	$h(x_n)$	x_{n+1}
1	2.000	16.000	36.000	56.000	1.357
2	1.357	2.208	10.095	26.245	0.972
3	0.972	-0.077	2.626	13.183	0.773
4	0.773	-0.376	0.550	7.815	0.703
5	0.703	-0.397	0.060	6.145	0.693
6	0.693	-0.397	0.001	5.925	0.693

Example using the Quasi-Newton method

$$f(x) = x^4 + x^3 - 2x^2$$



(x from -2 to 1.5)

We want to find the minimum of this function starting from x = 0.1

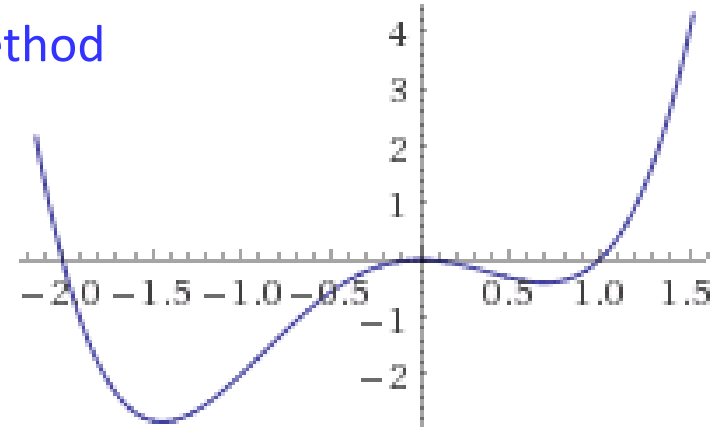
$$g(x) = 4x^3 + 3x^2 - 4x \qquad h(x) = 12x^2 + 6x - 4$$

$$x_{n+1} = x_n - \frac{g(x_n)}{h(x_n)}$$

n	x_n	$f(x_n)$	$g(x_n)$	$h(x_n)$	x_{n+1}
1	0.100	-0.019	-0.366	-3.280	-0.012
2	-0.012	0.000	0.047	-4.068	0.000
3	0.000	0.000	0.000	-4.001	0.000

Example using the Quasi-Newton method

$$f(x) = x^4 + x^3 - 2x^2$$



(x from -2 to 1.5)

We want to find the minimum of this function starting from x = -1

$$g(x) = 4x^3 + 3x^2 - 4x \qquad h(x) = 12x^2 + 6x - 4$$

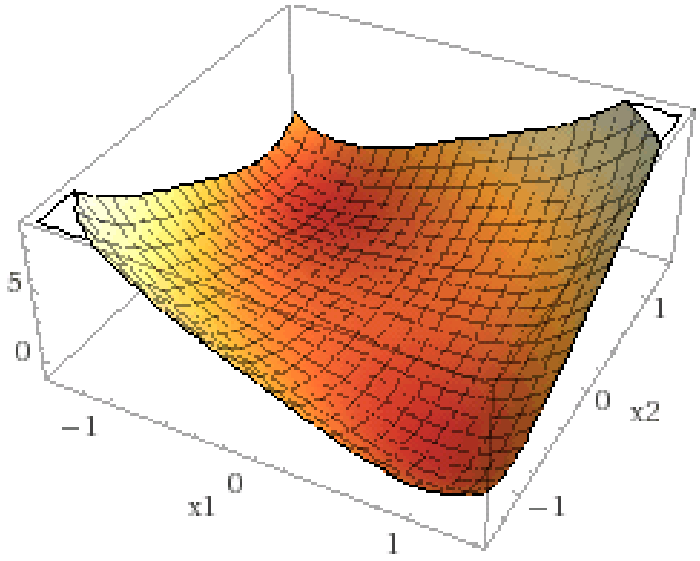
$$x_{n+1} = x_n - \frac{g(x_n)}{h(x_n)}$$

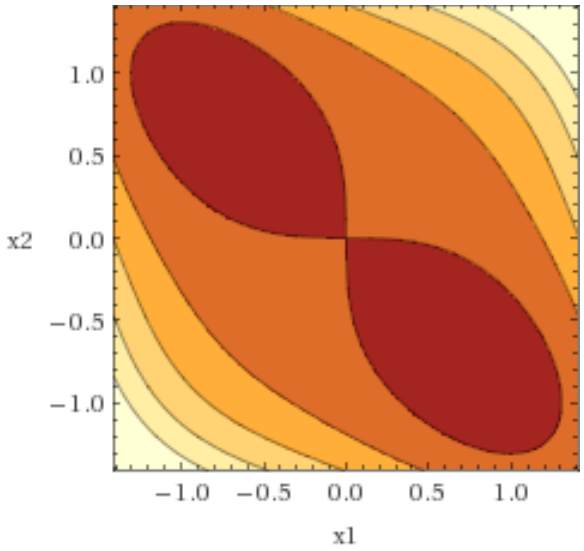
n	x_n	$f(x_n)$	$g(x_n)$	$h(x_n)$	x_{n+1}
1	-1.000	-2.000	3.000	2.000	-2.500
2	-2.500	10.938	-33.750	56.000	-1.897
3	-1.897	-1.071	-8.931	27.814	-1.576
4	-1.576	-2.712	-1.906	16.356	-1.460
5	-1.460	-2.832	-0.210	12.810	-1.443
6	-1.443	-2.833	-0.004	12.338	-1.443

Example of minimization with more than one variable

$$f(x_1,x_2)=x_1^4+3x_1x_2+x_2^4$$

$$\boldsymbol{g}(x_1,x_2)=\begin{bmatrix}4x_1^3+3x_2\\4x_2^3+3x_1\end{bmatrix} \qquad \boldsymbol{h}(x_1,x_2)=\begin{pmatrix}12x_1^2 & 3\\3 & 12x_2^2\end{pmatrix}$$





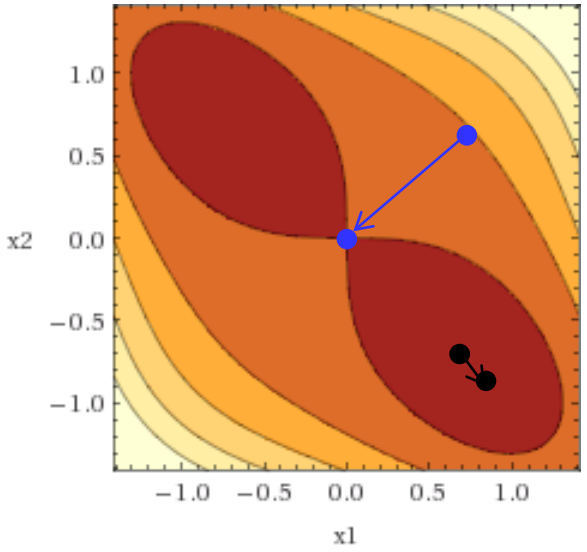
Example of minimization with more than one variable

$$\boldsymbol{x}_{n+1} = \boldsymbol{x}_n - \boldsymbol{h}^{-1} \cdot \boldsymbol{g}$$

$$\begin{pmatrix} x_{1,n+1} \\ x_{2,n+1} \end{pmatrix} = \begin{pmatrix} x_{1,n} \\ x_{2,n} \end{pmatrix} - \begin{pmatrix} 12x_1^2 & 3 \\ 3 & 12x_2^2 \end{pmatrix}^{-1} \cdot \begin{pmatrix} 4x_1^3 + 3x_2 \\ 4x_2^3 + 3x_1 \end{pmatrix}$$

n	$x_{1,n}$	$x_{2,n}$	Δx_1	Δx_2	$x_{1,n+1}$	$x_{2,n+1}$
1	0.700	0.700	-0.391	-0.391	0.309	0.309
2	0.309	0.309	-0.252	-0.252	0.057	0.057
3	0.057	0.057	-0.056	-0.056	0.000	0.000
4	0.000	0.000	0.000	0.000	0.000	0.000

n	$x_{1,n}$	$x_{2,n}$	Δx_1	Δx_2	$x_{1,n+1}$	$x_{2,n+1}$
1	0.700	-0.700	0.253	-0.253	0.953	-0.953
2	0.953	-0.953	-0.076	0.076	0.877	-0.877
3	0.877	-0.877	-0.010	0.010	0.866	-0.866
4	0.866	-0.866	0.000	0.000	0.866	-0.866

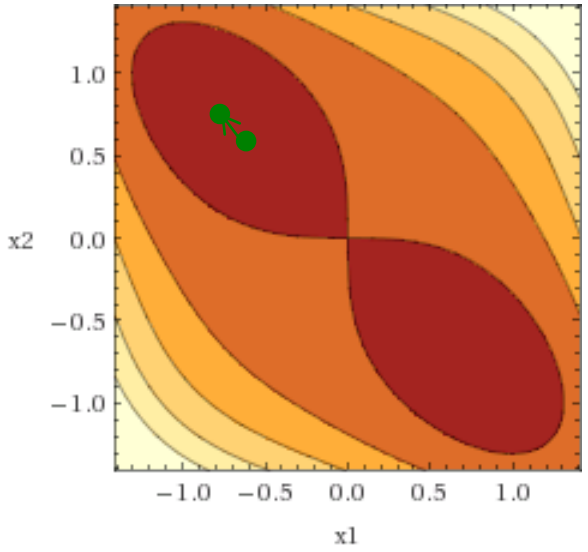


Example of minimization with more than one variable

$$\boldsymbol{x}_{n+1} = \boldsymbol{x}_n - \boldsymbol{h}^{-1} \cdot \boldsymbol{g}$$

$$\begin{pmatrix} x_{1,n+1} \\ x_{2,n+1} \end{pmatrix} = \begin{pmatrix} x_{1,n} \\ x_{2,n} \end{pmatrix} - \begin{pmatrix} 12x_1^2 & 3 \\ 3 & 12x_2^2 \end{pmatrix}^{-1} \cdot \begin{pmatrix} 4x_1^3 + 3x_2 \\ 4x_2^3 + 3x_1 \end{pmatrix}$$

n	$x_{1,n}$	$x_{2,n}$	Δx_1	Δx_2	$x_{1,n+1}$	$x_{2,n+1}$
1	-0.700	0.700	-0.253	0.253	-0.953	0.953
2	-0.953	0.953	0.076	-0.076	-0.877	0.877
3	-0.877	0.877	0.010	-0.010	-0.866	0.866
4	-0.866	0.866	0.000	0.000	-0.866	0.866

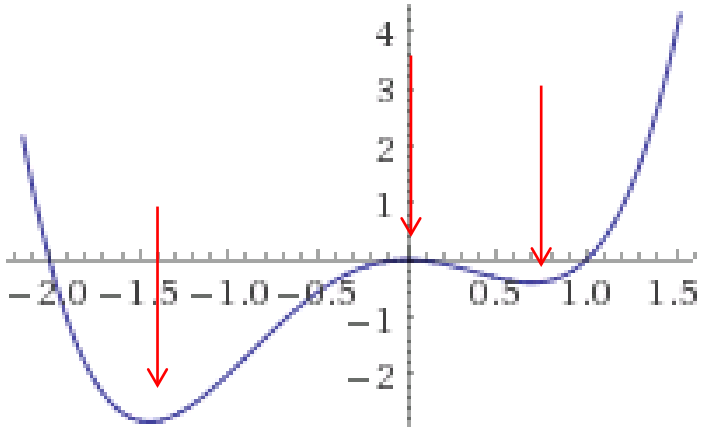


1.3. Stationary structures

In the previous examples we have found three solutions at which the gradient vanishes. They are stationary points (or structures) on the molecular potential energy surface.

Example

$$f(x) = x^4 + x^3 - 2x^2$$



(x from -2 to 1.5)

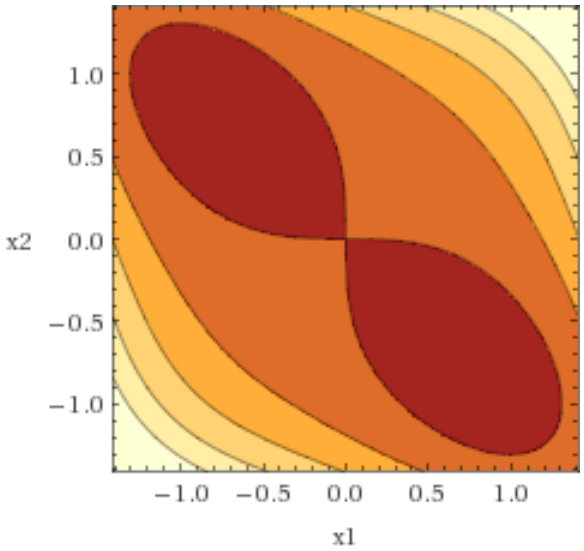
x	f(x)	g(x)	h(x)	Kind
0.000	0.0	0.0	-4.001	Maximum
0.693	-0.397	0.000	5.925	Minimum
-1.443	-2.833	0.000	12.338	Absolute minimum

The minima corresponds to **stable structures** (identified in the laboratory). Maxima in just one direction correspond to **transition structures** (that provide the minimum energy required to make the transition from one minimum to another)

1.3. Stationary structures

In cases of more than one dimension the procedure to identify the nature of the stationary points is similar, although somewhat more complex since the Hessian is now a matrix.

Example $f(x_1,x_2)=x_1^4+3x_1x_2+x_2^4$



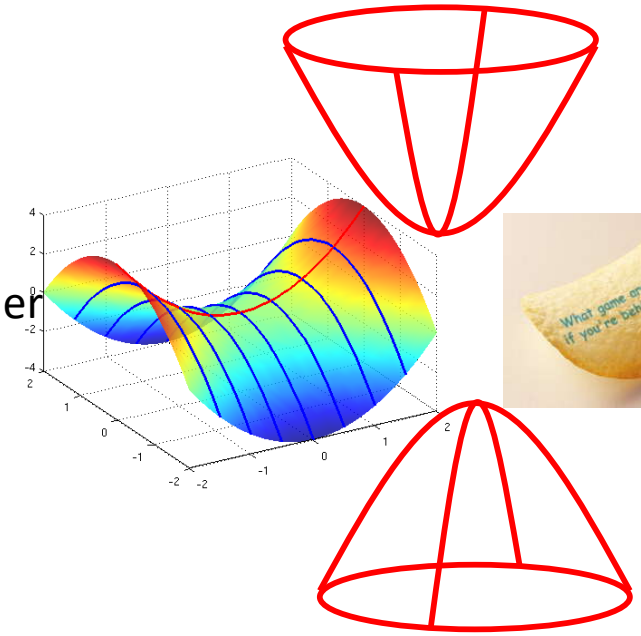
$$\boldsymbol{g}(x_1,x_2)=\begin{bmatrix}4x_1^3+3x_2\\4x_2^3+3x_1\end{bmatrix} \qquad \boldsymbol{h}(x_1,x_2)=\begin{pmatrix}12x_1^2 & 3\\3 & 12x_2^2\end{pmatrix}$$

(x_1,x_2)	$f(x_1,x_2)$	$\boldsymbol{g}(x_1,x_2)$	$\boldsymbol{h}(x_1,x_2)$
$(0,0)$	0	$\begin{pmatrix}0\\0\end{pmatrix}$	$\begin{pmatrix}0 & 3\\3 & 0\end{pmatrix}$
$(-0.866,0.866)$	-1.125	$\begin{pmatrix}0\\0\end{pmatrix}$	$\begin{pmatrix}9 & 3\\3 & 9\end{pmatrix}$
$(0.866,-0.866)$	-1.125	$\begin{pmatrix}0\\0\end{pmatrix}$	$\begin{pmatrix}9 & 3\\3 & 9\end{pmatrix}$

1.3. Stationary structures

Two-dimensional stationary points can be:

- Minima in both directions
- Minima in one direction and maxima in the other
- Maxima in both directions



In general, from a chemical point of view, on a surface of D dimensions we are interested in those structures that are **minima** in the D possible directions (stable structures) and those that are maximum in only 1 direction and minima in the remaining D-1 (**transition structures** between two minima).

To determine the nature of the stationary point we have to determine the curvature in all possible directions (the value of the Hessian). If it is always positive we will have a minimum, if it is negative only in one direction it will be a transition structure.

1.3. Stationary structures classification

In order to determine the curvature we have to transform the Hessian into a diagonal matrix. The curvature can then be determined by the sign of the values in this matrix.

$$\boldsymbol{h} = \boldsymbol{v}\cdot\boldsymbol{d}\cdot\boldsymbol{v}^{-1}; \quad \boldsymbol{d} = \begin{pmatrix} d_1 & 0 \\ 0 & d_2 \end{pmatrix}; \quad \boldsymbol{v} = \begin{pmatrix} v_{11} & v_{21} \\ v_{12} & v_{22} \end{pmatrix}$$

$v_{11}x_1 + v_{12}x_2$
$v_{21}x_1 + v_{22}x_2$

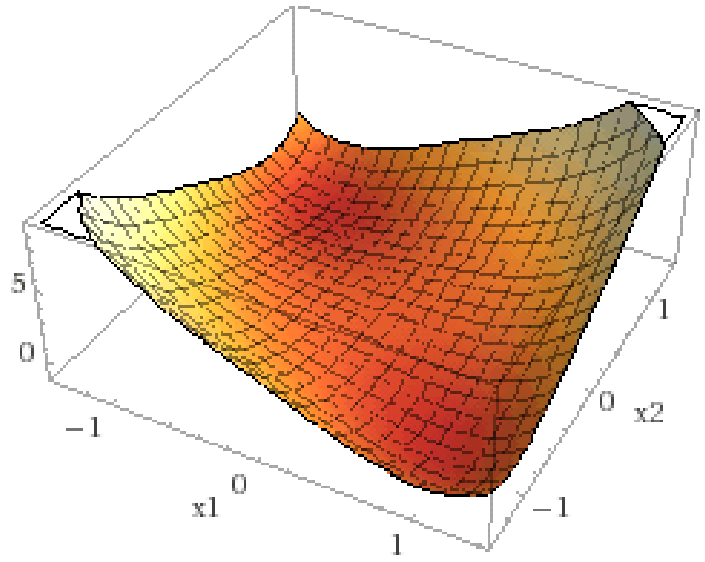
These are the so-called **eigenvectors**, the directions in which we can consider that the function varies independently, at least within a quadratic approximation. In a molecule these eigenvalues correspond to the **normal modes of vibration** .

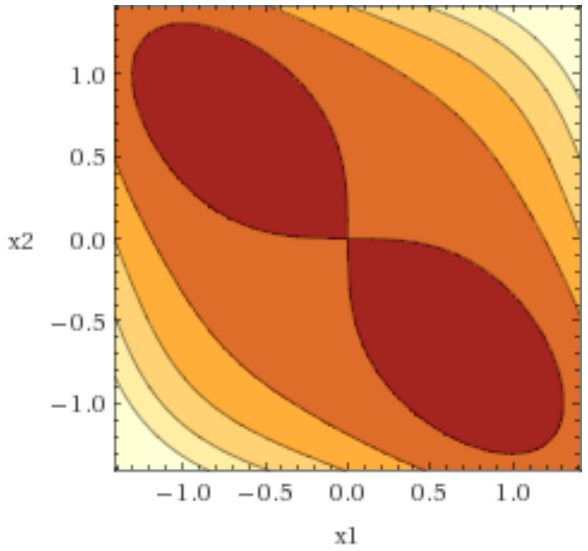
$d_1; d_2$

Eigenvalues inform us about the change of the function (the curvature) along the eigenvectors. In a molecule these values are the **force constants** associated to each normal mode

1.3. Stationary structures classification

(x_1,x_2)	$f(x_1,x_2)$	$\boldsymbol{g}(x_1,x_2)$	$\boldsymbol{h}(x_1,x_2)$	$\boldsymbol{d}$	
$(0,0)$	0	$\begin{pmatrix}0\\0\end{pmatrix}$	$\begin{pmatrix}0 & 3\\3 & 0\end{pmatrix}$	$\begin{pmatrix}-3 & 0\\0 & 3\end{pmatrix}$	⇒ Saddle point (maximum in one direction)
$(-0.866,0.866)$	-1.125	$\begin{pmatrix}0\\0\end{pmatrix}$	$\begin{pmatrix}9 & 3\\3 & 9\end{pmatrix}$	$\begin{pmatrix}6 & 0\\0 & 12\end{pmatrix}$	⇒ Minimum
$(0.866,-0.866)$	-1.125	$\begin{pmatrix}0\\0\end{pmatrix}$	$\begin{pmatrix}9 & 3\\3 & 9\end{pmatrix}$	$\begin{pmatrix}6 & 0\\0 & 12\end{pmatrix}$	⇒ Minimum





2. HF optimization. Basis set dependence

Exercise 1. Optimize the structures of HF, H_2O, NH_3 y CH_4 at Hartree Fock level using the following basis sets STO-3G, 3-21G, 6-31G* y 6-31G**. Compare the results with experimental distances, angles and dipolar moments.

```
#hf/sto-3g opt=z-matrix

Water molecule optimization sto-3g

0 1
o
h 1 r1
h 1 r1 2 a2

r1 0.96
a2 90.
```

#hf/sto-3g opt=z-matrix

Optimization NH3 sto-3g

0 1
N
H 1 r1
H 1 r1 2 a2
H 1 r1 2 a2 3 d3

r1 1.07
a2 107.
d3 100.

#hf/sto-3g opt=z-matrix

Optimization CH4 sto-3g

0 1
C
H 1 r1
H 1 r1 2 a2
H 1 r1 2 a2 3 d3
H 1 r1 2 a2 3 –d3

r1 1.30
a2 109.5
d3 120.

Exercise 1. Optimize the structures of HF, H_2O, NH_3 y CH_4 at Hartree Fock level using the following basis sets STO-3G, 3-21G, 6-31G* y 6-31G**. Compare the results with experimental distances, angles and dipolar moments.

Distances XH (Å)

	HF	H2O	NH3	CH4
STO-3G				
3-21G				
6-31G*				
6-31G**				
Exp.	0.917	0.957	1.012	1.085

Exercise 1. Optimize the structures of HF, H_2O, NH_3 y CH_4 at Hartree Fock level using the following basis sets STO-3G, 3-21G, 6-31G* y 6-31G**. Compare the results with experimental distances, angles and dipolar moments.

Angles HXH (degrees)

	H2O	NH3
STO-3G		
3-21G		
6-31G*		
6-31G**		
Exp.	104.5	106.7

Exercise 1. Optimize the structures of HF, H_2O, NH_3 y CH_4 at Hartree Fock level using the following basis sets STO-3G, 3-21G, 6-31G* y 6-31G**. Compare the results with experimental distances, angles and dipolar moments.

Dipole moments (debyes)

	HF	H2O	NH3
STO-3G			
3-21G			
6-31G*			
6-31G**			
Exp.	1.82	1.85	1.47

Exercise 1. Optimize the structures of HF, H_2O, NH_3 y CH_4 at Hartree Fock level using the following basis sets STO-3G, 3-21G, 6-31G* y 6-31G**. Compare the results with experimental distances, angles and dipolar moments.

Distances XH (Å)

	HF	H2O	NH3	CH4
STO-3G				
3-21G				
6-31G*				
6-31G**				
Exp.	0.917	0.957	1.012	1.085

Exercise 1. Optimize the structures of HF, H_2O, NH_3 y CH_4 at Hartree Fock level using the following basis sets STO-3G, 3-21G, 6-31G* y 6-31G**. Compare the results with experimental distances, angles and dipolar moments.

Angles HXH (degrees)

	H2O	NH3
STO-3G		
3-21G		
6-31G*		
6-31G**		
Exp.	104.5	106.7

Exercise 1. Optimize the structures of HF, H_2O, NH_3 y CH_4 at Hartree Fock level using the following basis sets STO-3G, 3-21G, 6-31G* y 6-31G**. Compare the results with experimental distances, angles and dipolar moments.

Dipole moments (debyes)

	HF	H2O	NH3
STO-3G			
3-21G			
6-31G*			
6-31G**			
Exp.	1.82	1.85	1.47

• In general, the inclusion of a second set of valence orbitals (split-valence bases) improves the description of the bond distances, allowing greater flexibility to place the electron density away from the core.

• The inclusion of d orbitals allows us to improve bond angles, providing more angular flexibility to describe the Electron density.

• HF exaggerates the dipole moments (overestimated Ionic contributions), but with an adequate basis set the correct values can be obtained.

3. Frequency calculations and thermochemistry

We can verify that the optimized structures correspond to energy minima by obtaining the Hessian: its diagonalization provides us with normal modes and force constants associated with each of them.

The frequencies associated to each normal mode may be easily obtained from the force constants:

$$\nu = \frac{1}{2\pi c}\sqrt{\frac{k_{f,i}}{m_i}}$$

Where m_i is the mass associated to each normal mode.

If all the force constants are positive the structure is a minimum. In this case, the frequencies will be positive.

If it is a saddle point one of the force constants will be negative, and one of the frequencies will be imaginary.

3. Frequency calculations and thermochemistry

Exercise 2. Using the optimized structure of a water molecule at the HF/6-31G* level, obtain its normal modes and frequencies

#hf/6-31G* freq

Water molecule frequencies

0 1
o
h 1 r1
h 1 r1 2 a2

r1 0.947
a2 105.47

Frequency calculations must be performed for the **optimized** geometry

#hf/6-31G* opt=z-matrix freq

Water molecule frequencies

0 1
o
h 1 r1
h 1 r1 2 a2

r1 0.947
a2 105.47

You can combine optimization and frequency calculation in a single input file. The program will run:

i) optimization
ii) frequencies for the optimized geometry.

Gaussian Output

Harmonic frequencies (cm**-1), IR intensities (KM/Mole), Raman scattering activities (A**4/AMU), depolarization ratios for plane and unpolarized incident light, reduced masses (AMU), force constants (mDyne/A), and normal coordinates:

	1	2	3
	A1	A1	B2
Frequencies --	1826.7838	4070.9366	4189.0013
Red. masses --	1.0823	1.0455	1.0828
Frc consts --	2.1280	10.2089	11.1951
IR Inten --	107.2284	18.2091	58.0127
Raman Activ --	5.7256	75.5150	39.1013
Depolar (P) --	0.5299	0.1829	0.7500
Depolar (U) --	0.6927	0.3093	0.8571

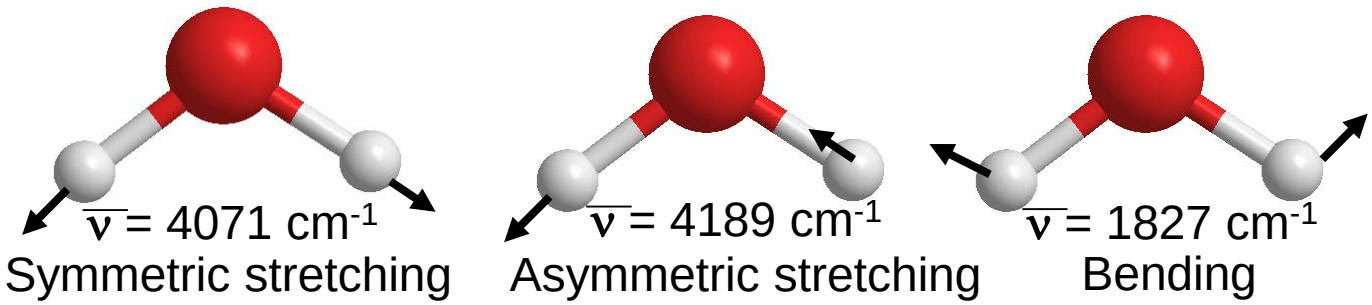
3x3 -6 = 3 vibrational modes

All are positive

Atom	AN	X	Y	Z	X	Y	Z	X	Y	Z
1	8	0.00	0.00	0.07	0.00	0.00	0.05	0.00	0.07	0.00
2	1	0.00	-0.43	-0.56	0.00	0.58	-0.40	0.00	-0.56	0.43
3	1	0.00	0.43	-0.56	0.00	-0.58	-0.40	0.00	-0.56	-0.43

Normal modes can be visualized with GaussView or Molden

GaussView: Results > Vibrations



Once the frequency values of the normal modes are known, along with the molecular mass and geometry, we can calculate the molecular partition function:

$$q(V,T) = q_{tras}(V,T)\,q_{rot}(T)\,q_{vib}(T)\,q_{ele}(T)$$

Translational $$q_{tras}(V,T) = \left(\frac{2\pi mkT}{h^2}\right)^{3/2}\frac{RT}{P^0}$$

Rotational (linear) $$q_{rot}(T) = \frac{T}{\sigma\theta_{rot}} \qquad \theta_i = \frac{h^2}{8\pi^2 kI_i}$$

Rotational (non-linear) $$q_{rot}(T) = \frac{\sqrt{\pi}}{\sigma}\frac{T^{3/2}}{\sqrt{\theta_a\theta_b\theta_c}}$$

Vibrational $$q_{vib}(T) = \prod_{i=1}^{3N-6}\frac{1}{1-e^{-\frac{\theta_{vib,i}}{T}}} \qquad \theta_{vib,i} = \frac{h\nu_i}{k}$$

Electronic $$q_{ele}(T) = \sum_{u}^{levels} g_{ele,u}e^{-\beta\varepsilon_{ele,u}}$$

Once the molecular partition function is known, one can then estimate the thermodynamic properties of the substance (assuming ideal gas behavior) from the relationships provided by statistical thermodynamics.

$$Q(N,V,T) = \frac{[q(V,T)]^N}{N!}$$ Partition function of an ideal gas

Internal energy $$U = U(0) + kT^2\left(\frac{\partial \ln Q}{\partial T}\right)_{N,V}$$

Enthalpy $$H = H(0) + kT^2\left(\frac{\partial \ln Q}{\partial T}\right)_{N,V} + kTV\left(\frac{\partial \ln Q}{\partial V}\right)_{N,T}$$

Free energy $$G = G(0) - kT\ln Q + kTV\left(\frac{\partial \ln Q}{\partial V}\right)_{N,T}$$

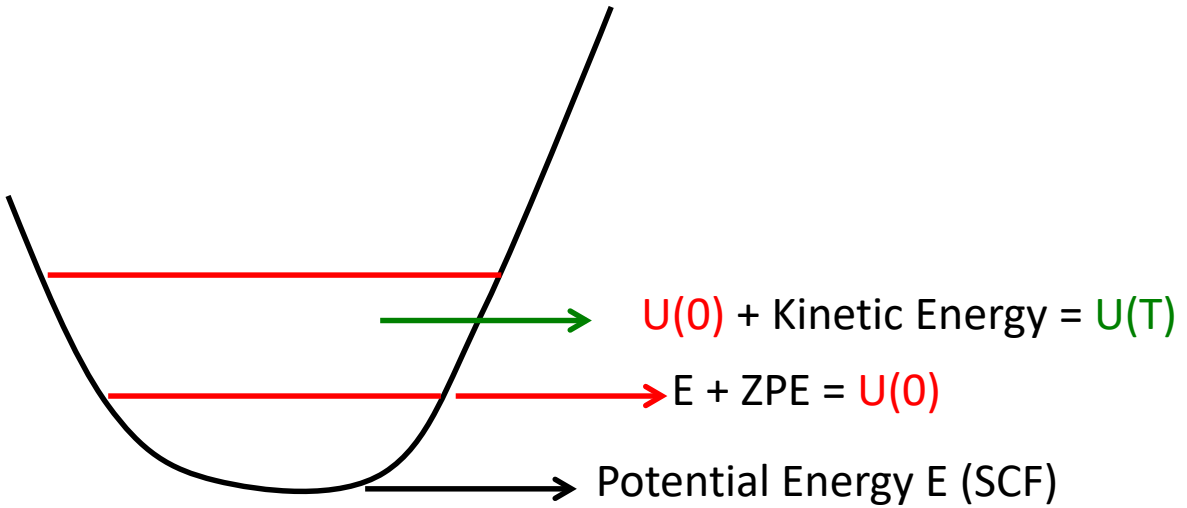
- Thermochemistry -

Temperature 298.150 Kelvin. Pressure 1.00000 Atm.
Atom 1 has atomic number 8 and mass 15.99491
Atom 2 has atomic number 1 and mass 1.00783
Atom 3 has atomic number 1 and mass 1.00783
Molecular mass: 18.01056 amu.
Principal axes and moments of inertia in atomic units:

	1	2	3
EIGENVALUES --	2.10296	4.09125	6.19421
X	0.00000	0.00000	1.00000
Y	1.00000	0.00000	0.00000
Z	0.00000	1.00000	0.00000

This molecule is an asymmetric top.
Rotational symmetry number 2.
Rotational temperatures (Kelvin) 41.18652 21.17053 13.98303
Rotational constants (GHZ): 858.18890 441.12268 291.35936
Zero-point vibrational energy 60332.0 (Joules/Mol)
14.41969 (Kcal/Mol)
Vibrational temperatures: 2628.33 5857.16 6027.03
(Kelvin)

Zero-point correction= 0.022979 (Hartree/Particle)
Thermal correction to Energy= 0.025813
Thermal correction to Enthalpy= 0.026757
Thermal correction to Gibbs Free Energy= 0.005382
Sum of electronic and zero-point Energies= -75.987767 U(0)
Sum of electronic and thermal Energies= -75.984933 U(T)
Sum of electronic and thermal Enthalpies= -75.983989 H(T) = U(T) + PV
Sum of electronic and thermal Free Energies= -76.005364 G(T) = H(T) + T·S(T)



Exercise 3. Using the molecular structure of NH_3 optimized at the HF/6-31G* level, obtain the normal modes and frequencies, and classify them (stretching, bending, etc.).

4. Density Functional Theory (DFT) methods

In principle, density functional theory (DFT) methods are not based on wave function, but rather on the electron probability density function, commonly called electron density or charge density, and designated by ρ(x,y,z). Apart from being experimentally observable, the electron density has a mathematical property particularly suitable for any method. It is a function of a unique position, i.e., only three variables (x, y, z), while a molecule of n-electrons' wave function is a function of 4n variables.

The Hohenberg-Kohn first theorem establishes that all the properties of a molecule in the ground electronic state are determined by the function of electron density of the fundamental state ρ0 (x, y, z).

$$E_0 = F[\rho_0] = E[\rho_0]$$

Hohenberg-Kohn second theorem is the analogous theorem to variational theorem in wave function theory. Any test function of the electron density will give an energy above (or equal if it is exactly the true function of Electron density) to the real energy of the system

$$E_v[\rho_t] \geq E_0[\rho_0]$$

$E_v[\rho_t] \geq E_0[\rho_0]$

4. Density Functional Theory (DFT) methods

The problem is that we do not know the function that returns energy from the Electron density. Therefore the DFT methods use approximate functionals.

The Kohn-Sahm strategy allows you to mitigate this problem through two basic ideas:

- The energy is expressed as a sum of terms, of which only a relatively small one is of an unknown shape.

- The electron density is calculated from a set of orbitals which are iteratively improved

$$\rho = \sum_{i=1}^{n} n_i / \Psi_i /^2$$

4. Density Functional Theory (DFT) methods

Kohn and Sham introduced the idea of a hypothetical reference system of independent electrons giving exactly the same distribution of electron density as the real system. Electronic energy is given by the sum of the kinetic and potential terms (electron-nucleo and electron-electron):

$$E_0 = < T[\rho_0] > + < V_{Ne}[\rho_0] > + < V_{ee}[\rho_0] >$$

The second term can be expressed exactly from the electron density:

$$< V_{Ne}[\rho_0] > = \sum_{nucleo\ A} Z_A \int \frac{\rho_0(\boldsymbol{r_1})}{\boldsymbol{r}_{1_A}} d\boldsymbol{r}_1$$

The kinetic term is expressed from the system of reference as:

$$< T[\rho_0] > \ = < T[\rho_0] >_{ref} + \Delta T[\rho_0] = \langle \Psi_r \left| \sum_{i=1}^{2n} -\frac{1}{2}\nabla_i^2 \right| \Psi_r \rangle + \Delta T[\rho_0]$$

As well as the electron-electron term:

$$< V_{ee}[\rho_o] > = < V_{ee}[\rho_o] >_{ref} + \Delta V_{ee}[\rho_0] = \iint \frac{\rho_0(\boldsymbol{r}_1)\rho_0(\boldsymbol{r}_2)}{r_{12}} d\boldsymbol{r}_1 d\boldsymbol{r}_2 + \Delta V_{ee}[\rho_0]$$

3. DFT methods

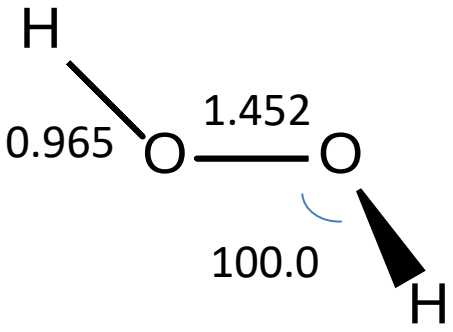
The energy is given by:

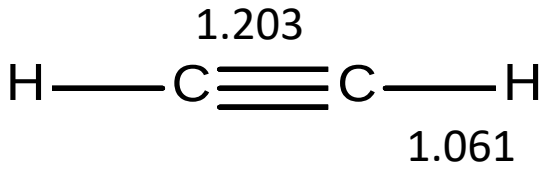
$$E_0 = \sum_{\text{nucleus}\ A} Z_A \int \frac{\rho_0(\boldsymbol{r_1})}{\boldsymbol{r}_{1_A}} d\boldsymbol{r}_1 + < T[\rho_0] >_{ref} + \frac{1}{2} \iint \frac{\rho_0(\boldsymbol{r}_1)\rho_0(\boldsymbol{r}_2)}{r_{12}} d\boldsymbol{r}_1 d\boldsymbol{r}_2 + E_{XC}[\rho_0]$$

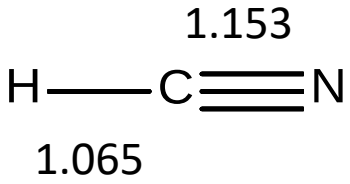
The fourth term corresponds to the correlation-exchange energy E_{xc} [ρ_o] and is the only term for which we have to devise a new method of calculation. It corresponds to the differences between the reference system and the real potential and kinetic energy.

- If it depends only on the electron density: local methods
- If it depends only on the electron density and its gradient: non-local methods
- If it includes part exchange HF energy: hybrid methods

Exercise 4. Optimize of HOOH, HCCH and HCN at the Hartree Fock and B3LYP levels using the 6 - 31 G * basis set. Compare the results of both methods with experimental values (in angstroms and degrees).





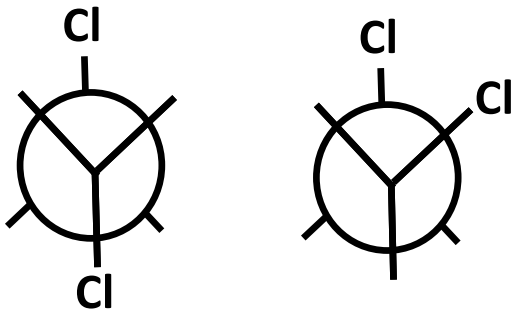


5. Conformational analysis

There are molecules for which there are more than one minimum energy structure. There are different possible conformers for a given molecule.

Sometimes one of the conformations is much more stable than the others and it is sufficient to describe and characterize this confomer to study the molecular properties. However, sometimes there are more than one low-energy conformations and it is necessary to characterize all of them. Also, on occasions, it is not easy to predict what will be the most stable conformation.

For example, 1,2-dichloroethane (DCE) may exists in two different conformations:



In this case, we can study the different conformations analyzing how the molecular energy changes with the Cl-C -C-Cl dihedral angle. It's a monodimensional potential energy surface.

Exercise 5. Obtain the variation of potential energy of a molecule of 1,2-dichloroethane depending on rotation internal around the C-C bond at the HF/6 -31* level

```
#hf/6-31G* opt=z-matrix

SEP DCE

0 1
c
c 1 r1
cl 1 r2 2 a2
cl 2 r3 1 a3 3 d3
h 1 r4 2 a4 3 d4
h 1 r5 2 a5 3 d5
h 2 r6 1 a6 4 d6
h 2 r7 1 a7 4 d7

r1 1.5
r2 1.8
r3 1.76
r4 1.1
r5 1.1
r6 1.1
r7 1.1
a2 109.
a3 110.
a4 109.
a5 110.
a6 109.
a7 110.
d3 180. S 10 -20.0
d4 120.
d5 240.
d6 120.
d7 240.
```

Carry out a relaxed scan of 10 steps changing -20 degrees each time the dihedral angle d3 (Cl-C-C-Cl)

In this case, we can study the different conformations analyzing how the molecular energy changes with the Cl-C -C-Cl dihedral angle. It's a monodimensional potential energy surface.

Exercise 5. Obtain the variation of potential energy of a molecule of 1,2-dichloroethane depending on rotation internal around the C-C bond at the HF/6 -31* level

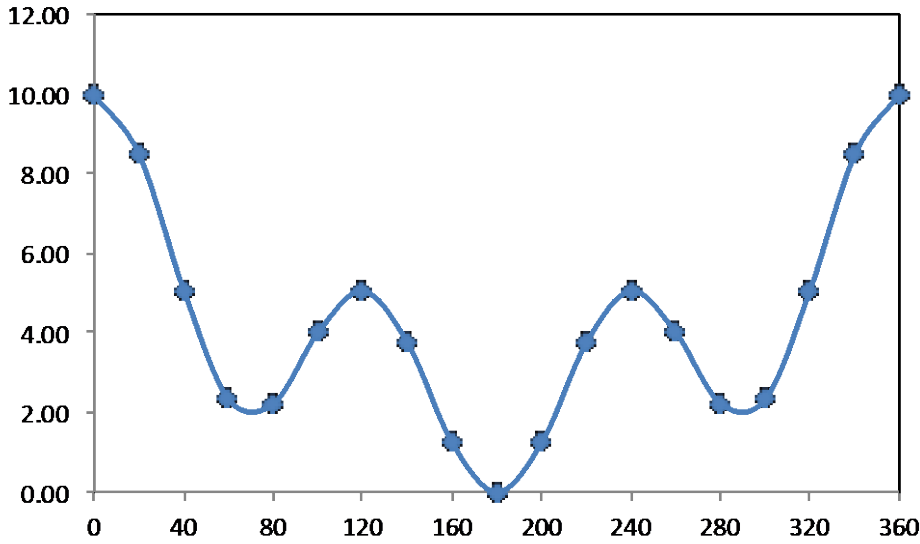
```
#hf/6-31G* opt=modredun

SEP DCE

0 1
c
c 1 r1
cl 1 r2 2 a2
cl 2 r3 1 a3 3 d3
h 1 r4 2 a4 3 d4
h 1 r5 2 a5 3 d5
h 2 r6 1 a6 4 d6
h 2 r7 1 a7 4 d7

r1 1.5
r2 1.8
r3 1.76
r4 1.1
r5 1.1
r6 1.1
r7 1.1
a2 109.
a3 110.
a4 109.
a5 110.
a6 109.
a7 110.
d3 180.
d4 120.
d5 240.
d6 120.
d7 240.

* 1 2 * R
3 1 2 4 S 10 -20.0
```

Summary of Optimized Potential Surface Scan

	1	2	3	4	5
EIGENVALUES --	-997.03094	-997.02897	-997.02497	-997.02289	-997.02456
R1	1.51604	1.52034	1.52925	1.53283	1.52552
R2	1.79160	1.79205	1.79260	1.79247	1.79007
R3	1.07763	1.07781	1.07776	1.07814	1.07917
R4	1.07764	1.07700	1.07625	1.07599	1.07656
R5	1.79165	1.79197	1.79252	1.79189	1.79048
R6	1.07761	1.07780	1.07776	1.07821	1.07918
R7	1.07762	1.07700	1.07623	1.07601	1.07648
A1	109.40947	109.92931	111.07052	112.07777	112.51490
A2	111.42683	112.01038	111.89299	110.88610	109.68234
A3	111.42285	110.91308	111.05862	111.64019	111.91642
A4	107.39194	107.85218	107.35259	106.30717	105.75027
A5	107.39357	106.42247	105.98449	106.64509	107.62339
A6	109.63083	109.51331	109.24092	109.03991	109.11804
A7	109.40506	109.93635	111.08193	112.05845	112.48170
A8	111.43079	111.99737	111.86282	110.89875	109.72604
A9	111.42661	110.90016	111.02954	111.64091	111.92503
A10	107.39105	107.86518	107.39239	106.31084	105.72796
A11	107.39247	106.44270	106.02726	106.64611	107.60721
A12	109.62912	109.50138	109.21162	109.04118	109.13606
D1	180.00000	160.00000	140.00000	120.00000	100.00000

	6	7	8	9	10
EIGENVALUES --	-997.02741	-997.02724	-997.02288	-997.01745	-997.01503
R1	1.51633	1.51589	1.52616	1.54015	1.54728
R2	1.78762	1.78442	1.78167	1.78101	1.78064
R3	1.08004	1.08012	1.07961	1.07866	1.07808
R4	1.07756	1.07848	1.07884	1.07832	1.07805
R5	1.78728	1.78384	1.78164	1.78087	1.78076
R6	1.08010	1.08018	1.07961	1.07867	1.07807
R7	1.07757	1.07859	1.07883	1.07836	1.07807
A1	112.60106	113.51663	115.40407	116.90909	117.44276
A2	109.07694	108.86983	108.66909	108.91766	109.33817
A3	111.68154	111.03674	110.35874	109.80699	109.36202
A4	106.25521	107.20684	107.65190	107.13041	105.97566
A5	107.73420	106.82033	105.70745	105.27758	105.98285
A6	109.31100	109.25253	108.85168	108.48692	108.37776
A7	112.60248	113.54988	115.40335	116.92997	117.45377
A8	109.08634	108.83464	108.67377	108.90383	109.33804
A9	111.66871	111.01700	110.35308	109.80243	109.34938
A10	106.26160	107.21263	107.66109	107.12023	105.99940
A11	107.73949	106.86376	105.70212	105.29563	105.96630
A12	109.30180	109.22472	108.84944	108.47512	108.37161
D1	80.00000	60.00000	40.00000	20.00000	0.00000

Dihedral Angle (degrees)	E(Hartrees)	Erel(Hartrees)	Erel (kcal/mol)
180	-997.03094	0	0.00
160	-997.02897	0.00197	1.24
140	-997.02497	0.00597	3.75
120	-997.02289	0.00805	5.05
100	-997.02456	0.00638	4.00
80	-997.02741	0.00353	2.22
60	-997.02724	0.0037	2.32
40	-997.02288	0.00806	5.06
20	-997.01745	0.01349	8.47
0	-997.01503	0.01591	9.98



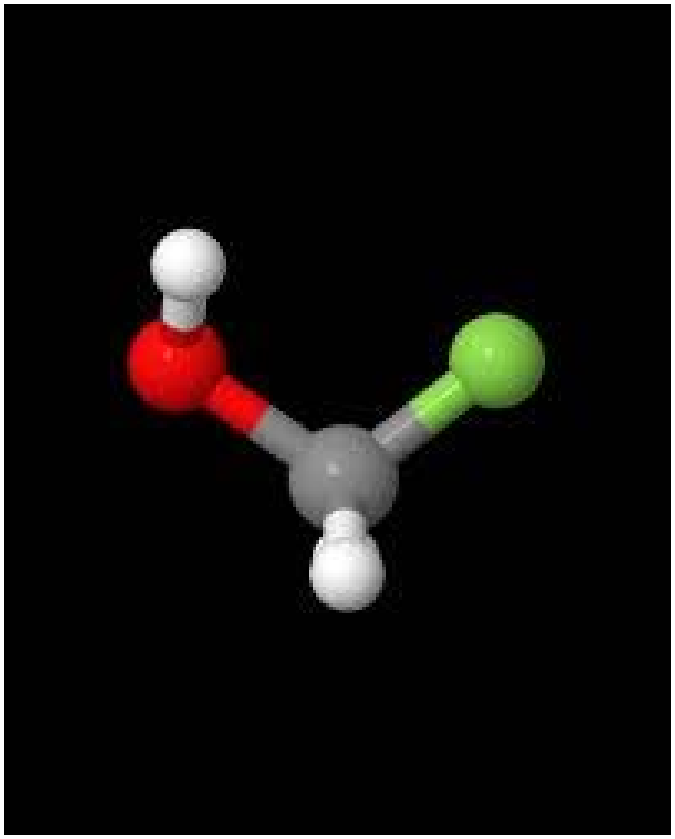
Exercise 6. Optimize the minimum energy structures corresponding to the conformations trans and gauche of 1,2-dichloroethane starting from structures close to the minima in the PES scan obtained at the HF/6-31G* level. Obtain the frequencies of the structures that are true minima and find the differences in the potential energy, enthalpy and free energy. From this last value, calculate the equilibrium constant for trans➔gauche at 298K.

	trans	gauche
Cl-C-C-Cl (degrees)		
E (Hartrees)		
H (Hartrees)		
G (Hartrees)		

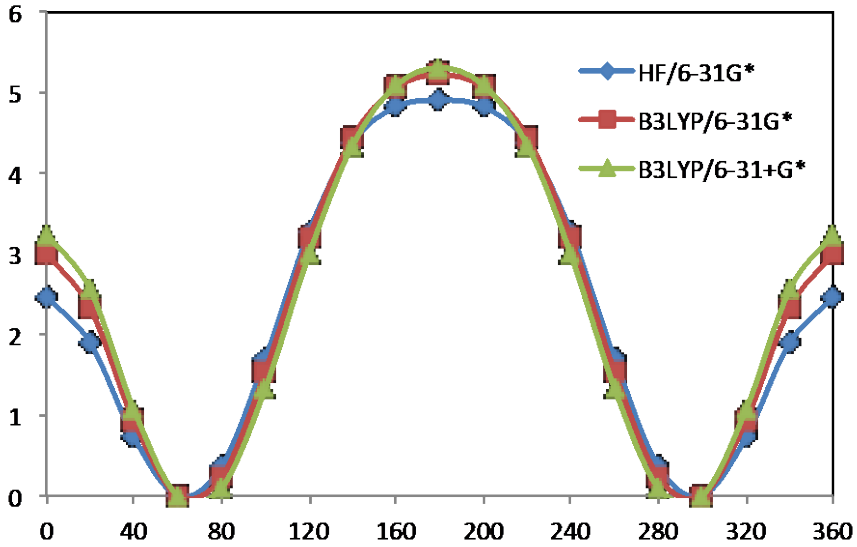
	(gauche – trans)
ΔE (kcal/mol)	
ΔH (kcal/mol)	
ΔG (kcal/mol)	

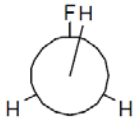
$$K = \exp(-\Delta G/RT)$$

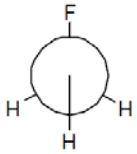
Exercise 7. Obtain the potential energy curve corresponding to the rotation of the alcohol hydrogen around the C-O bond in fluoromethanol. Explain the resulting curve depending on the different interactions that appear in the molecule when the rotation occurs.

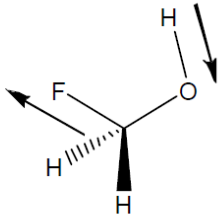


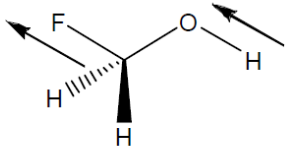
```
#b3lyp/6-31G* opt=z-matrix

Scan for fluoromethanol internal rotation

0 1
c
o 1 r1
f 1 r2 2 a2
h 2 r3 1 a3 3 z3
h 1 r4 2 a4 3 z4
h 1 r5 2 a5 3 z5

r1 1.4
r2 1.5
r3 0.95
r4 1.09
r5 1.1
a2 110.
a3 105.
a4 109.
a5 109.
z3 0. S 9 20.0
z4 120.
z5 -120.
```

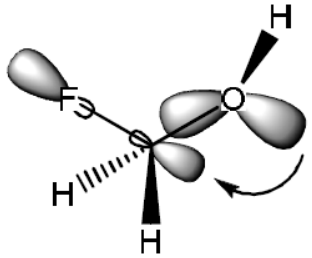

HF/6-31G*
B3LYP/6-31G*
B3LYP/6-31+G*
0
1
2
3
4
5
6
0
40
80
120
160
200
240
280
320
360

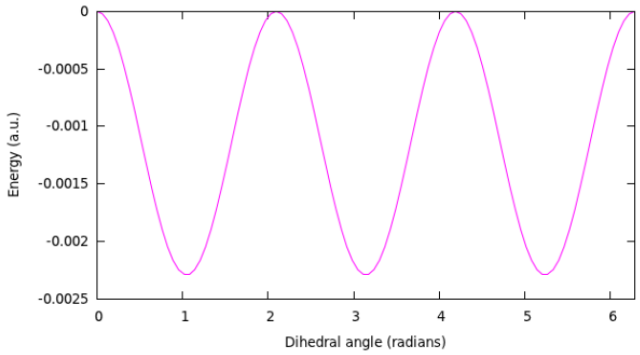

F
H
H
H

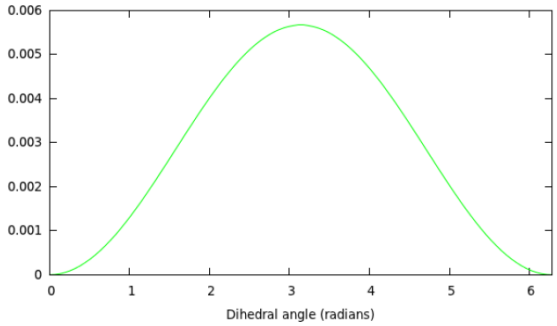

F
H
H
H

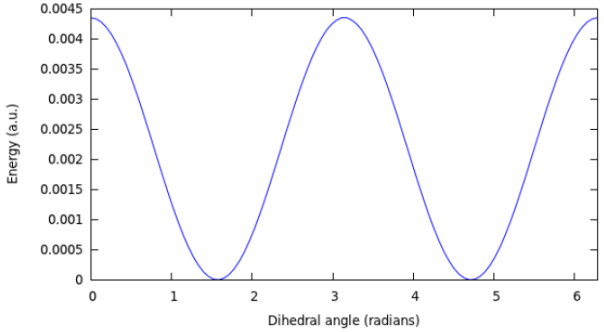

H
F
O
H
H


F
O
H
H
H


H
F
O
H
H

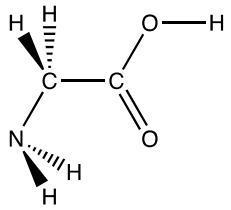

Energy (a.u.)
0
-0.0005
-0.001
-0.0015
-0.002
-0.0025
0
1
2
3
4
5
6
Dihedral angle (radians)


0.006
0.005
0.004
0.003
0.002
0.001
0
0
1
2
3
4
5
6
Dihedral angle (radians)

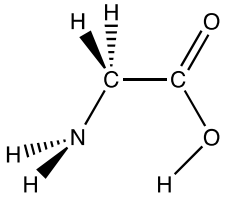

Energy (a.u.)
0.0045
0.004
0.0035
0.003
0.0025
0.002
0.0015
0.001
0.0005
0
0
1
2
3
4
5
6
Dihedral angle (radians)

Exercise 8. Some molecules can exist in many different conformations. One of these molecules is glycine. In the gas phase glycine exists as a neutral molecule and different conformations are possible. We are going to optimize the four lower-energy conformers of this molecule (denoted as A, B, C and D in the following figure).

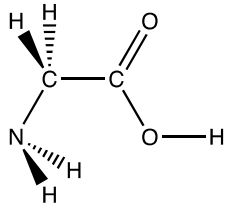
i) Optimize the conformations at the HF/6-31G* level.
ii) Confirm that all the obtained structures are real minima.
iii) Take note of the dipole moment of each conformer. We will use it in the next sessions.
iv) Calculate the free energy of the four conformers at 298 K and then their populations assuming that there are no other conformers available.
v) Repeat the calculations at the B3LYP/6-31G* level.



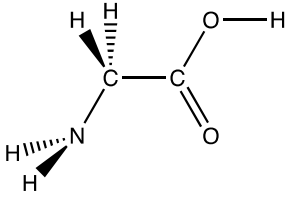
A



B



C



D

HF/6-31G*

	A	B	C	D
E (hartrees)				
ΔE (kcal·mol^{-1})				
μ (Debye)				
G (hartrees)				
ΔG (kcal·mol^{-1})				
P_i(%)				

$$p_i(\%) = \frac{e^{-\frac{\Delta G_i}{RT}}}{\sum_{j=A,B,C,D} e^{-\frac{\Delta G_j}{RT}}}$$

B3LYP/6-31G*

	A	B	C	D
E (hartrees)				
ΔE (kcal·mol^{-1})				
μ (Debye)				
G (hartrees)				
ΔG (kcal·mol^{-1})				
P_i(%)				

Studies
on
Chemical Reactivity

Computational Chemistry
Elective Course
Chemistry Degree
4th year

Content

1. Calculation of the potential energy surface
2. Transition Structures
3. Reaction path
4. Rate constants
5. Other Examples

1. Obtaining the PES for a S_N2 reaction:

$$F^- + CH_3Cl \rightarrow FCH_3 + Cl^-$$

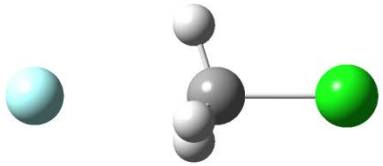
Exercise 1. Optimize the structures of reactants and products at the HF/6-31 + G * level and calculate their frequencies.

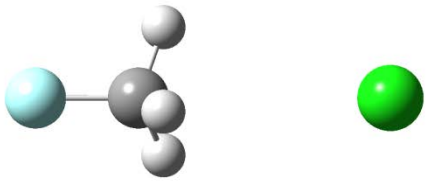
	fluoride	chloride	CH_3Cl	CH_3F
Energy (hartrees)				
E + ZPE				
Enthalpy				
Free Energy				
d(C-Cl) (Å)				
d(C-F) (Å)				

1. Obtaining the PES for a S_N2 reaction:

$$F^- + CH_3Cl \rightarrow FCH_3 + Cl^-$$

Exercise 2. Reactants and products may form an ion-dipole complex as shown in the figure. These complexes are the real reactants and products of the reaction since they associate in gas phase before reaction. Obtain their geometries and properties at the HF/6-31+G* level.



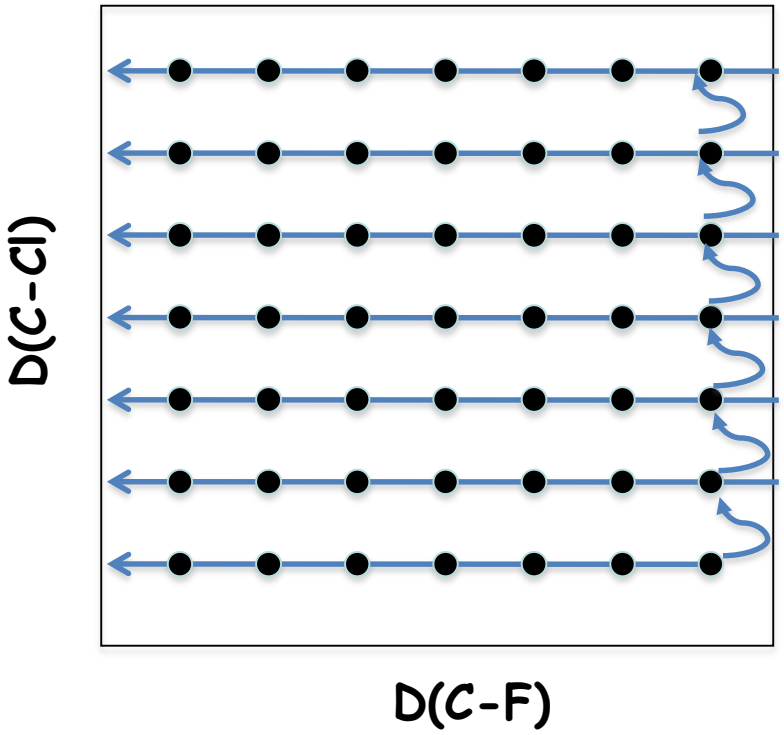


	F- ····CH_3Cl	FCH_3····Cl-
Energy (hartrees)		
E + ZPE		
Enthalpy		
Free Energy		
d(C-Cl) (Å)		
d(C-F) (Å)		
d(C-H) (Å)		

1. Obtaining the PES for a S_N2 reaction

Exercise 3. Obtaining the PES. Using the C-Cl y C-F distances as distinguished reaction coordinates obtain the PES at the HF/6-31+G* using the following input (or similar)

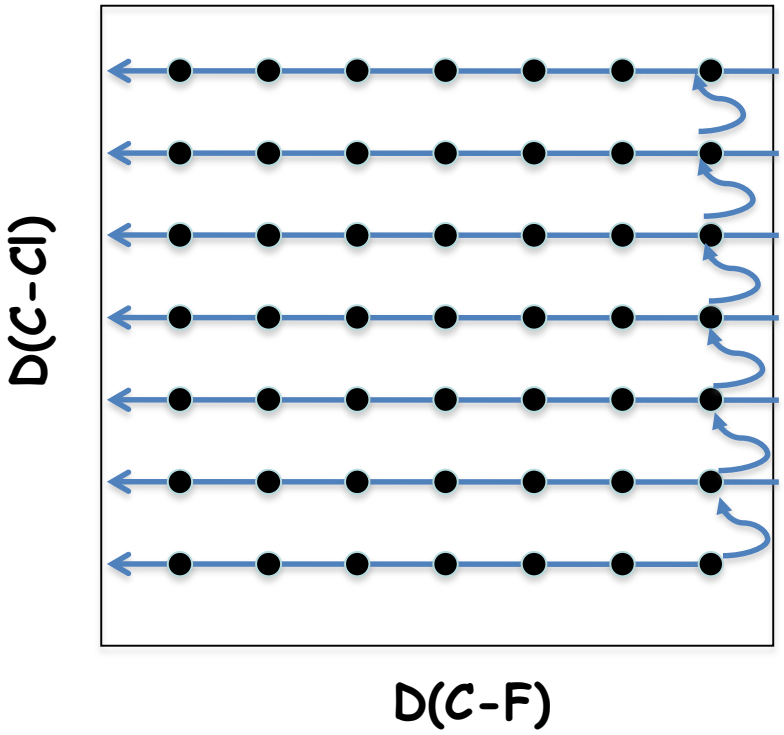
```
#HF/6-31+G* opt=modredun

Sep

-1 1
C
F 1 r1
H 1 r2 2 a2
H 1 r3 2 a3 3 z3
H 1 r4 2 a4 3 z4
X 1 1.0 2 90. 3 0.
Cl 1 r5 6 a5 2 z5

r1 1.3
r2 1.09
r3 1.08
r4 1.1
r5 3.6
a2 110.
a3 110.
a4 110.
a5 90.
z3 120.
z4 240.
z5 180.

1 2 S 8 0.2
1 6 S 10 -0.2
```



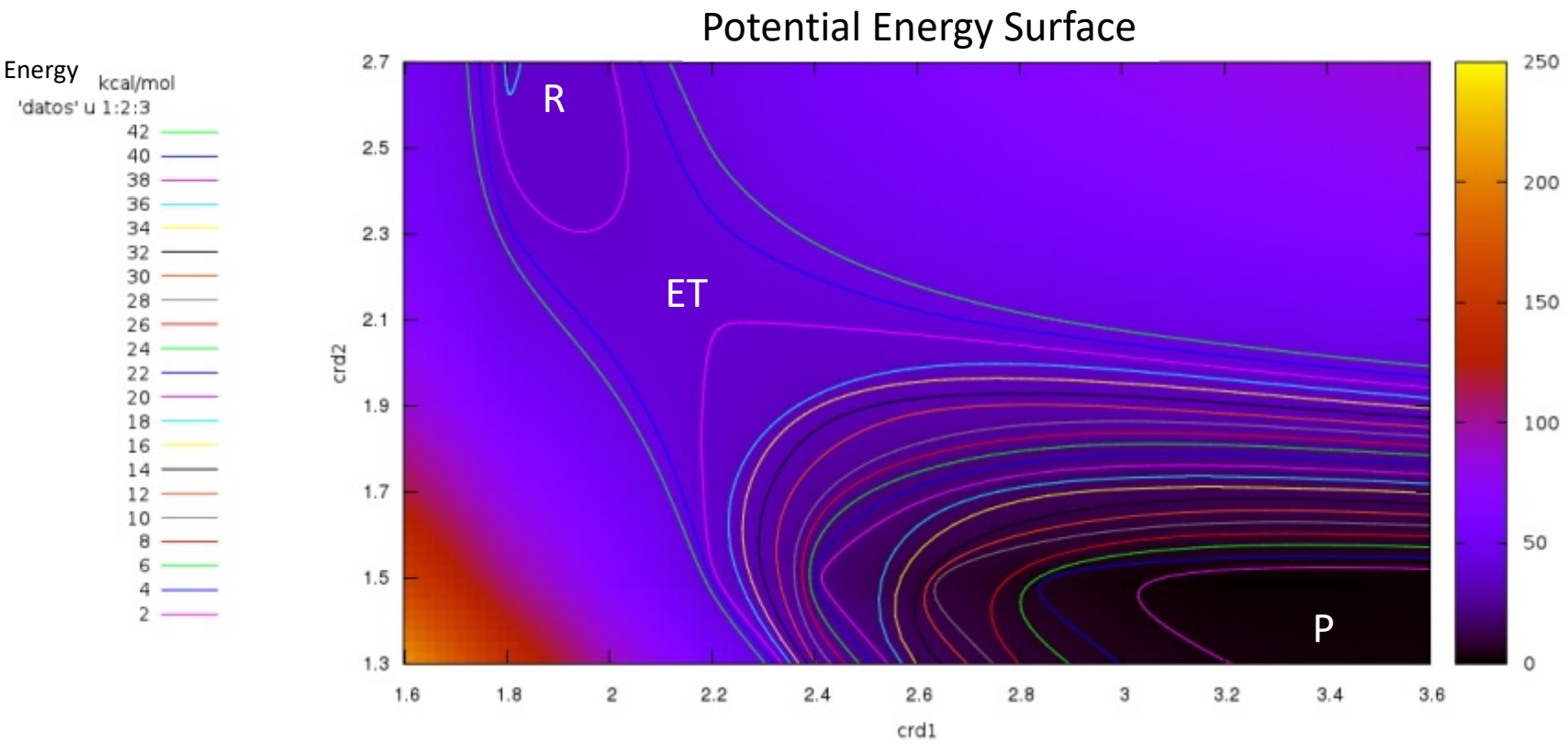
1. Obtaining the PES for a S_N2 reaction

Exercise 3. Obtaining the PES. Using the C-Cl y C-F distances as distinguished reaction coordinates obtain the PES at the HF/6-31+G* using the following input (or similar)

```
#HF/6-31+G* opt=z-matrix

Sep

-1 1
C
F 1 r1
H 1 r2 2 a2
H 1 r3 2 a3 3 z3
H 1 r4 2 a4 3 z4
X 1 1.0 2 90. 3 0.
Cl 1 r5 6 a5 2 z5

r1 1.3 S 8 0.2
r2 1.09
r3 1.08
r4 1.1
r5 3.6 S 10 -0.2
a2 110.
a3 110.
a4 110.
a5 90.
z3 120.
z4 240.
z5 180.
```



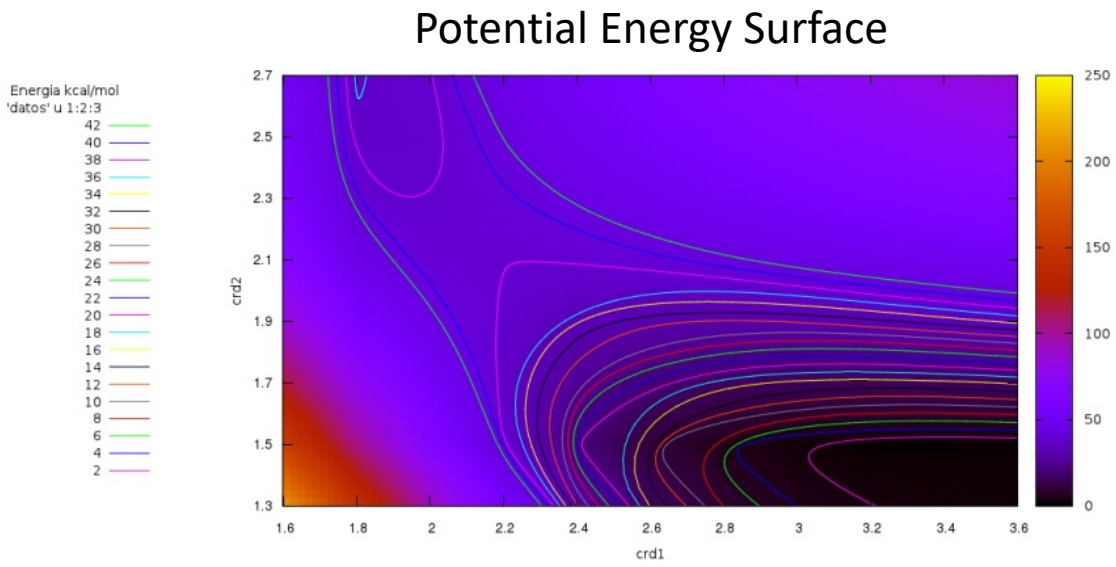
1. Obtaining the PES for a S_N2 reaction



2. Localizing transition structures

Exercise 4. Locate the TS and calculate the transition vector

Using the coordinates of a structure of the PES near where you assume that you will find the TS, perform a search using algorithms that use the Hessian matrix, hence the knowledge of the curvature of the PES around the structure.



#HF/6-31+G* opt=(z-matrix, calcfc,ts,noeigentest) freq

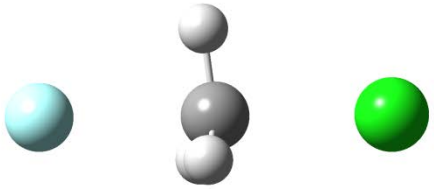
Ts

-1 1
C
F 1 r1
H 1 r2 2 a2
H 1 r3 2 a3 3 z3
H 1 r4 2 a4 3 z4
X 1 1.0 2 90. 3 0.
Cl 1 r5 6 a5 2 z5

r1 2.0
r2 1.09
r3 1.08
r4 1.1
r5 2.15
a2 110.
a3 110.
a4 110.
a5 90.
z3 120.
z4 240.
z5 180.

2. Localizing transition structures

Visualize the structure and the imaginary frequency using Gaussview or Molden

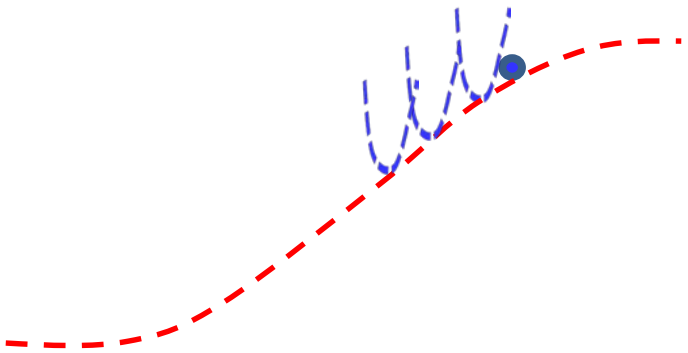


	$[F \cdot\cdot CH_3 \cdot\cdot Cl]^-$
Energy (hartrees)	
E + ZPE	
Enthalpy	
Free energy	
d(C-Cl) (Å)	
d(C-F) (Å)	
d(C-H) (Å)	
Imaginary freq (cm^{-1})	

3. Minimum energy path

Exercise 5. Obtain the minimum energy path

From the TS, we can follow the transition vector to reactants and products in search of the path of minimum potential energy. We seek the path that overrides the forces in any direction perpendicular to it, i.e. each structure on the way is a minimum for any orthogonal movement to the path.



#HF/6-31+G* irc=(maxpoints=5,calcfc,report=read)

Camino minima energia-

-1 1
C
F,1,r1
H,1,r2,2,a2
H,1,r3,2,a3,3,z3
H,1,r4,2,a4,3,z4
X,1,1.0,2,90.0,3,0.0
Cl,1,r5,6,a5,2,z5

r1=2.125
r2=1.062
r3=1.062
r4=1.062
r5=2.13370286
a2=82.74390542
a3=82.7316942
a4=82.7586379
a5=89.99905823
z3=119.99983145
z4=-120.00027513
z5=-179.98637128

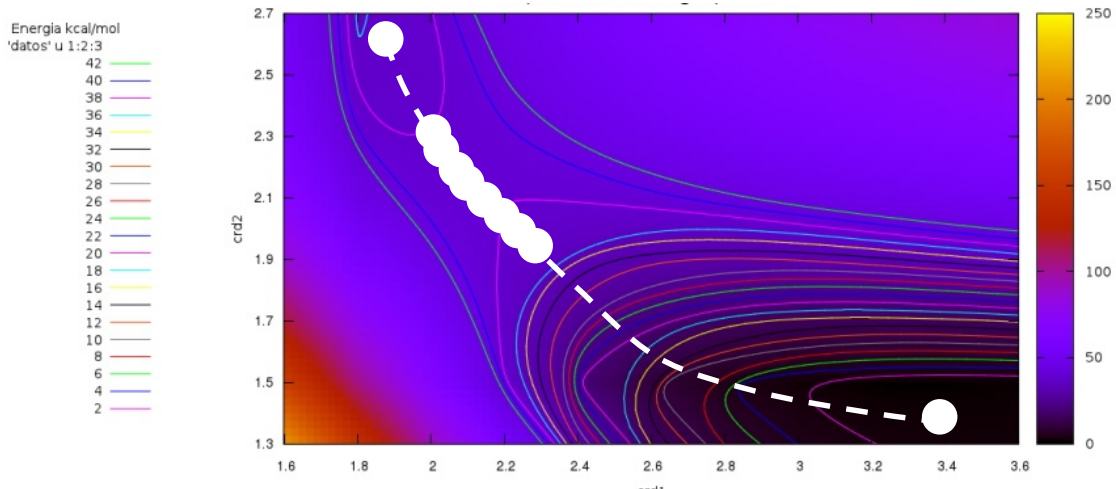
Report the bond distance between these atoms (C-F and C-Cl) at the end of the calculation →

1 2
1 6

Summary of reaction path following

	Energy	RxCoord	R1	R2
1	-0.01568	-1.77733	1.80279	2.41660
2	-0.00982	-1.42186	1.86834	2.36132
3	-0.00531	-1.06640	1.93338	2.30537
4	-0.00224	-0.71095	1.99784	2.24876
5	-0.00052	-0.35552	2.06167	2.19158
6	0.00000	0.00000	2.12500	2.13400
7	-0.00041	0.35545	2.18788	2.07624
8	-0.00142	0.71083	2.25069	2.01872
9	-0.00264	1.06621	2.31435	1.96223
10	-0.00368	1.42059	2.38022	1.91002
11	-0.00428	1.76782	2.44631	1.87606

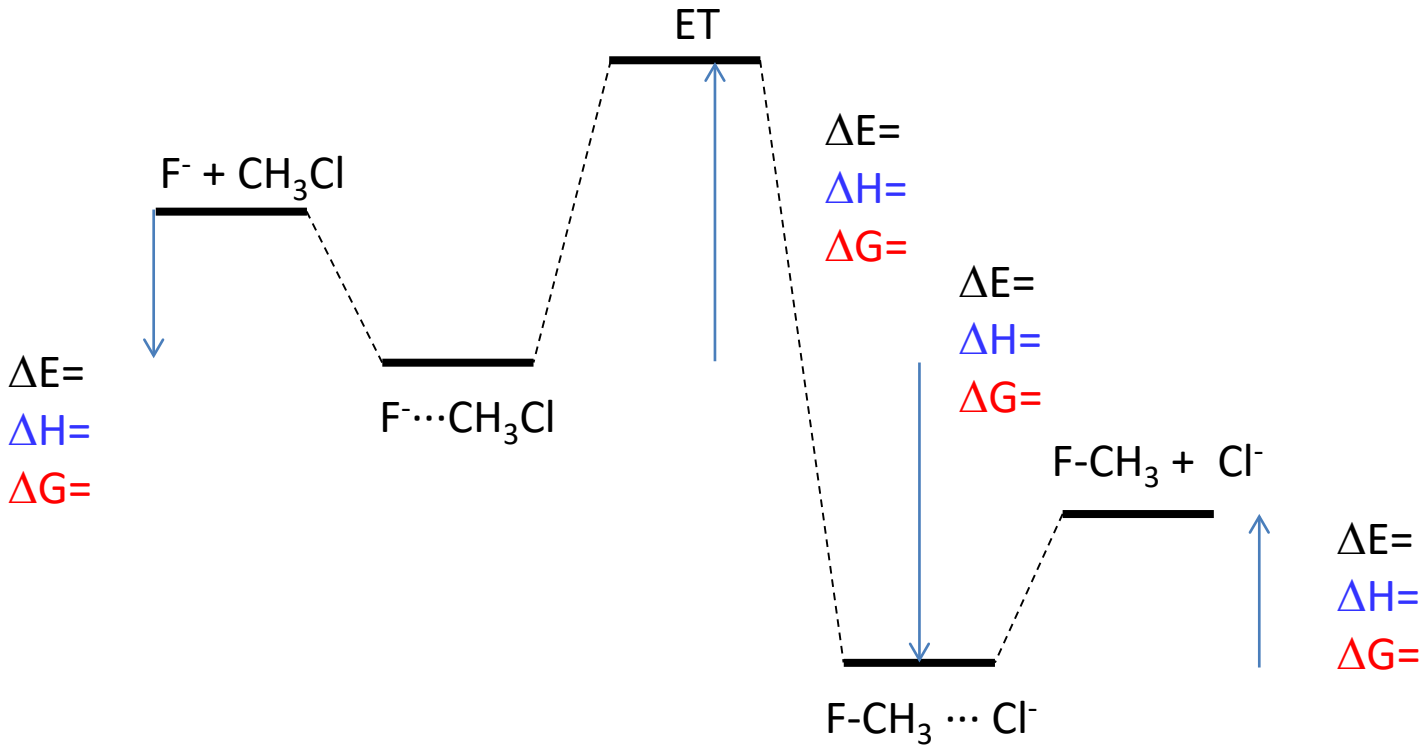
Potential Energy Surface



4. Calculating the rate constant

Exercise 6. Obtain the rate constant

Complete the following energy diagram for the reaction (in kcal·mol^{-1}) and obtain the rate constant at 298 K from the activation energy (the free energy difference between the TS and the reactants complex).



$$k_r = \frac{kT}{h}\left(\frac{RT}{P^0}\right)^{n-1} exp\left(-\frac{\Delta G_P^{0\ddagger}}{RT}\right)$$

5. Other Examples

Sometimes, a complete exploration of the PES is not required for the optimization of the TS. One can also try direct searches (assuming certain intermediate geometry between reactants and products) or looking for a minimum energy structure close to the real TS.

Exercise 7. Localize and characterize the TS for the following symmetric reaction at the HF/6-31+G* level

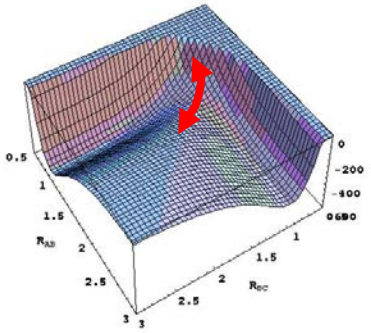
$CH_3F + F^- \rightarrow F^- + CH_3F$

First, you can optimize the minimum energy structure where the two C-F distances are equal. Then, you can use this structure as starting point to obtain the TS. Calculate the rate constant.

#HF/6-31+G* opt=z-matrix

TS-1

-1 1
C
F 1 r1
H 1 r2 2 a2
H 1 r3 2 a3 3 z3
H 1 r4 2 a4 3 z4
X 1 1.0 2 90. 3 0.
F 1 r1 6 a5 2 z5



Take the optimum value for r1 and use it as initial value for r5 and r6

#HF/6-31+G* opt=(calcfc,noeigentest,z-matrix,ts) freq

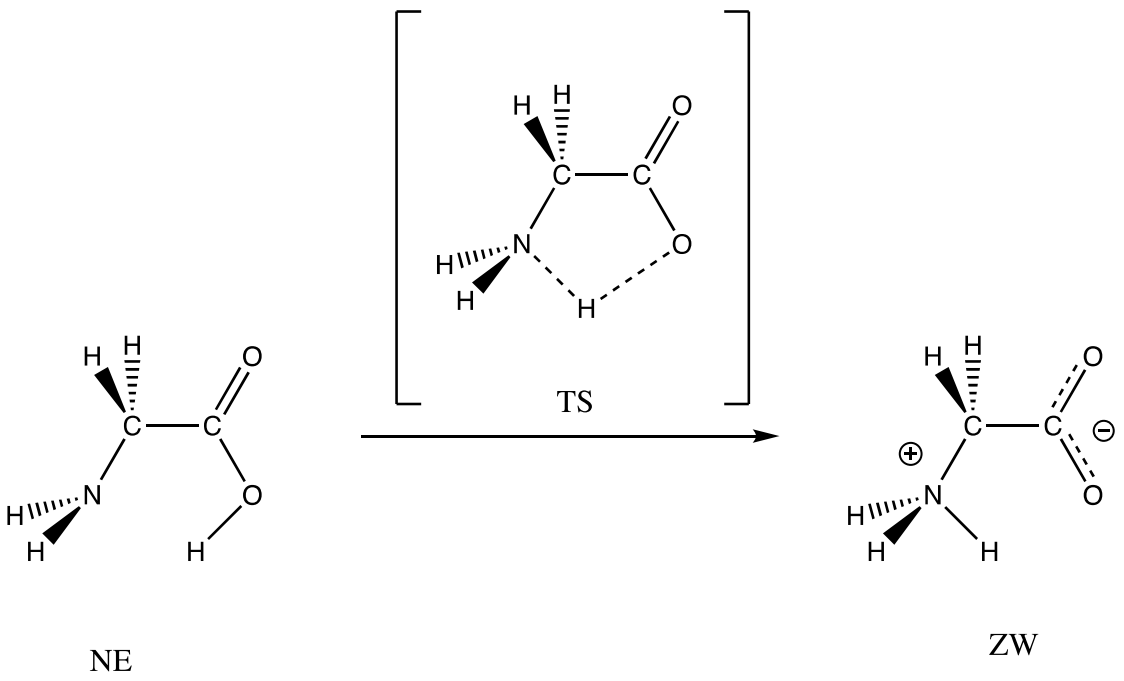
TS-2

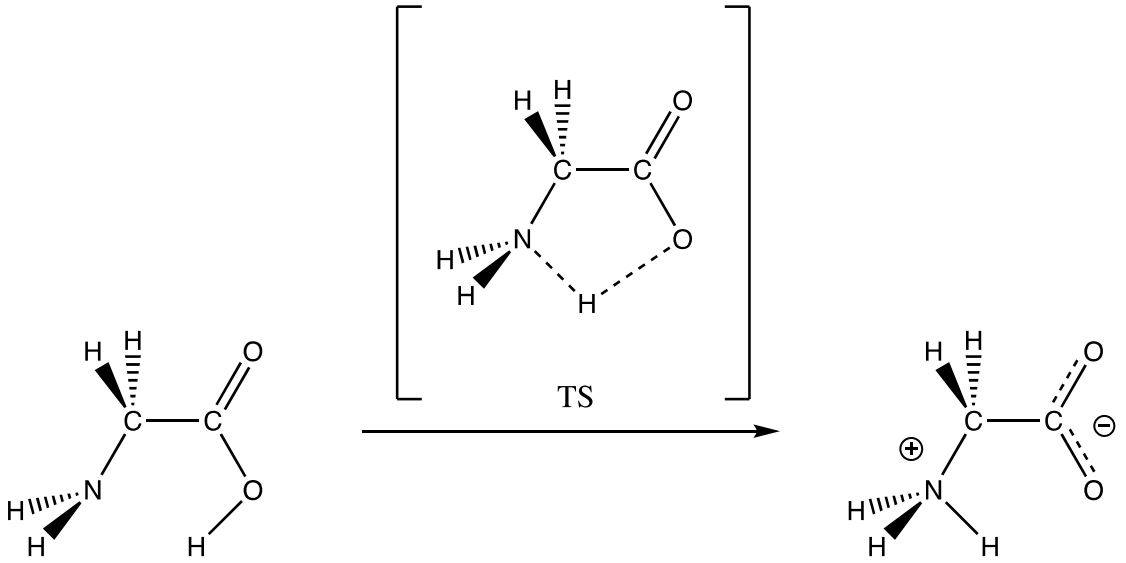
-1 1
C
F 1 r5
H 1 r2 2 a2
H 1 r3 2 a3 3 z3
H 1 r4 2 a4 3 z4
X 1 1.0 2 90. 3 0.
F 1 r6 6 a5 2 z5

Exercise 8. Localize and characterize (at the HF/6-31G* level) the TS for the reaction
H-CN→ CN-H
For that purpose, optimize first the reactants and products and then make an estimation of the TS structure. Calculate the rate constant.

Exercise 9. Glycine can be found in two tautomeric forms: as a neutral molecule (NE) and as a zwitterion (ZW). In addition, as we have already seen, the neutral form can be found in different conformations. Starting from the adequate conformation, the transformation from NE to ZW requires a simple proton transfer, as shown below.

i) Optimize the adequate NE and ZW structures at the HF/6-31G* level
ii) Confirm that these structures are real minima. Take note of the dipole moment
iii) Find the TS of the reaction
iv) Confirm that your structure is a saddle point. Take note of the dipole moment
v) Calculate the rate constant for the NE → ZW process at 298 K using TST





	NE	TS	ZW
E (hartrees)			
ΔE (kcal·mol^{-1})			
μ (Debye)			
G (hartrees)			
ΔG (kcal·mol^{-1})			

$$k_r = \frac{kT}{h}\left(\frac{RT}{P^0}\right)^{n-1} exp\left(-\frac{\Delta G_P^{0\ddagger}}{RT}\right)$$

Semiempirical Methods

Computational Chemistry
Elective Course
Chemistry Degree
4th year

Contents

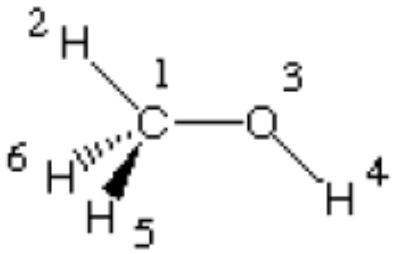
- ZDO – MNDO, AM1, PM3 (parametrization)
- CNDO (conformational studies)
- MNDO, AM1, PM3 (a critical case)
- PM3 (Calculating the rate constant)

MNDO, AM1, PM3 (parametrization)

- They are the most frequently used
- They belong to ZDO (they neglect all 2-electron integrals with more than 2 centers), and only valence electrons are considered
- They are based on NDDO, which must be distinguished from the older and simpler approximations CNDO and INDO
- Parametrized corrections are added to compensate for the approximations
- Each method is characterized by the parametrization
- In MNDO, AM1 and PM3 parametrizations, heats of formation are used rather than total energies

MNDO, AM1, PM3 (parametrization)

Exercise 1: (a) Compare the ***energies*** and the computational effort (***CPU time***) using B3LYP/6-31G* and AM1 for the dihedral scan of the methanol molecule



#AM1 OPT=(modredund) nosymm

Methanol dihedral scan at AM1 level

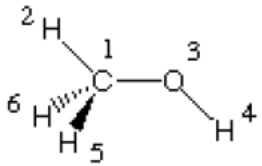
0 1
C1
H2 1 r2
O3 1 r3 2 a3
H4 3 r4 1 a4 2 d4
H5 1 r5 2 a5 3 d5
H6 1 r6 2 a6 3 d6

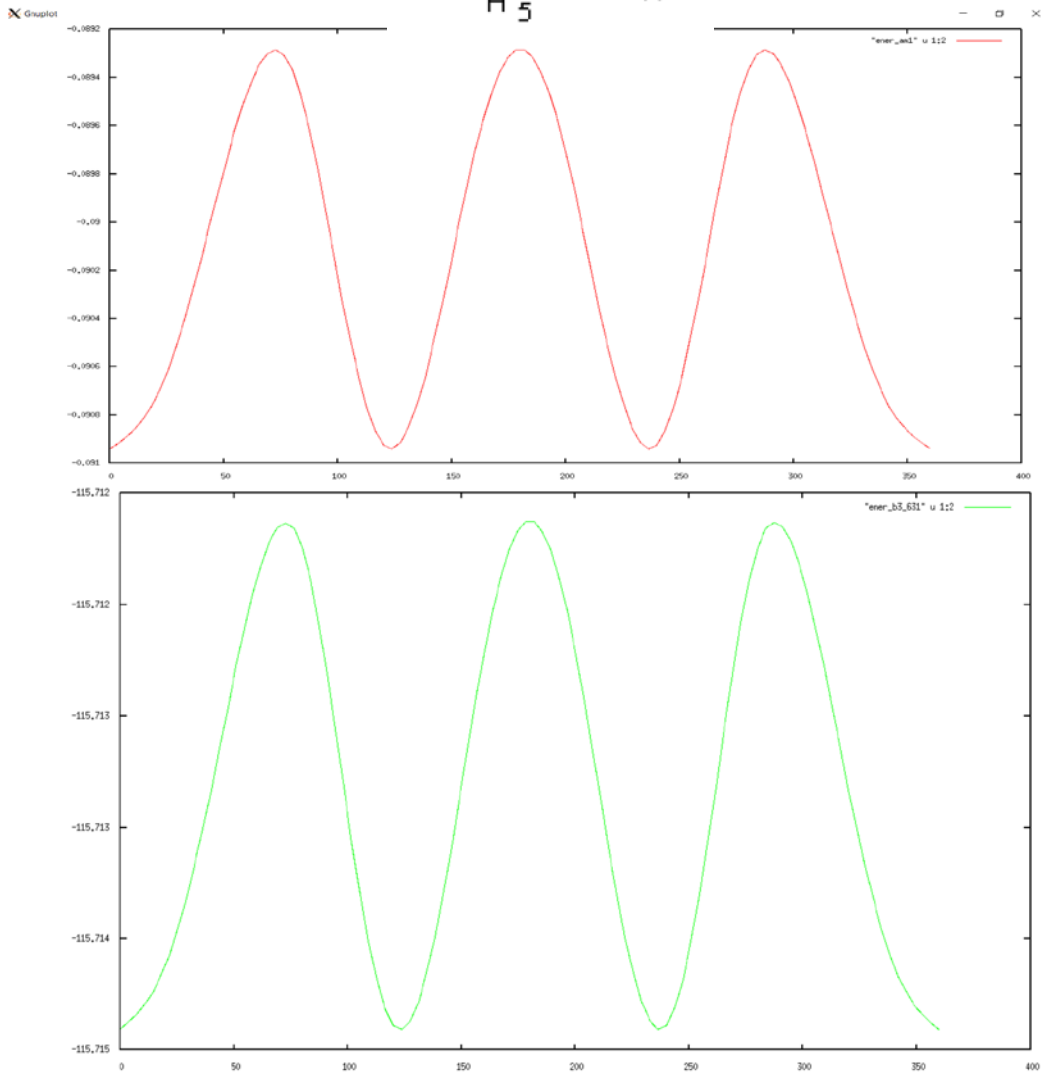
r2=1.12
r3=1.41
a3=109.0
r4=1.0
a4=109.0
d4=180.0
r5=1.0
a5=109.0
d5=120.0
r6=1.0
a6=109.0
d6=240.0

To remove predefined dihedral angles →

* 2 3 *
1 2 3 4 S 18 20.0

Exercise 1: (a) Compare the energies and the computational effort (CPU time) using B3LYP/6-31G* and AM1 for the dihedral scan of the methanol molecule





AM1
Job cpu time:
*B3LYP/6-31G**
Job cpu time:

AM1
D4=XXº→E= Hartrees
D4=YYº→E= Hartrees
ΔE = kcal mol^{-1}

*B3LYP/6-31G**
D4=XXº→E= Hartrees
D4=YYº→E= Hartrees
ΔE = kcal mol^{-1}

MNDO, AM1, PM3 (parametrization)

Exercise 1: (b) Compare how the core and valence electrons are treated by B3LYP/6-31G*, B3LYP/STO-6G and AM1 in the energy computations of the methanol molecule

AM1

Standard basis:
basis functions, primitive Gaussians, cartesian basis functions
alpha electrons beta electrons
nuclear repulsion energy Hartrees.

B3LYP/STO-6G

Standard basis:
basis functions, primitive Gaussians, cartesian basis functions
alpha electrons beta electrons
nuclear repulsion energy Hartrees.

B3LYP/6-31G*

Standard basis:
basis functions, primitive Gaussians, cartesian basis functions
alpha electrons beta electrons
nuclear repulsion energy Hartrees.

MNDO, AM1, PM3 (parametrization)

Exercice 1: (c) Analyze the difference between the model used for computing the nuclear repulsion energy in the B3LYP/6-31G* and NDDO methods in methanol .

- ***B3LYP/6-31G*:*** Simple point charge model (SPC)
  - $E_{AB} = Z_A Z_B / R_{AB}^2$

nuclear repulsion energy

- ***NDDO:*** Model SPC must be improved to compensate for the approximations
  - $E_{AB} = Z'_A Z'_B \langle s_A s_B | s_A s_B \rangle [1 + F_A + F_B]$
  - **MNDO:** $F_A = \exp(-\alpha_A R_{AB})$

    1 parameter: α_A

nuclear repulsion energy

  - **AM1:** $F_A = \exp(-\alpha_A R_{AB}) + \sum_i K_{Ai} \exp[L_{Ai} (R_{AB} - M_{Ai})^2]$

    4 parameters: α_A, K_{Ai}, L_{ai}, M_{Ai}

nuclear repulsion energy

MNDO, AM1, PM3 (parametrization)

Exercise 1: (d) Check the parameters used by AM1 (or PM3) using the command am1=print (or pm3=print) and consulting the section “Semi-empirical methods” in the Gaussian manual. Find parameters related to:

- 1-electron 1-center integrals
- 1-electron 2-center integrals
- 2-electron 1-center integrals
- 2-electron 2-center integrals
- core-core repulsion

CNDO (conformational studies)

Exercise 2: The CNDO semiempirical method significantly decreases the computational cost. Study its performance on the rotational energy barriers of one (or some) of the molecules from the table (review the section on conformational studies in the chapter "Optimization" to see how to prepare the computations).

	Barrier (kcal/mole)[b]		Barrier (kcal/mole)[b]	
Molecule	CNDO	AM1	B3LYP/6-31G*	Experiment
$CH_3–CH_3$				2.88
$CH_3–NH_2$				1.98
$CH_3–OH$				1.07
$CH_3–CH_2F$				3.33
$CH_3–CHF_2$				3.18
$CH_3–CF_3$				3.25
$CH_3–CH_2Cl$				3.68
$CH_3–CHCH_2$				1.99
cis–CH_3–CHCHCl				0.62
$CH_3–CHO$				1.16
$CH_3–NCH_2$				1.97

CNDO (conformational studies)

Exercise 2: The CNDO semiempirical method significantly decreases the computational cost. Study its performance on the rotational energy barriers of one (or some) of the molecules from the table (review the section on conformational studies in the chapter "Optimization" to see how to prepare the computations).

```
#CNDO OPT=(modredun) nosymm

Study of CH3CH3 rotation using CNDO

0 1
 c
 c   1 cc2
 h   2 hc3      1 hcc3
 h   1 hc4      2 hcc4      3 dih4
 h   2 hc5      1 hcc5      4 dih5
 h   2 hc6      1 hcc6      4 dih6
 h   1 hc7      2 hcc7      3 dih7
 h   1 hc8      2 hcc8      3 dih8

cc2      1.457447
hc3      1.120157
hcc3      112.091
hc4      1.120106
hcc4      112.093
dih4      0.000
hc5      1.120161
hcc5      112.094
dih5      59.979
hc6      1.120153
hcc6      112.093
dih6      179.978
hc7      1.120175
hcc7      112.089
dih7      59.975
hc8      1.120161
hcc8      112.092
dih8      179.971

* 2 1 *
3 2 1 4 S 18 20.
```

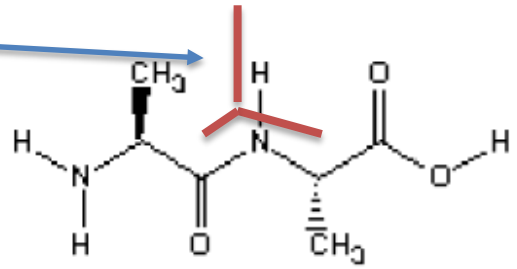
To remove predefined dihedral angles →

MNDO, AM1, PM3 (a critical case)

Exercise 3: The use of NDDO methods to determine heats of formation has been studied in a number of organic systems with C, H, N and O giving rise to relatively small errors (~50, 30 and 18 kJ/mol for MNDO, AM1 and PM3, respectively). However, the description is poorer for third-row elements like S or P, or hypervalent compounds. A critical case to analyze here is the geometric study of nitrogen compounds. For the inversion of the trivalent nitrogen, AM1 produces barriers that are too low, while PM3 gives rise to values that are too high.

3-a) Analyze the planarity of the central amide group in the ***alanine dipeptide*** with the semiempirical methods ***AM1*** and ***PM3***. To do that, build the molecule with GaussView, optimize it with both methods and take note of the inversion dihedral angle C(O)-N-H-C.

Method	Dihedral angle C(O)-N-H-C
AM1	
PM3	
B3LYP/6-31G*	



Exercise 3:

#AM1 OPT=(modredundan,maxcyc=200)
nosymm

Study of ALA_ALA dipeptide

0 1
c
c 1 cc2
n 2 nc3 1 ncc3
c 2 cc4 3 ccn4 1 dih4
n 4 nc5 2 ncc5 3 dih5
c 5 cn6 4 cnc6 2 dih6
c 6 cc7 5 ccn7 4 dih7
o 7 oc8 6 occ8 5 dih8
o 4 oc9 2 occ9 1 dih9
c 6 cc10 5 ccn10 4 dih10
o 7 oc11 6 occ11 5 dih11
h 3 hn12 2 hnc12 1 dih12
h 2 hc13 4 hcc13 9 dih13
h 1 hc14 2 hcc14 4 dih14
h 1 hc15 2 hcc15 4 dih15
h 1 hc16 2 hcc16 4 dih16
h 5 hn17 4 hnc17 2 dih17
h 6 hc18 5 hcn18 4 dih18
h 10 hc19 6 hcc19 5 dih19
h 10 hc20 6 hcc20 5 dih20
h 10 hc21 6 hcc21 5 dih21
h 3 hn22 2 hnc22 1 dih22
h 11 ho23 7 hoc23 6 dih23

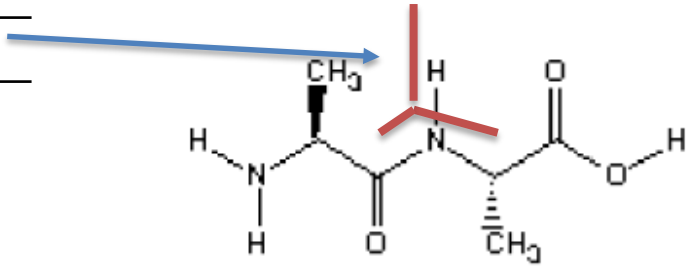
cc2 1.517085
nc3 1.506614
ncc3 111.455
cc4 1.566864
ccn4 109.123
dih4 119.604
nc5 1.366810
ncc5 116.320
dih5 157.660
cn6 1.469776
cnc6 118.384
dih6 138.958
cc7 1.582427
ccn7 110.221
dih7 -88.772
oc8 1.267963
occ8 117.130
dih8 -7.925
oc9 1.251321
occ9 116.775
dih9 103.493
cc10 1.518384
ccn10 111.850
dih10 149.933
oc11 1.248620
occ11 116.428
dih11 171.443
hn12 1.037176
hnc12 107.602
dih12 -89.760
hc13 1.128520
hcc13 110.756
dih13 -135.485
hc14 1.115206
hcc14 111.148
dih14 171.302
hc15 1.136064
hcc15 108.911
dih15 51.408
hc16 1.117450
hcc16 111.734
dih16 -65.681
hn17 1.008915
hnc17 114.248
dih17 7.889
hc18 1.126546
hcn18 108.543
dih18 29.214
hc19 1.115657
hcc19 111.588
dih19 -65.108
dih19 -65.108
hc20 1.119012
hcc20 107.149
dih20 175.237
hc21 1.116244
hcc21 111.427
dih21 56.780
hn22 1.019895
hnc22 110.215
dih22 29.851
ho23 0.950000
hoc23 109.471
dih23 180.000

MNDO, AM1, PM3 (a critical case)

Exercise 3: The use of the NDDO methods to determine heats of formation has been studied in a number of organic systems with C, H, N and O giving rise to relatively small errors (~50, 30 and 18 kJ/mol for MNDO, AM1 and PM3, respectively). However, the description is poorer for third-row elements like S or P, or hypervalent compounds. A critical case to analyze here is the geometric study of nitrogen compounds. For the inversion of the trivalent nitrogen, AM1 produces barriers that are too low, while PM3 gives rise to values that are too high.

3-b) Discuss how the error could propagate in the structure of oligopeptides and comment on the consequences regarding the determination of the three-dimensional structure of peptides. Build the system ALA-ALA-ALA (where ALA means Alanine). Analyze the planarity of the central amide group in the ***alanine tripeptide*** with the semiempirical methods ***AM1*** and ***PM3*** and compare them with B3LYP/6-31G*

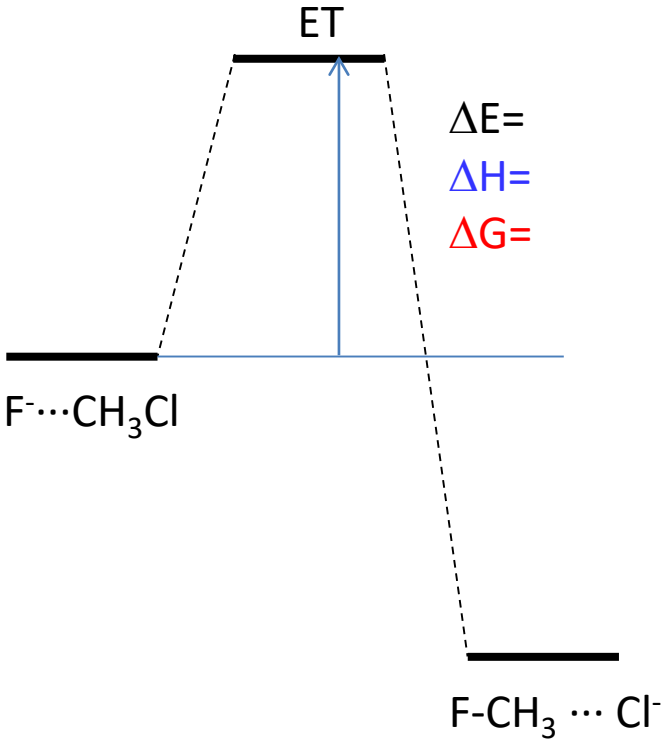
method	Dihedral angle C(O)-N-H-C
AM1	
PM3	
B3LYP/6-31G*	



PM3 (Calculating the rate constant)

Exercise 4. Obtain the rate constant

Complete the following energy diagram for the reaction (in kcal·mol^{-1}) and obtain the rate constant at 298 K from the activation energy (the free energy difference between the TS and the reactants complex) using PM3 and compare the results with the HF from the pass unit.



$$k_r = \frac{kT}{h}\left(\frac{RT}{P^0}\right)^{n-1} \exp\left(-\frac{\Delta G_P^{0\ddagger}}{RT}\right)$$

PostHF
and
Spectroscopy calculations

Computational Chemistry
Elective Course
Chemistry Degree
4th Year

Contents

- Spectroscopy calculations
- Rotational, vibrational, and electronic spectroscopy
- Normal modes
- **Concepts**: transitions between energy levels

Spectra: Types and regions of the light spectrum

- Two dimensions:
  - Abscissa: energy difference between two states
  - Ordinates: transition intensity
- Quantized molecular motions:
  - Rotation: microwave region
  - Rotation-vibration: infrared region
  - Rotation-vibration-electronic: UV-visible region, X-Ray
  - Nuclear states transitions: Gamma rays
  - Nuclear spin transitions: Microwaves? NMR
  - Electron spin transitions: Radiofrequencies? ESR

Rotational spectra

- Rotational levels (rigid rotor)

$$E_J = B\ J\ (J+1)$$

- State populations

$$N_J/N_0 = (2J + 1)\ \exp(-\Delta E_J/k_B T) / q$$

- Selection rules
  - Generic: $\mu \neq 0$
  - Specific: $\Delta J = +1$

EXERCISE 1: rotational spectra

- Identification of interstellar molecules: calculation of the moment of inertia of cyanopolyynes:

  HCN, HC_3N and HC_5N

  1. Build the molecules
  2. Perform a geometry optimization calculation HF/3-21G and B3LYP/6-31G* of the 3 molecules
  3. Build a table with the values of:
     - Rotational constants
     - Molecular geometries (bond lengths)
     - Total dipole moment

     And compare with experimental data that can be found in the literature (44.316; 4.5491; 1.3313 GHz for the rotational constants)

     Fit the experimental constants *vs* the theoretical ones using the two levels of theory.

HCN.com

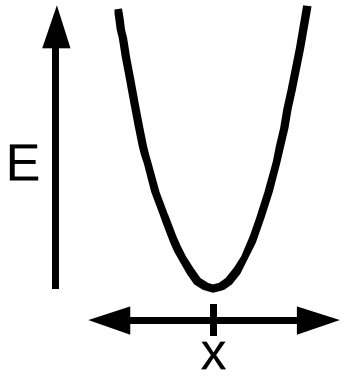
```
%chk=HCN.chk
# hf/3-21g  opt

Title Card Required

0 1
 C
 H             1       B1
 X             1       1.0    2    90.0
 N             1       B2     3     A1       2 180.0

   B1              1.050
   B2              1.137
   A1             90.00
```

Vibrational spectra

- Pure vibrational energy: $E_v = (v + 1/2)\, h\, \nu$
- State populations: normally only v=0 is significantly populated
- Selection rules:
  - Generic: $d\mu/dq \neq 0$
  - Specific: $\Delta v = +1(, +2, +3\ldots)$
- Normal modes of vibration: from the diagonalization of the Hessian matrix

Molecular vibrations

- A molecule has 3*N*-6 vibrational degrees of freedom



$$E = \frac{1}{2} k x^2 \quad \text{where} \quad k = \frac{\partial^2 E}{\partial x^2}$$

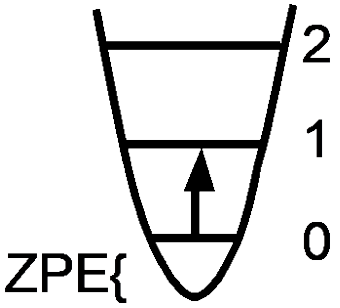
$$\nu = \frac{1}{2\pi} \sqrt{\frac{k}{\mu}}$$

Reduced mass

- The vibrations come from a parabolic potential energy surface
- Transition states have imaginary frequencies (k negative)

Vibrational quantum states

- Vibrational states are quantized



$$E = h\nu(v + 1/2)$$

- Ground state transition occurs from $0 \rightarrow 1$
- Zero point energy (ZPE) is always present

Infrared spectroscopy

- **FREQ** provides IR, Raman vibrational frequencies and intensities
- Frequency calculations must be performed on a stationary point
- Frequency calculations must use the same model and basis set as the optimization calculation
- *Hartree-Fock:* systematically overestimates frequencies by ~10% due to the lack of electron correlation and to anharmonicity → one can scale the calculated values by multiplying by ≈ 0.9

Normal modes

- **FREQ** provides atom displacements for each normal mode
- Gaussian *output* provides Cartesian displacements
- GaussView shows these displacements as 3D vectors

Stationary point characterization

- Minimum: all eigenvalues are positive
- Saddle point: imaginary eigenvalue (appears as negative)
- Top hill: more than one imaginary eigenvalue (negative)

EXERCISE 2: Normal modes of vibration of formaldehyde

1. Build formaldehyde, H_2CO. Optimize the geometry with a B3LYP/3-21G calculation
2. Visualize the result, modify the input to perform a B3LYP/3-21G calculation of the vibrational frequencies
3. Visualize the spectrum and each vibrational mode
4. Complete the following table of vibrational frequencies by visually identifying the description of each normal mode
5. Obtain an averaged scaling factor for the B3LYP/3-21G frequencies

Mode	Description	frequency B3LYP/3-21G (cm-1)	frequency Biblio. (cm-1)
n1	sym CH stretch		2811
n2	CO stretch		1756
n3	CH_2 bend		1500
n4	out-of-plane bend		1170
n5	antisym CH stretch		2861
n6	CH_2 rock		1251

EXERCISE 3: Stretching frequencies of carbonyl groups

- Calculate the C=O stretching frequency of formamide [1], acetaldehyde [2], and acetyl chloride [3]
- Optimize the geometry and obtain the vibrational frequencies at HF/3-21G level for each molecule. Calculated vibrational frequencies are systematically too high and are normally scaled to give scaled frequencies. The scale factor for HF/3-21G calculations is 0.9085
- Make a table with columns for:
  - The molecule
  - The calculated CO stretching frequency
  - The scaled CO stretching frequency
  - The experimental frequencies (1740, 1746, and 1822 cm^{-1} for [1], [2], and [3], respectively)
- Comment on the absolute and relative precision of the frequency calculations

Electronic spectra: Theory

- Planck equation:

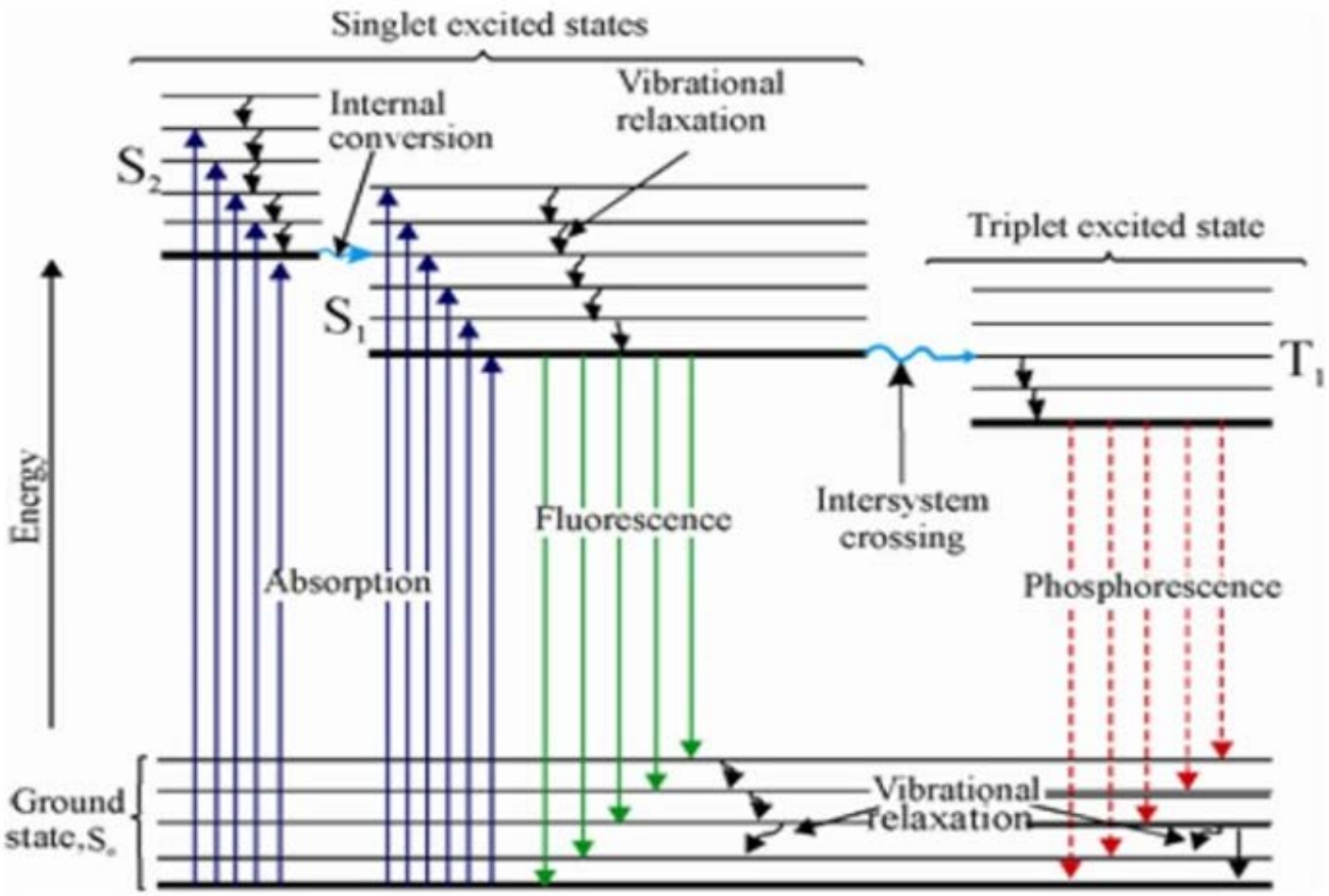
$$\Delta E = E_f - E_i = h\nu$$

- Energy states of the valence electrons ⇒ UV-Vis photons
- Transition probability → transition dipole moment:

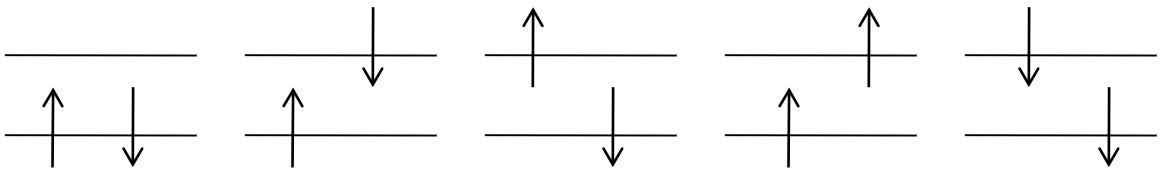
$$|\mu_{fi}|^2 = |\int \Psi_f \, \mu \, \Psi_i \, d\tau|^2$$

Selection rules

- $|\mu_{fi}| \neq 0$
- Selection rules:
  - Changes in orbital configuration (changes in electron distribution and orbital overlapping)
  - State multiplicity $\Delta S=0$ (if there is no spin-orbit coupling or other perturbations)


Singlet excited states
Internal conversion
Vibrational relaxation
S_2
S_1
Triplet excited state
T_1
Energy
Absorption
Fluorescence
Intersystem crossing
Phosphorescence
Ground state, S_0
Vibrational relaxation

Order of the states: 2 e^- in 2 MO



$$\left|{}^1\Psi_0\right\rangle \quad \left|\Psi^{\bar{2}}_{\bar{1}}\right\rangle \quad \left|\Psi^{2}_{1}\right\rangle \quad \left|{}^3\Psi^{2}_{\bar{1}}\right\rangle \quad \left|{}^3\Psi^{\bar{2}}_{1}\right\rangle$$

$$\left|1\,\bar{1}\right\rangle \quad \left|1\,\bar{2}\right\rangle \quad \left|2\,\bar{1}\right\rangle \quad \left|1\,2\right\rangle \quad \left|\bar{1}\,\bar{2}\right\rangle$$

Excited states methods

- Excited states: difficult of calculate → methods tend to find the ground state
- Options:
  - ZINDO: semiempirical
  - CIS (CI-Singles): fast, robust, qualitative
  - TDDFT (Time Dependent DFT): fast, qualitative
  - CASSCF (Complete Active Space SCF): slow, complicated, more accurate (FCI in an orbital subset)

CIS method

- Promotes a single electron from occupied orbitals to empty ones
- Wave functions are built from these interacting electronic configurations
- The wave function of lowest energy is the ground state; wave functions of higher energy are excited states
- Energy differences between states correspond to electronic transitions in the UV-Vis spectrum

EXERCISE 4: UV-Vis spectra of conjugated aldehydes

- Build and perform geometry optimization calculations at the HF/3-21G level:
  - Molecules: 2-propenal, 2-butenal, 2,4-pentadienal
- Using the resulting geometry, perform CIS calculations of the UV-Vis spectra for each molecule, with several basis sets, from 3-21G until 6-311++G**. Repeat the calculations with the ZINDO method.
- The first excited state (A^2) corresponds to a forbidden $n \rightarrow \pi^*$ transition and therefore is extremely weak. The second excited state (A′) corresponds to an allowed $\pi \rightarrow \pi^*$ transition and absorbs strongly.
- Compare the calculated absorption maximum with the experimentally observed maximum (experimental value of the absorption maximum (nm): 209, 221, 251) in a table with columns for: molecule, calculated maximum, experimental maximum and difference
- Comment on any trend that you can observe

propenal.com
1) Geometry optimization

```
%chk=Propenal.chk
# opt hf/3-21g  guess=save

Propenal opt

0 1
 O
 C         1        B1
 H         2        B2     1       A1
 C         2        B3     1       A2     3      D1
 …

   B1              1.21
   B2              1.08
   B3              1.47
   …
   A1            120.0
   A2            120.0
   …
   D1           -180.0
   …
```

propenal.com
2) CIS and ZINDO calculations

```
%chk=Propenal.chk
# rcis=(nstates=3)/3-21g geom=checkpoint guess=save

Propenal CIS 3-21g

0 1
```

```
%chk=Propenal.chk
# rzindo=(nstates=3) geom=checkpoint guess=save

Propenal CIS 3-21g

0 1
```

propenal.com
4) CIS/6-311++G(3df,3pd)

```
%chk=Propenal.chk
# rcis=(nstates=3)/6-311++g(3df,3pd) geom=ckeckpoint guess=save

Title Card Required

0 1
```

propenal.com
4) CIS/6-311++G(3df,3pd)

```
%chk=Propenal.chk
# rcis=(nstates=3)/6-311++g(3df,3pd) guess=save

Title Card Required

0 1
 O
 C                  1           B1
 H                  2           B2    1           A1
 C                  2           B3    1           A2    3
…

   B1              1.21004806
   B2              1.08752859
   B3              1.47390480

…
   A1            121.43213622
   A2            124.33428833
…
   D1           -179.97481689
…
```

Modeling Solvent Effects in Chemical Processes

Computational Chemistry
Elective Course
Chemistry Degree
4th year

Contents

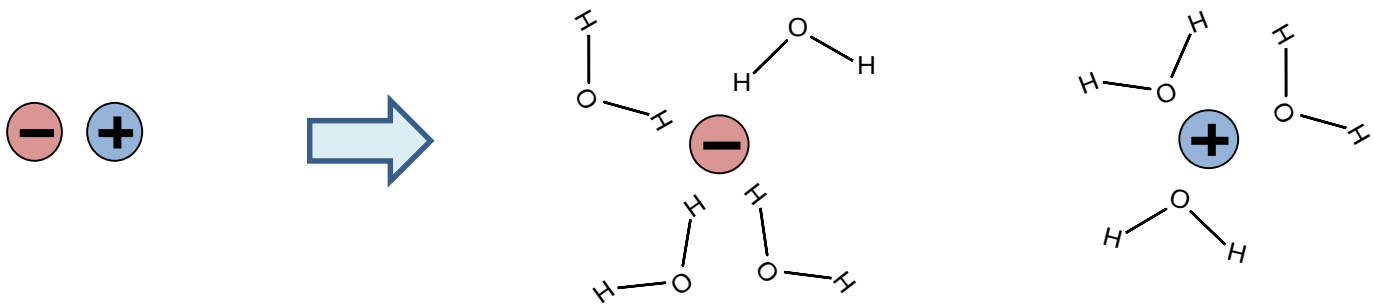
1. Introduction: the continuum model
2. Effect of solvent on molecular properties
3. Effect of solvent on intermolecular interactions
4. Effect of solvent on tautomeric equilibria
5. Effect of solvent on conformational equilibria
6. Effect of solvent on chemical reactivity

1. Introduction

So far we have studied molecular properties and chemical processes in the gas phase, meaning that our molecule or molecules are isolated.

However a large part of chemistry occurs in condensed phases: in solution or solid state. In particular, many organic or biochemical processes take place in solution (preferably water).

The solvent may have a decisive influence on molecular properties or on the energy of chemical processes. For example, a process of charge separation is very expensive in gas phase while the presence of a polar solvent may favor it:



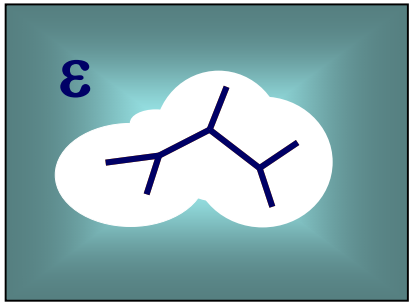
Problems:

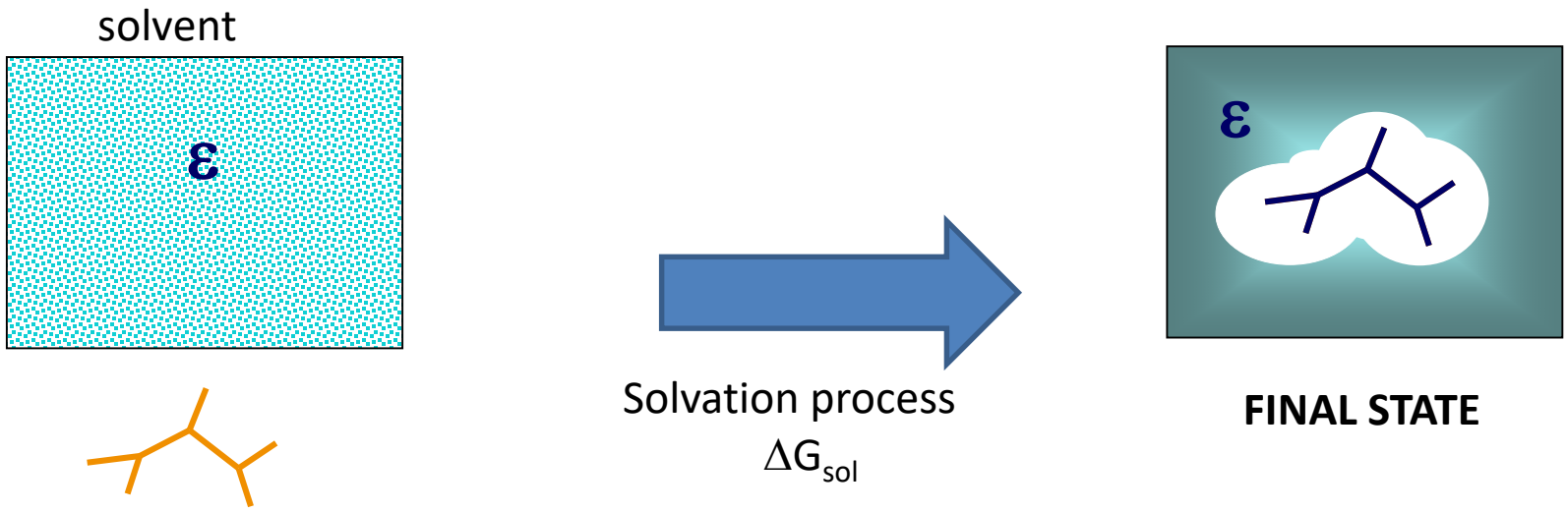
1) As electrostatic interactions decay smoothly with distance, simulations of solvent effects would require the consideration of a large number of solvent molecules, making it unviable from the point of view of computational calculations.

2) In addition, we must take into account thermal fluctuations that cause the movement of molecules, forcing us to consider many possible structures in our calculations and to obtain an average of them

Solution:

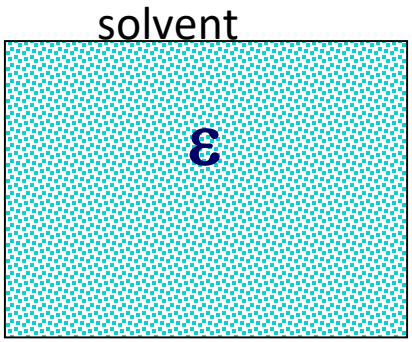
We can consider an 'average' infinite solvent: a continuous dielectric medium characterized by a few macroscopic properties: the dielectric constant, surface tension, the coefficient of thermal expansion... Molecules under study will be placed inside such an environment and its properties will be obtained solving the wavefunction in the presence of such an environment



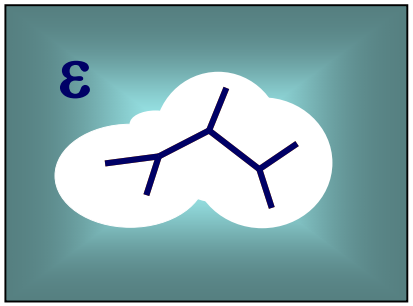


As we take an average representation of solvent, characterized by the value of the dielectric constant at the temperature of interest and constant pressure, we obtain the free energy associated to the process.

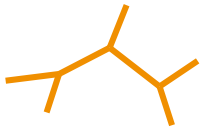
To calculate the free energy of the solute in the solvent the process is formally divided in 3 steps:



ΔG_{sol}



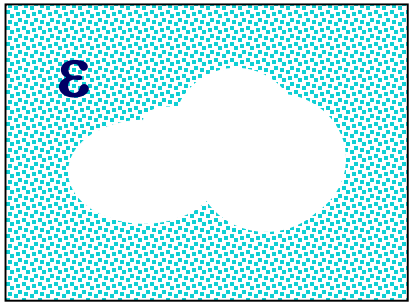
FINAL STATE



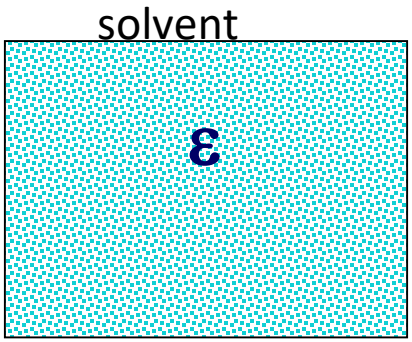
Gas phase solute

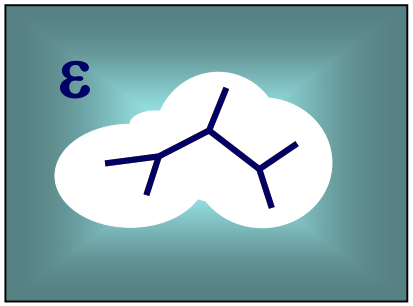
INITIAL STATE

Process of creation of a cavity in the solvent ΔG_{cav}

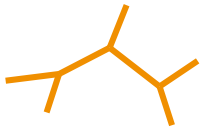


To calculate the energy of the solute in the solvent the process of dissolution of the solute is formally divided in 3 easy steps:



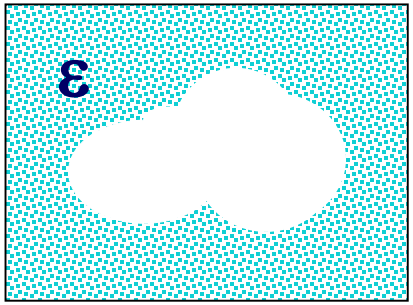


FINAL STATE

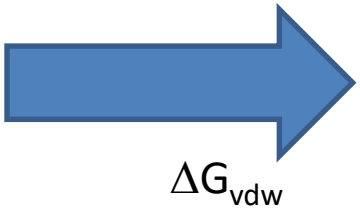


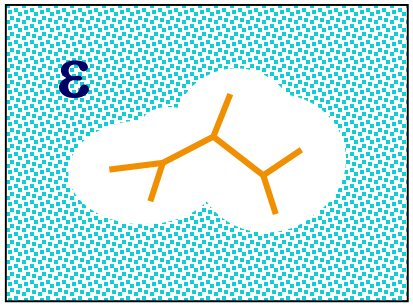
Gas phase solute

INITIAL STATE

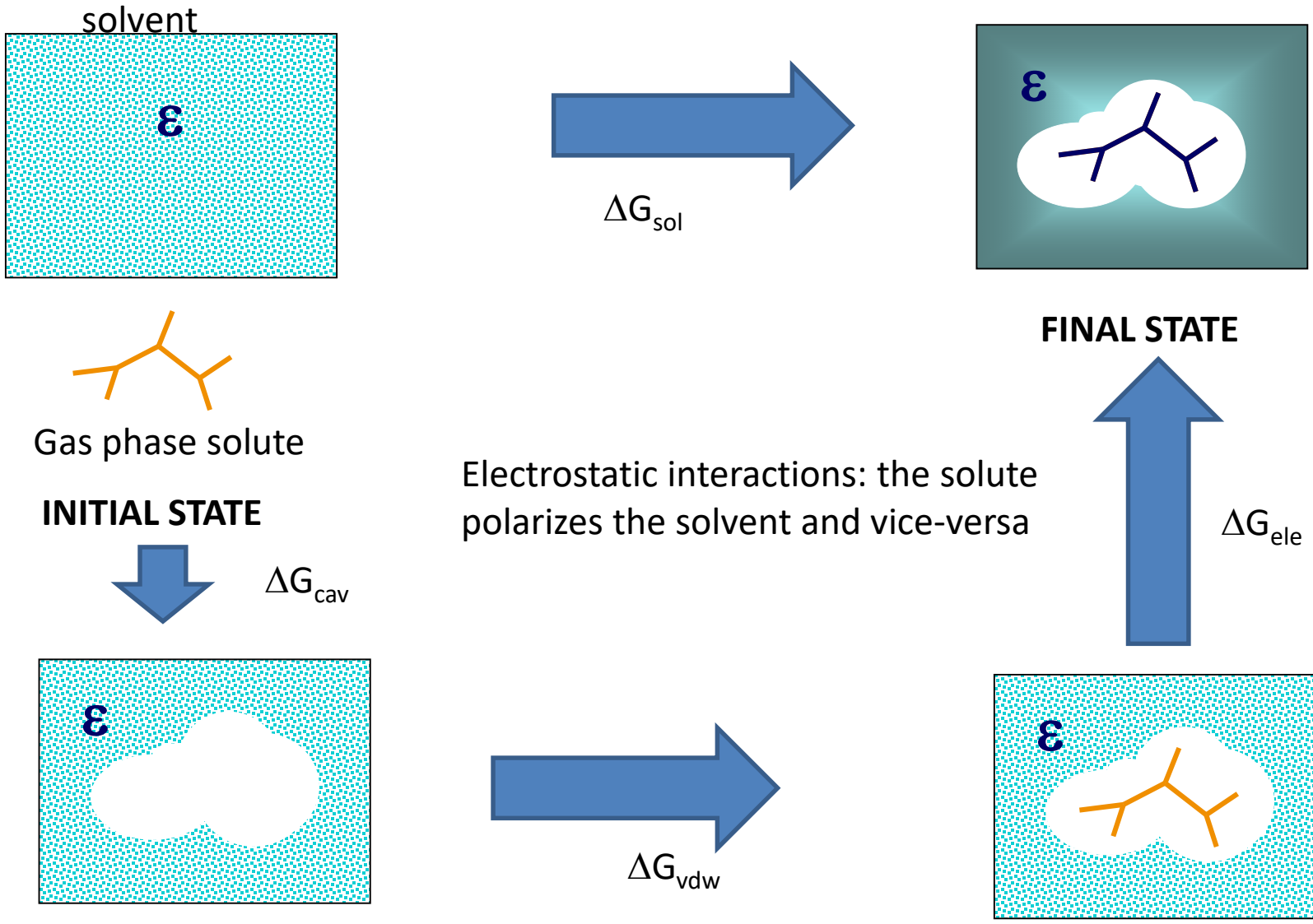


The solute enters the solvent, but first we 'connect' only dispersion-repulsion interactions and not electrostatic interactions

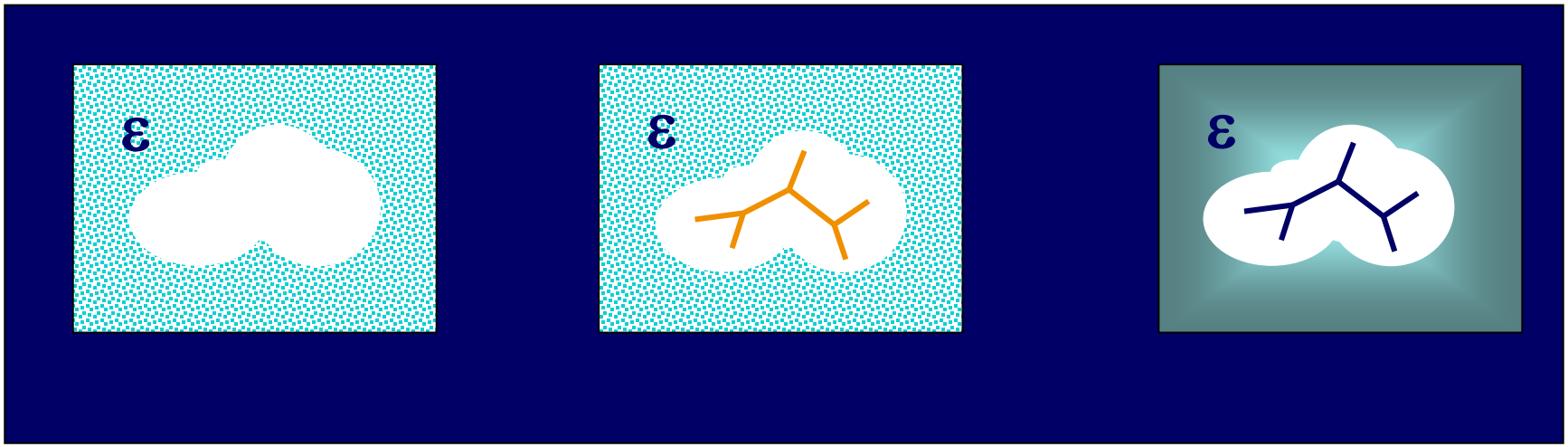




To calculate the energy of the solute in the solvent the process of dissolution of the solute is formally divided in 3 easy steps:



The energy balance of the solvation process therefore involves 3 different contributions in the continuous model



Cavity formation

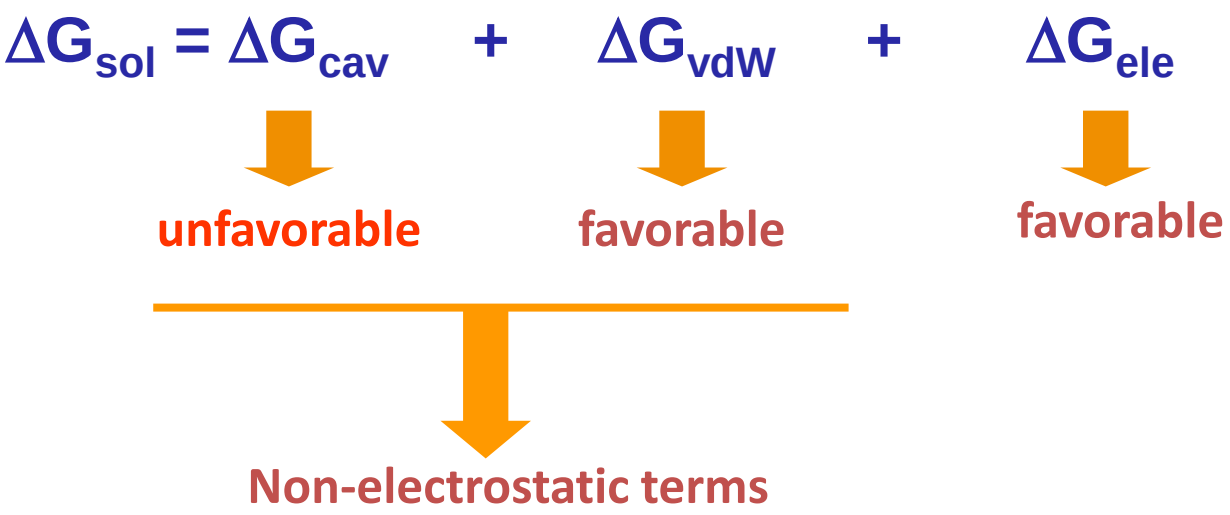
Dispersion-repulsion interactions (van der Waals)

Mutual polarization of solute and solvent

Non-electrostatic interactions

Electrostatic interactions

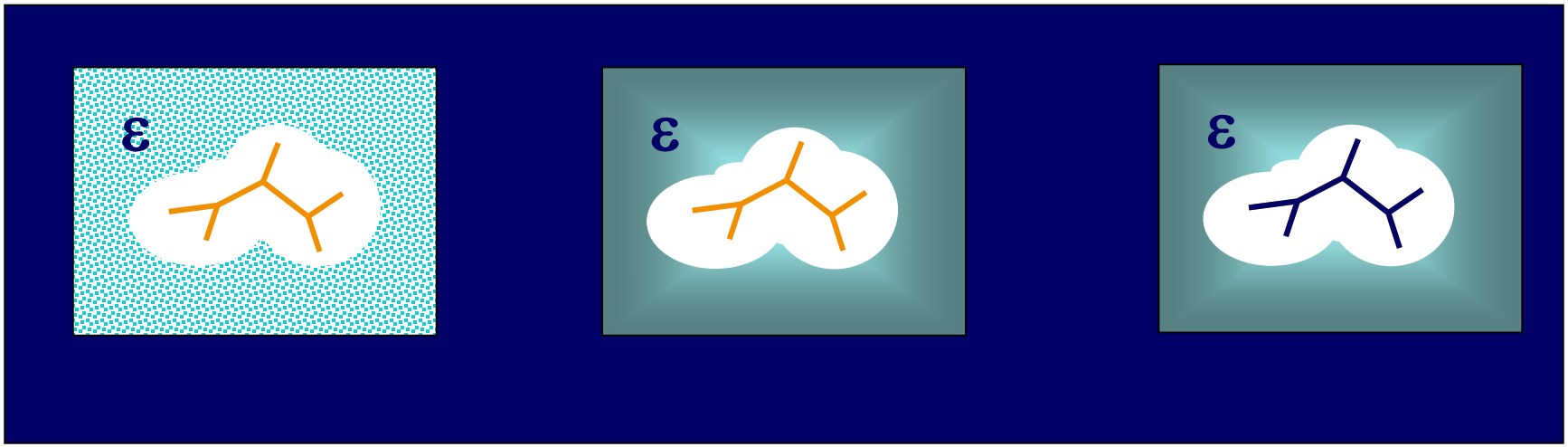
The energy balance of the solvation process therefore involves 3 different contributions in the continuous model.



The non-electrostatic term can be evaluated with different strategies. A simple way is to express this term as proportional to the size of the cavity enclosing the solute. One can then evaluate this as the sum of atomic contributions, expressed as the product of the exposed surface of the atom and an effective surface tension which can be parameterized from experimental free energies of solvation

$$\Delta G_{n-el} = \Delta G_{cav} + \Delta G_{vdw} = \sum_{k=1}^{N} \sigma_k S_k$$

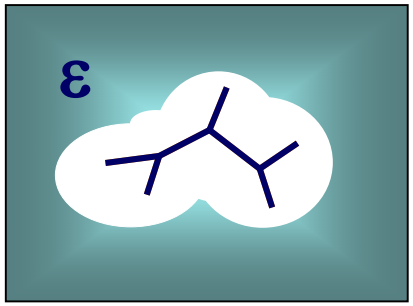
The last step involves a mutual polarization of solute and solvent to be solved iteratively



The solute polarizes the continuum → **The continuum creates an electrostatic potential on the solute:** ***Reaction Field*** → **The solute is polarized**

(Self Consistent)

We must take into account that the electronic distribution of the solute (and therefore its wavefunction) are modified (or polarized) by the presence of the solvent

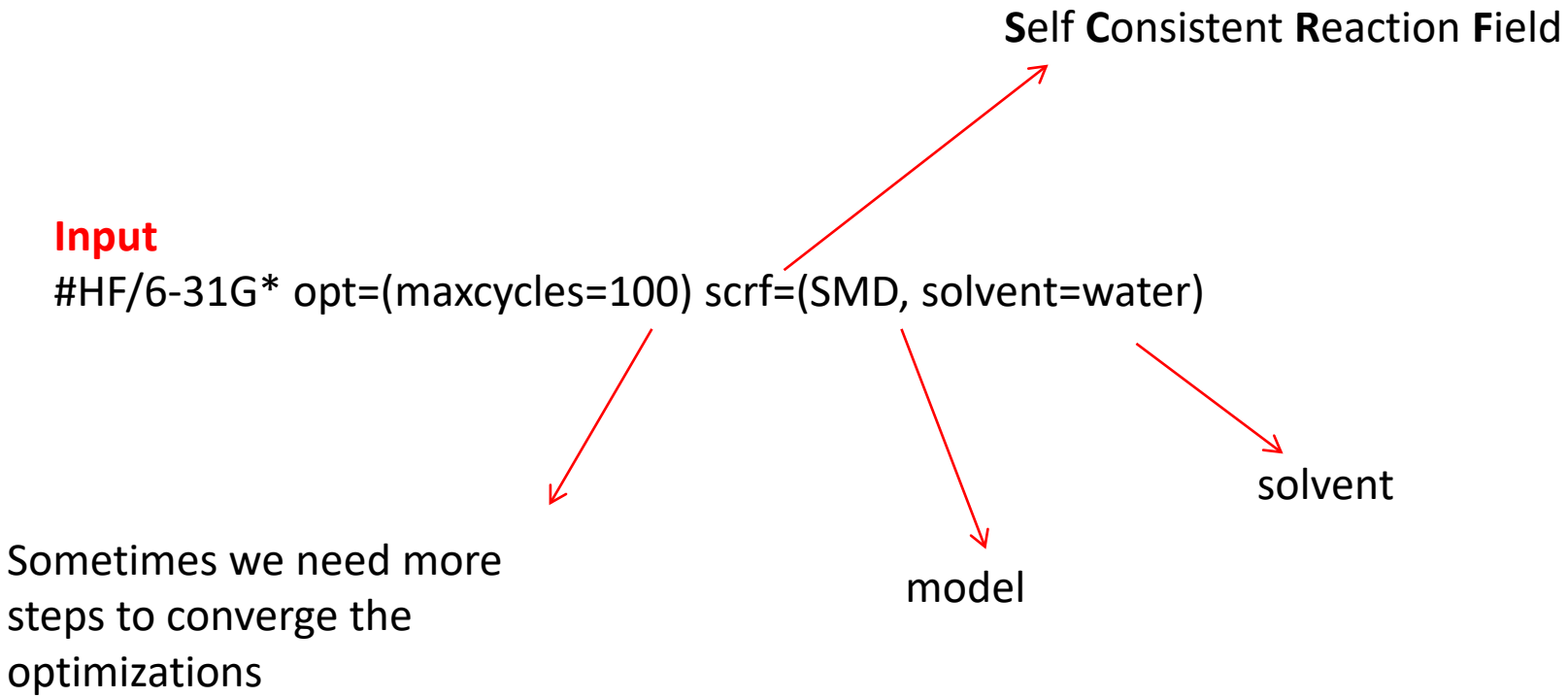


	Hamiltonian	Wavefunction	Energy
Gas phase	$\hat{H}_0$	Ψ_0	$E_0 = \langle \Psi_0 \vert \hat{H}_0 \vert \Psi_0 \rangle$
Continuum	$\hat{H} = \hat{H}_0 + \hat{V}_\varepsilon$	Ψ	$E + G_{ele} = \langle \Psi \vert \hat{H} \vert \Psi \rangle = \langle \Psi \vert \hat{H}_0 \vert \Psi \rangle + \langle \Psi \vert \hat{V}_\varepsilon \vert \Psi \rangle$

$\hat{V}_\varepsilon$ Solute-solvent electrostatic interaction operator. Depends on the dielectric constant, the size and shape of the cavity and the distribution of charges of the solute (this is why we must solve it in a self-consistent way).
It corresponds to an 'average' description of the solvent

Calculations will be run using Gaussian09

The basic command line is:



Output

```
Polarizable Continuum Model (PCM)
=================================
 Model              : PCM.
 Atomic radii       : SMD-Coulomb.
 Polarization charges : Total charges.
 Charge compensation  : None.
 Solution method    : Matrix inversion.
 Cavity type        : VdW (van der Waals Surface) (Alpha=1.000).
 Cavity algorithm   : GePol (No added spheres)
                      Default sphere list used, NSphG=    9.
                      Lebedev-Laikov grids with approx.  5.0 points / Ang**2.
                      Smoothing algorithm: Karplus/York (Gamma=1.0000).
                      Polarization charges: spherical gaussians, with
                                    point-specific exponents (IZeta= 3).
                      Self-potential: point-specific (ISelfS= 7).
                      Self-field    : sphere-specific E.n sum rule (ISelfD= 2).
 1st derivatives    : Analytical E(r).r(x)/FMM algorithm (CHGder, D1EAlg=3).
                      Cavity 1st derivative terms included.
 2nd derivatives    : Analytical E(r).r(xy)/FMM algorithm (CHGder, D2EAlg=3).
                      Cavity 2nd derivative terms included.
 Solvent            : Water, Eps=  78.355300 Eps(inf)=   1.777849
```

Spheres list:

ISph	on	Nord	Re0	Alpha	Xe	Ye	Ze
1	C	1	1.850	1.000	-0.121034	0.024816	-0.008263
2	Cl	2	2.380	1.000	1.678613	-0.008832	0.003393
3	N	3	1.890	1.000	-2.520759	-0.010300	0.001110
4	H	4	1.200	1.000	-0.476744	1.018545	-0.318060
5	H	5	1.200	1.000	-0.521091	-0.715673	-0.700890
6	H	6	1.200	1.000	-0.543655	-0.191562	0.983993
7	H	7	1.200	1.000	-2.869343	0.191187	-0.914254
8	H	8	1.200	1.000	-2.868402	-0.915736	0.282943
9	H	9	1.200	1.000	-2.885673	0.686592	0.650401

Atomic radii for non-electrostatic terms: SMD-CDS.

The cavity needed to place the solute within the continuum is obtained as a set of overlapping spheres centered on each of the atoms of the molecule

$$\langle \Psi | \hat{H} | \Psi \rangle + \Delta G_{n-el}$$

Includes all the energy components

SCF Done: E(RHF) = -555.266477111 A.U. after 12 cycles

Convg = 0.5457D-08 -V/T = 2.0006

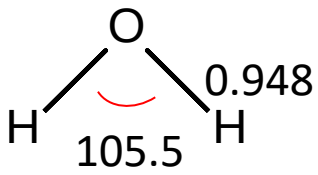
SMD-CDS (non-electrostatic) energy (kcal/mol) = 3.76

(included in total energy above)

2. Solvent Effects on Molecular Properties

The presence of a polar solvent polarizes the electron density of a molecule, causing electronic and structural changes. Thus, the dipole moment of a molecule of water in gas phase is lower than in aqueous solution. OH distances are also lengthened by solvent effects.

HF/6-31G*



μ = 2.20 D

Exercise 1. Optimize the structure of a water molecule in aqueous solution at level HF/6 - 31G* and compare it with the geometry, charges and dipole moment obtained in gas phase.

#hf/6-31G* opt SCRF=(SMD, Solvent=water)

Agua en dis. aq.

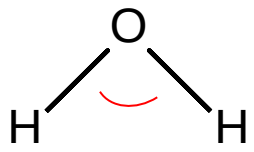
0 1
O
H 1 r1
H 1 r1 2 a2

r1 0.948
a2 105.5

2. Solvent Effects on Molecular Properties

The presence of a polar solvent polarizes the electron density of a molecule, causing electronic and structural changes. Thus, the dipole moment of a molecule of water in gas phase is lower than in aqueous solution. OH distances are also lengthened by solvent effects.

HF/6-31G*

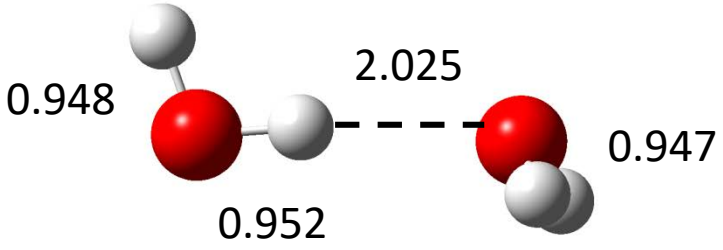


	Gas phase	Aq. Solution
D(O-H) (Å)		
HOH (degrees)		
Dipole Moment (D)		
Q(O) (a.u.)		
Q(H) (a.u.)		

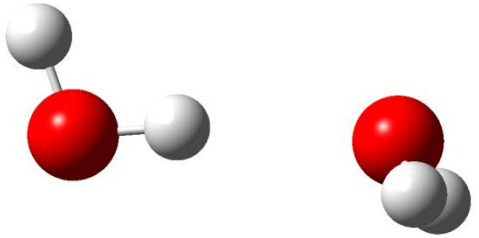
3. Solvent Effects on Intermolecular Interactions

The presence of a polar solvent can modify intermolecular interactions, such as hydrogen bonds. When polarizing a molecule, the solvent increases the charge separation within the molecule, but the interaction between charges is reduced by a factor equal to $1/\varepsilon$. The final result depends on each system in particular, and can increase or decrease the intensity of intermolecular interactions

Exercise 2. Optimize the structure of the dimer of water in aqueous solution at HF/6 -31 G * level and compare it with the geometry obtained in gas phase.



Gas phase



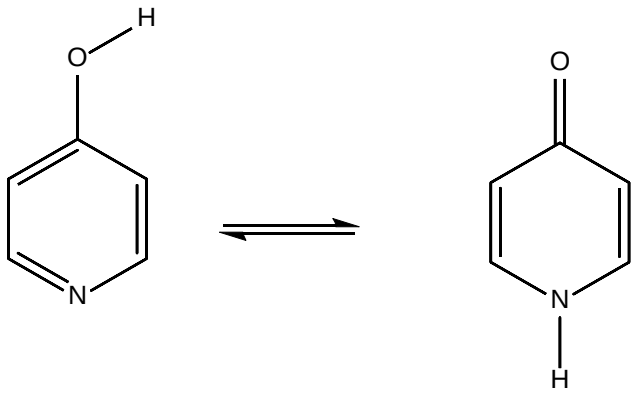
Aqueous Solution

3. Solvent Effects on Intermolecular Interactions

```
0 1
O
O 1 r1
X 1 1.0 2 90.0
H 1 r2 3 a2 2 0.0
H 1 r3 4 a3 3 z3
H 2 r4 1 a4 3 z4
H 2 r5 1 a5 3 z5

r1 3.0
r2 0.97
r3 0.95
r4 0.95
r5 0.95
a2 90.0
a3 105.0
a4 105.5
a5 106.0
z3 180.
Z4 60.0
z5 -61.0
```

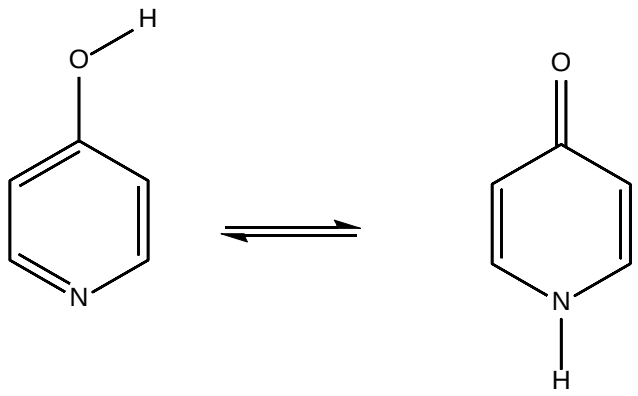
4. Solvent Effects on Tautomeric equilibria

The effect of a polar solvent can alter the tautomeric equilibrium between two species, changing the equilibrium constant by several orders of magnitude. This is the case for the 4-hydroxypyridine and 4-pyridone.



Exercise 3. Calculate the free energy difference between 4-hydroxypyridine and 4-pyridone at 298 K in gas phase and in aqueous solution at the HF/3-21G, HF/6-31G* and B3LYP/6-31G* levels.

4. Solvent Effects on Tautomeric equilibria

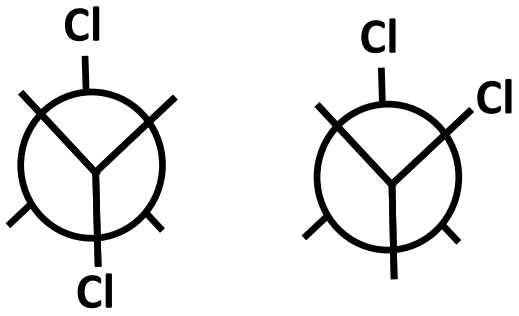


ΔG = G(4-pyridone) – G(4-hydroxipyridine)

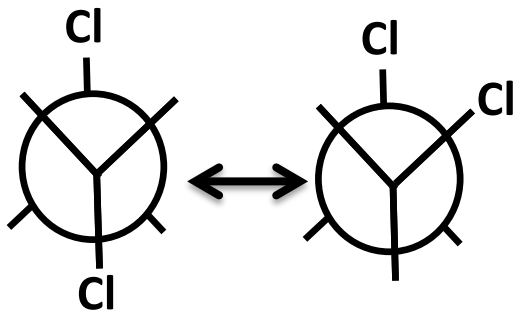
	ΔG_{gas} (kcal·mol^{-1})	ΔG_{sol} (kcal·mol^{-1})
HF/3-21G		
HF/6-31G*		
B3LYP/6-31G*		

5. Solvent Effects on Conformational Equilibria

The effect of a polar solvent can also affect conformational equilibria, favoring conformations that are more polar. An example is 1,2-dichloroethane. The more stable form in the gas phase is the trans conformer that has a dipole moment of zero. However, the gauche has a non-zero dipole moment and therefore interacts more strongly with the solvent.



Exercise 4. Calculate the conformational minima corresponding to the rotation around the C-C bond in dichloroethane in aqueous solution at the HF/6-31G* level and compare with the results obtained in session 3 in gas phase.
Optimize the two structures in solution, calculate frequencies and obtain the free energy difference between the two conformers and the equilibrium constant in solution at 298 K.

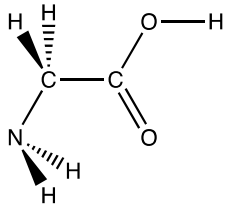


$\Delta G = G(\text{gauche}) - G(\text{trans})$

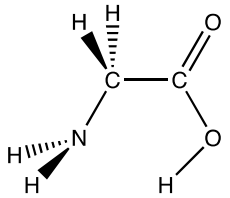
	Gas phase	Aq. solution
ΔG (kcal/mol)	1.85	
K	$4.4 \cdot 10^{-2}$	
% trans	95.8	

Exercise 5. Some molecules can exist in many different conformations and the solvent can change the populations of each with respect to the values found in gas phase. As seen previously, in the gas phase glycine exists as a neutral molecule and different conformations are possible. We are going to optimize the four lower-energy conformers of this molecule in solution and we will discuss if the solvent changes the populations of the neutral conformers.

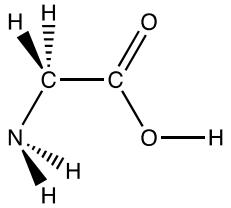
i) Optimize the conformations at the HF/6-31G* level in aqueous solution.
ii) Confirm that all the obtained structures are real minima
iii) Take note of the dipole moment in each conformer
iv) Calculate the free energy of the four conformers at 298 K, and then their populations assuming that there are no other conformers available



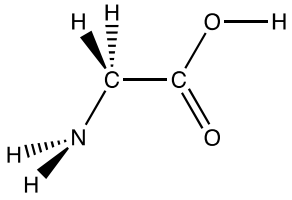
A



B



C



D

Gas Phase

	A	B	C	D
E (hartrees)				
ΔE (kcal·mol^{-1})				
μ (Debye)				
G (hartrees)				
ΔG (kcal·mol^{-1})				
P_i(%)				

$$p_i(\%) = \frac{e^{-\frac{\Delta G_i}{RT}}}{\sum_{j=A,B,C,D} e^{-\frac{\Delta G_j}{RT}}}$$

Aqueous Solution

	A	B	C	D
E (hartrees)				
ΔE (kcal·mol^{-1})				
μ (Debye)				
G (hartrees)				
ΔG (kcal·mol^{-1})				
P_i(%)				

5. Solvent Effects on Chemical Reactivity

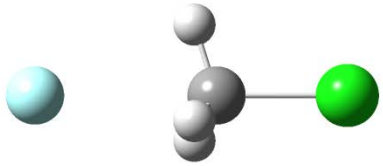
The Menshutkin reaction is a model example of solvent influence on the rate of a chemical reaction. This is an S_N2 type reaction leading to charged products.

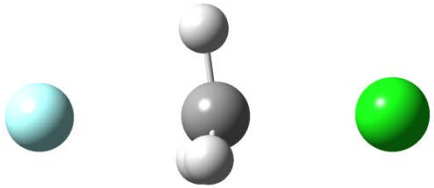
$$Et_3N + Met\text{-}I \rightarrow [Et_3N^{\delta+} \cdots Me \cdots I^{\delta+}]^{\ddagger} \rightarrow Et_3NMe^{+} + I^{-}$$

Solvent	**ε**	**k_{rel}**
Hexane	2.0	1.
CCl_4	2.2	31.
Chlorobenzene	5.6	1200.
Acetonitrile	37.5	12000.
Dimethylsulfoxide	46.7	50000.

The creation of charges is a process requiring more energy in the gas phase, and is carried out much more easily in an environment able to stabilize them.

Exercise 6. Starting from the TS and structures for the reactants complex obtained in gas phase for the $F^- + CH_3Cl \rightarrow FCH_3 + Cl^-$ reaction, reoptimize them in aqueous solution at the HF/6-31+G* level and recalculate the activation energy barrier. Compare the resulting barrier with that obtained in gas phase and rationalize the result.

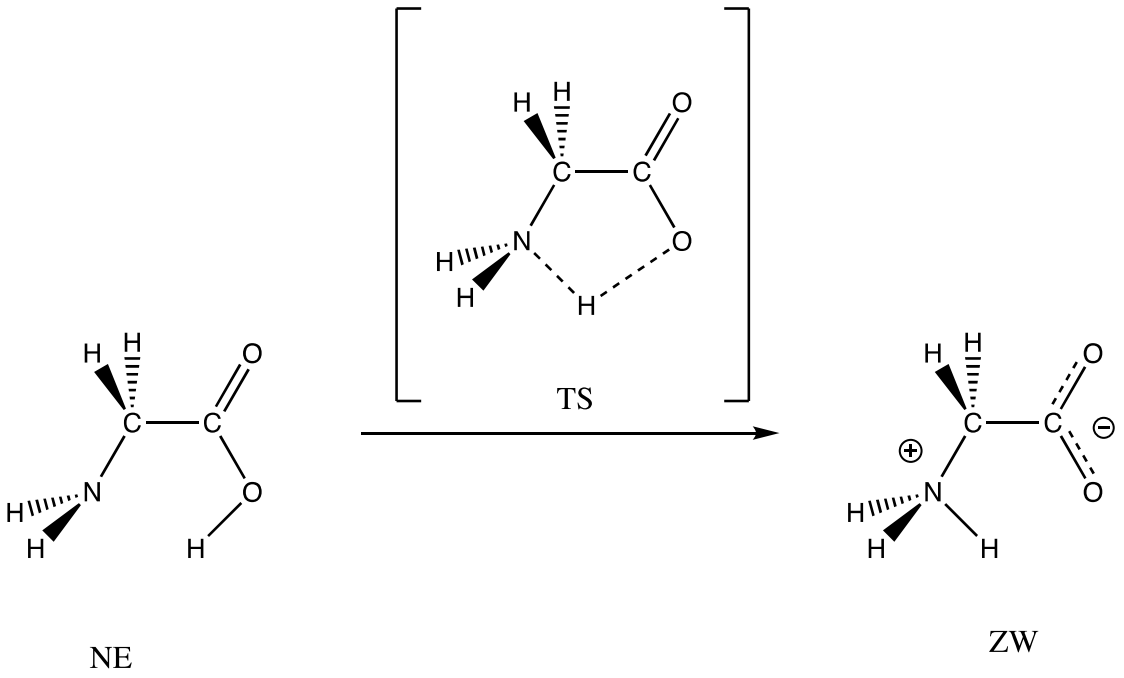




	$F^- \cdots CH_3Cl$ GAS PHASE	TS GAS PHASE	$F^- \cdots CH_3Cl$ SOLUTION	TS SOLUTION
Energy (hartrees)	-598.535868	-598.531104		
E + ZPE (hartrees)	-598.494855	-598.490328		
Enthalpy (hartrees)	-598.488752	-598.484991		
Free Energy (hartrees)	-598.523090	-598.517171		
$\Delta G^{\ddagger}$ (kcal·mol^{-1})	3.71			
d(C-Cl) (Å)	1.863	2.133		
d(C-F) (Å)	2.585	2.125		
d(C-H) (Å)	1.070	1.062		

Exercise 7. As previously discussed, glycine can be found in two tautomeric forms: as a neutral molecule (NE) and as a zwitterion (ZW). The ZW form is much more polar and so it is expected to be favored by the presence of polar solvents

i) Optimize the adequate NE and ZW structures at the HF/6-31G* level in aqueous solution
ii) Confirm that these structures are real minima. Take note of the dipole moment
iii) Find the TS of the reaction
iv) Confirm that your structure is a saddle point. Take note of the dipole moment
v) Calculate the rate constant for the NE → ZW process at 298 K in soluiton using TST



Gas Phase

	NE	TS	ZW
E (hartrees)			
ΔE (kcal·mol^{-1})			
μ (Debye)			
G (hartrees)			
ΔG (kcal·mol^{-1})			

$$k_r = \frac{kT}{h}\left(\frac{RT}{P^0}\right)^{n-1} exp\left(-\frac{\Delta G_P^{0\ddagger}}{RT}\right)$$

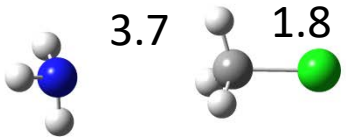
Aqueous Solution

	NE	TS	ZW
E (hartrees)			
ΔE (kcal·mol^{-1})			
μ (Debye)			
G (hartrees)			
ΔG (kcal·mol^{-1})			

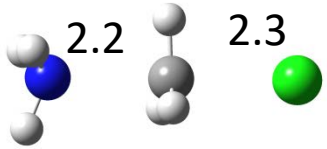
We are going to study a Menshutkin type reaction in different solvents, calculating the activation energy from the difference between reactants and transition structure.

$$H_3N + CH_3Cl \rightarrow H_3NCH_3^+ + Cl^-$$

The geometry of the reactants complex and the TS (in angstroms) in a solvent with a high dielectric constant are approximately:



REACTANTS



TS

Exercise 8. Obtain transition structures and reactant complexes for the Menshutkin reaction model at the HF/6-31+G* level for the following list of solvents:
Heptane (ε=1.9); Benzene (ε=2.3); Chloroform (ε=4.7); Acetonitrile (ε=35.6) and water (ε=78.4). Use previous structures as starting points.
Try to establish correlations between the N-C and C-Cl distances and the charge on the chloride anion in the TS with the dielectric constant. Obtain the free energy barriers and compare them with the dielectric constant.

Computational Chemistry

Molecular Dynamics

Scheme

1. Using NAMD and VMD

2. Molecular Dynamics of liquid water

3. Molecular Dynamics of Na^+ in liquid water

4. Molecular Dynamics of a protein

Using NAMD and VMD

Practical work on MD (classical MD) will be performed with the NAMD code

http://www.ks.uiuc.edu/Research/namd/

Results will be visualized using VMD:

http://www.ks.uiuc.edu/Research/vmd/

Both programs are distributed freely and are available for several operative systems (including Linux, Windows, MacOS,…).

Using NAMD and VMD

Download the file **Dinamica_agua.tar** **From Aula Virtual**

$ tar –xvf Dinamica_agua.tar

Input file: **dinamica_aguas.namd**:

```
## OUTPUT/INPUT

# Amber/(t,s,x)leap generated parameter file
# AMBER file containing parameters and topology (force field)
parmfile      aguas.prmtop


# Input
coordinates aguas_input.coor      #Initial coordinates
velocities  aguas_input.vel       #Initial velocities

# Output
binaryoutput       no              #Do not write output file in binary
restartfreq        500             #Frequency for writing restart files
dcdfreq            250             #Frequency for writing trajectory file
outputEnergies     100             #Frequency for writing energies in the output file
outputname         aguas_output_#Name of the output files
firsttimestep 0                    #First step starts at t=0
```

```
## FORCE FIELD PARAMETERS
amber on                        #Amber force field will be used
exclude       scaled1-4         #Interaction between atoms 1-4 are treated as non-bonding interactions
cutoff 12.                      #Cutoff radii for non-bonding interactions
switching on                    #Switching applied to non-bonding interactions
switchdist 10.                  #Inner cutoff radii for non-bonding interactions
pairlistdist 13.5                #Pair-list radii
readexclusions yes              #Pairs excluded for non-bonding interactions can be read from parm file
1-4scaling    0.83333333        #Factor used to scale down 1-4 non-bonding interactions

# Integrator parameters
timestep          1.00          #time step in fs
nonbondedFreq     1             #frequency for calculating non-bonded L-J interactions
fullElectFrequency  1           #frequency for calculating non-bonded electrostatic interactions



# DEFINITION OF PERIODIC BOUNDARY CONDITIONS
cellBasisVector1   32.2231  0.0 0.0
cellBasisVector2   0.0 32.3625  0.0
cellBasisVector3   0.0 0.0 32.2231
cellOrigin        0.0 0.0 0.0
wrapAll           on            #Output coordinates are always translated into the original simulation cell
```

```
# Constant Temperature Dynamics
langevin           on    # Langevin dynamics
langevinDamping     10.0   # damping coefficient for Langevin dynamics (gamma)
langevinTemp        300   # Target temperature



## Execution orders


numsteps 100000          # Number of steps (100000 x 1 fs = 100 ps)
```

Using NAMD and VMD

NAMD needs the following files:

Input file: contains program options and the name of the rest of files
Parm file: force field
Coordinates: initial coordinates
Velocities: initial velocities (these can be also automatically generated by the program)

NAMD generates the following output files

log: contains the more relevant information (options, energies, temperature, volume,…)
.vel: velocities
.coor: coordinates
.**xsc**: Boundary conditions (important if volume is not constant)
restart files: files containing coordinates(coor), velocities(vel) and boundary conditions (xsc) needed to restart the calculation if desired
dcd: contains the trajectory of the system, a movie where frames are saved with a fixed frequency specified in the input

Using NAMD and VMD

Coor file:

```
REMARK  FINAL COORDINATES WRITTEN BY NAMD AT TIMESTEP 100000
ATOM      1  O   WAT     1      0.516  13.743  -5.340  1.00  0.00
ATOM      2  H1  WAT     1     -0.419  13.646  -5.455  1.00  0.00
ATOM      3  H2  WAT     1      0.604  13.559  -4.394  1.00  0.00
ATOM      4  O   WAT     2    -13.812  10.569 -14.557  1.00  0.00
ATOM      5  H1  WAT     2    -14.786  10.381 -14.393  1.00  0.00
ATOM      6  H2  WAT     2    -13.422  10.233 -13.714  1.00  0.00
ATOM      7  O   WAT     3     14.207  12.314   8.056  1.00  0.00
ATOM      8  H1  WAT     3     13.689  12.286   8.919  1.00  0.00
ATOM      9  H2  WAT     3     14.500  13.243   8.111  1.00  0.00
ATOM     10  O   WAT     4     -3.420  -2.290   6.703  1.00  0.00
ATOM     11  H1  WAT     4     -2.931  -1.485   7.000  1.00  0.00
ATOM     12  H2  WAT     4     -3.701  -2.675   7.548  1.00  0.00
ATOM     13  O   WAT     5      9.141   1.166 -10.116  1.00  0.00
ATOM     14  H1  WAT     5     10.061   1.336 -10.356  1.00  0.00
ATOM     15  H2  WAT     5      9.068   1.401  -9.175  1.00  0.00
```

Header: when the file comes from x-ray or nmr it specifies the experimental conditions. If it comes from a previous MD run it specifies the timesteps

Cartesian coordinates

Additional information: x-ray beta-factor, probability of occupancy (in case of several conformations)

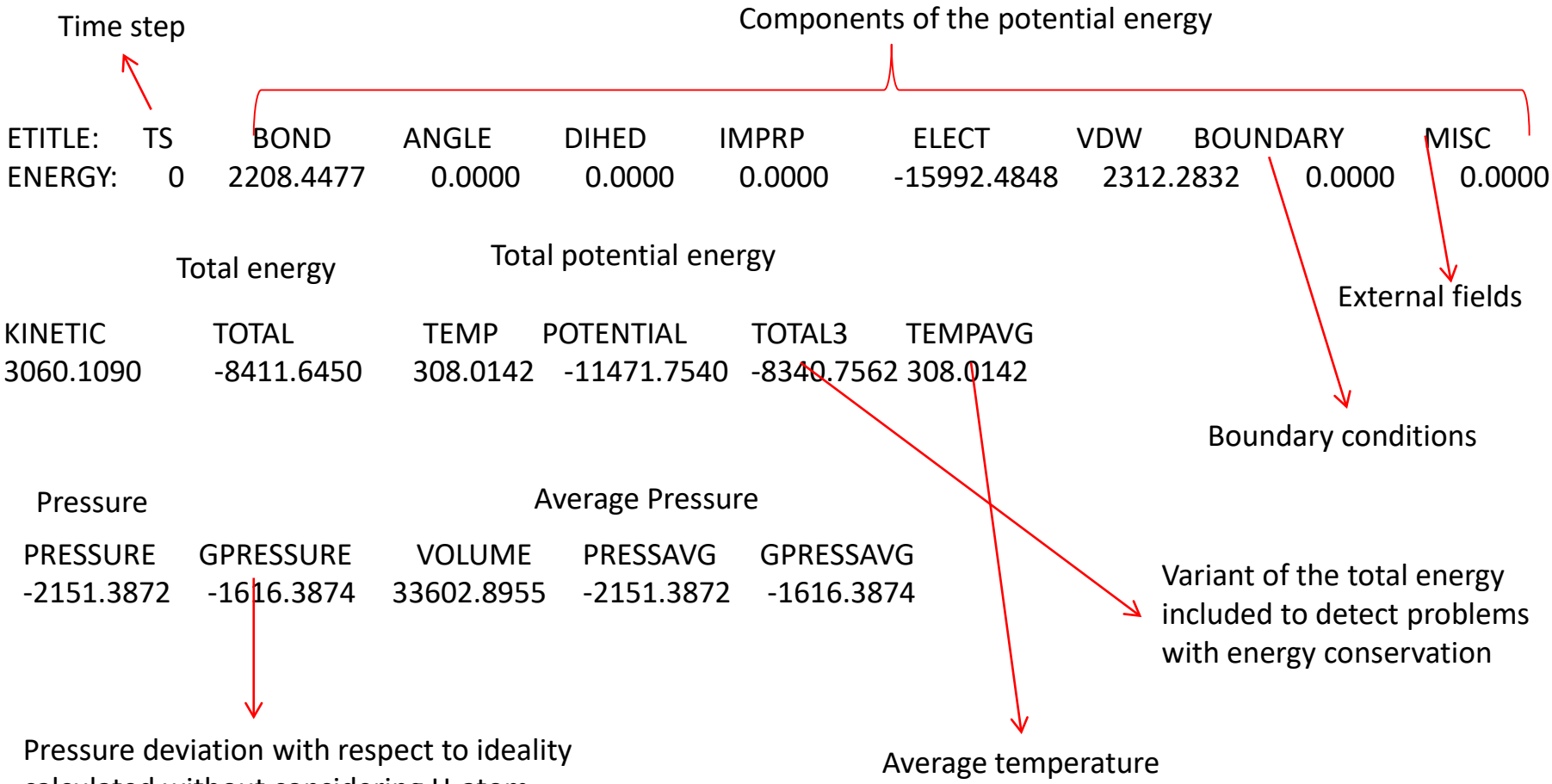
Atom number

Atom type: needed to identify their parameters in the force field

Molecule or residue to whom the atom belongs to

The output file:

This file contains the options used for the simulations and the result of reading the input files. This file contains the energy terms at different time steps, as well as other information needed to control the simulation.



MD of Liquid Water

Download and extract **“Dinamica_agua.tar”**

The folder contains the input and output files corresponding to a MD simulation of a box of water molecules described using the TIP3P potential. The box dimensions are 32.22 x 32.36 x 32.22 $Å^3$ and it contains 1111 molecules , resulting in a density of 0.988 $g \cdot cm^{-3}$

MD is performed in the NVT ensemble at 300K

The outputs corresponding to a 100ps long MD (100000 steps with a timestep of 1 fs) are located in the folder /outputs

Exercise 1. Prepare and perform a 100 ps long MD simulation of the system.

$namd2 input_file.namd > output_file.log

MD of Liquid Water

Analysis of the results

For the analysis we will use the VMD program

$vmd

File > New Molecule

Browse the file *.coor and in ‘file type’ select PDB and press ‘Load’. Click on the name of the file in the VMD Main box

File > Load Data into Molecule

Select the file outputs/*.dcd

It is possible to visualize the complete trajectory or frame by frame using the arrows that appear in the VMD Main box

With the mouse one can rotate the system (press R key), scale (press S) and translate (press T)

Graphics > Representations allows us to select different drawing and coloring methods

MD of Liquid Water

Analysis of the results

Extensions > Analysis > NAMD Plot

It allows us to plot different properties of the system obtained during the simulations

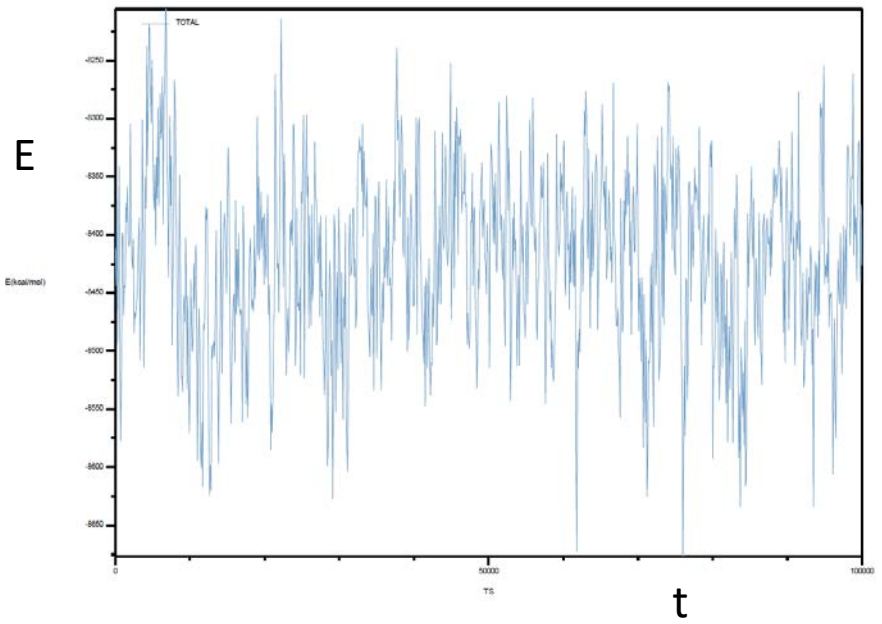
With the 'File' command select the file *.log, select the property to represent and plot it ('Plot Selected Data')
Plots can be saved (in postcript) and also the data (in ASCII)

1. Plot the total energy obtained during the simulation. Compare the fluctuations with the average value. If a simulation is stable, fluctuations must represent a small % of the total value of the property. Remember that according to Statistical Mechanics fluctuations diminish with the number of atoms of the system

2. Plot the instantaneous and averaged values of temperature. Compare the fluctuations with the average value.

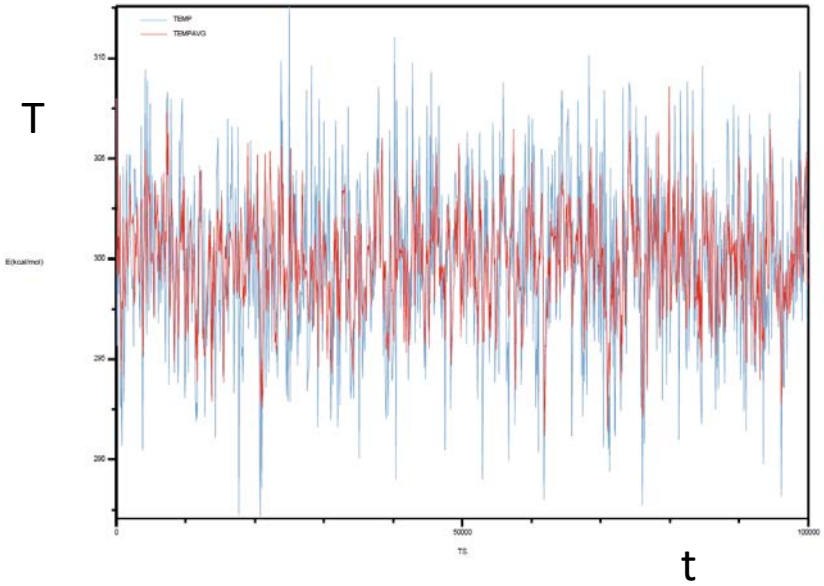
MD of Liquid Water

E



t

T



t

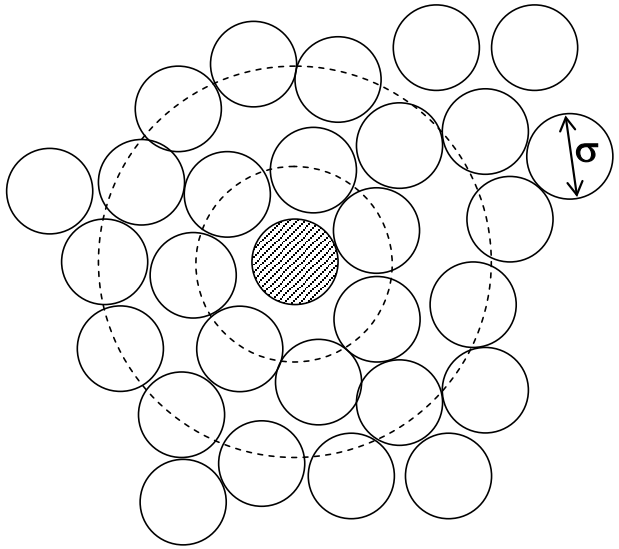
MD of Liquid Water

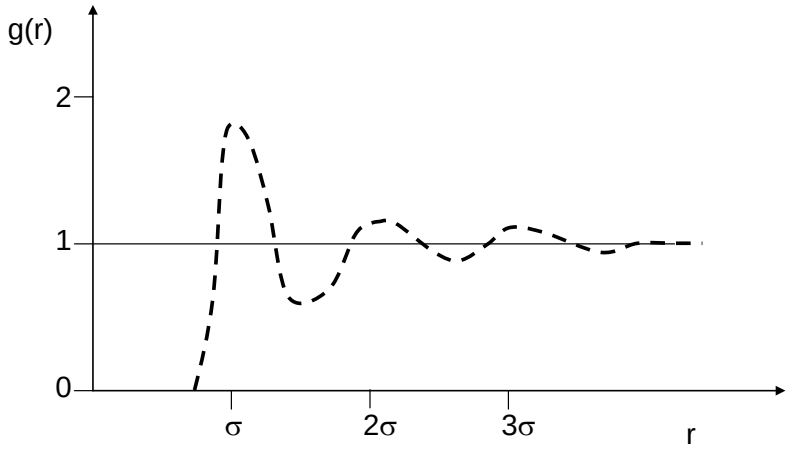
Radial distribution function

The radial distribution function g(r) allows us to study the structure of fluids, defined as a local normalized density. This is the density of molecules that are at a given distance r from one another, divided by the macroscopic density:

$$g(r) = \frac{\rho(0,r)}{\rho}$$

For a simple liquid made of spherical molecules, it looks like:





MD of Liquid Water

Radial distribution function

We are situated at a certain molecule, and we begin to move to increasing distances relative to that molecule:

- If we try to bring two molecules too close, the potential energy increases very fast and gives rise to a repulsive force between them. Therefore, the radial distribution function is equal to zero until we move to a distance $r \geq \sigma$
- At $r=\sigma$ we find the molecules that define the first solvation shell of our central molecule. The radial distribution function g(r) increases and the first peak appears.
- In turn, the molecules that make up this first solvation shell define a space that is inaccesible to others, therefore the radial distribution function decays when we move to greater distances.
- It then increases until reaching a distance of $r \approx 2\sigma$ where we would find the molecules that make up the second solvation shell.
- The disorder effect of the molecular movement increases as we move away from the central molecule (the first shell moves, the second shell also moves respect to the first one, and so on) and so the peak structure vanishes at longer distances.

MD of Liquid Water

Radial distribution function

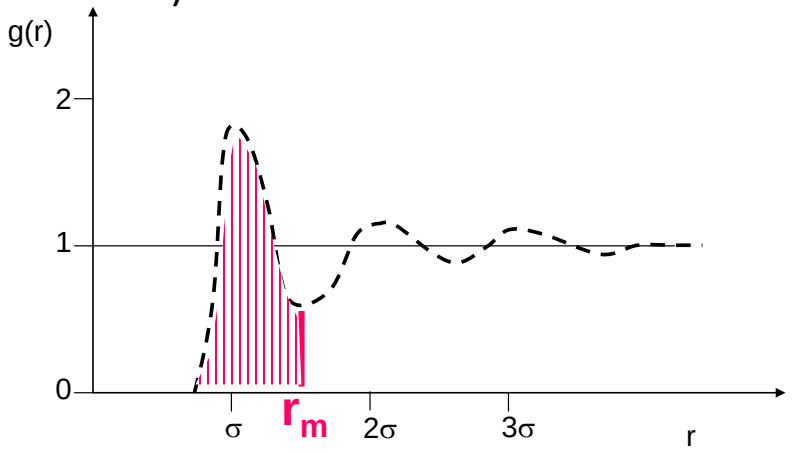
The radial distribution function is directly related to the probability of finding other molecules as a function of distance from the reference molecule

$$dp_r = \frac{dN_r}{N} = \frac{\rho(0,r)}{N} dV = \frac{\rho g(r)}{N} 4\pi r^2 dr$$

The coordination number can be obtained counting all the molecules in the first solvation shell (at distances smaller than the first minimum of the RDF)

$$N_C = \int_0^{r_m} dN_r = \int_0^{r_m} \rho(0,r) \cdot dV =$$

$$= \int_0^{r_m} \rho g\ (r) 4\pi r^2\, dr = 4\pi\rho \int_0^{r_m} g\ (r) r^2\, dr$$



MD of Liquid Water

Analysis of the results

3. Plot the radial distribution function between the water oxygen $g_{OO}(r)$ and compare it with that obtained experimentally (Chemical Physics 258 (2000) 121-137).
Obtain the coordination number of water

Extensions > Analysis > Radial Pair Distribution function

Selection > Selection1: name O Selection 2: name O
Frames > First: 1 Last: -1

MD of Liquid Water

Experimental RDF

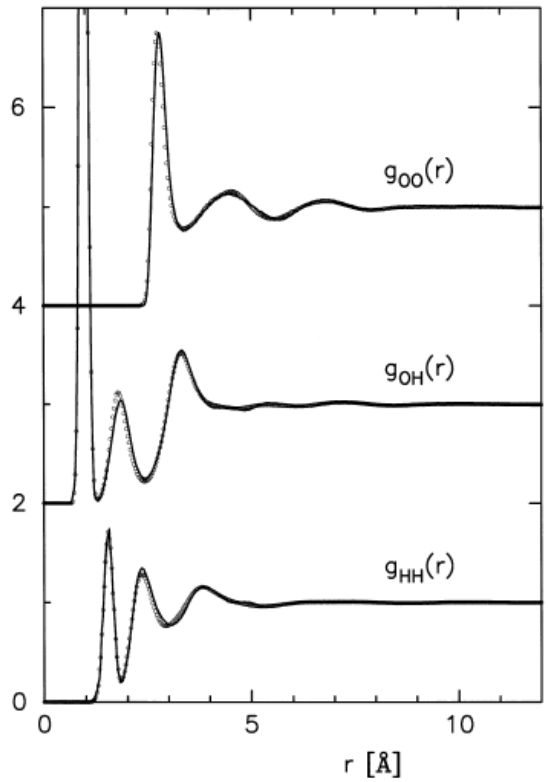
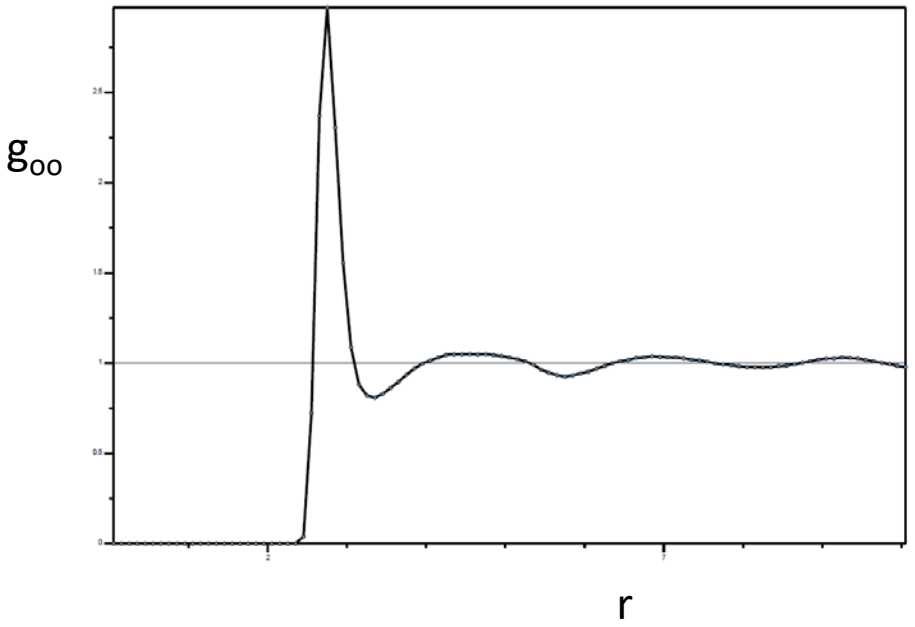


Fig. 6. Site–site radial distribution functions for water at 298 K, 1 bar, as derived from the two simulations used in Fig. 3. The simulations of the reactor data (Fig. 3(b)) are shown as dashed lines. It can be seen that although the systematic and truncation errors are different for the two datasets, the resulting radial distribution functions are closely similar.

MD RDF



MD of Na^+ in Liquid Water

Download the file "**Dinamica_sodio.tar"**
$tar –xvf Dinamica_sodio.tar

The folder contains the input files and a subfolder with output files corresponding to a MD simulation of a Na+ placed in a box of water molecules described by TIP3P. The dimensions of the box are 32.4009 x 32.3541x 32.4001 $Å^3$ and it contains 1136 water molecules.

MD is carried out using Langevin-Verlet (NVT) at 300K

Outputs for 100 ps (100000 steps with a timestep of 1 fs) are found in the folder /outputs

Exercise 2. Perform a 100 ps long MD simulation of Na-water at 300 K.

MD of Na^+ in Liquid Water

Analysis of the simulation

For the analysis we will use the facilities provided by the VMD program

$vmd

File > New Molecule

Select the file *.coor and in the option ‘file type’ select PDB.
Click on the name of the file in the box VMD Main

File > Load Data into Molecule
Select the file *.dcd

Graphics > Representations > Create Rep
Selections. Selected Atoms write resname ‘Na+’
Draw Style. Select ‘Drawing Method’ VDW

Visualize the trajectory

MD of Na^+ in Liquid Water

Analysis of the simulation

To visualize the first solvation shell

Graphics > Representations > Create Rep
Selections. 'Selected Atoms' write: same residue as within 4.0 of resname 'Na+'
& in Trajectory press
Draw Style. Select 'Drawing Method' CPK

1. Plot the total energy obtained during the simulation. Compare the fluctuations with the average value. If a simulation is stable, fluctuations must represent a small % of the total value of the property. Remember that according to Stat Mech fluctuations diminish with the number of atoms of the system

2. Plot the instantaneous and averaged values of temperature. Compare the fluctuations with the average value.

MD of Na^+ in Liquid Water

Analysis of the simulation

3. Plot the RDFs $g_{NaO}(r)$ y $g_{NaH}(r)$. Obtain the number of water molecules in the first solvation shell of the ion and compare it with experimental data.

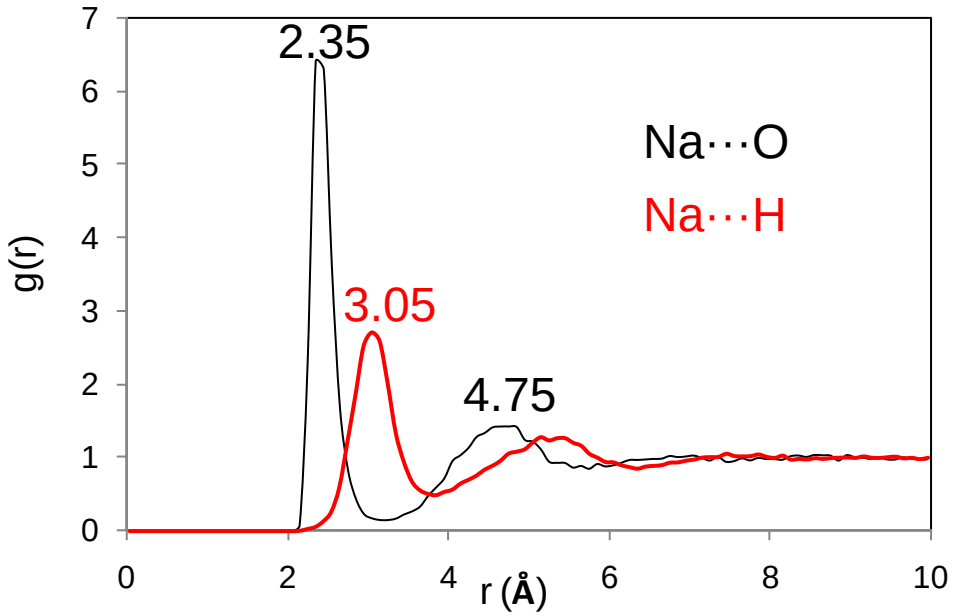
Extensions > Analysis > Radial Pair Distribution function

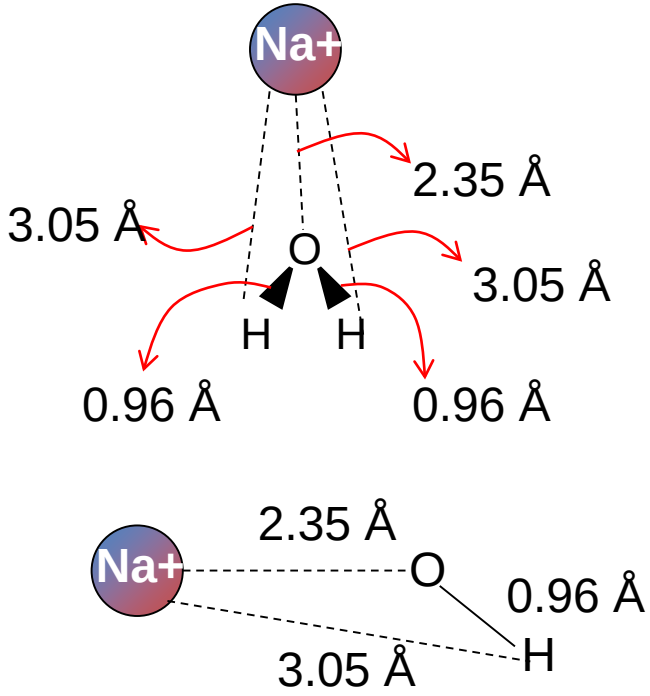
Selection > Selection1: name 'Na+' Selection 2: name O
Frames > First: 1 Last: -1

Selection > Selection1: name 'Na+' Selection 2: name H1 or name H2
Frames > First: 1 Last: -1

MD of Na^+ in Liquid Water

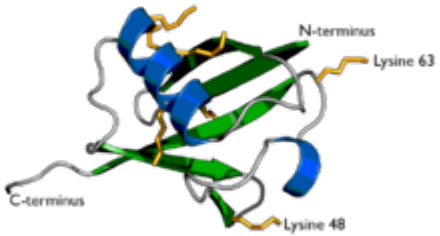
Radial Distribution Functions Na^+-O_w and Na^+-H_w





MD of a protein

Ubiquitin is a small protein (76 residues). Its functional role is to bind to other proteins in a post-translational modification. The addition of ubiquitin signals proteins for their degradation.



Download from Aula Virtual the file "**Dinamica_ubiquitin.tar**"

The folder contains the input and output files for an MD simulation of ubiquitin in water. The force field for the protein is CHARMM and TIP3P for water.

The MD is carried out in NVT ensemble at 300 K

The outputs for 100 ps (100000 steps of 1 fs) are found in the folder /outputs

```
#############################################################
## JOB DESCRIPTION                                         ##
#############################################################

# Ubiquitin in a Water Box

#############################################################
## ADJUSTABLE PARAMETERS                                   ##
#############################################################

structure          ubq_wb.psf
coordinates        initial.coor
velocities         initial.vel

firsttimestep      0

#############################################################
## SIMULATION PARAMETERS                                   ##
#############################################################

# Input
paraTypeCharmm          on
parameters         par_all27_prot_lipid.inp
```

Contains the topology of the protein

Initial coordinates and velocities

CHARMM force field parameters

```
# Force-Field Parameters
exclude           scaled1-4
1-4scaling         1.0
cutoff            12.0
switching          on
switchdist         10.0
pairlistdist       14.0


# Integrator Parameters
timestep           1.0  ;# 1fs/step


# Constant Temperature Control
langevin           on    ;# do langevin dynamics
langevinDamping     10    ;# damping coefficient (gamma) of 1/ps
langevintemp       300

# Periodic Boundary Conditions
cellBasisVector1    38.665    0.   0.0
cellBasisVector2     0.0   40.5062  0.0
cellBasisVector3     0.0    0   43.268
cellOrigin         31.0   29.0  17.5

wrapAll            on
```

```
# Output
outputName         ubq_output

restartfreq        500
dcdfreq            250
outputEnergies     100


#####################################################################
## EXTRA PARAMETERS                                ##
#####################################################################


#####################################################################
## EXECUTION SCRIPT                                ##
#####################################################################


run 100000 ;# 100ps
```

Exercise 3. Perform a 100 ps MD (100000 steps of1 fs) of ubiquitin in water at 300 K

Move to the folder outputs and run VMD

File > New Molecule Load the file psf
File > Load Data into Molecule Load the file dcd

Graphics > Create Rep In selected atoms write protein
Deactivate select all
Try different Drawing Methods (CPK, Tube, NewCartoon)
Try different Coloring Methods (Name, ResType, Secondary Structure,…)

1. Represent the total energy and compare the maximum fluctuation with the average value .

2. Represent the instantaneous and average temperatures and compare the maximum fluctuation with the average value.

3. Analyze the stability of the protein structure

The stability of the protein structure during the simulation can be monitored through the root of the mean squared displacement

$$RMSD(i,t) = \sqrt{r_i^2(t) - r_i^2(t=0)}$$

Extensions > Analysis > RMSD Trajectory Tool

Selection protein
Trajectory Frame ref **1** (0 is the first structure loaded, the pdb)
Select the option 'noh' (RMSD calculated for all non-hydrogen atoms)
Select the option Plot
Click on Align
Click on RMSD

The RMSD determined in this way takes into account the internal fluctuations of the protein atoms. In order to isolate the contributions of the internal fluctuations, one must first superimpose the structure analyzed and the original, to remove contributions from center of mass displacement and rotation around principal axes.

4. Color residues according to their mobility

We are going to color the residues of the protein according to the value of the RMSD

Extensions > Tk Console

We will calculate the RMSD of each residue. Using 'cd' move to the current directory where the routine is found:

% **source residue_rmsd.tcl**

We define a variable (sel_resid) to save the residue numbers [1, 2, ….76]

% **set sel_resid [[atomselect top "protein and alpha"] get resid]**

We calculate the rmsd of each residue and we assign the values to the list sel_resid

% **rmsd_residue_over_time top $sel_resid**

A file called residue_rmsd.dat will have been generated that can be opened with any program (excel), and contains the rmsd versus the residue number

Graphics > Representations
Selected atoms protein

Draw Method → NewCartoon or Tube
Coloring Method → trajectory > user > user

In Trajectory

Color scale data range 0.4 and 1.00. Click set

Open the file residue_rmsd.dat in excel
Plot de rmsd versus residue number
Analyze the mobility of each residue considering the different motifs of the structure (alpha helix, beta sheets, loops,….)