

Chemistry Lab: Computational Chemistry 2

Nakhyun Kim (kimnh97@snu.ac.kr),
Jeonghyeon Kim (jasonkjh@snu.ac.kr),
Hong Suk Kim (khs01654@snu.ac.kr),
Byunghyun Bae (bbh9955@snu.ac.kr)

October 19, 2023

1 Objectives of the experiment

- Understand the basic theories of computational chemistry and calculation algorithms.
- Conduct Hartree-Fock (HF) calculations for monomers of small molecules (H₂O, NH₃, CH₄) and find the optimized geometry and energy.

2 Introduction

Computational chemistry is a branch of chemistry that lies between theory and experiment and solves and interprets the equations of various systems with the help of computers, providing various microscopic insights. In this lecture, we will perform quantum chemical (QC) calculation, specifically using Hartree-Fock method, to determine the stable structures of several molecules.

2.1 Quantum mechanics

According to quantum mechanical interpretation, all materials have the duality of waves and particles. In particular, microscopic materials have significant wave properties, so they are mainly expressed as a wave function. Therefore, mathematical and physical descriptions with calculus and linear algebra are required to deal with the wave function. The basic postulates are as follows. For additional information, we recommend that refer to the textbook.

Postulate 1. The state of quantum-mechanical system is completely specified by a function ψ of the coordinates and the time. This function ψ called the wave function or the state function, contains all the information that can be determined about the system.

$$|\psi(x, t)|^2 dx: \text{probability that the particle lies in the interval } dx \quad (1)$$

$$|\psi(x, y, z, t)|^2 dx dy dz: \text{probability that the particle lies in the interval } dx, dy, \text{ and } dz \quad (2)$$

Postulate 2. To every observable in classical mechanics, there corresponds an operator in quantum mechanics. Observable can be obtained by calculating operators on the wave function.

$$\text{Momentum operator in x direction: } \hat{P}_x = -i\hbar \frac{\partial}{\partial x} \quad (3)$$

$$\text{Total energy operator: } \hat{H} = -\frac{\hat{P}^2}{2m} + V(\mathbf{r}, t) \quad (4)$$

Postulate 3. In any measurement of the observable associated with the operator, the only values that will ever be observed are the eigenvalues.

$$\hat{A}\psi_n = a_n\psi_n \quad (5)$$

Postulate 4. If a system is in a state described by a normalized wave function ψ , then the average value of the observable corresponding to \hat{A} is given by

$$\langle a \rangle = \int_{-\infty}^{\infty} \Psi \hat{A} \Psi dx. \quad (6)$$

Postulate 5. The wave function of a system evolves in time according to the time-dependent Schrödinger equation.

$$i\hbar \frac{\partial \Psi}{\partial t} = \hat{H} \Psi \quad (7)$$

2.2 Atomic Orbitals, LCAO-MO and Basis sets.

The goal in quantum chemistry is to mathematically obtain the wave function or state function $\psi_n(r)$, which is an eigenfunction of the Hamiltonian operator, and to know the physical and chemical properties of molecules with this function. The Schrödinger equation for electrons using the Born-Oppenheimer approximation is as follows. Starting from the first term, it represents the kinetic energy of the electron, the interaction between the electron and the nucleus, and the interaction between electrons, respectively.

$$\hat{H}_{el}\psi_n(\mathbf{r}) = \left[-\frac{\hbar^2}{2m_e} \nabla^2 - \sum_{i=1}^N \sum_{A=1}^M \frac{Z_A e^2}{4\pi\epsilon_0 r_{iA}} + \sum_{i=1}^N \sum_{j>i}^N \frac{e^2}{4\pi\epsilon_0 r_{ij}} \right] \psi_n(\mathbf{r}) = E_n \psi_n(\mathbf{r}) \quad (8)$$

However, it is impossible to obtain a mathematically accurate solution for atoms and molecules with two or more electrons due to interactions between electrons in atoms or molecules (the third term in the middle). The only cases in which it can be solved directly are systems with only one electron, such as a hydrogen atom. The wave function for one electron obtained in this system is called an orbital, and the orbitals obtained for one atom are called atomic orbitals (AO).

In the case of atoms or molecules more complex than hydrogen, an exact solution cannot be obtained, so instead, the wave function is constructed by introducing the linear combination of atomic orbitals (LCAO) approximation. The orbitals of the actual system can be approximately obtained by optimizing the coefficients of linearly combined atomic orbitals using variational principles or perturbation theory.

The functions called the basis functions are introduced to calculate easier and the set of functions is called the basis set. There are many different types of basis function sets such as STO-nG (STO: Slater type orbital; G: Gaussian function) and Pople basis set. For example, s orbital with STO-3G is written as follows.

$$\psi_{\text{STO-3G}}(s) = c_1\phi_1 + c_2\phi_2 + c_3\phi_3 \quad (9)$$

$$\phi_1 = \left(\frac{2\alpha_1}{\pi}\right)^{\frac{3}{4}} e^{-\alpha_1 r^2} \quad (10)$$

$$\phi_2 = \left(\frac{2\alpha_2}{\pi}\right)^{\frac{3}{4}} e^{-\alpha_2 r^2} \quad (11)$$

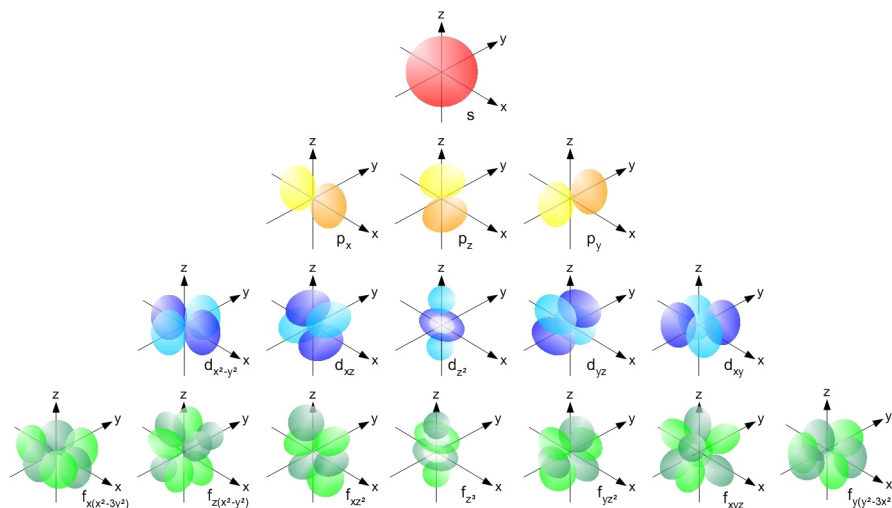


Figure 1: Shapes of hydrogen-like orbitals

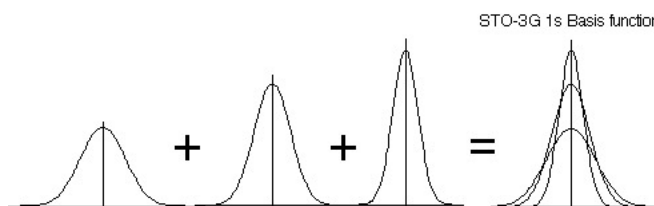


Figure 2: 1s basis function with 3 Gaussian primitives

$$\phi_3 = \left(\frac{2\alpha_3}{\pi}\right)^{\frac{3}{4}} e^{-\alpha_3 r^2} \quad (12)$$

The choice of basis set is important for successful and efficient calculation. If the basis set is too heavy, the result would be accurate but calculation time would be very long. On the other hand, if your basis set is too light, the result might fall short of your expectation while calculation time would be short. We should consider this trade-off relation between accuracy and computational cost when conducting calculations.

Followings are some keywords to describe basis sets.

- **Split valence:** Split-valence (SV) basis sets classify electrons in atoms into core and valence electrons. Valence electrons take part in the formation of molecule, so it is natural to describe valence electron wavefunction in detail with two or more basis functions (called 'double-zeta', 'triple-zeta', ... and so on) while the core electrons are approximated by just one basis function.
- **Polarization function:** Adding polarization function extends basis set beyond the valence orbitals. It makes our calculation consider the polarization effect. Since light atoms(H and He) have small polarization effect, they are sometimes excluded from adding polarization functions.
- **Diffuse function:** Diffuse functions are very broad Gaussian functions, which can be used to represent far-from-nuclei portion of atomic orbitals. Similar to polarization functions, light atoms can be excluded from adding diffuse functions.

We will use 6-31g(d) basis set in this experiment. 6-31g is a Pople type basis set, which represents core electrons with 1 basis function composed of 6 Gaussian primitive functions, valence electrons with 3 basis functions composed of 3 and 1 Gaussian primitives each. The diffuse function is included while the diffuse

polarization function is not.

2.3 Hartree-Fock method

In order to describe molecules quantum mechanically, one must transform the Schrödinger equation into a solvable mathematical form (closed form). To achieve this, several assumptions are introduced in the Hartree-Fock method. 1) The interactions between electrons are treated in an average manner to obtain the wave function for a single electron. The potential energies that other electrons exert on electron 1 are approximated as average potential energies. To compute this average potential energy, the wave functions of the other electrons are required, and the calculations proceed by sequentially computing them for electrons 2, 3, ..., N while checking the error with respect to the previous step. This approach is known as the self-consistent field (SCF) method.

$$\mathbf{H}_1 = \mathbf{F}(1) = -\frac{1}{2}\nabla_1^2 + V_1^{eff} \quad (13)$$

$$V_1^{eff} = -\sum_{A=1}^M \frac{Z_A}{r_1} + \sum_{j=1}^N \int d\mathbf{r}_2 \frac{|\psi_j(r_2)|^2}{r_{12}} - \sum_{j=1}^N \delta_{\sigma_i, \sigma_j} \int d\mathbf{r}_2 \frac{\psi_j^*(r_2)\psi_1(r_2)}{r_{12}} \quad (14)$$

$$\mathbf{F}(1)\boldsymbol{\psi}(1) = \boldsymbol{\epsilon}_1 \boldsymbol{\psi}(1) \quad (15)$$

2) Expressing the wave function as a Slater determinant allows it to incorporate the spin states of electrons and satisfy the Pauli exclusion principle.

$$\Psi(\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_N) = \begin{vmatrix} \chi_1(\mathbf{X}_1) & \chi_2(\mathbf{X}_1) & \cdots & \chi_N(\mathbf{X}_1) \\ \chi_1(\mathbf{X}_2) & \chi_2(\mathbf{X}_2) & \cdots & \chi_N(\mathbf{X}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_1(\mathbf{X}_N) & \chi_2(\mathbf{X}_N) & \cdots & \chi_N(\mathbf{X}_N) \end{vmatrix} \quad (16)$$

The specific derivation of the Hartree-Fock equation is referred to in other literature. With the mentioned approximations and assumptions, it can be expressed as a matrix composed of coefficients of the basis functions, known as the Hartree-Fock-Roothaan equation.

$$\mathbf{FC} = \mathbf{SC}\boldsymbol{\epsilon} \quad (17)$$

F represents the Fock matrix, S is the overlap matrix, C is the matrix of coefficients of the basis functions, and $\boldsymbol{\epsilon}$ represents the energies of possible electronic states (orbitals). The Fock matrix and overlap matrix vary based on the basis function set or the coefficients of the basis functions. The process of obtaining the appropriate orbitals using the Hartree-Fock-Roothaan equation is as follows:

1. First, guess the form of the orbitals and determine the coefficients.
2. Calculate the elements of the Fock matrix with these coefficients, solve the Hartree-Fock-Roothaan equation, and obtain new coefficients and energies.
3. Calculate the elements of the Fock matrix again with the obtained coefficients and solve the newly created Hartree-Fock-Roothaan equation to obtain the coefficients and energies for the next step.
4. Repeat this process, and the energy difference gradually becomes smaller. If the energy difference falls below an acceptable threshold, the iteration stops, and the coefficients and energy at this point become the final result.

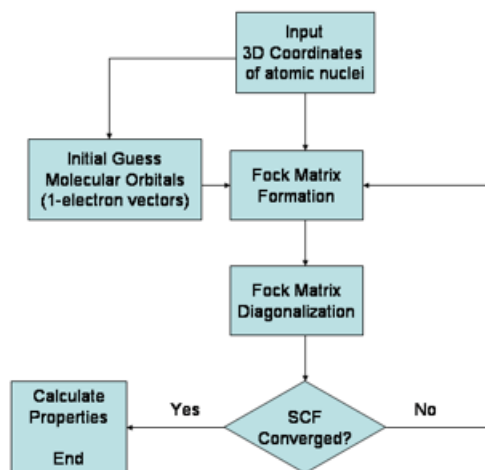


Figure 3: Procedure for Hartree-Fock Self-Consistent Field method

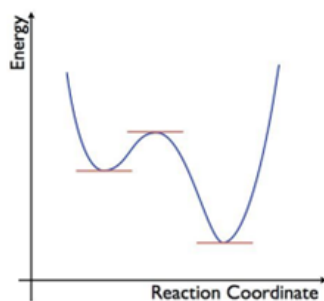


Figure 4: Potential energy curve for the system

2.4 Geometry Optimization

The Figure 4. represents the energy of a molecule as a function of reaction coordinates, such as bond lengths and bond angles. Molecular energies can vary depending on their structures. Among the various possible structures a molecule can adopt, it tends to favor the one with the lowest energy because it is usually the most stable. For example, in the case of a water molecule, it adopts a structure with a bond angle of 104.5° , which has the lowest energy among various possible bond angles. Therefore, to find the most stable structure of a molecule, one needs to vary the molecular structure while seeking the point at which the energy is minimized.

Energy minima are points on the energy graph where the first derivative of energy is zero, and the second derivative is positive. A single molecule can have multiple local minima depending on its structure, and the goal of structure optimization is to find the minimum with the lowest energy. Various methods, such as Steepest Descent (SD), Conjugate Gradient (CG), Newton-Raphson, and quasi-Newton methods, are available for optimization calculations. As explained earlier, even for a single molecule, there can be multiple local minima, and depending on how the initial conditions are set, the molecule may end up in a local minimum near the initial conditions rather than the global minimum. Therefore, to correctly proceed with structure optimization, it is necessary to set appropriate initial structures that do not violate physical/chemical intuition, such as arranging three or more atoms in a linear configuration or placing four or more atoms in the same plane.

3 Experimental Details

We will conduct our calculation on the EDISON platform (<https://www.edison.re.kr/>), using ORCA package. ORCA is an *ab initio* quantum calculation software package developed by Frank Neese's group, which contains various *ab initio* and semi-empirical electronic structure methods.

3.1 Pre-lab assignments

- Investigate the bond lengths and bonding energies of $C-H$, $N-H$, and $O-H$
- Explain the bond angle of H_2O , NH_3 , and CH_4 in the sense of VSEPR, valence bond theory.

⚠ Do not forget to cite the reference.

3.2 Goals

- Optimize the geometry of 3 single molecules H_2O , NH_3 , and CH_4 and find the bonds and most stable structures.
- Optimize the geometry of dimers and find the difference between the monomer calculations.

3.3 Requirements

- Avogadro: a software for building input files and visualization of output files of ORCA and other computational chemistry programs. Download from <https://avogadro.cc/>
- ORCA: An *ab initio* quantum chemistry software, free for academic purpose. Already installed in EDISON platform.

3.4 ORCA input file formats

ORCA file format is composed of keyword line and geometry block. Keyword line contains the calculation options, and geometry block contains the information of the system of interest.

- To construct your own ORCA input file, please follow these basic rules.

You can start the ORCA input script with ! symbol at the front. Everything behind ! symbol is considered as the calculation option. Keywords can be listed in any order. You can also set specific options starting with % symbol.

- Geometry block seems like "block" of multiple lines between two * symbols. You should define your own system information (total charge, multiplicity, coordinates of atoms) between two * symbols.

- You may insert a comment line starting with # symbol, which does not affect the calculation, to any location.

Please refer to the [ORCA Manual](#) if you have any problem or want to study further.

Fig. 5 shows the structure of ORCA input file for DFT calculation. Keyword line shows that the ORCA will conduct a DFT-based single point energy calculation with B3LYP functional, SV(p) basis set and small tolerance for judging the convergence. In the geometry block, we can find that the simulation is for CO molecule with no charge and excitation (by its multiplicity) and the geometry of atoms are represented as Cartesian coordinate.

Keywords essential to execute a HF calculation are listed below. You must contain these keywords in the keyword line of your ORCA input.

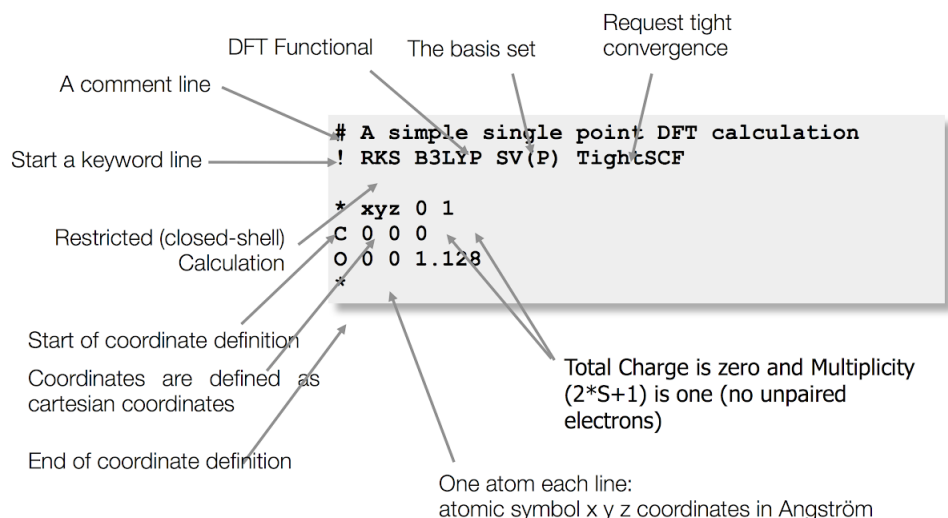


Figure 5: Input file format of ORCA package

- HF or UHF: keywords that make ORCA execute HF calculation. HF for systems with no single electron (multiplicity = 1), UHF for systems with 1 or more single electrons (multiplicity > 1)
- Job type: Determines what type of calculation to be conducted with the selected method; single point energy (by default, keyword: Energy or SP), geometry optimization (keyword: Opt), vibrational frequency (keyword: Freq) ... , Some options can be conducted simultaneously.
- Basis sets: we will use 6-31g(d) basis set. There are numbers of basis sets. If you want to use another basis set, see ORCA manual.
- Print option: There are four general printing keywords; "MiniPrint", "SmallPrint", "NormalPrint", and "LargePrint." We use "LargePrint" to visualize the calculated orbitals. You can add this keywords next to the other options in the same line.

Geometry block starts and ends with asterisk(*). Behind the opening asterisk, options determining the coordinate system (xyz for Cartesian coordinate), charge and multiplicity of the system. Geometry of the system is represented in the rest of the geometry block. Lines between the opening and closing asterisk have format of 'X coord1 coord2 coord3' in Cartesian coordinate system. X represents the atomic symbol, coord 1,2 and 3 represent the x, y, z position of atom.

⚠ If you set the distance between atoms too long, geometry optimization may fail to form a molecule.

3.5 How to execute a job on EDISON platform?

TA will provide your own ID/PW for EDISON. You can enter the ORCA execution page by the following route; Log in - 통합검색 - search 'ORCA' - run.

Then you can see the screen that looks like Fig. 6 (c) on your display. Copy and paste your ORCA input file text on the highlighted box (or upload your ORCA input file with "menu"). Check your input file again and click "submit" at the top of the page, and your job will be uploaded on the server. If your job is successfully completed, you can see and download the output file at other boxes in the page.

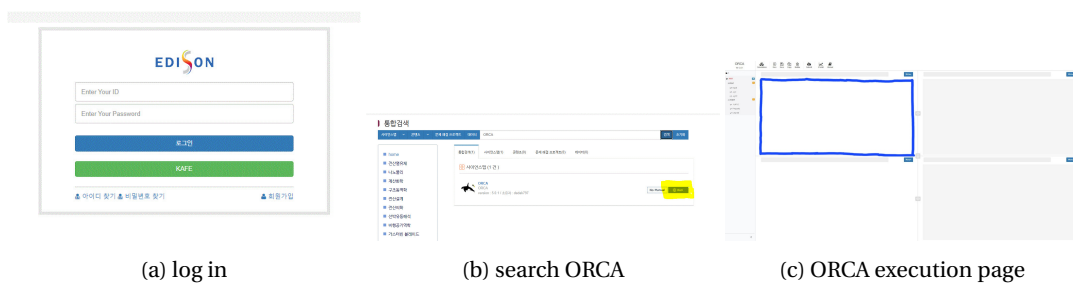


Figure 6: ORCA execution on EDISON

3.6 Experimental Procedures

- 1) Make your own input files for H_2O , NH_3 , CH_4 , and HF , with keyword "Opt" for the purpose of geometry optimization.
- 2) Find ORCA package in EDISON; and run your own job by submitting input file. You can monitor your job at the side bar on the left side.
- 3) Download output files after the job is finished.
- 4) You can get the results including bond angle, bond length and energies from the output files. You can also visualize your output via Avogadro.

3.7 Precautions Before Practice

- 1) Computational chemistry practice is conducted as an individual practice rather than a group practice, so each person should be prepared and familiar with the manual. In particular, unlike other "laboratory" practices, unfamiliar theoretical knowledge may be needed, and we cannot handle all of them in class, so it is recommended to study the manual in advance.
- 2) Some students are caught doing irrelevant things using lab computer during the class. In this case, there will be a disadvantage in your final score of the practice, so you should focus on the practice.

4 Guidance for the Report and Assignments

- 1) Basically, you should follow standard report format from the class.
- 2) You don't have to study too hard about the background or introduction part.
- 3) TA will give you two post-lab assignments. Write down your answers for the assignments on your report.
- 4) You must include your own abstract, conclusion and key information of your calculation.
- 5) Please don't include too much information such as full output files, figures of optimized molecules, and so on. Your report has only to contain information essential for the result analysis. If you really want to contain any data, figure or table that makes your report volume exceeds the maximum, you can consider to make the supporting information section.
- 6) If you need any help or further information regarding the class (e.g. report, theoretical background, software ...), feel free to contact TA.

5 References

- [1] Neese, F. (2012). The ORCA program system. Wiley Interdisciplinary Reviews: Computational Molecular Science, 2(1), 73-78.

- [2] Hanwell, M. D., Curtis, D. E., Lonie, D. C., Vandermeersch, T., Zurek, E., & Hutchison, G. R. (2012). Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *Journal of cheminformatics*, 4(1), 1-17.
- [3] Bode, B. M., & Gordon, M. S. (1998). MacMolPlt: a graphical user interface for GAMESS. *Journal of Molecular Graphics and Modelling*, 16(3), 133-138.
- [4] Levine, I. N., Busch, D. H., & Shull, H. (2009). *Quantum chemistry* (Vol. 6). Upper Saddle River, NJ: Pearson Prentice Hall.
- [5] Szabo, A., & Ostlund, N. S. (2012). *Modern quantum chemistry: introduction to advanced electronic structure theory*. Courier Corporation.
- [6] Atkins, P., Jones, L., & Laverman, L. (2018). *Chemical Principles. The Quest for Insight*, 6th.