République Algérienne Démocratique et Populaire

Ministère de l'Enseignement supérieur et de la Recherche scientifique

Université de Médéa

Faculté des Sciences

Polycopié de cours : Handout : Materials Modeling

Selon le programme pédagogique officiel du ..... (L3, Spécialité : Chimie des Matériaux, Filière : Chimie, Domaine : Sciences de la Matière)

Dr. BITAM Said

Maître de conférences A

Département : Sciences de la Matière

Validé par le conseil scientifique de la faculté des sciences à **26 Octobre 2025.**

Ce polycopié a été expertisé par :

1. Dr. BOUKAOUD Abdelali, ... (MCA), ... (Université de Médéa).
2. Dr. TANDJAOUI Nassima, ... (MCA), ... (Université de Médéa).

Contents

Preface 4

Chapter I: 5

I. Introduction: 6

I.1. Schrödinger Equation: 6

I.1.1. Born-Oppenheimer Approximation: 8

I.1.2. Hartree-Fock Theory: 8

I.1.3. Slater Determinants: 9

I.1.4. Variational Principle: 9

I.1.5. Correlation: 9

I.2. Ab Initio 10

I.2.1. Hartree-Fock-Roothaan Equations (HFR) 10

I.2.2. The Classical Form of the Hartree-Fock Equations: 12

I.2.3. Roothaan Equations 12

I.3. Slater Orbitals and Gaussian Orbitals 14

I.3.1. Slater Functions 14

I.3.3. Gaussian Functions 14

I.3.4. Classification of the Basis Sets 18

1.4. Semi-Empirical Methods: 20

I.5. Density Functional Theory (DFT): 21

I.6. Molecular Mechanics 22

I.6.1. Force Field in Molecular Mechanics 23

I.6.2. Interaction Energy Between Bonded Atoms 24

I.6.3. Commonly used Force Fields 26

I.7. Molecular Dynamics (MD): 28

I.8. Potential Energy Surface 29

I.8.1. Stationary Points 29

I.8.2. Stationary Points and vibrations 30

I.8.3. Transition State: 31

I.8.4. Geometry Optimization 31

I.8.5. Thermochemistry from ab initio MO methods 32

I.8.6. Some Basic Physical Constants 37

Chapter II: 38

II. Introduction: 39

II.I. HYPERCHEM SOFTWARE 39

II.1.1. BASIC CONCEPTS ON HYPERCHEM: 39
II.2. GaussView: 41
II.3. Avogadro: 42
II.3.1. The Avogadro Interface 43
II.3.2. Tools 43
II.4. ChemBio 3D 45
II.4.1. ChemBio3D Interface 45
II.4.1. Tools Palette 46
II.5. MoCalc2012 48
II.6. MolCalc: https://molcalc.org 50
Chapter III: 52
III. Introduction 53
III.1. HyperChem: 53
III.2. Gaussian: 53
III.2.1. Spin Multiplicity: 56
III.2.2. MOLECULAR ORBITALS AND POPULATION ANALYSIS 56
III.3. GAMESS: 58
III.3.1. Input Description: 59
III.4. MOPAC: 61
III.4.1. Some keywords used in MOPAC 61
III.4.2. Example of data for ethylene 62
Chapter IV: 64
Practical Exercise 01: Geometric Optimization 65
Practical Exercise 02: Molecular Mechanics 68
Practical Exercise 03: Molecular Dynamics (MD) 73
Practical Exercise 04: Analyzing the Structural Properties of Cyclohexane Conformation 77
Practical Exercise 05: Analyzing Molecular Orbitals 79
Practical Exercise 06: Vibrational Analysis 81
Practical Exercise 07: NMR Chemical Shift Calculations 83
Practical Exercise 08: Predicting Melting Point 84
Practical Exercise 09: Estimating Viscosity of a Small Organic Molecule 85
Practical Exercise 10: Reaction Coordinate Calculation 86
Practical Exercise 11: Studying the Effect of External Electric and Magnetic Fields on a Molecule Using Hyperchem and Gaussian 89
Practical Exercise 12: The Effect of water on methanol Using Avogadro and GAMESS 92

Practical Exercise 13: Calculating the thermodynamics properties of molecules 94
Practical Exercise 14: Predicting the Refractive Index using MOPAC 96
Step 1: Calculate Molecular Polarizability 97
Step 2: Use the Lorentz-Lorenz Equation 97
Step 3: Calculate Number Density 97
Step 4: Combine Values 98
Example Calculation 98
Practical Exercise 15: Estimating Elastic Modulus of a Polymer 100
REFERENCES: 102

Preface

This handout is designed to introduce third-year bachelor's degree students in materials chemistry to the fundamental concepts of computational chemistry. As the field of materials science increasingly relies on computational tools to predict and understand material properties, a solid grasp of these techniques is essential for aspiring chemists.

The content is divided into three main sections. The first section covers the basic principles of computational chemistry, providing a foundational understanding of how computational methods can be applied to solve complex chemical problems. The second section offers a detailed description of the most widely used graphical interface software in the field. Understanding these tools is crucial for practical applications, as they enable users to visualize molecular structures, perform simulations, and analyze results efficiently. The final section focuses on practical applications, where theoretical knowledge is put into practice through exercises and real-world examples.

This handout aims to equip students with the knowledge and skills necessary to utilize computational chemistry tools effectively, preparing them for more advanced studies and professional work in materials chemistry. I hope that this guide will not only serve as a valuable educational resource but also inspire further exploration into the fascinating world of computational chemistry.

Chapter I:

Theoretical Background

I. Introduction:

Theoretical chemistry (or computational chemistry) is a modern branch of chemistry. Its first use was by Heitler and London in 1927 to calculate the bond in the hydrogen molecule. However, its practical application only grew and accelerated with the advent of computers.

The importance of theoretical chemistry lies in several factors. It is a safe and easy method, requiring only a computer, and is relatively inexpensive. Through computational chemistry, one can predict or estimate the course of chemical reactions before carrying them out. Additionally, it allows for the understanding of the nature of the substances involved in a reaction and their stability, as well as predicting the mechanism of the reaction.

Unfortunately, computational chemistry is not entirely complete, as there is no equation or function that can fully describe the position and state of an electron. Therefore, the field of computational chemistry is entirely based on approximate theories, some of which are better than practical methods, while others are quite poor.

Historically, classical physics helped scientists understand many everyday phenomena, but it was inadequate at very high speeds (the speed of light), in extremely massive objects (black holes), or in very small particles (atoms, molecules, or electrons). Hence, classical physics was replaced by quantum physics, which allows for the description, measurement, and even prediction of the behavior of a wide range of physical systems, from tiny particles to nuclei and atoms to extremely massive objects.

Quantum physics mathematically describes the state of electrons, but once again, there is no ideal equation or perfect solution for any system except for the hydrogen atom.

I.1. Schrödinger Equation:

The famous Schrödinger equation is considered the fundamental building block of quantum physics and, consequently, theoretical and computational chemistry. It was formulated by Erwin Schrödinger in 1925 and published in 1926. It provided the first description of the electron as a wave function.

$$\hat{H}\Psi(r) = E\Psi(r)$$

Where:

- $\hat{H}$: The Hamiltonian operator.
- Ψ: The wave function.
- E: The energy, eigenvalue of the Hamiltonian operator.

The Hamiltonian operator $\hat{H}$ can be expressed as the sum of the kinetic and potential energies of the nuclei and electrons as follows:

$$\hat{H} = \hat{T}_n + \hat{T}_e + \hat{V}_{en} + \hat{V}_{ee} + \hat{V}_{nn}$$

Where:

- $\hat{T}_n$, $\hat{T}_e$: Operators representing the kinetic energies of the nuclei ($\hat{T}_n$) and electrons ($\hat{T}_e$).
- $\hat{V}en + \hat{V}ee + \hat{V}nn$: Represent the interactions between the nuclei and electrons ($\hat{V}_{en}$), the electron-electron Coulomb interaction ($\hat{V}_{ee}$), and the nucleus-nucleus repulsion term ($\hat{V}_{nn}$).

It can be expressed mathematically as follows:

$$\hat{H} = -\frac{1}{2}\sum_{i}^{electrons}\nabla_i^2 - \frac{1}{2}\sum_{A}^{nuclei}\frac{1}{M_A}\nabla_A^2 - \sum_{i}^{electrons}\sum_{A}^{nuclei}\frac{Z_A}{r_{iA}} + \sum_{i<}^{electrons}\sum_{j}\frac{1}{r_{ij}} + \sum_{A<}^{nuclei}\sum_{B}\frac{Z_A Z_B}{R_{AB}}$$

Where:

- Z: The nuclear charge.
- M_A: The ratio of the mass of nucleus A to the mass of an electron.
- R_{AB}: The distance between nuclei A and B.
- r_{ij}: The distance between electrons *i* and *j*.
- r_{iA}: The distance between electron iii and nucleus A.

Unfortunately, the Schrödinger equation cannot be solved exactly for multi-electron systems, even for the simplest systems like the helium atom or the hydrogen molecule. Therefore, it was necessary to develop some approximate theories to achieve practical and solvable theories.

I.1.1. Born-Oppenheimer Approximation:

One way to simplify the Schrödinger equation is to assume that the nucleus is stationary. Of course, it does move, but its speed is negligible compared to the speed of the electrons, which move at near-light speeds. This approximation is known as the Born-Oppenheimer approximation, and it leads us to the electronic Schrödinger equation:

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

where the electronic Hamiltonian is given by:

$$\hat{H}_{el} = -\frac{1}{2}\sum_{i}^{electrons}\nabla_i^2 - \sum_{i}^{electrons}\sum_{A}^{nuclei}\frac{Z_A}{r_{iA}} + \sum_{i<}^{electrons}\sum_{j}\frac{1}{r_{ij}}$$

I.1.2. Hartree-Fock Theory:

Hartree-Fock theory is one of the simplest theories based on wavefunctions, where the primary approximation is the central field approximation. This means that the electron-electron repulsion is calculated based on the average interaction of all electrons, which gives an overall measure of electron-electron repulsion without considering the mutual effects between electrons.

This type of calculation is known as variational calculations, meaning the approximate energy will be equal to or greater than the true energy. Thus, the energy calculated by Hartree-Fock is always greater than or equal to the true energy.

One feature of this theory is that it divides the multi-electron Schrödinger equation into several simpler single-electron equations. Solving each single-electron equation yields a wavefunction called an orbital, and the corresponding energy is known as the orbital energy.

The unit of calculated energy is the Hartree (1 Hartree = 27.2116 electron volts).

I.1.3. Slater Determinants:

According to the Pauli exclusion principle [9], no two electrons can have the same set of quantum numbers. Since electrons are fermions, the exchange of electrons is represented by changing the sign of the wavefunction Ψ:

$$\Psi(\overrightarrow{x_1}, \overrightarrow{x_2}) = -\Psi(\overrightarrow{x_2}, \overrightarrow{x_1})$$

Considering both spin and position:

$$\overrightarrow{x_i} = (\overrightarrow{r_i}, r_i)$$

Thus, a Slater determinant ϕ_{SD} can be used to express the appropriate wavefunction for the ground state (Ψ_0) of multi-electron systems with *N* electrons and spin orbitals $\chi_1, \chi_2, \ldots, \chi_N$ to satisfy the antisymmetry principle:

$$\Psi_0 \approx \phi_{SD} = \frac{1}{N!}\begin{vmatrix} \chi_1(\overrightarrow{x_1}) & \chi_2(\overrightarrow{x_1}) & \cdots & \chi_N(\overrightarrow{x_1}) \\ \chi_1(\overrightarrow{x_2}) & \chi_2(\overrightarrow{x_2}) & \cdots & \chi_N(\overrightarrow{x_2}) \\ \vdots & \vdots & & \vdots \\ \chi_1(\overrightarrow{x_N}) & \chi_2(\overrightarrow{x_N}) & & \chi_N(\overrightarrow{x_N}) \end{vmatrix}$$

The functions $\chi_i(\overrightarrow{x_i})$ are single-electron wavefunctions known as spin orbitals. Each spin orbital consists of the spatial part of the orbital and one of two spin functions, either α(s) or β(s), as shown in the following equation:

$$\chi(\vec{x}) = \phi(\vec{r})\sigma(s) \;\; ; \quad \sigma = \alpha, \beta$$

I.1.4. Variational Principle:

Since the ideal wavefunction is unknown, an approximate wavefunction (Ψ_{approx}) is used. Consequently, the energy derived from this wavefunction is the approximate energy (E_{approx}) rather than the true energy (E_{Exact}). According to this principle, the calculated energy is greater than or equal to the true energy.

I.1.5. Correlation:

When using Hartree-Fock theory, the probability of finding an electron around the nucleus is determined by its distance from the nucleus, without considering its distance from other

electrons. This results from the central field approximation, which is the foundation of this theory.

The importance of including correlation lies in increasing the accuracy of energy calculations and the molecular spatial geometry.

I.2. Ab Initio

The term "Ab Initio" is a Latin phrase that means "from the beginning". This is the foundation of this group of methods, as they rely on the Schrödinger equation, meaning they are based entirely on quantum mechanical theories and their solutions from scratch.

It is well known that the Schrödinger equation is only solvable exactly for single-electron systems, such as hydrogen or hydrogen-like atoms. Therefore, some mathematical approximations must be introduced to solve this equation. One of the simplest of these approximations is the **Hartree-Fock method**.

I.2.1. Hartree-Fock-Roothaan Equations (HFR)

We assume the possible definition of single-electron functions ϕ_k, called molecular orbitals (MO) in the case of a molecule, or atomic orbitals (AO) in the case of an atom. The many-electron wavefunction is taken in the form of an antisymmetrized product (Slater determinant) of spin orbitals. These are the products of a spatial function ϕ_k and a spin function α or β, i.e., $\phi_k\alpha$ and $\phi_k\beta$. This form chosen for ϕ respects the Pauli principle; it vanishes when two electrons with the same spin occupy the same point in space. For a closed-shell system (non-radical), denoting ϕ_k and $\bar{\phi}_k$ as the two spin orbitals $\phi_k\alpha$ and $\phi_k\beta$, and using the simplified notation $\phi_k(\vec{r_1}) = \phi_k(1)$, the many-electron wavefunction ϕ is written as:

$$\Phi = \frac{1}{\mathrm{N!}}\begin{vmatrix} \phi_1(1)\bar{\phi}_1(1) & \phi_2(1)\bar{\phi}_2(1) & \dots\dots & \phi_{\mathrm{N}/2}(1)\bar{\phi}_{N/2}(1) \\ \phi_1(2)\bar{\phi}_1(2) & \phi_2(2)\bar{\phi}_2(2) & \dots\dots & \phi_{\mathrm{N}/2}(2)\bar{\phi}_{N/2}(2) \\ \vdots & \vdots & & \vdots \\ \phi_1(N)\bar{\phi}_1(N) & \phi_2(N)\bar{\phi}_2(N) & & \phi_{\mathrm{N}/2}(1)\bar{\phi}_{N/2}(1) \end{vmatrix}$$

or, in a simplified form:

$$\Phi = \left|\phi_1\bar{\phi}_1\phi_2\bar{\phi}_2\phi_3\bar{\phi}_3 \dots \phi_{\mathrm{N}/2}\bar{\phi}_{N/2}\right|$$

If we denote h (core operator) as the single-electron operator describing an electron μ evolving in the field of the nuclei:

$$h(\mu) = -\frac{1}{2}\Delta_\mu - \sum_{i=1}^{M}\frac{Z_i}{r_{ij}}$$

$$H_e(\vec{r},\vec{R}) = \sum_{\mu=1}^{N} h(\mu) + \sum_{\mu<\nu}\frac{1}{r_{\mu\nu}}$$

For the following, H_e will be denoted as H. The energy of the system is approximated by the integral (in Dirac notation) $E = \langle \Phi \mid H \mid \Phi \rangle E$, where Φ is normalized. The expression for the energy in terms of the molecular orbitals is then

$$E = \sum_{k=1}^{n/2}\left[2h_{kk} + \sum_{l=1}^{n/2}(2J_{kl} - K_{kl})\right]$$

h_{kk}, J_{kl}, K_{kl} being respectively the core, Coulomb, and exchange molecular integrals, defined as:

$$h_{kk} = \langle \Phi_k | h | \Phi_k \rangle$$

$$J_{kl} = \langle \Phi_k | J_l | \Phi_k \rangle$$

$$K_{kl} = \langle \Phi_k | K_l | \Phi_k \rangle$$

J_l and K_l represent the Coulomb and exchange operators defined by the integrals:

$$J_l(1) = \int |\Phi_l(2)|^2 \frac{1}{r_{12}} dr$$

$$K_l(1) = \int \Phi_l(2)\frac{1}{r_{12}} d\tau_2 P_{12}\Phi_l(2)$$

where P_{12} is the permutation operator for electrons 1 and 2.

The Hartree-Fock theory does not forbid two electrons with opposite spins from being located at the same point in space. The Coulomb operator $J_l(1)$ represents the classical energy of the repulsion exerted by electron 2, described by the molecular orbital Φ_l, on electron 1. The exchange operator $K_l(1)$ has no classical equivalent and appears in the theory due to the consideration of the Pauli principle; when applied to the function $\Phi_k(1)$, this operator permutes electrons 1 and 2 between the functions Φ_k and Φ_l

I.2.2. The Classical Form of the Hartree-Fock Equations:

$$F\Phi_k = \varepsilon_k \Phi_k$$

where

$$F = h + \sum_l (2J_l - K_l)$$

The solution of the Hartree-Fock equations is performed using an iterative process called the SCF (**Self-Consistent Field**) method.

I.2.3. Roothaan Equations

In practice, the Φ_k are taken as a linear combination (LC) of a finite number *n* of atomic orbitals (AO) χ_r (CLOA approximation):

$$\Phi_k = \sum_r^n c_{rk}\chi_r$$

Under these conditions, the equations take the following matrix form, constituting the Roothaan equations:

$$FC_k = \varepsilon_k S C_k$$

where *F* and *S* represent the Fock matrix and the overlap integrals, respectively, with elements given by:

$$S_{rs} = \langle \chi_r | \chi_s \rangle$$

$$F_{rs} = h_{rs} + \sum_{p}^{n}\sum_{q}^{n} P_{pq}\left[\langle rs|pq\rangle - \frac{1}{2}\langle rq|ps\rangle\right]$$

$$h_{rs} = \langle \chi_r|h|\chi_s\rangle$$

$$P_{pq} = 2\sum_{k}^{N/2} c_{pk}c_{qk}$$

$$\langle rs|pq\rangle = \langle \chi_r(1)\chi_p(2)\frac{1}{r_{12}}\chi_s(1)\chi_q(2)\rangle$$

The solution of the Roothaan equations, in terms of eigenvalues, leads to n $\Phi_k$and n ε_k, with the matrices *F* and *S* being of dimension *n*.

For a closed-shell system, the Hartree-Fock method used is said to be spin-restricted and denoted RHF (**Restricted Hartree-Fock**).

For open-shell systems (radicals and biradicals), the UHF (**Unrestricted Hartree-Fock**) formalism can be used, which describes α and β spin electrons using different spatial functions. Such systems can also be studied using the ROHF (**Restricted Open-shell Hartree-Fock**) method. In this case, there is a double occupation of levels, where the two electrons of a doubly occupied level are described using the same spatial function, and the single electrons are described by different molecular orbitals. The principles of these different methods are schematically illustrated in Figure 1.

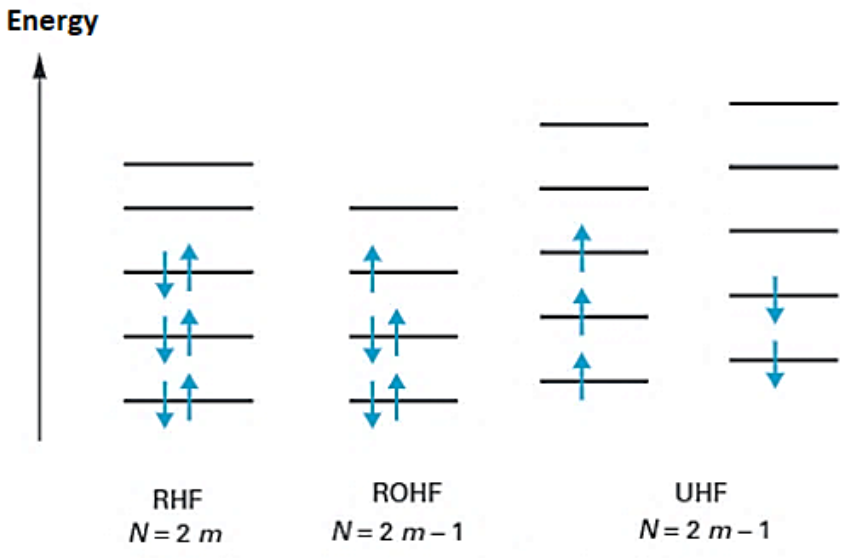


Figure 1: Representation of the different Hartree-Fock methods

I.3. Slater Orbitals and Gaussian Orbitals

Two types of atomic functions can be used to express molecular orbitals: Slater functions, more commonly called STO (Slater Type Orbitals), and Gaussian functions, known as GTO (Gaussian Type Orbitals).

I.3.1. Slater Functions

The general expression of a Slater function is given by the following relation:

$$\chi(r_A, \theta_A, \varphi_A, n, l, m, \zeta) = \frac{(2\zeta)^{n+\frac{1}{2}}}{(2n!)^{\frac{1}{2}}} r_A^{n-1} e^{-\zeta r_A} Y_l^m(\theta_A, \varphi_A)$$

where ζ is the Slater exponent; it is a positive number that depends on the atomic number and can be determined using Slater's empirical rules or by optimization. n, l, m are the principal, azimuthal, and magnetic quantum numbers, respectively. $r_A, \theta_A, and\ \varphi_A$ are the spherical coordinates that locate the electron relative to the center *A* carrying the Slater function, and $Y_l^m(\theta_A, \varphi_A)$ represents a spherical harmonic. In Cartesian coordinates, the above expression is written as:

$$\chi\ (x_A, y_A, z_A, n, i, j, k, \zeta)\ =\ N\ (\zeta, n, i, j, k)\ r_A^{n-1-i-j-k} x_A^i\ y_A^j z_A^k e^{-\zeta r_A}$$

where $N\ (\zeta, n, i, j, k)$ is a normalization coefficient, *i*, *j*, and *k* are non-negative integers such that their sum $l\ =\ i\ +\ j\ +\ k$ defines the symmetry of the Slater function (e.g., $l\ =\ 0$ gives an s-type Slater function, $l = 1$ gives a p-type Slater function, etc.); the coordinates $x_A, y_A\ \ and\ z_A$ of the electron (x, y, z) relative to the center $A\ (A_x, A_y\ , A_z)$, are given by: $x_A = x - A_x\ ;\ y_A = y - A_y;\ z_A = z - A_z$.

Such functions, which are very suitable for atomic calculations due to their good behavior at both very short and very long distances from the nucleus, remain less used in the study of molecular systems due to the complexity of calculating the two-electron integrals $\langle rs|pq\rangle$ that they generate.

I.3.3. Gaussian Functions

Boys was the first to propose the use of Gaussian functions in quantum calculations. Their general normalized form is:

$$g\ (x_A, y_A, z_A, n, i, j, k, \alpha) = \left(\frac{2\alpha}{\pi}\right)^{3/4} \left[\frac{(8\alpha)^{i+j+k}\, i!\, j!\, k!}{(2i)!\,(2j)!\,(2k)!}\right]^{\frac{1}{2}} x_A^i\, y_A^j z_A^k e^{-\alpha(x_A^i + y_{A+}^j z_A^k)}$$

where α is a positive number representing the Gaussian exponent, and the parameters $i, j, k, x_A, y_A\ and\ z_A$ have the same meaning as in the case of a Slater function.

These functions have a key property: the product of two Gaussians $\boldsymbol{g_A}\ (\boldsymbol{\alpha_A})$ and $\boldsymbol{g_B}\ (\boldsymbol{\alpha_B})$ centered on two different atoms A and B is also a Gaussian $\boldsymbol{g_C}\ (\boldsymbol{\alpha_C})$, located between the centers A and B, with an exponent given by:

$$\alpha_C = \alpha_A + \alpha_B$$

Thus, four-center two-electron integrals can be reduced to two-center two-electron integrals, making the calculations simpler and faster. However, unlike Slater functions, these Gaussian functions do not have satisfactory physical behavior very close to the nucleus (Figure 2), as their derivative is zero at this point, nor at long distances because they decay too quickly.

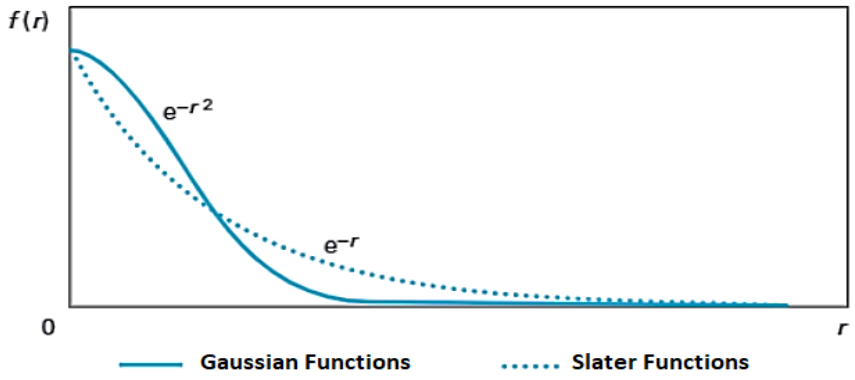


Figure 2. Representation of a Slater function and a Gaussian function

Huzinaga had the idea to combine the strengths of different types of functions by suggesting that a Slater function be considered as a linear combination of Gaussians. This approach, on the one hand, facilitates the integral calculation and, on the other hand, retains a good behavior of the function near the nucleus. We can write:

$$\chi\,(n,l,m,\zeta) = \sum_{i=1}^{L} c_i g_i(\alpha_i)$$

where L is the number of Gaussian functions g_i with exponent α_i involved in the linear combination. When a basis function is a linear combination of Gaussians, it is called a **contracted** basis function, with the Gaussians that comprise it being referred to **as primitives.** In a contracted Gaussian basis set, each basis function is characterized by the coefficients and exponents of its primitive Gaussians.

Pople and his collaborators were the first to establish sets of coefficients and exponents of Gaussians describing Slater functions for many atoms in the periodic table. In such bases, the structure of the shell model is preserved, meaning that the exponents of s-type and p-type functions are identical.

They developed different types of basis sets:

- **Single zeta (SZ) or minimal basis sets**, in which each atomic orbital of a given symmetry is represented by a Slater function, a linear combination of N Gaussians with the same symmetry. These are denoted as STO-NG, where N is an integer. For example, in an STO-3G basis, each function is a linear combination of three Gaussians (the notation (3s)/[1s], which is equivalent, means that a contracted s-type basis function is a linear combination of three s-type primitive Gaussians).
- **Double, triple, or quadruple zeta bases**, where each atomic orbital is represented by two, three, or four linear combinations of Gaussians, respectively.
- **Split-valence double zeta (SVDZ), triple zeta (SVTZ), or quadruple zeta (SVQZ) bases**, where core orbitals are represented by one function and valence orbitals by two, three, or four functions. As an example, Table 1 presents a 4-31G basis used for the carbon atom.

Tableau 1. 4-31G basis sets of carbon

α_{1s}	C_{1s}	α_{sp}	C_{2s}	C_{2p}
486,9669	0,017726	8,673525	-0,121384	0,063545
73,3711	0, 123478	2,096619	- 0,227339	0,298268
16,4135	0,433875	0,604651	1,18174	0,762103
4,3450	0,561504	0,696709	1	1

This table indicates that, in the 4-31G basis, the functions describing the atomic orbitals of carbon are expressed as follows:

$\chi_{1s} = 0{,}017726\, e^{-486{,}9669r^2} + 0{,}123478\, e^{-73{,}3711r^2} + 0{,}433875\, e^{-16{,}4135r^2} + 0{,}5615046\, e^{-4{,}3450r^2}$

$\chi_{2s} = -0{,}121384\, e^{-8{,}673525r^2} - 0{,}227339\, e^{-2{,}096619r^2} + 1{,}18174 e^{-0{,}604651r^2}$

$\chi'_{2s} = e^{-0{,}696709r^2}$

$\chi_{2p_x} = 0{,}063545 x e^{-8{,}673525r^2} + 0{,}298268 x e^{-2{,}096619r^2} + 0{,}762103 x e^{-0{,}604651r^2}$

$\chi'_{2p_x} = x e^{-0{,}696709r^2}$

$\chi_{2p_y} = 0{,}063545 y e^{-8{,}673525r^2} + 0{,}298268 y e^{-2{,}096619r^2} + 0{,}762103 y e^{-0{,}604651r^2}$

$\chi'_{2p_y} = y e^{-0{,}696709r^2}$

$\chi_{2p_z} = 0{,}063545 z e^{-8{,}673525r^2} + 0{,}298268 z e^{-2{,}096619r^2} + 0{,}762103 z e^{-0{,}604651r^2}$

$\chi'_{2p_z} = z e^{-0{,}696709r^2}$

- **STO-3G:** Uses a total of four STO functions (one for 1s, three for 2s and 2p).
- **3-21G:** Uses a total of four Gaussian functions (one for 1s, three for 2s and 2p).
- **6-31G:** Uses a total of five Gaussian functions (one for 1s, three for 2s and 2p, one diffuse Gaussian function).

- *6-31G:** Uses a total of ten Gaussian functions (one for 1s, four for 2s and 2p, one diffuse Gaussian function, and four additional p-polarization functions).
- **6-31G**: Uses a total of thirteen Gaussian functions (one for 1s, four for 2s and 2p, one diffuse Gaussian function, four additional p-polarization functions, and three additional d-polarization functions).

The expression for a diffuse Gaussian function is as follows:

$$f(x) = A\, exp(-\alpha(x - x_0)^2/2)$$

Where:

- A: is the amplitude of the function
- α: is the diffusion parameter (smaller than for a normal Gaussian)
- x_0 is the center of the function

The expression for a polarization Gaussian function is:

$$f(x) = A\, exp(-\alpha(x - x_0)^2/2)\, cos(\theta)$$

Where:

- A: is the amplitude of the function
- α: is the diffusion parameter
- x_0 is the center of the function
- θ: is the polarization angle

I.3.4. Classification of the Basis Sets

The molecular wavefunction is approximated as a linear combination of atomic orbitals (LCAO), and through this approximation, molecular orbitals can be synthesized by combining simpler functions. This set of functions is called the basis sets. Therefore, the basis set is a set of functions used to express the shape of the molecular orbitals through the linear combination of these functions.

We can find the electronic distribution around the nucleus using several methods, such as using hydrogenic wavefunctions that depend on the solutions of the Schrödinger equation for the hydrogen atom, or using adjustable polynomial functions, as well as Slater functions and Gaussian functions.

- The orbitals of the Slater type are called STOs.
- The orbitals of the Gaussian type are called GTOs.

Both Gaussian functions and Slater functions are mathematically simple, and it can be said that Slater-type orbitals have good results compared to Gaussian-type orbitals, but on the other hand, Gaussian-type orbitals have a simpler and faster solution compared to Slater-type orbitals. As a solution to this problem, Slater-type orbitals (with good results) can be expressed by approximating them using Gaussian-type orbitals (with simpler and faster solutions). For example, the STO-3G basis set is a Slater-type function approximated using three Gaussian-type orbitals. Here are some classifications:

I.3.4.1. Minimal Basis Sets

These basis sets are used to take an initial look at the property or properties of a molecule. The weakness of this method is that it classifies the electrons as being of the same importance, and the general form of these basis sets is STO-nG.

STO-3G is an example of this type of basis set, where a Slater-type orbital is approximated using three Gaussian-type orbitals.

I.3.4.2. Split-Valence Basis Sets

The nG basis sets are examples of split-valence basis sets. Here, Gaussian-type orbitals are used to approximate Slater-type orbitals, and they have two sets of numbers, e.g. "FirstNumber-SecondNumberG".

The first part, where the number is 3, means that the Slater-type orbitals will be approximated using three Gaussian-type orbitals.

The second part, 21, does not mean 21 times. It represents a "Double Zeta" basis set, meaning the calculation is done in two steps. The first step approximates using two Gaussian-type orbitals, and the second step uses one Gaussian-type orbital. So, each electron in the outer shell is calculated twice, first using STO-2G and then using STO-1G for the same electron.

One important example of a triple zeta basis set is the 6-311G basis set, and its extension is the 6-311G++ basis set.

I.3.4.3. Diffuse Functions Basis Sets:

In general, the probability of finding an electron near the nucleus is very high, and this probability decreases as we move away from the nucleus. At large distances, we may not find any electrons. However, in some systems, especially for radicals and atoms in an excited state, we need to use the diffuse function to ensure an increased distance for calculating the presence of the electron. The symbols used to indicate the addition of the diffuse function are:

- + for heavy atoms except hydrogen.
- ++ for heavy atoms and hydrogen.

Examples:

6-31+G is a split-valence basis set for heavy atoms.

6-31++G is a split-valence basis set for heavy atoms and hydrogen.

I.3.4.4. Polarization Functions Basis Sets:

In some cases, the electron may be polarized (distorted) from its orbital to another orbital, and in this case, we need to take into account the polarization property. The polarization function provides a more accurate explanation of the location of the electron or the location where it may exist. The following symbols are added to indicate polarization:

for all atoms except hydrogen

** for all atoms including hydrogen

Or we can simply add the orbital name d or d,p

Examples:

(6-31G)* or (6-31G(d)) is a split-valence basis set with polarization for all atoms except hydrogen.

(6-31G)** or (6-31G(d,p)) is a split-valence basis set with polarization for all atoms including hydrogen.

6-311G++(d,p) Basis set.

1.4. Semi-Empirical Methods:

This type of calculations used in computational chemistry is somewhat similar to the Ab Initio methods, as it is based on the Schrödinger equation. However, the difference is that this type of calculations replaces the complex integrals used in the Ab Initio methods with empirical

parameters and integrals. These empirical integrals are obtained from a library (database) of integrals that have been shown to produce good results and high accuracy, often taken from high-level theoretical calculations such as Density Functional Theory (DFT). In other words, the Schrödinger equation is solved using empirical parameters, and this is the main difference between the different types of this group of calculations, which vary in the database of parameters and empirical integrals they use.

Relatively, these methods are faster than the Ab Initio methods, but slower than the Molecular Mechanics methods.

Some examples of these methods are:

1. CNDO: Complete Neglect of Differential Overlap.
2. INDO: Intermediate Neglect of Differential Overlap.
3. MNDO: Modified Neglect of Differential Overlap.
4. AM1: Austin Model 1.
5. PM3: Parametric Model number 3.

I.5. Density Functional Theory (DFT):

This method is also similar to the Ab Initio and semi-empirical methods in terms of its reliance on the Schrödinger equation. However, the main difference lies in the wave function. Instead of using the wave function, it is replaced by the electron distribution directly (i.e., the electron density function), and the functional is a function that produces another function, i.e., the function of the function. Consequently, the energy is considered a functional of the density.

The ground-state energy, $E[\rho(\vec{r})]$, of a system can be expressed using the following equation:

$$E[\rho(\vec{r})] = T_s[\rho(\vec{r})] + J[\rho(\vec{r})] + E_{xc}[\rho(\vec{r})] + E_{Ne}[\rho]$$

Where:

- $T_s[\rho(\vec{r})]$: Kinetic energy of non-interacting system.
- $J[\rho(\vec{r})]$: Electron-Electron repulsion energy.
- $E_{xc}[\rho(\vec{r})]$: External potential.
- $E_{Ne}[\rho]$: Exchange-Correlation energy functional.

The Exchange-Correlation energy functional, $E_xc[\rho(\vec{r})]$, includes all the information that cannot be obtained, and it is usually calculated as the sum of the Exchange energy functional, $E_x[\rho(\vec{r})]$, and the Correlation energy functional, $E_c[\rho(\vec{r})]$.

$$E_{xc}[\rho(\vec{r})] = E_x[\rho(\vec{r})] + E_c[\rho(\vec{r})]$$

The main difference between the various methods in Density Functional Theory (DFT) is the approach used to choose the form of the exchange-correlation functional, $E_{xc}[\rho(\vec{r})]$. While this functional is considered unique and suitable for all systems, the exact and explicit form of $E_{xc}[\rho(\vec{r})]$ is unknown. Therefore, several approximations have been developed, including:

1. **Local Density Approximation (LDA):** This approach assumes that the exchange-correlation energy at a point in space depends only on the electron density at that point.
2. **Generalized Gradient Approximation (GGA):** This method takes into account not only the electron density but also its gradient, providing a better approximation than LDA.
3. **Hybrid** Functionals:
   In this approach, DFT is combined with Hartree-Fock (HF) theory to describe the exchange-correlation functional. Examples include:
   - **B3LYP:** Becke's 3-parameter exchange-correlation functional, which uses 3 parameters, and the Lee-Yang-Parr (LYP) correlation functional.

In addition to these, there are dozens of other hybrid and non-hybrid functional forms, each with its own advantages and disadvantages. The choice of the appropriate functional depends on the specific system and the desired level of accuracy.

I.6. Molecular Mechanics

Molecular mechanics emerged in 1930 but has evolved since the 1960s with advancements in computer performance. Through this method, we can determine thermodynamic properties such as thermal content, entropy, or free energy. Additionally, we can calculate the molecular energy based on atomic coordinates and search for the minimum energy corresponding to the stable state of the molecule.

Molecular mechanics considers the system's energy only as a function of atomic positions. This approximation, known as the Born-Oppenheimer approximation, assumes that by separating the movements of nuclei and electrons in a molecule, the latter can instantly adapt to the position

of the nuclei. Consequently, ignoring nuclear movements provides a substantial reduction in computation time. Specifically, molecular mechanics allows for the prediction of the system's structure and the study of a wide range of properties by describing the system's energy as the sum of various contributions representing intra- and intermolecular interactions. The minimum spatial energy is calculated as the sum of contributions such as stretching energy, bending energy, Van der Waals energy, etc. This can be mathematically expressed as:

$$E_{Total} = E_{stretch} + E_{bending} + E_{torsion} + E_{van\ der\ Waals} + \ldots..$$

Each of these terms has a preferred equilibrium state (e.g., bond length, bond angle, etc.).

The search for minimum energy by optimizing geometry plays a crucial role. The molecular energy is expressed as the sum of contributions related to deviations of the structure from reference structural parameters. In this regard, molecular mechanics is similar to ball-and-stick models but more quantitative. The evaluation of the total energy of the molecular system relies on various parameters generally imported from experimental data or high-level quantum computations (e.g., HF, DFT). This set of necessary parameters for MM calculations is known as force fields.

I.6.1. Force Field in Molecular Mechanics

A force field in molecular mechanics is a mathematical model that represents the potential energy variations with molecular geometry. It is the fundamental tool of molecular mechanics, allowing for the prediction of vibrational movements or intermolecular forces. To find the optimal geometry for a set of atoms, it's necessary to use three Cartesian coordinates for each atom. For example, a protein with 1000 atoms would require 3000 Cartesian coordinates. Therefore, the goal is to find the minimum energy in a space of several thousand variables. The challenge lies in choosing a simple potential function that can accurately and quickly calculate these molecular coordinates. This function must be simple enough to be computed with sufficient speed and accuracy to simulate the structural and thermodynamic properties of large molecules in an acceptable manner.

A force field consists of several potential energy functions that capture the molecular interactions between bonded and non-bonded atoms:

$$\underbrace{(E_{Total} = E_{stretch} + E_{bend} + E_{tors})}_{E_{Bond}} + \underbrace{(E_{van\ der\ Waals} + E_{electrostatic} + E_{bond\ H})}_{E_{no-Bond}}$$

One of the main challenges of molecular mechanics (MM) is selecting the appropriate force field for modeling the molecular system of interest. The specificity of each force field depends on the number of terms in the general equation, which generally increase with the complexity of the force field.

I.6.2. Interaction Energy Between Bonded Atoms

This energy includes the following components:

I.6.2.1. Interaction Energy Between Bonded Atoms:

- **Stretching Energy**

Stretching energy refers to the extension of bonds, and it is a term that aims to regulate the distance between two covalently bonded atoms. Initially, this contribution was expressed using the Morse potential (to simplify calculations) or by using a harmonic oscillator (x2x^2x2) derived from Hooke's law, which describes energy as the deformation of a spring. By analogy, force fields are often compared to models where atoms are represented as balls connected by springs.

$$E(L) = \frac{1}{2} K_r (L - L_0)^2$$

Where:

- K_r: is the stretching constant (Hooke's constant).
- L: is the bond length in the model.
- L_0: is the reference bond length.

- **Bending Energy**

Bending energy arises from the slight movements of atoms around their equilibrium positions, resulting in the distortion of bond angles. The bending energy, $E_{bend,}$ is typically represented by an expression similar to that of bond stretching and usually involves the term θ^2:

$$E(\theta) = \frac{1}{2} K_\theta (\theta - \theta_0)^2$$

Where:

- K_θ: is the bending constant.
- θ: is the bond angle in the molecule.
- θ_0 is the reference bond angle.

➢ **Torsional Energy**

Torsional energy relates to the angle between two planes formed by atoms 1, 2, 3, and 4. It is particularly important for estimating the three-dimensional structure of a molecule. This energy is typically estimated using a simplified Fourier series function:

$$E(\phi) = \frac{1}{2}[V_1(1 + cos\ \phi) + V_2(1 - cos\ 2\phi) + V_3(1 + cos\ 3\phi)]$$

Where:

- V_1, V_2, V_3 : are the torsional potential energy constants.
- ϕ :is the dihedral angle.

I.6.2.2. Interaction Energy Between Non-Bonded Atoms

Non-bonded interactions refer to those between atoms that do not share a covalent bond. Only atoms separated by more than three bonds are considered capable of having non-bonded interactions. This is because the interactions between atoms separated by one, two, or three bonds are sufficiently well described by terms related to the deformation of bond lengths and angles formed by two covalent bonds.

➢ **Van der Waals Energy**

Van der Waals forces are attractive forces that play a crucial role in the interaction between molecules. This energy is related to non-bonded atoms that are not directly connected to each other or the shared atom. Van der Waals energy consists of two components: a repulsive part and an attractive part. It can be expressed by the following equation:

$$E(vdw) = \epsilon * \times [-C_1(\frac{r*}{r})^6 + C_2 exp\ (-C_3(\frac{r}{r*})]$$

Where:

- $\epsilon *$: is a variable energy constant, also known as the hardness constant.
- $r *$: is the sum of the Van der Waals radii of the interacting atoms.
- r: is the distance between the atoms.
- C_1, C_2, C_3 :are constants specific to the force field.

- **Electrostatic Interactions**

Electrostatic interactions occur between two atoms that each carry a charge. Coulomb's law describes the energy of this interaction:

$$E_{elect} = \sum \frac{q_i q_j}{D r_{ij}}$$

Where:

- $q_i q_j$: are the atomic charges of the atoms.
- r_{ij}: is the distance between the atoms.
- D: is the dielectric constant in the equilibrium state.

I.6.3. Commonly used Force Fields

Numerous studies have proposed various force fields, distinguished by the terms used to develop the energy expression for a molecule. Each force field has a specific application range, and the choice of force field depends on the properties of the system to be studied, such as the type of compound (e.g., carbohydrates, metal complexes), the environment (e.g., gas, solution), and the type of interaction being studied (e.g., hydrogen bonding).

Some of the most commonly used force fields are:

- **MM2, MM3, MM4**: MM2 is the most commonly used force field by organic chemists. It was originally designed for simple organic molecules such as alkanes, alkenes, alkynes, carbonyl compounds, and amines. However, the more refined version, MM3, allowed for the handling of complex organic molecules.
- **MM+**: An extension of MM2, with additional parameters. It can be applied to more complex molecules, including inorganic compounds.
- **AMBER (Assisted Model Building with Energy Refinement)**: A molecular mechanics force field developed by Kollman. It can be applied to proteins, nucleic acids, polymers, and other small molecules.
- **Chemistry at Harvard Macromolecular Mechanics (CHARMM Bio+):**

CHARMM was developed in 1983 by R. Brooks, M. Karplus, and others. Its concept is similar to that of AMBER, using an empirical energy function to study molecular systems and biological molecules such as proteins and nucleic acids.

- **MMFF94:**

  MMFF94 is one of the modern molecular mechanics force fields, developed by Tom Halgren. The stretching, bending, and torsional energies in MMFF94 are similar to those in MM3.

- **Universal Force Field (UFF):**

  The universal force field capable of simulating molecules containing any combination of atoms from the periodic table.

Finally, to determine whether a molecule has the optimal geometric representation (Geometry optimization), it must be in its most stable state with the lowest energy. Achieving this involves adjusting the spatial configuration of the molecule (bond lengths, bond angles, dihedral angles) until the lowest energy is obtained. Searching for this stable configuration produces a vast amount of energies and values, resulting in what is known as the potential energy surface, represented by a set of points: Hilltops, Global Minimum, Saddle Point, etc. If the study is conducted using two independent variables (e.g., two dihedral angles), this surface can be represented in three dimensions. Figure 5 describes the potential energy surface.

I.7. Molecular Dynamics (MD):

Molecular Dynamics is one of the most widely used methods for simulating real systems. It involves computational simulations of molecules and compounds based on Newton's laws of motion, predicting the movement, speed, and positions of molecules over very short time intervals (femtoseconds to picoseconds) under specific conditions (pressure, volume, and temperature). In this approach, each atom is treated as a particle that follows Newton's law of motion:

$$\sum \vec{F} = m\vec{a}$$

The successive integrations of this equation with respect to time yield the trajectory of each atom. Each atom i with position vector X_i and mass m_i will experience an acceleration a_i according to Newton's law:

$$Fi = m_i a_i = \frac{\partial^2 r_i}{\partial t^2}$$

Where:

r : represents the Cartesian coordinates of the atom (position vector).

Molecular Dynamics simulations require defining a potential function, or a description of the terms that represent how particles interact during the simulation. In chemistry and biology, this is usually referred to as a force field, while in material physics, it's called an interatomic potential. Potentials can be defined at various levels of physical accuracy, with the most common in chemistry being those based on molecular mechanics. These embody a classical mechanical treatment of particle interactions that can reproduce structural and conformational changes but generally cannot account for chemical reactions.

The reduction from a full quantum description to classical potentials involves two major approximations. The first is the Born-Oppenheimer approximation, which assumes that electron dynamics are so fast that they can be considered to instantly respond to nuclear movements, allowing for separate treatment. The second treats the nuclei, much heavier than electrons, as point particles following classical Newtonian dynamics. When more detailed levels of description are needed, quantum mechanics-based potentials are used. Some methods attempt to create hybrid classical/quantum potentials where most of the system is treated classically,

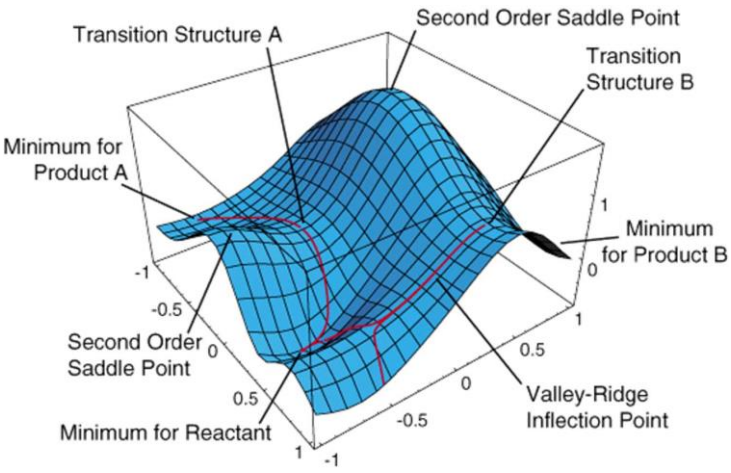
but a small region, typically undergoing a chemical transformation, is treated quantum mechanically.

MD simulations using empirical force fields are limited in the same way as Molecular Mechanics (MM) and, therefore, cannot be applied, for example, to the study of enzymatic reactions.

MD simulations are crucial for investigating the structural behavior of biomolecules as a function of time. Using MD, one can study the flexibility or rigidity of biomolecules, measure interactions between molecules and between proteins and ligands (or other biomolecules), calculate the Helmholtz free energy, or study the solvent's effect on biomolecular structure.

I.8. Potential Energy Surface

A **Potential Energy Surface (PES)** is a theoretical concept used in chemistry and physics to describe the energy of a molecular system as a function of its atomic positions. It represents the relationship between the molecular geometry (i.e., the arrangement of atoms) and the potential energy of the system. It's often visualized as a multidimensional landscape, where the valleys represent stable configurations (energy minima) and the peaks represent unstable configurations (energy maxima).



I.8.1. Stationary Points

The stationary points on the potential energy surface are points where the surface is flat, i.e., parallel to the horizontal line responsible for the geometric structure parameters (or parallel to the horizontal plane responsible for the geometric structure). Mathematically, at a critical point, the first derivative of the energy with respect to each of the geometric structure parameters is equal to zero.

$$\frac{\partial E}{\partial q1} = \frac{\partial E}{\partial q2} = \ \ldots \ = 0$$

The important and distinctive critical points on the potential energy surface are of three types:

- **Minima**

The local minima on the potential energy surface represent the stable states. These points determine the minimum energy of the reactants, products, and intermediate compounds. The reaction pathway can be determined by identifying the path of lowest energy connecting two local minima, which is called the Intrinsic Reaction Coordinate (IRC). This is the expected path the compound will follow when transitioning from one local minimum to another.

I.8.2. Stationary Points and vibrations

The stationary points (minimum, transition state, saddle points) can be distinguished by performing a vibrational frequency calculation. When performing a geometry optimization, it is preferable to accompany it with a vibrational frequency calculation, as this can be used to identify the type of stationary point (in addition to the importance of vibrations for obtaining some thermal properties) as follows:

Minima: All vibrational frequencies have positive values.

Transition state: There is only one negative vibrational frequency value.

Hilltops: There is more than one negative vibrational frequency value.

Practically, there is no such thing as a negative frequency, and this "frequency" is actually an **imaginary frequency**. This is because the frequency calculation involves taking the square root of the force constant, and the square root of a negative number is an imaginary number (there is no real square root of a negative number).

A single negative vibrational frequency does not necessarily mean the structure is a **transition state**. When searching for a transition state, it is important to note:

- The transition state must be a valid point between the reactants and products.
- There must be only one and only one negative vibrational frequency.
- The imaginary frequency must clearly and explicitly contain the reaction pathway (reaction mechanism) through the vibrational motion. For example, the HCN/HNC isomerization must have only one negative frequency, and the hydrogen atom migration between the carbon and nitrogen atoms must be clearly visible in the vibrational motion.
- The transition state energy must always be higher than the energies of the reactants and products.

I.8.3. Transition State:

This is also a stationary point on the potential energy surface, but it represents the highest point on the line connecting two local minima. Therefore, it takes the shape of a saddle, and is thus called a Saddle Point

I.8.4. Geometry Optimization

As we mentioned earlier, the important stationary points on the potential energy surface can be local minima, transition states, Therefore, the process of finding these points can be described as an Energy Minimization or Minimization process. The search for the transition state is then a Transition State Optimization.

The geometric structure can be optimized by varying the geometric structure parameters of the input structure (the closer the structure, the faster the result). This involves changing the bond lengths, bond angles, and calculating the energy corresponding to each set of parameters until one of the lowest points (the closest) or the lowest energy point is obtained.

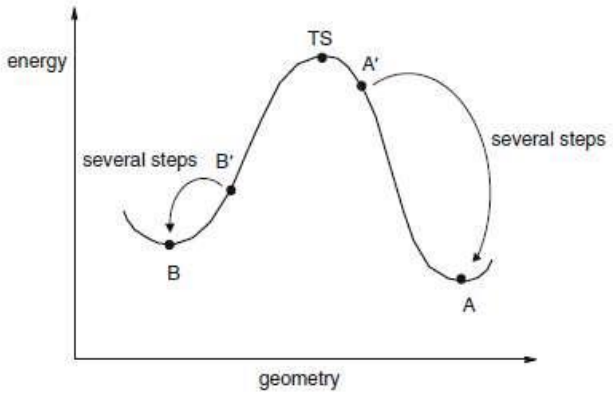


Figure 3. Reaction profile showing energy barriers and intermediates

I.8.5. Thermochemistry from ab initio MO methods

Ab initio molecular orbital (MO) methods yield total energies, E_{eq}, as the sum of electronic energy and nuclear-nuclear repulsion energies for molecules that are isolated in a vacuum and assumed to be at 0 K without any vibrational motion:

$$E_{eq} = E_{el} + E_{nuclear-nuclear}$$

From the 0 K potential energy surface and applying the harmonic oscillator approximation, we can determine the vibrational frequencies ν_i of the normal modes of vibration. These frequencies allow us to compute the vibrational, rotational, and translational contributions to thermodynamic quantities such as the partition function Q and heat capacity C as the system is heated from 0 to T K.

Q: partition function

E: energy

S: entropy

C: heat capacity

Vibration

The vibrational partition function $\boldsymbol{Q_{vib}}$ is given by:

$$Q_{vib} = \sum_i \frac{1}{1 - exp(-\frac{h\nu_i}{kT})}$$

The vibrational energy E_{vib} for a molecule at temperature T can be expressed as:

$$E_{vib} = \sum_i \left(\frac{h\nu_i}{2} + \frac{h\nu_i exp(-\frac{h\nu_i}{kT})}{21 - exp(-\frac{h\nu_i}{kT})} \right)$$

where:

- h :is the Planck constant,
- ν_i :is the $i^{\text{-th}}$ normal mode frequency,
- k :is the Boltzmann constant.

For one mole of molecules, E_{vib} must be multiplied by Avogadro's number $N_A = \frac{R}{k}$:

$$E_{vib} = N_A \sum_i \left(\frac{h\nu_i}{2} + \frac{h\nu_i \exp\left(-\frac{h\nu_i}{kT}\right)}{21 - \exp\left(-\frac{h\nu_i}{kT}\right)} \right)$$

The first term in this equation represents the zero-point vibrational energy, while the second term accounts for the additional vibrational contribution due to temperature increase from 0 K to T K:

$$E_{vib} = E_{zero} + E_{vib}(0 \rightarrow T)$$

where:

$$E_{zero} = N_A \sum_i \frac{h\nu_i}{2}$$

$$E_{vib}(0 \rightarrow T) = N_A \sum_i \frac{h\nu_i \exp\left(-\frac{h\nu_i}{kT}\right)}{21 - \exp\left(-\frac{h\nu_i}{kT}\right)}$$

The vibrational entropy S_{vib} is given by:

$$S_{vib} = R\sum_i \left[\frac{h\nu_i}{kT}\frac{\exp\left(-\frac{h\nu_i}{kT}\right)}{1-\exp\left(-\frac{h\nu_i}{kT}\right)} - ln\left(1-\exp\left(-\frac{h\nu_i}{kT}\right)\right)\right]$$

The vibrational heat capacity C_{vib} is expressed as:

$$C_{vib} = R\sum_i \left[\left(\frac{h\nu_i}{kT}\right)^2 \frac{\exp\left(-\frac{h\nu_i}{kT}\right)}{1-\exp\left(-\frac{h\nu_i}{kT}\right)^2}\right]$$

At temperatures $T > 0$, a molecule rotates about the x, y, and z axes and translates in the x, y, and z directions. Assuming equipartition of energy, the energies for rotation and translation, E_{rot} and E_{tr}, can be calculated.

Rotation

Let σ be the symmetry number and I be the moment of inertia. The moments of inertia about the A, B, and C axes are denoted as IA, IB, and IC.

Linear Molecule

The rotational partition function Q_{rot} for a linear molecule is given by:

$$Q_{rot} = \frac{8\pi^2 IkT}{\sigma h^2}$$

The rotational energy E_{rot} is:

$$E_{rot} = RT$$

The rotational entropy S_{rot} can be expressed as:

$$S_{rot} = Rln\left(\frac{8\pi^2 IkT}{\sigma h^2}\right) + R$$

This simplifies to:

$$S_{rot} = RlnI + RlnT - Rln\sigma - 4.349203$$

Where

$$-4.349203 = Rln\left[\frac{8\times10^{-16}\pi^2 k}{N_a h^2}\right]$$

The rotational heat capacity C_{rot} is:

$$S_{rot} = R$$

Non-linear Molecule

For a non-linear molecule, the rotational partition function is:

$$Q_{rot} = \left(\frac{\sqrt{\pi}}{\sigma}\right)\left(\frac{8\pi^2 IkT}{\sigma h^2}\right)\sqrt{I_A I_B} I_C$$

The rotational energy is:

$$E_{rot} = \frac{3}{2}RT$$

The rotational entropy becomes:

$$S_{rot} = \frac{R}{2} ln\left(\frac{\sqrt{\pi}}{\sigma}\left(\frac{8\pi^2 cI_A}{h}\right)\left(\frac{8\pi^2 cI_B}{h}\right)\left(\frac{8\pi^2 cI_C}{h}\right)\left(\frac{kT}{hc}\right)^3\right) + \frac{3}{2}R$$

This simplifies to:

$$S_{rot} = \frac{R}{2} ln(I_A I_B I_C) + \frac{3}{2} RlnT - Rln\sigma - 5.3863921$$

where -5.3863921is calculated as:

$$Rln\left(\frac{1}{h^3}(10^{-16}N_a)^{3/2}\sqrt{\frac{3\times 2}{9\times \pi^7 k}}\right) + \frac{3}{2}R$$

The rotational heat capacity C_{rot} is:

$$C_{rot} = \frac{3}{2}\mathrm{R}$$

Translation

Let M be the molecular weight. The translational partition function Q_{tra} is given by:

$$Q_{tra} = \left(\frac{\sqrt{2\pi MkT}}{N_a h}\right)^3$$

The translational energy E_{tra} is:

$$E_{tra} = \frac{3}{2}RT$$

The translational entropy S_{tra} is expressed as:

$$S_{tra} = R\left(\frac{5}{2} + \frac{3}{2}ln\left(\frac{2\pi k}{h^2}\right) + lnk + \frac{3}{2}ln\left(\frac{M}{N_a}\right) + \frac{5}{2}lnT - lnp\right)$$

This simplifies to:

$$S_{tra} = \frac{5}{2}RlnT + \frac{3}{2}RlnM - Rlnp - 2.31482$$

The translational heat capacity C_{tra} is:

$$C_{tra} = \frac{5}{2}R$$

The enthalpy H_{tra} is:

$$H_{tra} = \frac{5}{2}R \ \ (due\ to\ the\ pV\ term, H = U + pV)$$

The internal energy U at temperature T is expressed as:

$$U = E_{eq} + [E_{vib} + E_{rot} + E_{tra}]$$

Alternatively, it can be written as:

$$U = E_{eq} + [(E_{zero} + E_{vib}(0 \rightarrow T) + E_{rot} + E_{tra}]$$

The enthalpy H for one mole of gas is defined as:

$$H = U + pV$$

Assuming ideal gas behavior ($pV = RT$), we have:

$$H = U + pV = U + RT$$

Thus, the Gibbs free energy G can be calculated as:

$$G = H - TS(0 \rightarrow T)$$

I.8.6. Some Basic Physical Constants

Speed of light, c = 2.99792458 × 10^{10} cm/s

Boltzmann constant, $k = R/N_A$ = 1.380658 × 10^{-23} J/K = 1.380658 × 10^{-16} erg/K

Planck constant, h = 6.6260755 × 10^{-34} J s = 6.6260755 × 10^{-27} erg s

Gas constant, R = 8.314510 J/mol/K = 1.987216 cal/mol/K

Avogadro number, N_A = 6.0221367 × 10^{23} /mol

Volume of 1 mol of gas, V_0 = 22.41410 l/mol (at 1 atm, 25 C) 1 J = 1. × 10^7 erg

1 $kcal$ = 4.184 kJ (Definition)

1 eV = 23.0606 kcal/mol

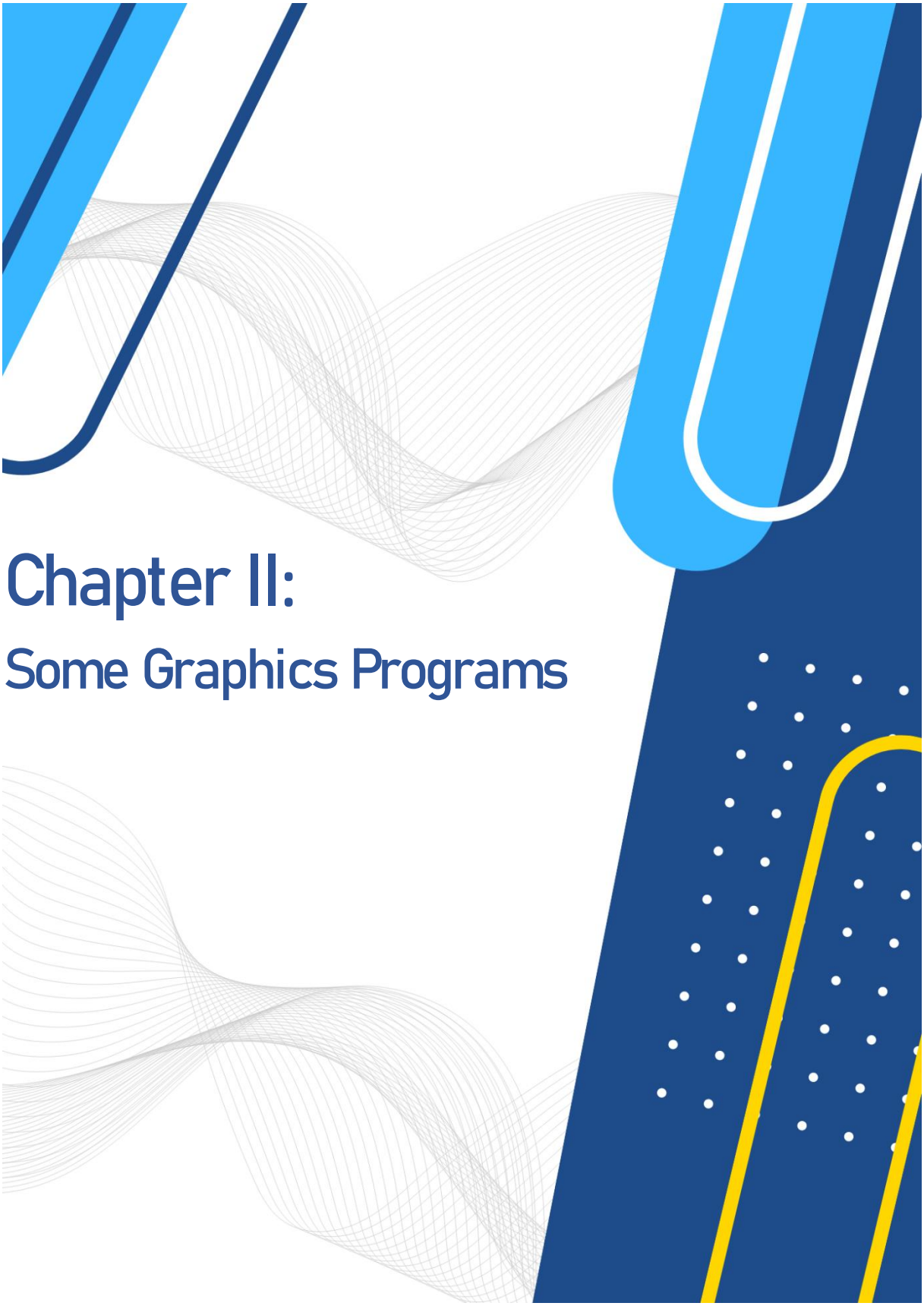
1 $a.u.$ = 27.21135 eV/mol = 627.509 6 kcal/mol

1 cm^{-1} = 2.859144 cal/mol = $N_A$hc/4.1847

1 atm = 1.01325 × 10^5 Pa = 1.01325 × 10^6 dyn/cm^2

Moment of inertia: I 1 amu angstrom^2 = 1.660540 × 10^{-40} g cm^2 .

Rotational constants: A, B, and C (e.g. A = h/(8π^2 I)) With I in amu angstroms^2 then: A (in MHz) = 5.053791 × 10^5/I A (in cm^{-1}) = 5.053791 × 10^5/cI = 16.85763/I



Chapter II:

Some Graphics Programs

II. Introduction:

Graphical software is essential in computational chemistry, enabling users to interact with molecular models and input data intuitively. Users can manually enter instructions and computational commands or utilize specialized graphical tools like Avogadro and ChemBio3D for input and output. Another powerful option is GaussView, part of the Gaussian suite developed by the same company. This program allows users to input critical calculation data, including spatial structure information, charge, the computational method employed, and the type of calculation to be performed.

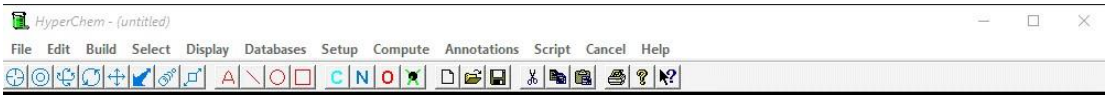
Additionally, these programs primarily serve to display results effectively. They provide visualizations such as graphs, spatial representations, and even kinetic information related to vibrations and reaction mechanisms, among other outputs.

II.I. HYPERCHEM SOFTWARE

HyperChem 8.0 software was initially developed by Autodesk, Inc. and is now distributed by Hypercube, Inc. (Ontario, Canada). In addition to performing molecular mechanics and molecular dynamics on small molecules, this software also allows for semi-quantitative and even simple quantum calculations. IBM provides excellent support for the calculation of wave functions, electronic potentials of small molecules, and charge distributions.

This section summarizes how to use the software, and the practical part of the course will focus on semi-empirical calculations of small molecules and evaluations of charge densities and representations of molecular orbitals.

II.1.1. BASIC CONCEPTS ON HYPERCHEM:

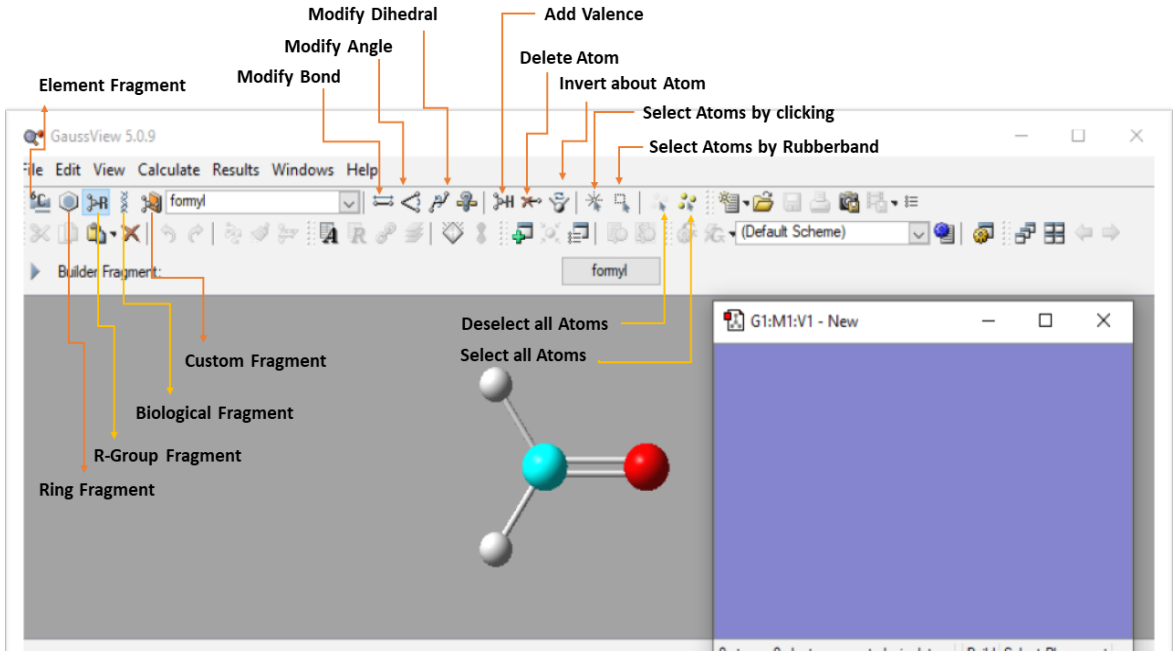


- You launch the software by clicking on the HyperChem icon.
- To save your files, use "**Save As**" and place them in your own directory under Users.
- **Under the File menu**, you will find the usual options found in all software (e.g., Open, Save, Print, Merge, etc.). However, there is a "**Start Log**" command used for storing numerical information during calculations. You must activate this function before submitting work because HyperChem does not create an Output file.

- **Under Edit**, in addition to the standard commands, there is "**Align Viewer**" to view the molecule from a desired axis, and "**Set Bond**," "**Set Angle**," and "**Set Torsion**" to set known parameters.
- **In Display**, under "**Show Initial Axes**," you can visualize the axes.
- You construct a molecule using the first cursor on the left (each cursor can lead to a window by double-clicking it). Modifications are made via the "**Build**" mode and "**Add Hydrogens**" (adding H), "**Add H & Model Build**" (standard 3D arrangement of the molecule after construction), "**Set Charge**" (to set a charge), "**Constrain**..." (add constraints), and "Default Element" (choose other atoms via the periodic table).
- **Under Select**, you can set parameters to **Atoms**, **Residues**, and they will be retained or placed in the "Working Set" (in memory).
- **Under Display**, there are graphic options like **Color**, **Label**, **Rendering**, **Recompute HBonds** (search for hydrogen bonds), **Scale to Fit** (to center the work on the screen).
- **In Databases**, there is a list of **Amino Acids** (with N-Me and N-Ac), **Nucleic Acids**, and the ability to add a counter-ion with "**Add Counter Ions**" or **create a Zwitt**erion (to end a peptide chain).
- **Calculations are performed using Setup**. In **Molecular Mechanics**, there are AMBER, MM+, OPLS, BIO+, and Options.
- **With Semi-empirical**, you have AM1, INDO, MNDO, CNDO, ZINDO, PM3, etc., and in Options, there are Charges, Spin Multiplicity, UHF/RHF, SCF, Config. Inter. (CI).
- **With Ab initio**, you have STO-3G, 3-21G, 6-31G*, and under Options: Charges, UHF/RHF, MP2 (an electronic correlation with exchange integral, unlike the Semi-empirical mode).
- **You can also choose "Periodic Box"** to construct a solvent box around the molecule. If you select Molecules in the "Select" menu and choose the molecule with the second icon, it becomes active in green. With the third icon (translate) and right-clicking the mouse, only the selected molecule moves within the solvent box. With a left-click, everything moves.
- **Under "Compute,"** you can submit the calculation via **Single Point**, **Geometry Optimization** (Steepest descent, Conjugate gradient), **Molecular Dynamics**, **Langevin Dynamics**, or **Monte Carlo**. Each mode has Options: vacuum, solvent box, etc. You start the calculation with "Proceed."

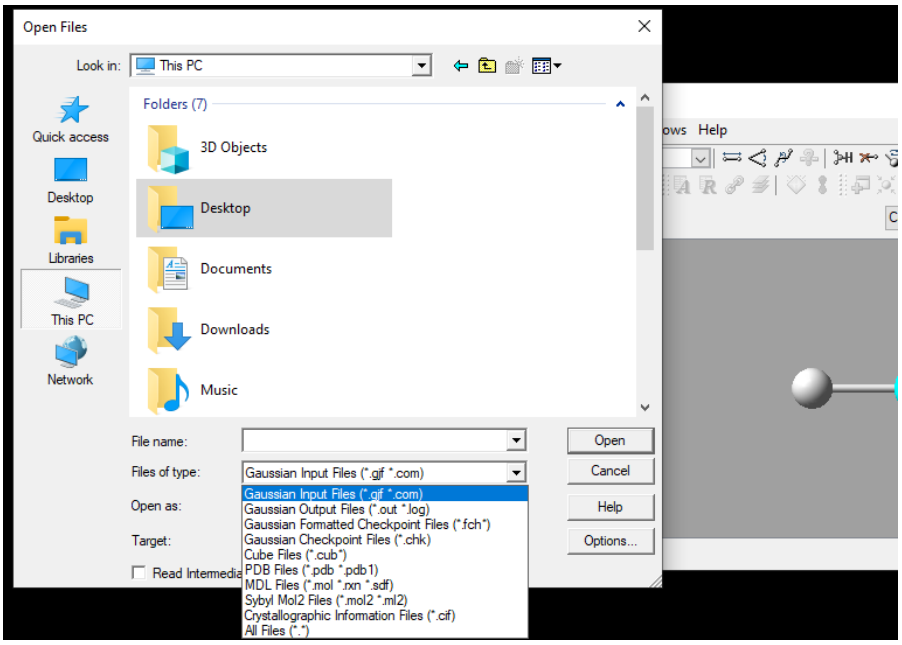
- **The "Compute" menu** also includes the **Vibration mode**, which automatically calculates **the IR vibration spectrum**. You can animate it with **"Vibrational Spectrum"** and the "Animate" option by selecting the frequency on the spectrum beforehand.
- **With the Electronic Spectrum mode**, you obtain the corresponding UV spectrum.
- **In Transition State mode**, you calculate the eigenvalues (eigenvectors of the transition state).
- **In Orbitals**, you get the complete energy diagram of the molecule, and you can then visualize the orbitals in 2D (plane) or 3D with presentation options.
- **The "Cancel" menu** is used to stop a calculation (a job in progress).
- **In "File" -> "Preferences,"** you can change the color of the screen and the isosurfaces.

II.2. GaussView:

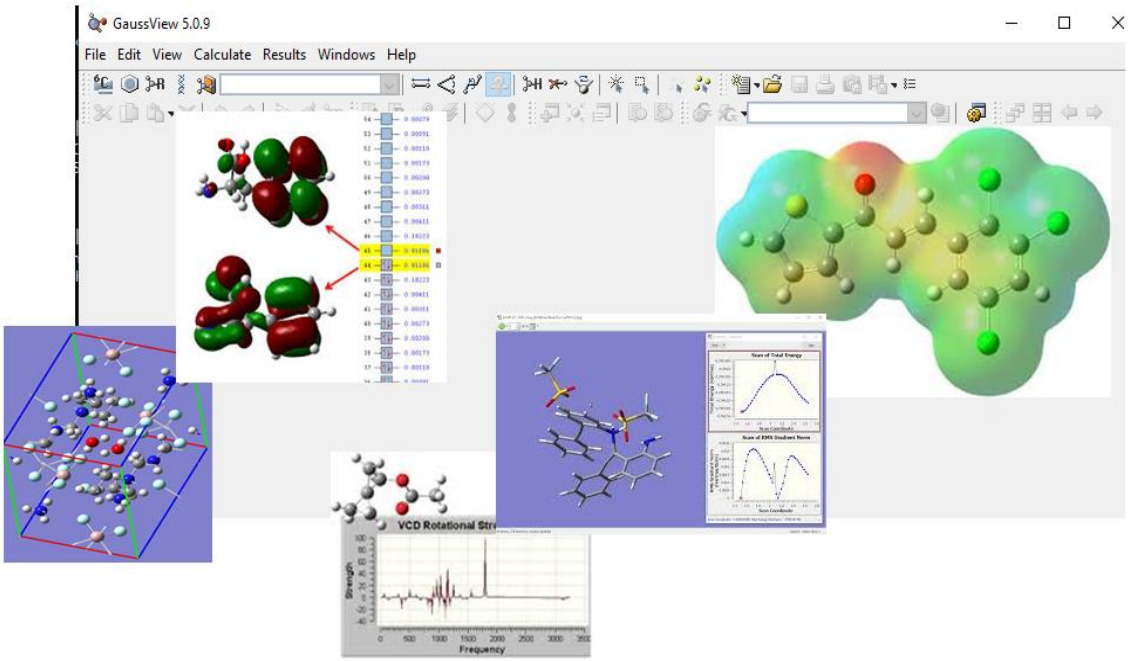


GaussView is the most advanced and powerful graphical user interface available for Gaussian. With GaussView, you can import or build the molecular structures of interest to you, configure, launch, monitor and control Gaussian calculations, and retrieve and visualize the results, all without ever leaving the application.

This software supports multiple formats.



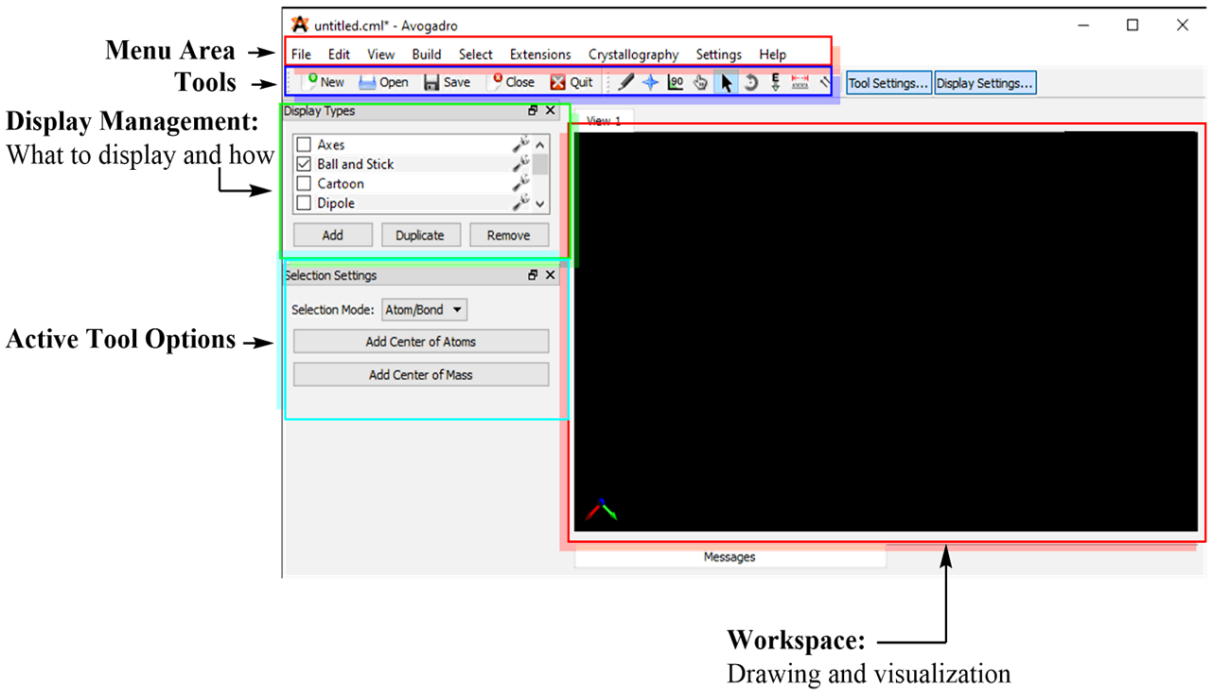
It helps to visualize crystallographic structures, polymers, and IR, Raman, UV spectra, and more.



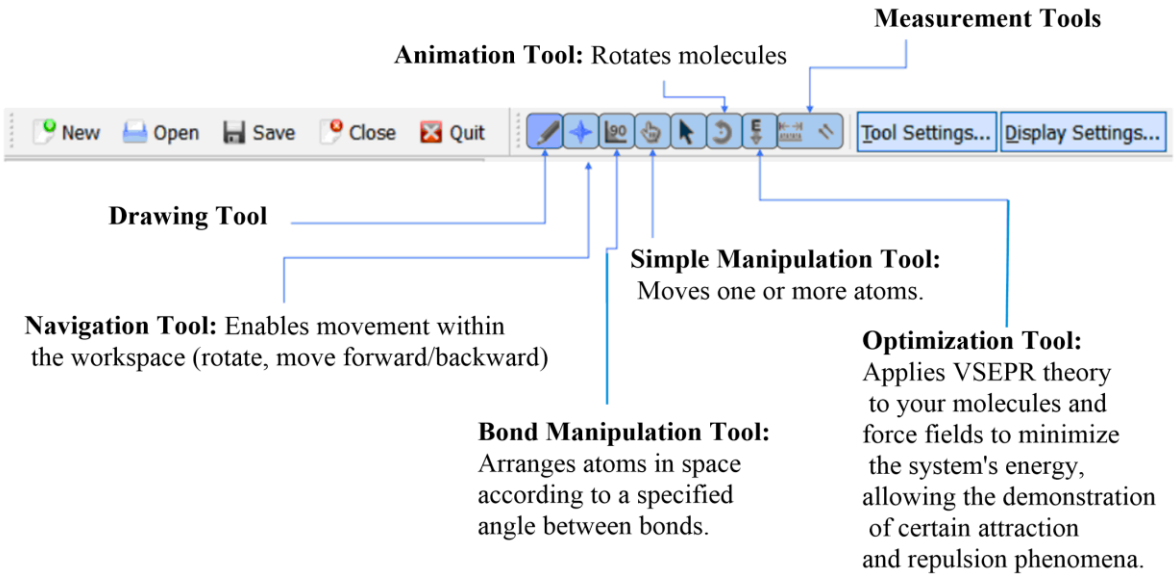
II.3. Avogadro:

Avogadro is a free and open-source molecular editor and visualization tool compatible with Mac, Windows, and Linux. It is designed for applications in computational chemistry, molecular modeling, bioinformatics, materials science, and related fields. The software features flexible high-quality rendering and a robust plugin architecture.

II.3.1. The Avogadro Interface

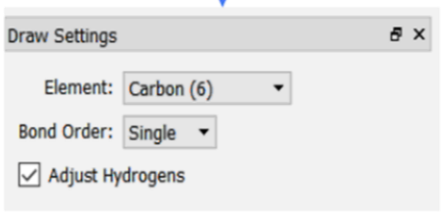


II.3.2. Tools



If the drawing tool is not activated, click on it. At the bottom left, the drawing options will appear





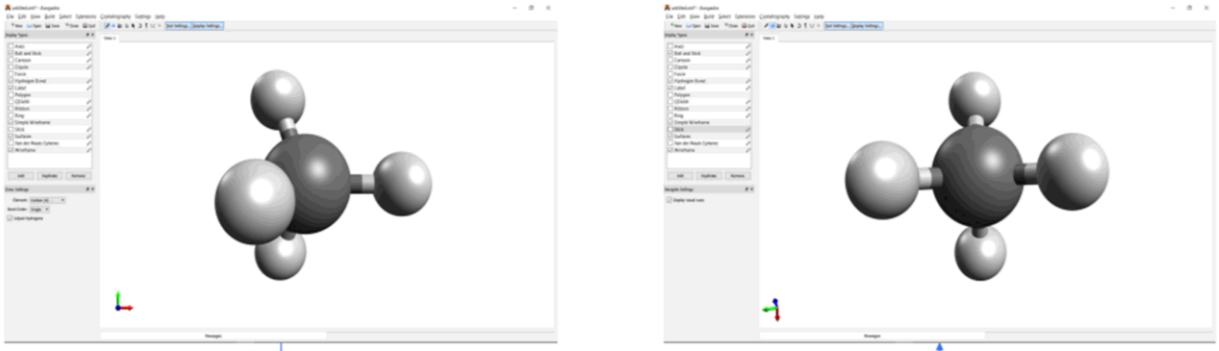
You can choose the atom you want to add:

* Whether it will form a single bond or another type.
* Adjust Hydrogens will automatically add H atoms as needed to the structure.



Click the LEFT mouse button in the middle of the workspace.
And there you go, you've placed a C atom, and the tool automatically adds the H atoms.

You can **zoom in** on your work using the scroll wheel of your mouse.

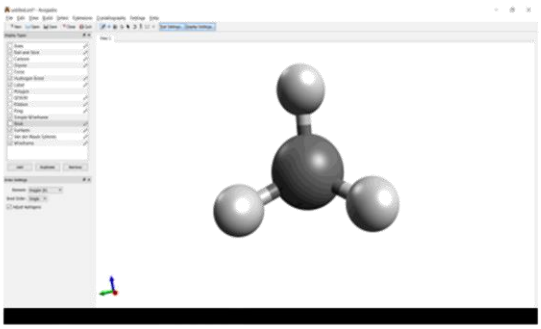


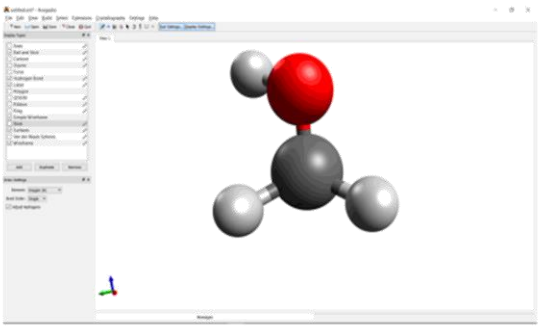
To rotate your molecule and view it from all angles, activate the **navigation tool**.

Left click held down: rotation
Right click held down: movement

Let's add an OH group to this:

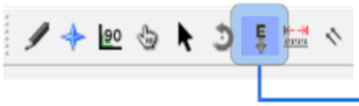
* Position your methane to clearly see one of its H atoms.
* Then switch back to Drawing mode (press F8 for quicker access).
* In the tool options, select the O element with a single bond.
* Once that's done, simply click left on one of the H atoms of the methane; that H will be replaced by an OH group.



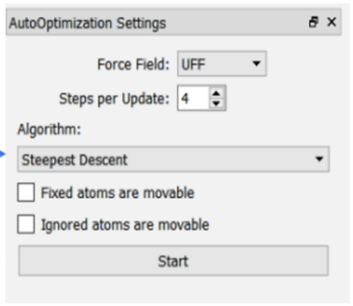


Let's minimize the energy

Select the optimization tool:



The options for this tool are:



Steps per Update: The number of calculation steps that occur between two refreshes of the display. The higher this number, the smoother the movement will be (useful for interacting with the simulation).

Algorithm:

Steepest Descent: An iterative method seeking a minimum energy step by step.

Conjugate Gradients: Another iterative method, more refined.

Molecular Dynamics: Accounts for oscillations and elongations of bonds (varies with temperature and thus the kinetic energy of the system) and is VERY DEMANDING ON COMPUTATIONAL POWER!

Force Field:

UFF: Universal force field (will not account for any hydrogen bonds).

MMFF94 and **MMFF94s**: General force fields that take hydrogen bonds into account.

Ghemical: uses its own force field primarily designed for organic molecules. It allows for the simulation of molecular interactions and can handle various types of molecular systems.

GAFF (General AMBER Force Field): designed for general organic molecules. It extends the capabilities of the AMBER force field to include a wider range of chemical functionalities. GAFF is suitable for use in molecular dynamics simulations and can model small organic molecules effectively by providing parameters for various atom types and bonding interactions.

II.4. ChemBio 3D

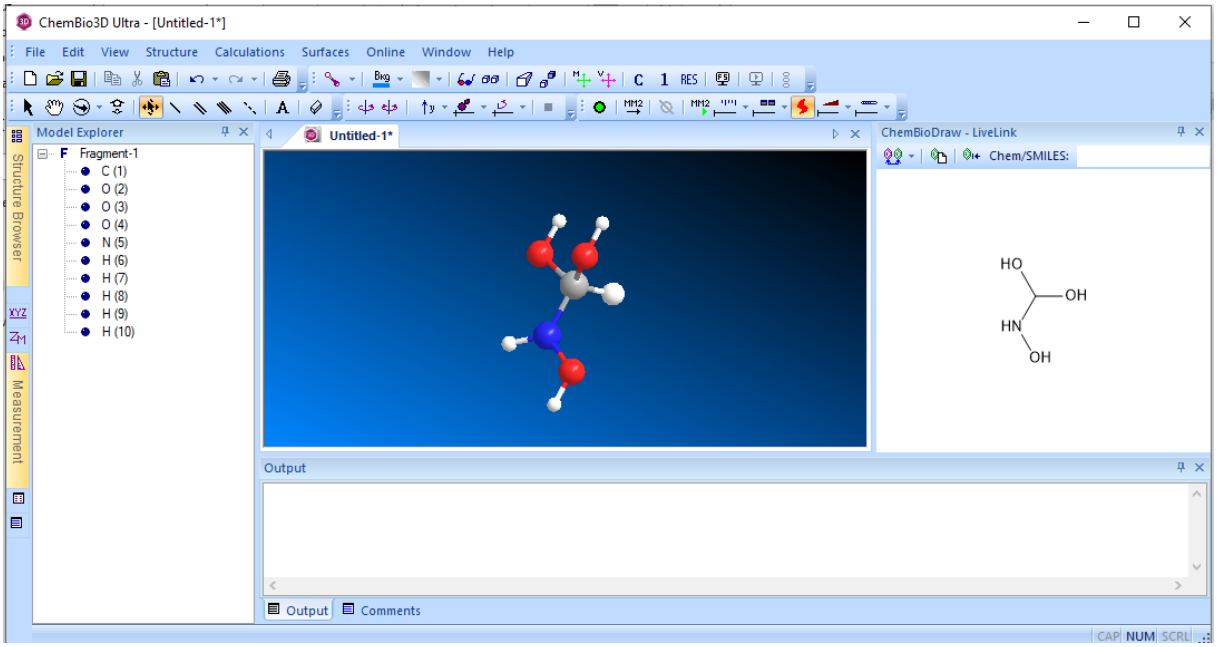
II.4.1. ChemBio3D Interface

Upon launching Chem 3D, you'll notice two main sections: the model window on the left and ChemDraw on the right. You can close the **ChemDraw** window by clicking the X in the upper

right corner, although you may leave it open if you prefer. The interface resembles other Windows applications, with familiar menus like File, Edit, Window, and Help.

To customize your workspace, you can add useful tools to the toolbar. If the desired toolbar is not visible, select **View > Toolbars**.

For common tasks, open the **Output Box** and **Measurement Table** by selecting **View > Output Box** and **View > Measurement Table**. These windows can be detached and moved around as needed.



II.4.1. Tools Palette

Chem 3D includes various tools for building and editing molecular models, accessible from the Tools palette.

You can create molecular structures directly in the model area, though the drawing functionality has limitations. To input a simple molecular structure, select the **Text tool** and type the structure using proper case (e.g., (CNO3H5). The Text tool also allows you to modify atoms easily.

Chem 3D supports importing several file formats, including ChemDraw (chm), Protein DB (pdb), and MDL (mol) files. ChemDraw and Chem 3D are part of the ChemOffice suite, while PDB and MDL files are mainly used in visualization programs like DS Viewer and ChemSketch.

- **Rotating Molecules**

Use the **Trackball tool** to rotate molecules freely. You can also rotate around specific axes by moving your cursor to the window edges and dragging with the left mouse button or by selecting an axis and specifying the degrees of rotation.

- **Selecting Atoms and Bonds**

To select atoms or bonds, use the **Selection tool**. Click to select individual atoms or bonds. For multiple selections, hold down the Shift key. To select a section of a molecule, click and drag to create a rectangle around the desired area.

- **Moving Atoms or Molecules**

Use the **Translation tool** to move the entire molecule, or the Move Objects tool for selecting and moving specific atoms.

- **Energy Calculations**

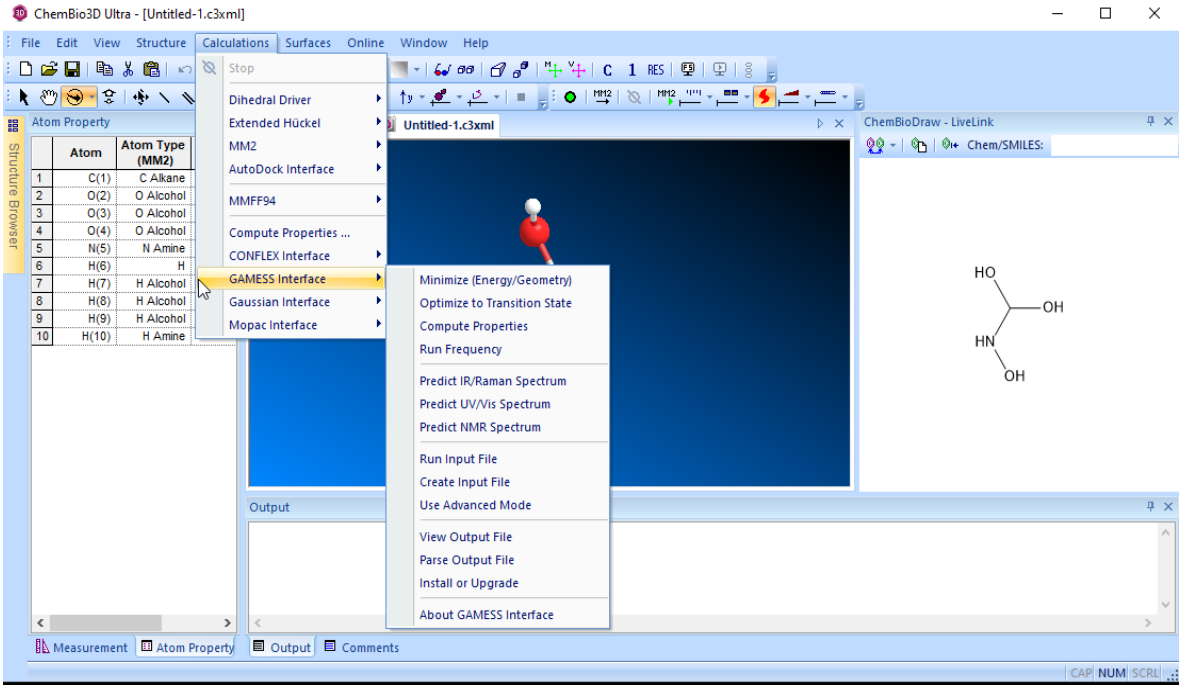
To perform a **Single Point Energy Calculation**, navigate to **Calculations > MM2 > Compute Properties**. Under the Properties tab, choose **Steric Energy Summary** and ensure Compute Properties is selected before clicking Run.

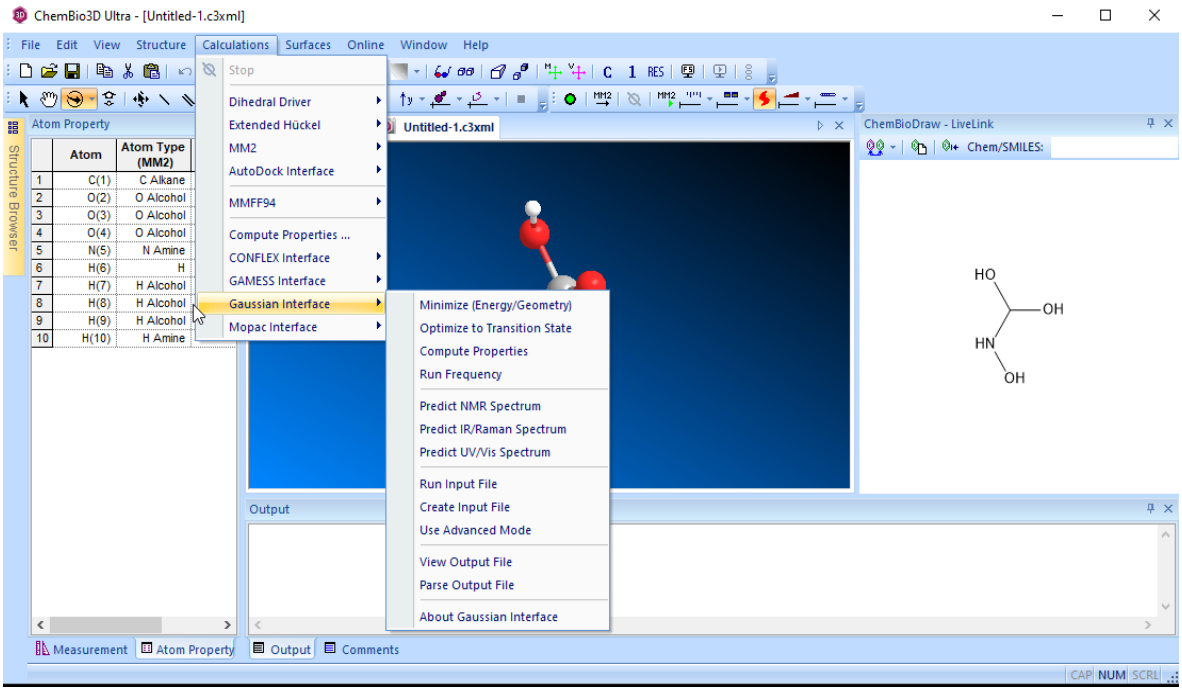
The output will include values for various energy terms, such as Stretch, Bend, and Torsion, all measured in kcal/mole.

- **Geometry Optimization**

For energy minimization, select Calculations > MM2 > Minimize Energy. Choose Minimize Energy as the Job Type, check Display Every Iteration, and click Run. The output will indicate the total steric energy after optimization.

You can also generate input files for external software such as GAMESS, Gaussian, and MOPAC to perform tasks like geometry optimization, vibrational analysis, and property calculations.



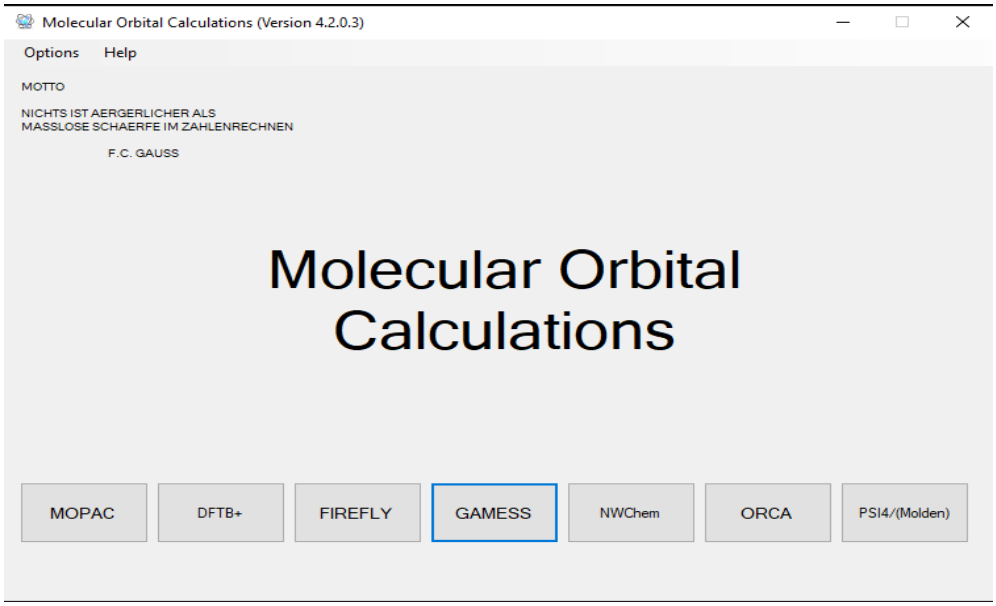


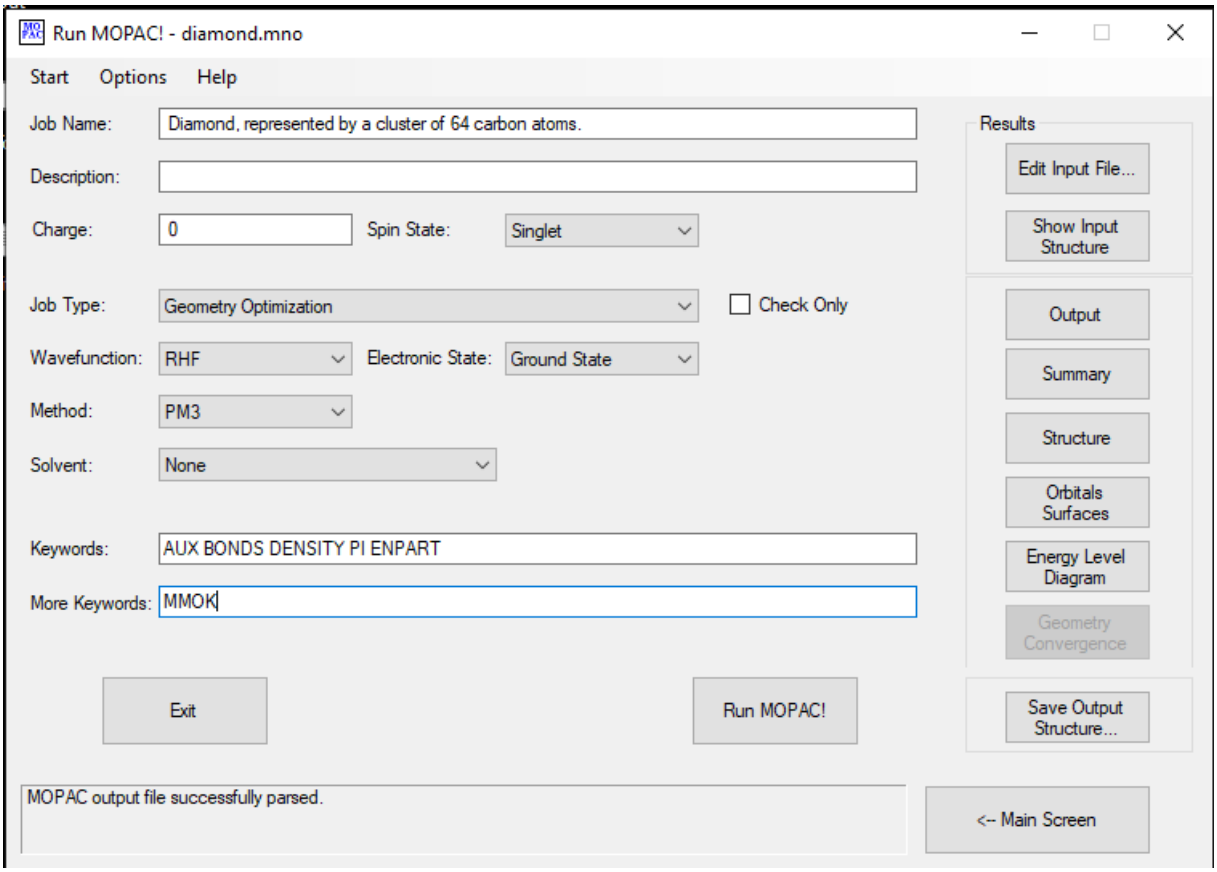
II.5. MoCalc2012

MoCalc2012 is a straightforward and efficient graphical user interface compatible with MOPAC, GAMESS(US), Firefly, and ORCA.

Key Features:

- Integrate with your preferred molecule editor, such as 'Avogadro' or ACD's 'ChemSketch,' for input generation.
- Load molecules from local files or import them directly from online databases.
- Supports input in various formats, including 2D, 3D, SMILES/InChI structures, and crystallographic data.
- Offers extensive filtering options for files containing multiple structures.
- Provides MM-based optimization and conformer search before conducting molecular orbital calculations.



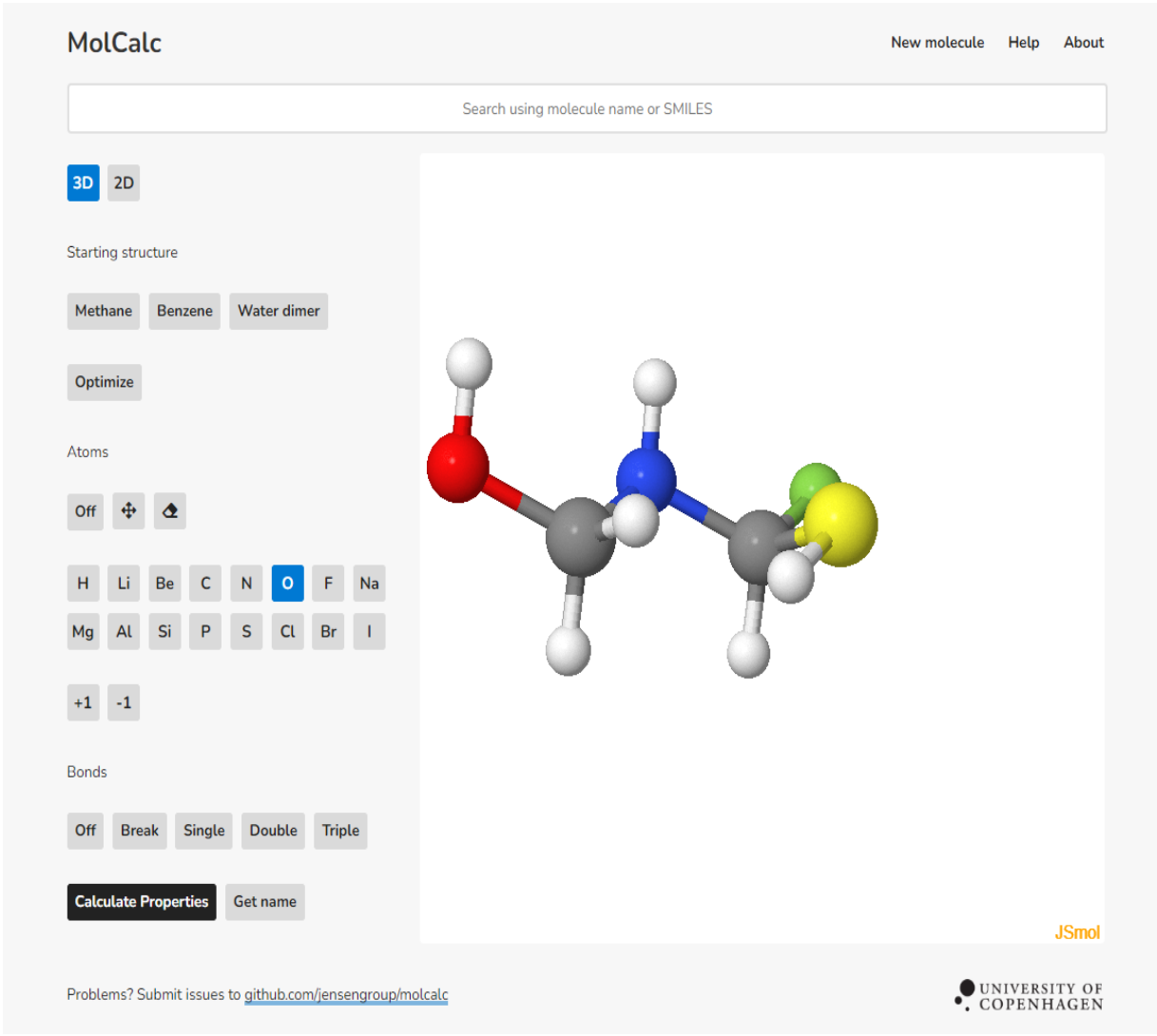


II.6. MolCalc: https://molcalc.org

MolCalc is a web-based interface that enables chemists to construct small molecules and estimate various molecular properties, including:

- Molecular structure
- Heats of formation and other thermodynamic properties
- Vibrational frequencies and modes
- Molecular orbitals and their energies
- Dipole moments and solvation surfaces

These estimations can be completed in seconds or minutes, depending on the size of the molecule.


MolCalc
New molecule
Help
About
Search using molecule name or SMILES
3D
2D
Starting structure
Methane
Benzene
Water dimer
Optimize
Atoms
Off
H
Li
Be
C
N
O
F
Na
Mg
Al
Si
P
S
Cl
Br
I
+1
-1
Bonds
Off
Break
Single
Double
Triple
Calculate Properties
Get name
JSmol
Problems? Submit issues to github.com/jensengroup/molcalc
UNIVERSITY OF COPENHAGEN

Chapter III:

Computational Software

III. Introduction

Computational software is a vital tool in the field of chemistry and materials science, facilitating the modeling, simulation, and analysis of molecular structures and reactions. These programs enable researchers to study complex systems at the atomic level, predict molecular behavior, and visualize intricate interactions that would be challenging to observe experimentally. By leveraging advanced algorithms and computational methods, such software allows for the exploration of various phenomena, including thermodynamics, kinetics, and electronic properties.

III.1. HyperChem:

HyperChem is a powerful computational chemistry software package that provides a user-friendly interface for molecular modeling and simulations. As described in the previous section, it offers a range of tools for optimizing molecular structures, performing quantum mechanical calculations, and visualizing molecular interactions. HyperChem's versatility makes it suitable for both educational purposes and advanced research applications, allowing users to explore complex chemical systems with ease.

III.2. Gaussian:

Gaussian 09W is one of the famous programs used in computational chemistry It can operate on both the Windows and Linux operating systems, and there are versions for both 32-bit and 64-bit architectures.

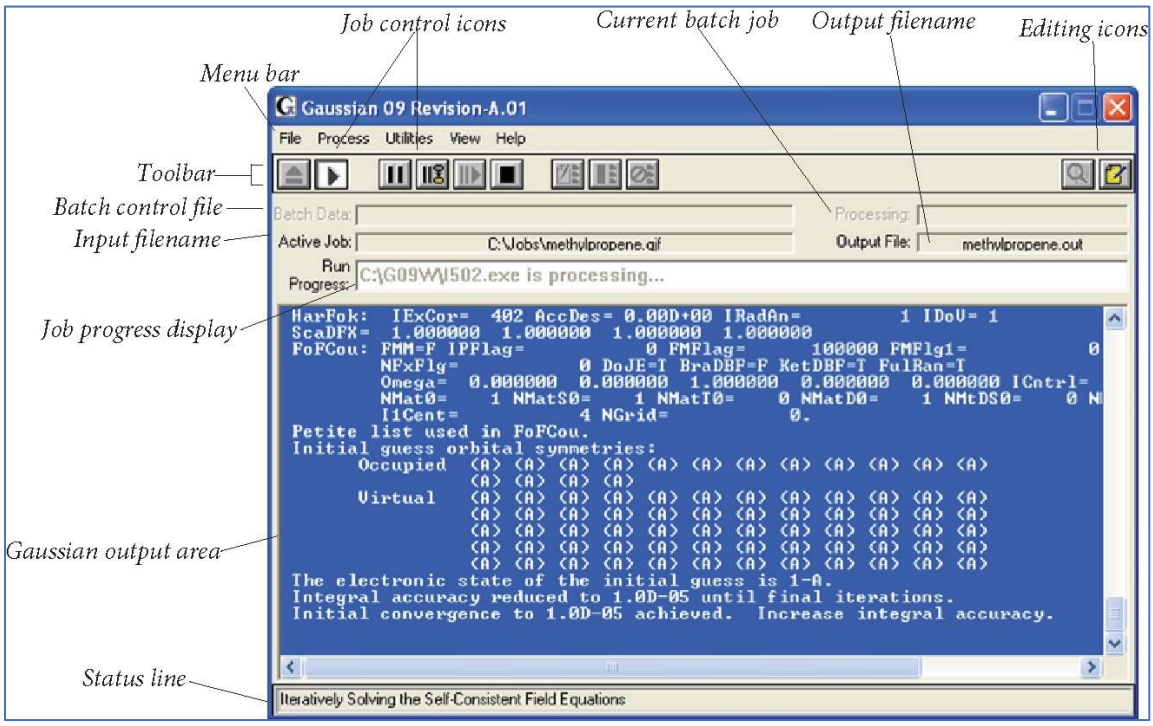
Some key points:
• Gaussian 09W is a widely used computational chemistry software package. • It is capable of running on both Windows and Linux operating systems. • There are versions of Gaussian 09W available for both 32-bit and 64-bit computer architectures. • As one of the leading programs in the field of computational chemistry, Gaussian 09W provides a range of advanced simulation and modeling capabilities for chemists and researchers.

The **Gaussian 09W** program can recognize input files with the following extensions:

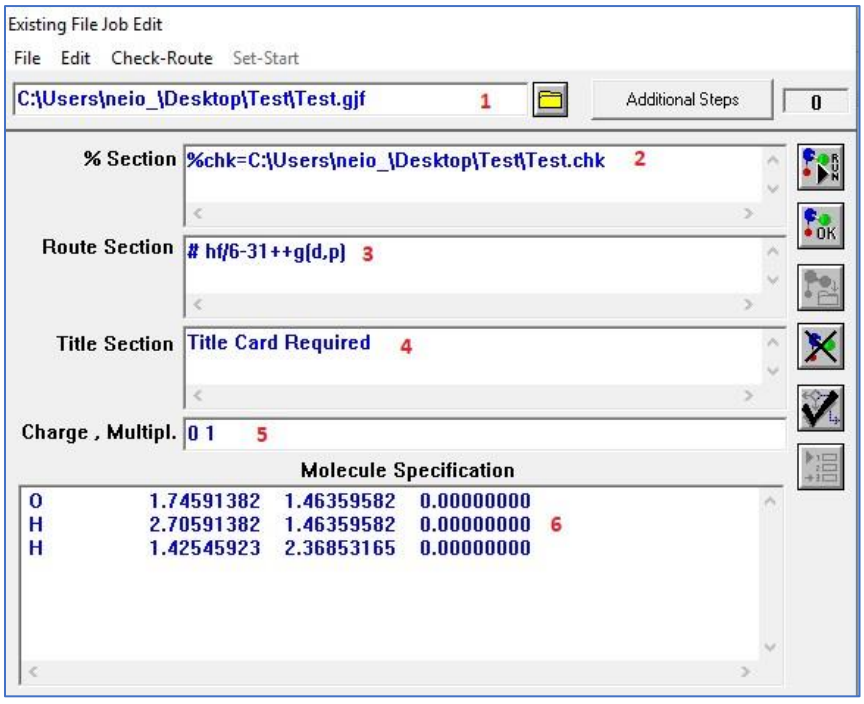
Gaussian Input File: **.gjf**

Batch Control File: **.bcf**

The main window "Job Processing" is the place from which Gaussian 09W jobs are controlled and executed, and where their output is displayed. Its main parts are described in the following illustration:



The



inputs can be entered manually or through a specific program. In general, regardless of the program used, the inputs have the following format.

This window can be opened by following the path: **File**...**Open**...

This window contains the following:

- **File path:** for the input file.
- **% Section:** contains the file path for the Checkpoint file and is called this. link 0 line.
- **Route Section:** contains information about the work to be performed and consists of:

#X Theoretical model/Basis set Type of calculation Options

 # : mandatory sign to begin route section.

X: Represents the amount of information generated and has the following options:

- (X=T) terse output,
- (X=N) normal output,
- (X=P) max output.

Method: The type of work to be performed, whether it is a **zero-point energy calculation** or **structure and vibration optimization**, using keywords such as **SP for zero-point energy** calculation.

Theoretical model: keyword that tells G09 which theoretical model to use (ex: RHF)

• Basis set: specifies the basis set to use (ex: 6-31G(d)).

• Type of calculation: specifies one or more keywords for G09 jobs to do

separated by a space.

hf/6-31g*

#p M06/cc-pvdz opt

#MP2/cc-pvtz maxdisk=12gb

HF/6-311g(d,p) freq scf=direct nosymm opt

#CCSD=(T,maxcyc=200)/6-31G* opt=ts

- **The title** and name of the work.

- **The charge and spin.**
- **Cartesian coordinates** (or other coordinates). These can be calculated and entered manually, but for large compounds, the process becomes complex and almost impossible, which is the function of the input programs used.

III.2.1. Spin Multiplicity:

The spin multiplicity, or simply multiplicity, measures the degeneracy of the electronic wave functions, that is, the number of those that differ only in the orientation of their spin. If S represents the resulting spin of the electrons in an atom, the multiplicity is **2S+1**. It results from Hund's rules, which favor the filling of degenerate orbitals each by a single unpaired electron, and measures the number of electronic configurations of the same energy.

III.2.2. MOLECULAR ORBITALS AND POPULATION ANALYSIS

Population analysis is a mathematical method for partitioning a wave function or electron density into charges on the atomic nuclei, bond orders, and other related information. These results are widely used, even though the calculated atomic charges are not directly observable experimentally.

Atoms do not actually have discrete partial charges, as they consist of a positive nucleus surrounded by a negatively charged electron cloud. However, breaking down the electron density and nuclear charges into partial atomic charges provides a way to understand the distribution of electron density. These partial charges are not whole number formal charges, but rather represent the fraction of an electron associated with each nucleus, based on the percentage of time an electron is near that nucleus.

Although this is an artificial assignment, it is very useful for predicting sites susceptible to nucleophilic or electrophilic attack, and for understanding molecular interactions in terms of ionic or covalent bonding, polarity, and so on. The partial charges correspond well to a chemist's intuitive view of these concepts. Only the most ionic compounds will have nearly whole number charges, while organometallic and organic compounds typically have smaller partial charges on the order of +/- 0.5 or less.

Pop=X keyword

X=NONE no orbital information displayed

X=REG HOMO-5 up to LUMO+5 orbital information displayed

X=FULL all orbitals information displayed

X=NBO Mulliken analysis is replaced by Natural Bond-Order analysis

X=MK, CHELP, OR CHELPG produce charges fit to electrostatic potential (ESP)

Note: The HF method is chosen as indicated in Table below.

Method	**RHF**	**ROHF**	**UHF**
Hatree-Fock	Restricted Hartree-Fock	Restricted Open shell Hartree-Fock	Unrestricted Hartree-Fock
For systems with multiplicity	1	All multiplicites	All multiplicites
Do the alpha and beta electronic orbitals have the same spatial part?	Yes	Yes	Pour la multiplicité 1, oui. Pour les autres multiplicités, non
Relative energy for multiplicity 1	Identical energy	Identical energy	Identical energy
Relative energy for a multiplicity different from 1	Not relevant	Greater than E(UHF)	Lower than E(ROHF)
Advantages for multiplicity different from 1	Not relevant	• The orbital analysis is simpler than UHF. • The total spin calculated for the system is accurat.	The calculated energy is lower than E(ROHF)
Disadvantages for multiplicity different from 1	Not relevant	The calculated energy is Greater than E(UHF)	The total spin calculated for the system is not accurate.

III.3. GAMESS:

GAMESS (General Atomic and Molecular Electronic Structure System) is a powerful program for ab initio molecular quantum chemistry. It can compute Self-Consistent Field (SCF) wave functions, including Restricted Hartree-Fock (RHF), Restricted Open Hartree-Fock (ROHF), Unrestricted Hartree-Fock (UHF), Generalized Valence Bond (GVB), and Multi-Configuration Self-Consistent Field (MCSCF) methods.

For correlation corrections to these SCF wave functions, GAMESS employs techniques such as Configuration Interaction (CI), second-order perturbation theory, Coupled-Cluster methods, and Density Functional Theory (DFT) approximations. Excited states can be calculated using CI, Equation of Motion (EOM), or Time-Dependent DFT (TD-DFT) procedures.

The program supports nuclear gradients, enabling automatic geometry optimization, transition state searches, and reaction path following. Additionally, the computation of the energy Hessian allows for the prediction of vibrational frequencies, along with infrared (IR) or Raman intensities.

Solvent effects can be modeled using discrete Effective Fragment potentials or continuum approaches like the Polarizable Continuum Model. GAMESS also offers a range of relativistic computations, including infinite-order two-component scalar relativity corrections and various spin-orbit coupling options. The Fragment Molecular Orbital (FMO) method makes it possible to apply advanced treatments to large systems by dividing them into smaller fragments. Furthermore, nuclear wave functions can be computed using Variational Self-Consistent Field (VSCF) methods or through explicit treatment of nuclear orbitals via the NEO code.

A wide variety of molecular properties can be calculated, from simple dipole moments to frequency-dependent hyperpolarizabilities. GAMESS includes numerous basis sets, effective core potentials, and model core potentials, allowing for the consideration of nearly the entire periodic table.

Most computations can be carried out using direct techniques or run in parallel on suitable hardware. For visualization of results, graphics programs such as MacMolplt (compatible with Macintosh, Windows, and Linux) are available, and the **Avogadro** program can assist with input preparation.

III.3.1. Input Description:

Let's Consider Some Examples:

Example 01:

```
!  input file for formaldehyde
$CONTRL SCFTYP=RHF RUNTYP=OPTIMIZE COORD=CART
   NZVAR=0 MULT=1 ICHARG=0 $END
 $SYSTEM TIMLIM=20000 MEMORY=10000000 $END
 $STATPT NSTEP=1000 $END
 $BASIS GBASIS=STO NGAUSS=3 $END
 $GUESS GUESS=HUCKEL $END
 $DATA
```

The input file consists of several groups, each beginning with a $GroupName and ending with $END. Even the DATA section specifies the nuclear framework and symmetry for the molecule.

It's important to note that the first column of any line (except comments) should be left blank. Comment lines in a GAMESS input file typically start with a !. Additionally, there should be no blank lines above or below the input text.

In the **$CONTRL** group, the following keywords are important:

SCFTYP: Possible values include RHF (Restricted Hartree-Fock), ROHF (Restricted Open-Shell Hartree-Fock), UHF (Unrestricted Hartree-Fock), and MCSCF (Multi-Configuration Self-Consistent Field).

RUNTYP: Options include ENERGY (for fixed-geometry single-point calculations), OPTIMIZE (to optimize nuclear geometry), and GRADIENT (to calculate gradients around the single-point geometry).

COORD: Options are CART (Cartesian coordinates) and ZMT.

NZVAR: Determines the method of geometry optimization, with NZVAR=0 indicating the use of Cartesian coordinates.

MULT: Represents the spin multiplicity of the molecular electronic state (1 for ground states of H_2, N_2, etc., and 3 for O_2).

ICHARG: Indicates the net charge of the molecule/ion (0 for HCHO, +1 for NH_4^+).

In the other groups, the following keywords are relevant:

TIMLIM: Sets the maximum time limit for calculations in minutes (use a large value).

MEMORY: For computers with >100 MB RAM, a value of ten million is often effective.

NSTEP: Specifies the number of steps for geometry optimization; a larger value (e.g., 1000) prevents premature termination of the calculation.

GBASIS: Specifies the Gaussian basis set, with options like STO, N21, N31, N311, DZV, etc.

NGAUSS: Indicates the number of Gaussians, applicable only to certain basis sets like STO, N21, N31, or N311. For example, NGAUSS=2 or 3 corresponds to the STO-2G or STO-3G basis sets, respectively. The combination GBASIS=N31 and NGAUSS=6 refers to the 6-31G basis set. To specify the 6-31G** (i.e., 6-31G(d,p)), add NDFUNC=1 and NPFUNC=1.

GUESS: Defines the type of initial guess for the molecular orbital. GUESS=HUCKEL is commonly used, while GUESS=MOREAD is recommended for MCSCF calculations, which requires an additional $VEC ... $END group.

This example focuses exclusively on Hartree-Fock (HF) methods, which do not account for electron correlation and thus provide approximate results.

Example 02:

```
! File created by the GAMESS Input Deck Generator Plugin for Avogadro
$BASIS GBASIS=AM1 $END
$CONTRL SCFTYP=RHF RUNTYP=HESSIAN $END
$FORCE METHOD=NUMERIC VIBANL=.TRUE. $END

$DATA
Title
C1
C 6.0 -3.72125 1.61155 -0.00008
H 1.0 -2.91624 1.16517 -0.57277
H 1.0 -3.36481 1.77649 1.01041
H 1.0 -3.95174 2.57633 -0.43758
H 1.0 -4.73438 -0.26554 0.43761
$END
```

Explanation of Input Sections:

$BASIS GBASIS=AM1 $END: This section defines the basis set for the calculation. The AM1 semi-empirical method is selected here.

$CONTRL SCFTYP=RHF RUNTYP=HESSIAN $END:

RUNTYP=HESSIAN: Indicates that the calculation will focus on computing the Hessian matrix, which is crucial for analyzing the curvature of the potential energy surface and determining vibrational frequencies.

$FORCE METHOD=NUMERIC VIBANL=.TRUE. $END:

METHOD=NUMERIC: Specifies that force calculations will be performed using numeric methods.

VIBANL=.TRUE.: This flag enables vibrational analysis, allowing for the calculation of vibrational frequencies.

III.4. MOPAC:

MOPAC is a versatile semi-empirical molecular orbital software designed for analyzing chemical structures and reactions. It employs semi-empirical Hamiltonians such as MNDO, MINDO/3, AM1, and PM3 in the electronic calculations to determine molecular orbitals, the heat of formation, and its derivative concerning molecular geometry. With these calculations, MOPAC can compute vibrational spectra, thermodynamic properties, isotopic substitution effects, and force constants for a wide range of molecules, radicals, ions, and polymers. Additionally, MOPAC includes routines for locating and optimizing transition states, making it valuable for studying chemical reactions. To effectively use MOPAC, users need to understand how the software functions, how to input data correctly, how to interpret the results, and how to troubleshoot any issues that arise.

III.4.1. Some keywords used in MOPAC

AIGIN	use a Gaussian zmatrix geometry specification
BIRADICAL	system has two unpaired electrons
BONDS	print bond order matrix
C.I.	configuration interaction (useful for excited states)
CHARGE=N	charge for ions
DFP	Davidson
DIPOLE	fit the ESP to the calculated dipole
DRC	dynamic reaction coordinate calculation
ESP	electrostatic potential calculation
ESR	calculate rhf unpaired spin density

EXCITED	optimize first excited singlet state
FORCE	force calculation (vibrational frequencies)
GEO	ok
HYPERFINE	hyperfine coupling constants to be calculated
IRC	intrinsic reaction coordinate calculation
K=(N,N)	one dimensional Brillouin zone structure (polymer)
LOCALIZE	print localized orbitals
MMOK	use molecular mechanics correction to CONH bonds
MULLIK	mulliken population analysis
NOMM	do not use molecular mechanics correction to CONH bonds
OPEN	open shell rhf calculation
PI	resolve density matrix into sigma and pi bonds
POLAR	calculate first, second and third order polarizablilities
PRECISE	criteria to be increased by 100 times
PULAY	use Pulay's converger to obtain a scf
RESTART	calculation restarted
ROOT=N	root n to be optimized in a C.I. calculation
ROT=N	symmetry number of the system is n
SADDLE	optimize transition state
SCALE	scaling factor for Van der Waals distance in ESP
SYMMETRY	impose symmetry conditions
UHF	unrestricted calculation
VECTORS	print final eigenvectors (molecular orbital coeficients)
XYZ	do all geometric operations in cartesian coordinates

III.4.2. Example of data for ethylene

```
Line 1 :    UHF PULAY MINDO3 VECTORS DENSITY LOCAL T=300
Line 2 :        EXAMPLE OF DATA FOR MOPAC
Line 3 :            MINDO/3 UHF CLOSED-SHELL D2D ETHYLENE
Line 4a:        C
Line 4b:        C  1.400118    1
Line 4c:        H  1.098326   1  123.572063  1
Line 4d:        H  1.098326   1  123.572063  1  180.000000   0   2   1   3
Line 4e:        H  1.098326   1  123.572063  1  90.000000    0   1   2   3
Line 4f:        H  1.098326   1  123.572063  1  270.000000   0   1   2   3
```

Line 5 :

As shown, the first three lines are textual. The first line contains keywords (in this example, seven keywords are used) that control the calculation. The following two lines are for comments or titles, where the user might want to include the molecule's name and the purpose of the run. These three lines are mandatory. If no name or comment is needed, leave these lines blank. Similarly, if no keywords are specified, ensure the first line is blank. A common mistake is inserting an extra blank line before the keyword line; this can be a tricky error to identify. Therefore, make sure not to have four lines before the geometric data begins (lines 4a-4f in the example). Regardless of content, these three lines, whether blank or filled, are required.

Chapter IV:
Practical Applications

Practical Exercise 01: Geometric Optimization

Background:

➢ **Quantum and Semi-Empirical Methods:** In quantum mechanics, the focus is on the distribution of electrons (orbitals) in space. The best programs include geometry optimization processes. The primary goal of quantum mechanics is to determine the energy and electron distribution. Molecular energies are calculated using the Schrödinger equation with molecular orbital (MO) formalism. The Schrödinger equation for a molecular system can be solved without approximation (ab initio) or with approximations (semi-empirical).

Examples:

- **Ab initio methods:** 3-21G, 6-31G, 6-31G (d, p) …
- **Semi-empirical methods:** AM1, PM3, CNDO, RM1, INDO…

➢ **Geometric Optimization:** The geometry of a molecule is stabilized by a set of forces acting on and between each atom. Starting from a very approximate geometry, the goal is to find the set of Cartesian coordinates that minimizes the sum of all energy contributions due to internal coordinate distortions and interactions between non-bonded atoms. In principle, it is sufficient to take the first derivative (Gradient) of the steric energy with respect to each degree of freedom of the molecule and find the point on the energy hypersurface where, for each coordinate r_i, $\left(\frac{\partial E}{\partial r}\right) = 0$. The procedures to achieve this are of two types:

- Some use only the slope of the surface (first derivative).
- Others use both the slope and the curvature of the surface (first and second derivatives).

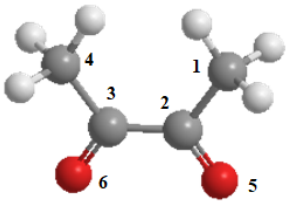
Almost all minimization methods have at least one common feature: they start at a given point on the hypersurface and descend towards the nearest minimum, without knowing if this minimum is local or absolute. Therefore, several starting conformations, in the form of internal coordinates, must be presented to the computer, drawing inspiration from molecular models. These methods use iterative algorithms.

Different algorithms are distinguished:

- The **"steepest descent"** method
- The **conjugate gradient** method
- The **Newton-Raphson** method
- The **simulated annealing** method ("anneal")

Practical Application:

1. **Construct the Molecule:** Using the HyperChem software, build the molecule depicted below.



2. **Energy Calculation:**
   - Calculate the energy of the molecule before optimization using a single point calculation.
3. **Geometric Optimization:**
   - Optimize the geometry using a semi-empirical calculation with the AM1 method and the Polak-Ribiere algorithm.
   - Optimize the geometry again using an ab initio calculation with the large 6-31G* basis set.
4. **Fill in the Tables:**
   - Complete the tables below using the output file data.
   - What do you observe?

	AM1	6-31G*
R(C2-C3)		
R(C2-O5)		
θ(C1-C2-C3)		
θ(O5-C2-C3)		
Φ(O5-C2-C3-O6)		
Φ(C1-C2-C3-C4)		
E_{Total}(u.a)		

5. **Determine the Optimal Structure:**
    - Define a dihedral angle composed of the two oxygen atoms and the two carbon atoms. The value of the dihedral angle will appear in the status bar.
    - Click on Compute / Potential. Enter -180, 180, 10, and click OK. After a few seconds, an energy graph will appear.
6. **Analysis:**
    - Which conformer is more stable?

Practical Exercise 02: Molecular Mechanics

Background:

The Force Field:

The total potential energy E_{tot} is calculated using an empirical model called a "force field," where the system is composed of "balls" (atoms) and "sticks" (bonds).

The total energy is expressed as follows:

$$E_{tot} = E_{stretching} + E_{bending} + E_{torsion} + E_{non-bonded}$$

AMBER Force Field (Assisted Molecular Building and Energy Refinement):

a) Bond Stretching Deformation

The bond stretching energy is given by:

$$E_{stretching} = \sum \frac{1}{2} K_r (r - r_0)^2$$

where *r* is the instantaneous bond length (in Å), r_0 is the equilibrium bond length, and K_r is the force constant (in kcal·mol^{-1}·Å^{-1}).

b) Angle Bending Deformation

The angle bending energy is given by:

$$E_{bending} = \sum \frac{1}{2} K_\theta (\theta - \theta_0)^2$$

where θ is the instantaneous angle (in °), θ_0 is the equilibrium angle (in °), and K_θ is the force constant associated with the angle.

c) Torsional Deformation

The torsion term is defined as:

$$E_{torsion} = \sum V_n(1 + \cos(n\phi - \gamma))$$

where V_n is the rotational barrier of order n (in kcal·mol^{-1}), γ is the phase (in °), and ϕ is the instantaneous dihedral angle (in °).

d) Van der Waals Interactions

Van der Waals interactions are represented by a 6-12 potential, expressed as:

$$E_{vdW} = \epsilon_{ij}\left[\left(\frac{R*_{ij}}{R_{ij}}\right)^{12} - 2\left(\frac{R*_{ij}}{R_{ij}}\right)^{6}\right]$$

where ϵ_{ij} and $R*_{ij}$ are the Lennard-Jones parameters for interactions between atoms i and j. ϵ_{ij} represents the depth of the Van der Waals potential well at the distance $R*_{ij}$. The term $\frac{1}{R_{ij}^6}$ represents the attractive London dispersion between two atoms, and the term $\frac{1}{R_{ij}^{12}}$ represents the repulsion due to Pauli exclusion.

e) Electrostatic Interactions

Electrostatic interactions are represented by a Coulomb potential:

$$E_{elec} = \frac{q_i q_j}{R_{ij}}$$

where $q_i q_j$ are the partial charges on atoms i and j.

Example: The MM+ Method

A MM+ calculation on H_2O provided the following results:

HyperChem log start -- Wed Apr 04 21:09:11 2018. Geometry optimization, Molecular Mechanics, molecule = (untitled). mmplus.

Polak-Ribiere optimizer

Energy=0.191862 kcal/mol Gradient=3.708928 Converged=NO (1 cycles 5 points).

Energy=0.000283 kcal/mol Gradient=0.252524 Converged=NO (2 cycles 8 points).

Energy=0.000000 kcal/mol Gradient=0.004079 Converged=NO (3 cycles 10 points). Energy=0.000000 kcal/mol Gradient=0.004079 Converged=YES (3 cycles 10 points). Bond=1.11918e-008 Angle=2.17996e-007 Dihedral=0 Vdw=0 Stretch-bend=-1.23592e-008 Electrostatic=0.

Bond Stretching:

The bond stretching term is calculated using the cubic stretch term:

$$E_{bond} = 143.88 \sum \frac{1}{2} k_r (r - r_0)^2 [1 + CS(r - r_0)]$$

Where CS=−2.0, k_r=4.6, and r_0=0.942 A˚

The calculation gives E_{bond}=1.0589E−08 (Bond=1.11918e-08).

Angle Bending:

The angle bending term is calculated as:

$$E_{bond\ angle} = 0.043828 \sum \frac{1}{2} k_\theta (\theta - \theta_0)^2 [1 + SF(\theta - \theta_0)^4]$$

Where SF is a scale factor = 7.0E-8, k_θ=0.4, and θ_0=104.5. The result is $E_{bond\ angle}$= =2.19E−07 (Angle=2.17996e-07).

Bond Stretch and Angle Bending:

This refers to the coupling between bond stretching and angle bending. If the angle is defined to include atoms *i*, *j*, and *k* (where k is the central atom), MM+ couples the bond stretching of *ik* and *jk* with the angle:

$$E_{stretch\,bend} = 2.51118 \sum_{angles} k_{sb}(\theta - \theta_0)_{ijk}[(r - r_0)_{ik} + (r - r_0)_{jk}]$$

If atom *i* or *j* is a hydrogen, then$(r - r_0)$= 0, so no stretch-bend interaction is defined for the XH_2 group.

For an atom other than hydrogen:

- k_{sb}= 0.120K_{sb} = 0.120Ksb=0.120 for XR_2 (X = first-row element of the periodic table)
- k_{sb}= = 0.090 for XRH
- k_{sb}= 0.250 for XR_2 (X = second-row element of the periodic table)
- k_{sb}= −0.400 for XRH

In our case: No stretch-bend interaction is defined for the XH_2 group. The expression simplifies to:

$$E_{stretch\,bend} = 2.51118 \sum_{angles} k_{sb}(\theta - \theta_0)_{ijk}[(r - r_0)_{ik}]$$

With X = O (second-row element of the periodic table), Ksb = 0.25, giving

$E_{stretch\,bend}$=−1.255E−08 (Stretch-bend=-1.23592e-08).

Dihedrals:

$$E_{dihedral} = \sum_{dihedral} \frac{V_1}{2}(1 + cos\ \phi) + \frac{V_2}{2}(1 + cos\ 2\phi) + \frac{V_3}{2}(1 + cos\ 3\phi)$$

In our case: $E_{dihedral}$= 0, Vdw= 0, and $E_{electrostatic}$ = 0

Practical Application:

- Build a molecular structure of methane (CH_4), using the molecule-building tools.
- Go to Setup > Molecular Mechanics and choose the MM+ force field.
- Optimize the geometry by selecting Compute > Geometry Optimization.
- Record the final total energy after the optimization.
- After the geometry optimization, open the output file with Notepad++
- Record each energy component for the molecule.
- Select one of the C-H bonds
- Go to Setup > Edit Parameters, and record the values of k_r and r_0
- Select an angle H-C-H
- Go to Setup > Edit Parameters, and record the values of k_θ and θ
- Repeat the same steps as in the previous example

Practical Exercise 03: Molecular Dynamics (MD)

Background:

Molecular dynamics (MD) is a simulation method used to calculate the movement of atoms and molecules over time. It is widely employed to simulate the properties of solids, liquids, and gases. MD allows for the calculation of physical and chemical properties of a system, assuming the interaction potential is known.

Principle of Molecular Dynamics (MD)

Molecular dynamics uses Newton's second law to describe the movement of a molecule over time:

$$\vec{F_i} = m_i . \vec{a_i}$$

Where:

- $\vec{F_i}$:is the force vector acting on atom iii.
- m_i : is the mass of atom iii.
- $\vec{a_i}$: is the acceleration vector of atom iii.

This equation indicates that the velocity and direction of atomic movement depend on the forces acting between atoms.

The force $\vec{F_i}$ acting on an atom iii at position $r_i(t)$ is determined by:

$$\vec{F_i} = \frac{dE(r_i)}{dr_i}$$

The total potential energy of the system is calculated using molecular mechanics techniques.

Typical MD Simulation Protocol:

1. **Creation of the initial molecular structure**
2. **Energy minimization**

3. **Heating of the system**
4. **Equilibration of the system to the study temperature**
5. **Trajectory production**
6. **Trajectory analysis**

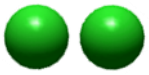
HyperChem can perform molecular dynamics simulations under periodic boundary conditions. The simplest type of simulation involves a periodic box containing water molecules, in addition to one or more solute molecules.

MD calculations can automatically average and record the following values:

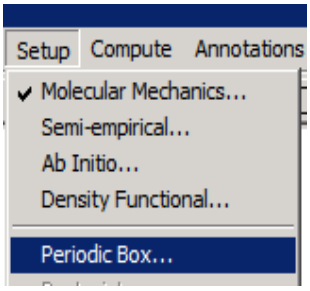
- **Kinetic energy (EKIN)**
- **Potential energy (EPOT)**
- **Total energy (ETOT)**
- **Temperature (TEMP)**

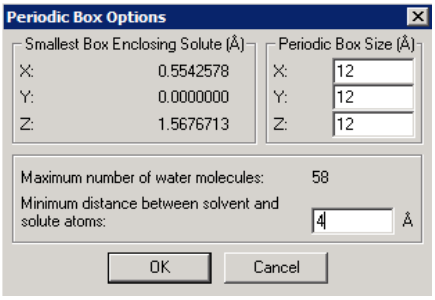
Practical Application:

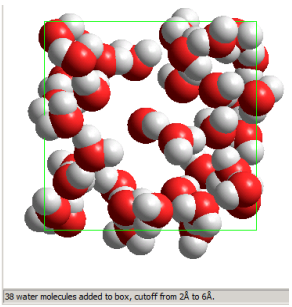
1. **Creating an Argon Dimer:** Start with the creation of an Argon dimer in an empty workspace.



2. **Geometry Optimization:** Use the MM+ force field to optimize the geometry of the dimer.
3. **Setting up a Periodic Box:** Place a periodic box around the solute molecule. The dialog box will prompt for the dimensions (x, y, z) of the periodic box and a "minimum distance" between water and solute molecules.

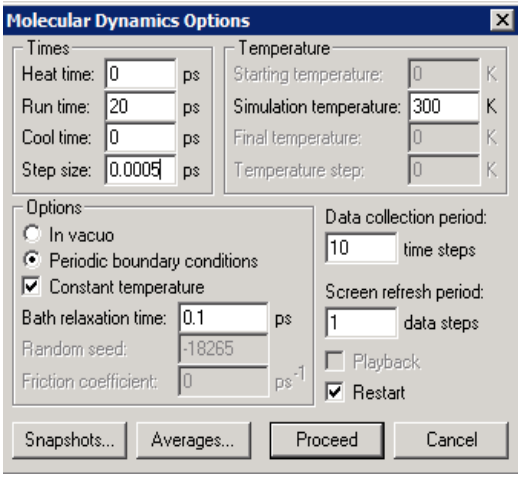


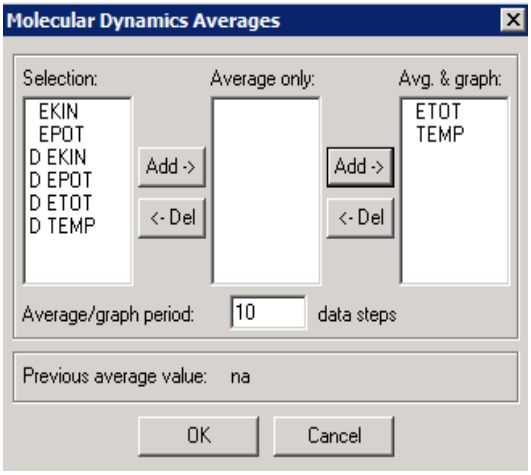




4. **Periodic Box Display:** The periodic box with added water molecules is displayed. In this case, 38 water molecules were added.

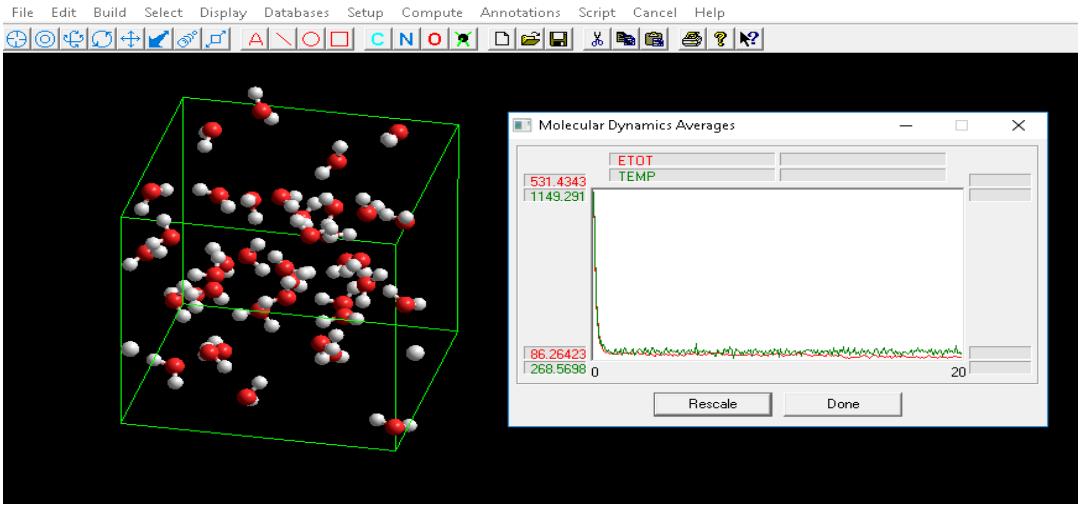
5. **Control Screens:** Use control screens to specify recording parameters during the simulation. A time step of 0.5 fs is typically small enough to ensure stability in most systems, but a larger time step may be advantageous.





After Calculation:

- **Rapid Relaxation:** Rapid relaxation near the initial state.
- **Equilibration:** Equilibrium is reached before 20 ps.



Note: To extract data for analysis in an external program (e.g., Excel), select either “Average Only” or “Average and Graph” in the “Averages” tab of “Molecular Dynamics Options.” At the end of the run, a file with a ".csv” extension is created in the HyperChem directory, which can then be imported into Excel for averaging and statistical uncertainty calculations.

Practical Exercise 04: Analyzing the Structural Properties of Cyclohexane Conformation

Background:

Cyclohexane, a six-membered ring structure, can adopt several conformations to minimize strain and achieve stability. The most important conformations of cyclohexane are:

1. Chair Conformation:

The most stable conformation of cyclohexane, where the carbon atoms are staggered, minimizing torsional strain. The chair conformation is free from angle strain (all C-C-C bond angles are approximately 109.5°) and torsional strain (all hydrogen atoms are staggered).

2. Boat Conformation:

The boat conformation is less stable than the chair due to torsional strain (from eclipsing interactions) and steric strain (from the flagpole interactions between the hydrogen atoms at the "bow" and "stern" of the boat). The boat conformation is typically an intermediate in the conversion between two chair conformations.

3. Twist-Boat (or Skew-Boat) Conformation:

The twist-boat conformation is a slightly more stable form of the boat conformation. It reduces some of the steric strain by twisting the ring, which decreases the interaction between the flagpole hydrogens and alleviates some torsional strain.

Summary of Stability:

Chair Conformation: Most stable.

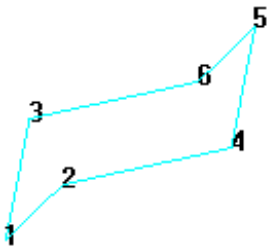
Twist-Boat Conformation: Less stable than the chair, but more stable than the boat.

Boat Conformation: Less stable, higher energy due to steric and torsional strain.

The chair conformation is the most energetically favorable and is the conformation that cyclohexane predominantly adopts under standard conditions.

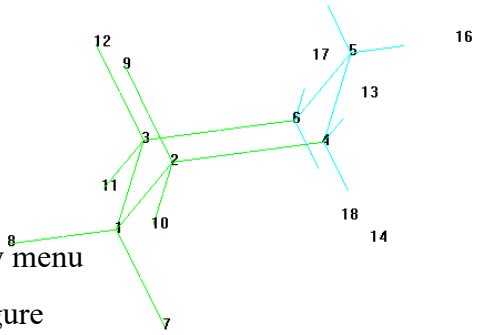
Practical Application:

Building Chair Cyclohexane:



- Build a six-membered ring structure
- Choose Add H & Model Build on the Build menu
- Turn off Show Hydrogens on the Display menu
- Use Molecular Mechanics to Optimize this structure and performing vibrational analysis at MM+ Level

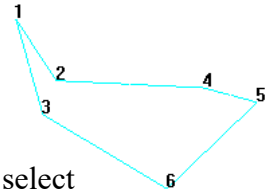
Transforming from Chair to Boat Cyclohexane:



- Select bonds 2-4 and 3-6
- From Name Selection, choose PLANE
- Turn on Show Hydrogens on the Display menu
- Extend the selection as shown in this Figure
- Choose Reflect on the Edit menu
- Optimize this structure and perform vibrational analysis at MM+ Level



Creating Twist Boat Cyclohexane



- Select bonds 1-2, 2-4, and 4-5
- Choose Constrain Bond Torsion on the Build menu, select
- Other, and enter 30
- Double-click on the Selection tool to invoke the Model Builder

Compare relative energies with experiment:

	Calculated	expt. Estimate (Kcal/mol)
E(chair)		1.33
E(boat)		8.31
E(Twist-boat)		7.22
ΔE (boat-chair)		6.9
ΔE (twistboat_chair)		5.3

Practical Exercise 05: Analyzing Molecular Orbitals

Application 01:

- Use GaussView to create CH_3CHO.
- Exit GaussView and modify the input file with a text editor to perform an energy optimization calculation using the 6-31g(d,p) basis set.
- Open the *.chk file in GaussView to visualize the HOMO and LUMO orbitals.

Application 02:

- Set up an HF calculation with the STO-3G basis set for the H_2 molecule with an internuclear distance of 0.74 Å.
- Save this file as H2.gjf.
- Launch and save the calculation using the Submit icon.
- Once the calculation is complete, use GaussView to open the H2.log file.
- Use Notepad to open the same H2.log file.

Results: Analysis and Significance of Calculated Quantities

Total Energy: -1.1175059 a.u.

Molecular Orbitals

The expression of the orbitals is given in the form:

Molecular Orbital Coefficients:			
	1	2	OMs (OM1=HOMO, OM2=LUMO)
	(SGG)—O	(SGU)—V	(σ_g *Simetry*)-Occupied (σ_u Simetry)- Vacant
Eigenvalues--	-0.59022	0.70064	Energy (u.a)
1 1 H 1S	0.54586	1.24623	OA Coefficients
2 2 H 1S	0.54586	-1.24623	

Energy of the MOs

The energy εi of each MO is the eigenvalue of the corresponding operator $\hat{F}$.

$$\hat{F}\varphi i \ = \ \varepsilon_i \varphi_i$$

The orbital OM1 is occupied, with energy -0.59022 (a.u.), and OM2 is empty with $\varepsilon 2 = 0.70064$ (a.u.).

The energy gap between HOMO and LUMO is

$\Delta E = |\varepsilon LUMO - \varepsilon HOMO| = 0.70064 - (-0.59022) = 1.29086$ a.u.

Wave Function of the MOs

For MOs of the CLOA type resulting from an HF calculation:

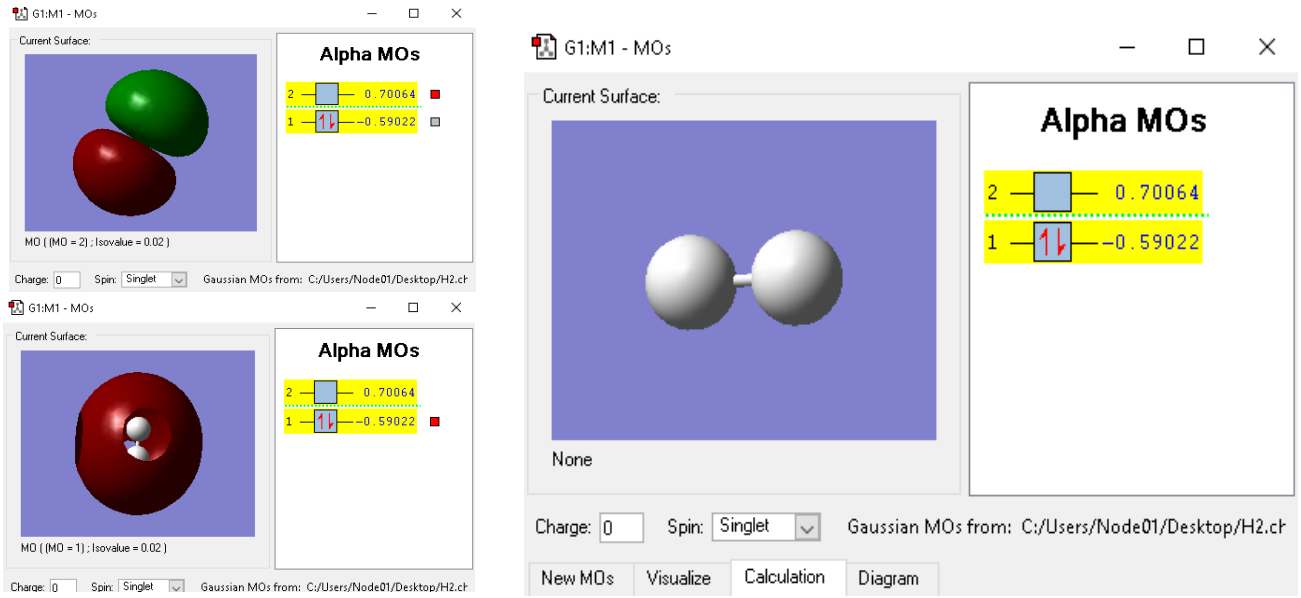
$$\Psi_i = \sum_q c_{qk} |\chi_q\rangle$$

The mathematical expression for OM1 is:

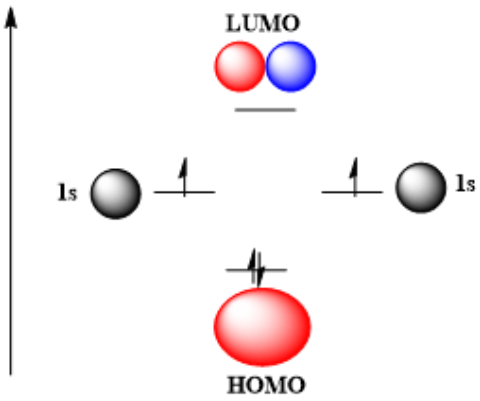
$\Psi 1 = 0.54586\ (1s)_1 + 0.54586\ (1s)_2$

Visualization of the MOs

To visualize the MOs, open the Checkpoint file: H2.chk with GaussView.



Energy Diagram



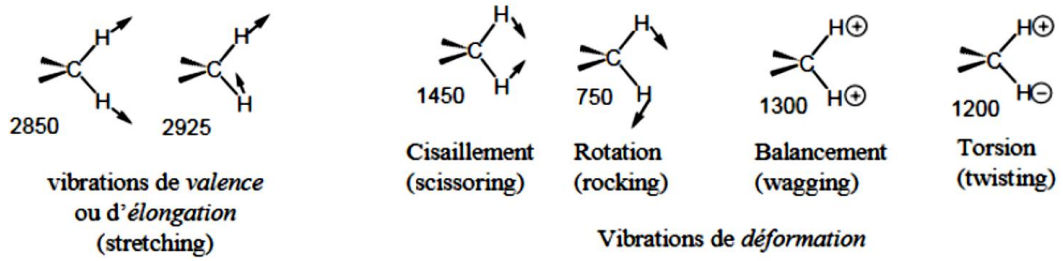
Practical Exercise 06: Vibrational Analysis

Background:

A molecular vibration occurs when the atoms in a molecule undergo periodic motion while the molecule as a whole experiences translational and rotational movement. The frequency of this periodic motion is called the vibration frequency. A non-linear molecule consisting of n atoms has $3n-63n-6$ normal modes of vibration, while a linear molecule has $3n-53n-5$ normal modes, as rotation around its molecular axis cannot be observed. A diatomic molecule thus has only one normal mode of vibration. The normal modes of polyatomic molecules are independent of each other, with each mode involving simultaneous vibrations of different parts of the molecule.

Vibrational Coordinates

The internal coordinates include the following types, as illustrated by the methane molecule, often referred to by English terms:



Application 1

Calculation of vibration frequencies for H_2

Job:

opt freq hf/sto-3g geom=connectivity

Compare the experimental and theoretical entropy of H_2

Experimental data: S = 31.26 cal/mol/K.

Application 2

a. Determine the equilibrium geometry of the formaldehyde molecule.

b. Complete the table.

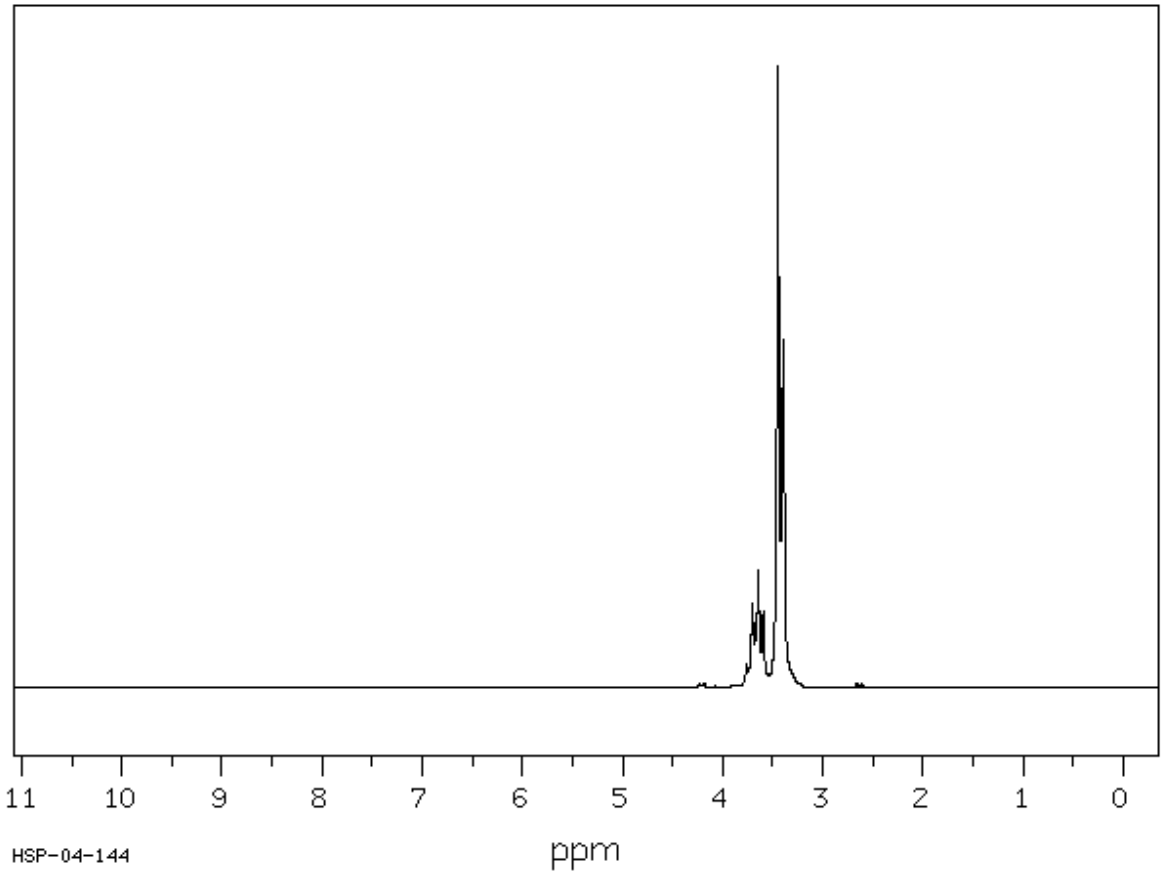
c. Compare the obtained vibration frequencies with the experimental values.

Type	Calculated Values (cm^{-1})	Experimental Values (cm^{-1})
		683
		898
		1290
		1173
		2847
		2968

Practical Exercise 07: NMR Chemical Shift Calculations

Application:

- Build the methanol structure using a graphical interface GaussView.
- Perform a geometry optimization using a B3LYP/6-31G*) basis set.
- run NMR calculation to obtain shielding tensors.
- Extract the chemical shift values from the output file and **compare them to experimental values to validate the calculation.**



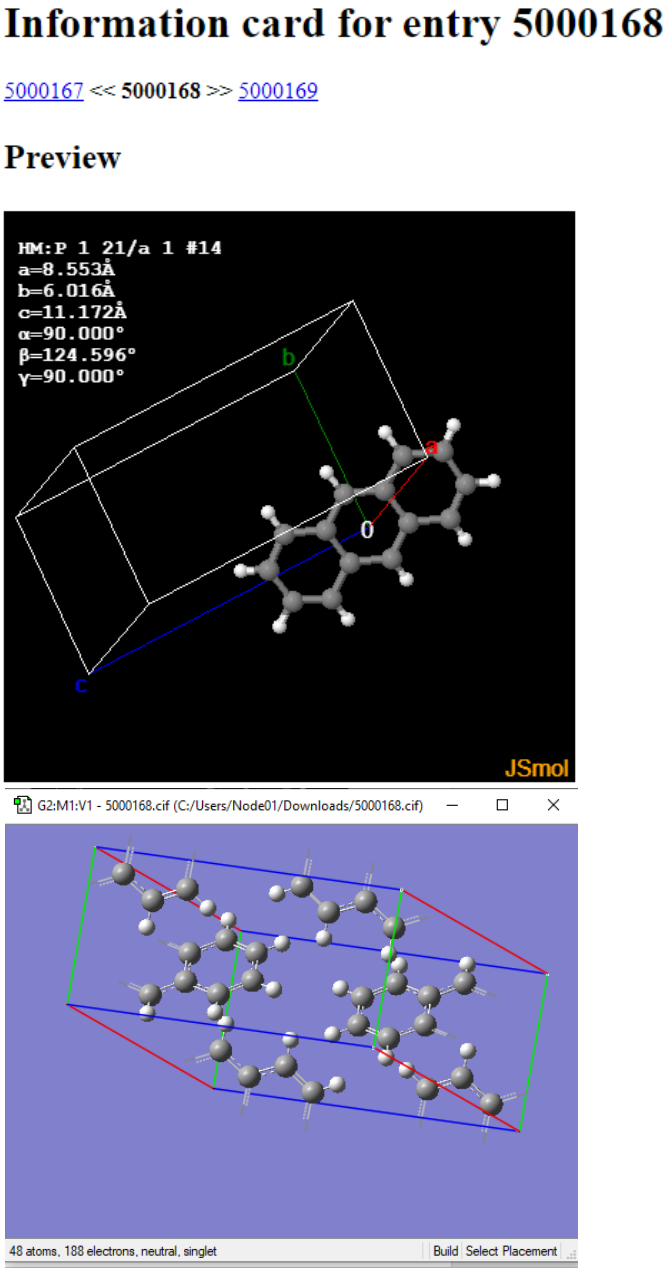
Peak DATA

Hz	ppm	Int.
336.81	3.761	36
331.25	3.699	135
326.13	3.642	188
321.75	3.593	124
308.63	3.447	1000
304.00	3.395	562

Practical Exercise 08: Predicting Melting Point

Application:

- Go to the Crystallography Open Database and search for the cristaline structure of Antracene by introducing the code ID: **5000168**, as show bellow.



- Download the file **5000168**.cif
- Open it using Gaussview
- Use Gaussian to perform a geometry optimization
- Run a frequency calculation to obtain thermodynamic data.
- From the output file, find the Gibbs free energy (G) and entropy (S) at a specified temperature (usually 298 K).
- Use the Gibbs-Helmholtz equation to estimate the melting point:

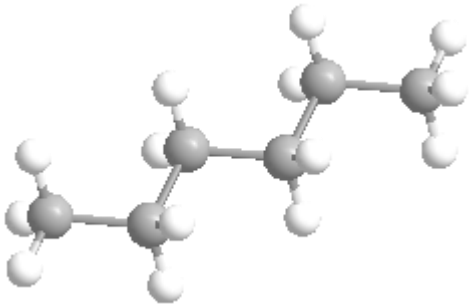
$$T_m = \frac{\Delta H}{\Delta S}$$

- Compare the predicted melting point with experimental values for validation

Practical Exercise 09: Estimating Viscosity of a Small Organic Molecule

Application:

- create the structure of the Hexane using the graphical interface GaussView



- perform a geometry optimization and frequency analysis.
- analyze the output files for:

  Molecular Weight: Determine the molecular weight of hexane.

  Density: Use literature values or other methods to estimate the density of hexane (typically around 0.655 g/cm³).
- Use empirical correlations or models to estimate viscosity based on molecular weight and density. A common empirical relationship for estimating viscosity ($\eta\eta$) of small alkanes at room temperature is:

$$\eta \approx 0.1 \times MW^{0.5} \times Density$$

Where MW is the molecular weight in g/mol and density in g/cm³.

- Compare your calculated viscosity with experimental viscosity data for hexane (which is approximately 0.36 mPa·s at 25°C) to assess the accuracy of your estimation.

Practical Exercise 10: Reaction Coordinate Calculation

Background:

When studying a reaction using IRC (Intrinsic Reaction Coordinate) calculations, it is essential to ensure that the calculated transition state (TS) leads to the correct reactants and products. Specifically, the following should be true:

1. The IRC calculation should proceed smoothly from the transition state to both reactants and products.

2. The transition state should connect the reactants and products through a single, negative curvature along the reaction coordinate, reflecting the correct transition path.

3. The transition state should represent the highest energy point on the reaction pathway, both towards the reactants and towards the products. If this condition is not met, the calculated transition state might not correctly represent the transition state of the desired reaction, potentially leading to an incorrect conclusion.

In some cases, it is possible to obtain a different transition state (TS') instead of the correct one (TS), leading to the formation of an unintended product rather than the desired product.

For example, starting from complex A and passing through the transition state TS, the reaction should ideally proceed to form complex B. However, if the reaction passes through an alternative transition state (TS'), it might instead lead to complex C.

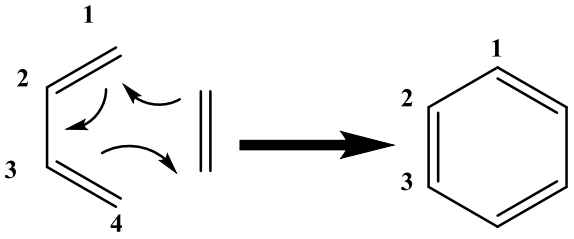
This highlights the importance of accurately performing and interpreting IRC calculations to ensure the correct understanding of the reaction mechanism.

Performing IRC Calculations

To perform Intrinsic Reaction Coordinate (IRC) calculations, it is necessary to obtain the transition state. Then, we conduct reverse calculations or calculations in both directions. The

calculations for the transition state can be carried out either towards the forward reaction or in the reverse direction.

We will study the Diels-Alder reaction, which is one of the well-known organic reactions and is typically of the cycloaddition type, where a cyclic compound is formed. The reaction occurs between a diene (a double-bonded alkene) and a substituted alkene (dienophile), resulting in a cyclic hexene. The mechanism of the reaction is as follows:



diene dienophile

Input for Intrinsic Reaction Coordinate (IRC) Calculations:

The input file for transition state calculations can be created using the keyword TS=Opt.

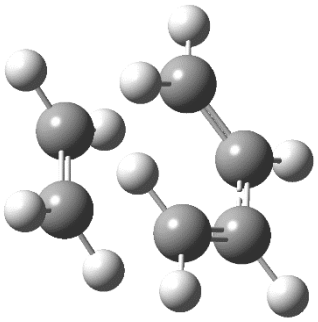
%chk=C:DieTS.chk

#p opt=(calcall,ts) freq=noraman hf/3-21g

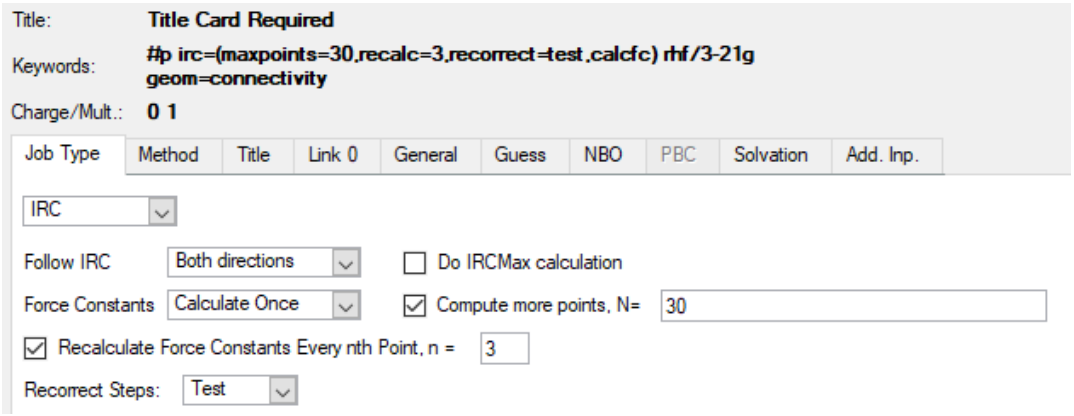
Title Card Required

0 1

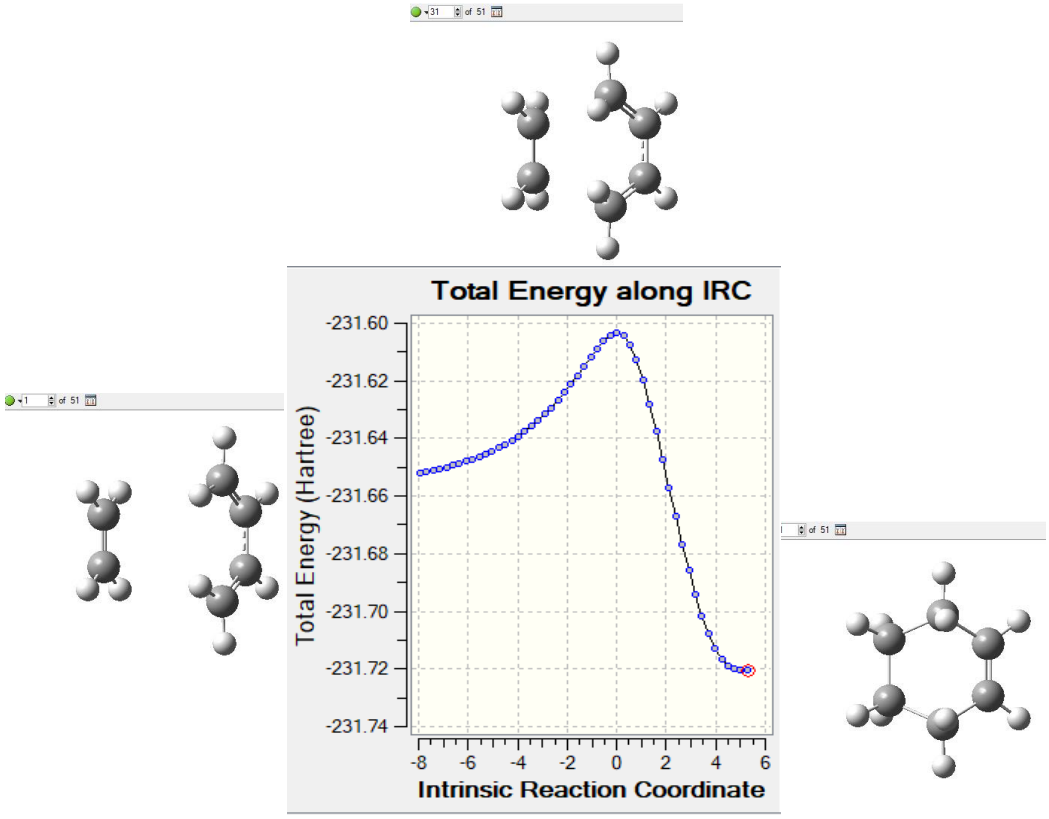
C	1.29370384	-0.69730288	-0.29061299
C	1.29378382	0.69715026	-0.29066736
H	1.82906798	-1.20632982	-1.07084495
H	1.82918834	1.20605187	-1.07095402
C	0.42785187	-1.41245742	0.49410264
H	0.35674899	-2.47840074	0.38093209
H	0.12348455	-1.04374717	1.45223531
C	0.42802902	1.41246439	0.49400852
H	0.35702912	2.47840338	0.38072827
H	0.12371178	1.04389355	1.45220210
C	-1.52968271	0.68799874	-0.23026678
H	-1.42394232	1.22313419	-1.15171182
H	-2.03755268	1.22079260	0.55150322
C	-1.52971709	-0.68787806	-0.23043324
H	-2.03772050	-1.22086007	0.55112087
H	-1.42382779	-1.22278800	-1.15199585



After performing the calculations, open the .chk file and enter the calculation as follows:



Although the calculation did not finish due to insufficient RAM, we obtained the following results:



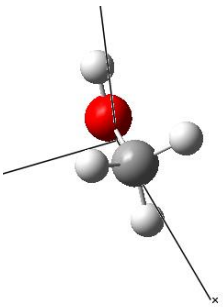
Practical Exercise 11: Studying the Effect of External Electric and Magnetic Fields on a Molecule Using Hyperchem and Gaussian

Effect of an External Electric Field using Gaussian:

- Use GaussView or another molecular editor to build the structure of the methanol (CH_3OH)
- Save the structure as a Gaussian input file (.gjf).
- Perform an initial geometry optimization without any external field:

opt rhf/3-21g

- Record the optimized geometry, dipole moment, and HOMO-LUMO gap from the output file.



Applying an External Electric Field

- Display the Cartesian axis from view
- Adjust the Gaussian input file to apply an external electric field along a specified axis (e.g., the x-axis). Add the Field keyword to specify the field strength:

opt rhf/3-21g nosymm geom=connectivity field=x+0.001

- The Field=x+0.001 keyword applies an electric field of 0.001 **atomic units** in the positive x direction. (514 V/nm =1 a.u)
- Compare the dipole moment, HOMO-LUMO gap, and optimized geometry with those obtained without the field.
- Perform additional calculations by varying the electric field strength (e.g., X+0.002, X+0.003, X+0.004, etc.).
- Observe how increasing or decreasing the field strength affects the molecular properties.
- Plot the dipole moment and energy gap as a function of the field strength.

Effect of an External Magnetic Field using Gaussian:

- Magnetic effects on the molecule, can be studied by using specific Gaussian keywords that affect magnetic properties, such as NMR

- Compare the NMR chemical shifts and other magnetic properties with those calculated under different conditions.

Effect of an External Electric Field using Hyperchem:

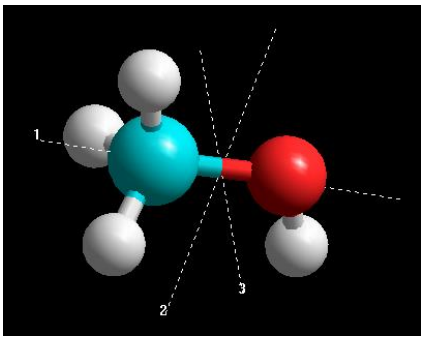
Follow the previous steps

- To display the Cartesian axes by navigating to **Display > Show Inertial Axes**.

To Set Up the External Field:

- Go to **Setup** and select **Set Finite Field**. A window will pop up where you can specify the parameters for either an electric field or a magnetic field.

- Choose to apply an electric field in the x direction with a magnitude of 0.01 a.u.



Select Basis Set and Perform Geometry Optimization:

- Choose the 3-21G basis set for your calculations.

- Perform a geometry optimization with the applied electric field.

Analyze and Compare Results:

- Compare the dipole moment, HOMO-LUMO gap, and optimized geometry with the values obtained from the calculation without the external field.

Varying the Electric Field:

- Perform additional calculations by varying the electric field strength (e.g., X+0.002, X+0.003, X+0.004, etc.).

- Plot the dipole moment and energy gap as a function of the field strength to observe the trends.

Repeat with Magnetic Field:

- Follow the same procedure using a magnetic field instead of an electric field. Observe and compare how the external magnetic field influences the molecular properties.

Practical Exercise 12: The Effect of water on methanol Using Avogadro and GAMESS

Application

- Use Avogadro to create the structure of methanol.

Optimize the Structure

- Perform a molecular mechanics (MM) calculation to optimize the methanol structure.

Generate the GAMESS Input File

- Export the optimized structure as a GAMESS input file.

Edit the Input File

- Open the input file using Notepad++. Modify it as follows:

```
$CONTRL SCFTYP=RHF RUNTYP=OPTIMIZE NZVAR=12 $END
$SYSTEM TIMLIM=2 MWARDS=2 $END
$BASIS GBASIS=N31 NGAUSS=6 NDFUNC=1 $END
$GUESS GUESS=HUCKEL $END
$ZMAT IZMAT(1)=1,1,2, 1,2,3, 1,3,4, 1,3,5, 1,3,6,
      2,1,2,3, 2,2,3,4, 2,2,3,5, 2,2,3,6,
      3,1,2,3,4, 3,1,2,3,5, 3,1,2,3,6 $END
$STATPT OPTTOL=1D-5 $END
$DATA

Cs

H 1.0   -1.0616171503   0.8036449245   0.0000000000
O 8.0   -0.6870131482  -0.0653470836   0.0000000000
C 6.0    0.7093551399   0.0291827007   0.0000000000
H 1.0    1.0836641283   0.5408321444   0.8835398105
H 1.0    1.0975386849  -0.9797829903   0.0000000000
$END
```

Save the File

- Save the modified input file as "methanol_gas.inp".

Run the Batch File Maker

- Go to the WinGAMESS directory and run the batch file maker. Select your input file and save the job.

Execute the Job File

- Run the job file to generate the output.

Modify the Input File for Solvent Effects

- Edit the input file again to include solvent effects as follows:

```
$CONTRL SCFTYP=RHF RUNTYP=OPTIMIZE NZVAR=12 $END
$SYSTEM TIMLIM=2 MWARDS=2 $END
$PCM SOLVNT=WATER $END
$BASIS GBASIS=N31 NGAUSS=6 NDFUNC=1 $END
$GUESS GUESS=HUCKEL $END
$ZMAT IZMAT(1)=1,1,2, 1,2,3, 1,3,4, 1,3,5, 1,3,6,
      2,1,2,3, 2,2,3,4, 2,2,3,5, 2,2,3,6,
      3,1,2,3,4, 3,1,2,3,5, 3,1,2,3,6 $END
$STATPT OPTTOL=1D-5 $END
$DATA

Cs

H 1.0   -1.0616171503   0.8036449245   0.0000000000
O 8.0   -0.6870131482  -0.0653470836   0.0000000000
C 6.0    0.7093551399   0.0291827007   0.0000000000
H 1.0    1.0836641283   0.5408321444   0.8835398105
H 1.0    1.0975386849  -0.9797829903   0.0000000000
$END
```

Save the New Input File

- Save this modified input file as "methanol_water.inp".

Repeat the Steps

- Follow the same steps to generate the output file for the new input.

Compare Results

- Compare the results from "methanol_gas.out" and "methanol_water.out" to analyze the differences between the gas-phase and solvent-phase calculations.

Practical Exercise 13: Calculating the thermodynamics properties of molecules

Application

1. Use a web browser to access The Molecule Calculator (MolCalc) website at https://molcalc.org/.

2. Explore the molecule-building tools available on the entry page.

3. Create a diatomic molecule (O_2) using the molecule-building function.

4. Click on "Calculate" to determine the molecule's minimum energy. Two types of calculations are performed using GAMESS:

(a) GAMESS calculates molecular orbitals and orbital energies at the RHF/STO-3G level of theory (RHF = restricted Hartree-Fock, STO-3G is the basis set).

(b) GAMESS computes thermodynamic properties, vibrational frequencies, and modes using the semi-empirical PM3 method.

5. Click on "Thermodynamics" and record the following values:

(a) Heat of formation (compare with the accepted value).

(b) Enthalpy.

(c) Entropy.

(d) Translational enthalpy.

(e) Rotational enthalpy.

(f) Vibrational enthalpy.

(g) Translational entropy.

(h) Rotational entropy.

(i) Vibrational entropy.

6. Click on "Molecular Orbitals" and "View" the orbitals:

(a) Create an energy level diagram for the valence molecular orbitals.

(b) Record the energies of the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO).

(c) Calculate the wavelength of light required to induce a HOMO to LUMO transition.

(d) Sketch the shapes of the HOMO and LUMO.

7. Click on "Vibrational Frequencies" and view the animated vibrational motions:

(a) Record the vibrational frequencies and compare them to accepted values.

8. Repeat steps 4-7 using Avogadro and GAMESS softwares.

Practical Exercise 14: Predicting the Refractive Index using MOPAC

Background:

The refractive index (often denoted as nn) is a dimensionless number that describes how light propagates through a medium. It is defined as the ratio of the speed of light in a vacuum to the speed of light in the medium. Mathematically, it can be expressed as:

$$n = \frac{c}{v}$$

where:

- c: is the speed of light in a vacuum (approximately 3.00×1083.00×108 m/s),
- v: is the speed of light in the medium.

The refractive index of a vacuum is defined as n=1. For most materials, the refractive index is greater than 1. Common values for glass, water, and air are approximately 1.5, 1.33, and 1.0003, respectively.

Typical Values: The refractive index of common materials varies:

Air: ~1.0003

Water: ~1.33

Glass: ~1.5 to 1.9

Diamond: ~2.42

2. Comparative Analysis

Material Comparison: Compare the calculated refractive index with known values for similar materials. This helps contextualize the results.

Example: If you calculate a refractive index of 1.55 for a polymer, compare it with:

Polycarbonate: ~1.58

PMMA (Acrylic): ~1.49

Material Properties and Applications

Optical Materials: Assess the suitability of the material for optical applications (e.g., lenses, coatings). A low refractive index may be desirable for anti-reflective coatings, while a high index may be beneficial for high-index lenses.

Biocompatibility and Transparency: In applications like medical devices or coatings, ensure the material's refractive index aligns with desired transparency and biocompatibility.

Here's a step-by-step example of how to calculate the refractive index.

Step 1: Calculate Molecular Polarizability

1. **Obtain the Polarizability**: After running a calculation in MOPAC, extract the polarizability (α) from the output. This value may be given directly or can be calculated from the atomic coordinates and charge distribution.

   For example, let's say the polarizability obtained is:

   α=10.0 A˚3α=10.0A˚3

Step 2: Use the Lorentz-Lorenz Equation

The refractive index nn can be estimated using the Lorentz-Lorenz equation:

$$n^2 = \frac{1 + 2 \cdot \eta}{1 - \eta}$$

where:

- η: is the volume fraction of the polymer in the mixture.
- The volume fraction can be calculated as:

$$\eta = \frac{N \cdot \alpha}{V}$$

where:

- N: is the number density of the molecules (number of molecules per unit volume).
- V: is the molar volume of the polymer.

Step 3: Calculate Number Density

1. **Determine Number Density**: If you know the molecular weight (M) of the polymer and its density (ρ), you can find the number density N:

$$N = \frac{\rho}{M} \cdot N_A$$

where N_A is Avogadro's number (6.022×10^{23} mol^{-1}).

Step 4: Combine Values

1. **Calculate the Volume of the Polymer**: If you have the molar volume V_m of the polymer, then:

$$V = \frac{M}{\rho}$$

2. **Plug values into the Lorentz-Lorenz equation**.

Example Calculation

Assuming:

- Molecular weight M=1000 g/mol
- Density ρ=1.2 g/cm^3
- Polarizability α=10.0 $A^{\circ 3}$=10.0×10^{-24} cm^3

Calculate Number Density:

V_m=1000/1.2≈833.33 cm^3/mol

N=’1.2 g/cm^3/1000 g/mol) ·6.022×10^{23}≈7.27×10^{21} molecules/cm^3

Calculate Volume Fraction:

η= $(7.27 \times 10^{21}) \cdot (10.0 \times 10^{-24})/833.33 \approx 8.73 \times 10^{-3}$

1. **Calculate Refractive Index**:

n^2≈1.026

n≈1.013

Practical Application:

Assuming:



- Molecular weight of PVC M= 250 g/mol
- Density ρ=1.4 g/cm^3

1. Using Chem3D software, perform a geometry optimization of PVC using the PM7 approximation.
2. Perform a properties calculation, including the polarizability calculation: use the following keywords: AUX PM7 1SCF CHARGE=0 STATIC SHIFT=80 (use STATIC instead POLAR)
3. Calculate the refractive index.

Practical Exercise 15: Estimating Elastic Modulus of a Polymer

Background:

Young's modulus (E) measures a material's stiffness or resistance to elastic deformation when a tensile or compressive load is applied. It is defined as the ratio of tensile stress to tensile strain.

It is typically expressed in pascals (Pa), megapascals (MPa), or gigapascals (GPa)

Comparative Analysis

Material Comparison: Compare the obtained elastic modulus with known values for similar materials (e.g., other polymers, metals, composites). This helps to understand where the material stands in terms of stiffness.

Example: If the calculated Young's modulus for a polymer is 6.49 GPa, you can compare it to common materials:

Polyethylene: ~ 0.2–1.0 GPa

Polystyrene: ~ 3–4 GPa

Steel: ~ 200 GPa

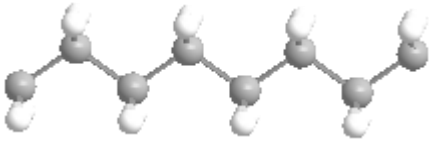
Analyzing the Material Behavior

Elastic vs. Plastic Behavior: A higher elastic modulus indicates that the material will deform less under a given load, remaining in the elastic region longer. If the modulus is low, the material may be more prone to plastic deformation under stress.

Temperature Dependence: Consider how temperature might affect the modulus. Polymers typically exhibit lower Young's moduli at higher temperatures due to increased molecular mobility.

Practical Application:

Let’s consider the polyethylene (PE), and outline the steps to calculate the elastic modulus.



1. Optimize the structure of the polymer at PM7 level using MOPAC software.

2. After optimization, create a new input file for vibrational frequency calculations.
3. Open the output file from the frequency calculation and look for the normal modes and their corresponding frequencies (the lowest one).
4. Calculate the Force Constants: The force constants can be derived from the vibrational frequencies. If ν is the frequency in cm−1, the force constant k can be estimated by:

$$k = \left(\frac{2\pi\nu}{c}\right)^2 \cdot \mu$$

where:

- c is the speed of light (approximately 2.998×10^{10} cm/s,
- μ is the reduced mass of the vibrating atoms.

The reduced mass μμ for two atoms is calculated as:

$$\mu = \frac{m1 \cdot m2}{m1 + m2}$$

5. Estimate Young's Modulus using a simplified approach based on the relationship between force constant and Young's modulus:

$$E = \frac{k}{L}$$

where:

E: is Young's modulus,

L: is the bond length or the distance over which the force is applied.

k (N/m), L(m).

6. Conclude your findings.

REFERENCES:

1. Applications Of Molecular And Quantum Mechanics 2nd, Errol G. Lewars, 2011.
2. HyperChem Release 7.0 for Windows: Reference Manual. (2002). Canada: Hypercube, Incorporated.
3. Frisch, M. J. (2009). Gaussian 09 User's Reference. États-Unis: Gaussian, Incorporated.
4. MOPAC Manual. (Seventh Edition). A General Molecular Orbital Package. (1993). États-Unis: Defense Technical Information Center.
5. Chem3D User Guide. 1998-2023 PerkinElmer Informatics, Inc.
6. Avogadro: an open-source molecular builder and visualization tool. Version 1.XX. http://avogadro.cc/
7. Marcus D Hanwell, Donald E Curtis, David C Lonie, Tim Vandermeersch, Eva Zurek and Geoffrey R Hutchison; "Avogadro: An advanced semantic chemical editor, visualization, and analysis platform" *Journal of Cheminformatics* **2012**, 4:17.
8. GAMESS,Recent developments in the general atomic and molecular electronic structure system,152, 0021-9606,doi :10.1063/5.0005188. website: https://www.msg.chem.iastate.edu/index.html
9. MoCalc 2012: Depizzol DB, Paiva MH, Dos Santos TO, Gaudio AC. MoCalc: a new graphical user interface for molecular calculations. J Comput Chem. 2005 Jan 30;26(2):142-4. doi: 10.1002/jcc.20151. PMID: 15584073.
10. MolCalc: J. H. Jensen and J. C. Kromann, The Molecule Calculator: A Web Application for Fast Quantum Mechanics-Based Estimation of Molecular Properties, J. Chem. Educ., 2013, 90 (8), pp 1093–1095. DOI: 10.1021/ed400164n. website: https://molcalc.org/
11. Crystallography Open Database: https://www.crystallography.net
12. NMR Database: https://sdbs.db.aist.go.jp/SearchInformation.aspx
13. MOPAC2016: http://openmopac.net/home.html
14. https://qm-tutorial.readthedocs.io/en/latest/index.html
15. https://mason.gmu.edu/~sslayden/Chem350/manual/docs/chem3d-pro10.pdf