ISSN: 2710-0588 DOI: 10.29556/FSE.202110_1(7).0012

Analysis of Small Organic Molecules with Quantum Chemistry

Yuan Chang

Shanghai Soong Ching Ling School, No.2, Yehui Road, Qingpu District, Shanghai, China.

Abstract

In this paper, computational chemistry with the quantum-mechanics-based model is employed for analyzing the molecular properties of the small organic molecules, which consist as strong potential candidates for important drug design and material development issues. It is shown that, the highest occupied molecular orbital (HOMO), lowest unoccupied molecular orbital (LUMO) and the band gaps have some underlying relationship, which might act as the evidence for their macro-scope experimental performances.

Keywords

Organic Molecules; Quantum Chemistry; Hatree-fock Theory; DFT; Molecular Properties.

1. Introduction

Computational chemistry is a branch of chemistry utilizing simulation tools for analysis. Among it, quantum chemical calculation is usually used for analyzing electronic structure of molecules based on the quantum mechanics models. After quantum chemical calculation, electronic state information of molecules such as orbital energy and wavefunction can be acquired. After processing the draw information, other more important properties such as polarizability and optical constants can be obtained. In analysis of quantum chemistry, approximated Schrödinger equation is solved with several computational methods.

2. Hatree-Fock Theory

Hartree-Fock theory [1-2] is fundamental to much of electronic structure theory. It is the basis of molecular orbital (MO) theory, which posits that each electron's motion can be described by a single-particle function (orbital) which does not depend explicitly on the instantaneous motions of the other electrons. The ubiquity of orbital concepts in chemistry is a testimony to the predictive power and intuitive appeal of Hartree-Fock MO theory. However, it is important to remember that these orbitals are mathematical constructs which only approximate reality. Only for the hydrogen atom or other one-electron systems like He+ are orbitals exact eigenfunctions of the full electronic Hamiltonian. As long as we are content to consider molecules near their equilibrium geometry, Hartree-Fock theory often provides a good starting point for more elaborate theoretical methods such as many-body perturbation theory and single-reference configuration interaction, which are better approximations to the electronic Schrödinger equation [3].

Hartree-Fock theory is solving Schrodinger equation for multi-electron atoms or molecules based on the Born-Oppenheimer approximation and Hartree-Fock approximation. In Born-Oppenheimer approximation, the motion of atomic nuclei and electron in a molecule can be separated. One can assume that the nucleus is fixed and then solve equation on electrons separately. In Hartree-Fock approximation, all orbital of electron is defined as a certain type of wavefunction and the molecular wavefunction is the product of them. In order to make agreement with Pauli's principle, the molecular wavefunction is expressed as Slater determinant, considering antisymmetric nature of wavefunction. The electric interaction to one

ISSN: 2710-0588 DOI: 10.29556/FSE.202110_1(7).0012

electron from other electrons is expressed in mean field approximation [4-5]. Electrons not only originate electric field, but also receive electric interaction at the same time. Thus, wavefunctions that generate electric field and receive influence should coincide.

3. Density Functional Theory

Density-functional theory (DFT) is a computational quantum mechanical modelling method used in physics, chemistry and materials science to investigate the electronic structure (or nuclear structure) (principally the ground state) of many-body systems, in particular atoms, molecules, and the condensed phases. Using this theory, the properties of a many-electron system can be determined by using functionals, i.e. functions of another function [6]. In the case of DFT, these are functionals of the spatially dependent electron density. Due to the calculation accuracy and modest calculation cost, DFT is widely used in many application studies. As mentioned before, condensed matter is made of positively charged nuclei and negatively charged electrons. Electrons behave as point-like particles and, to a very good approximation, the nuclei can be considered to be point-like particles also. The complete system of electrons and nuclei can be described by the many-body Schrödinger equation [7].

4. Quantum Chemical Dataset of Small Organic Molecules

Computational de novo design of new drugs and materials requires rigorous and unbiased exploration of chemical compound space. However, large uncharted territories persist due to its size scaling with molecular size. QM9 is famous for reporting computed geometric, energetic, electronic, and thermodynamic properties for 134k stable small organic molecules made up of C, H, O, N and F. These molecules correspond to the subset of all 133,885 species with up to nine heavy atoms (CONF) out of the GDB-17 chemical universe of 166 billion organic molecules. Geometries minimal in energy, corresponding harmonic frequencies, dipole moments, polarizabilities, along with energies, enthalpies, and free energies of atomization are reported. All properties were calculated at the B3LYP/6-31G(2df,p) level of quantum chemistry. This database serves the benchmarking of existing methods, development of new methods, such as hybrid quantum mechanics/machine learning, and systematic identification of structure-property relationships [8-9].

Most recently, researchers in Tecent AI Lab introduced a new molecular dataset, named Alchemy, for developing machine learning models useful in chemistry and material science. The dataset comprises of 12 quantum mechanical properties of 119,487 organic molecules with up to 14 heavy atoms, sampled from the GDB MedChem database. The Alchemy dataset expands the volume and diversity of existing molecular datasets [10].

Table 1. QM9 and Alchemy Molecular Properties.

Property	Unit	Description
μ	D	Dipole moment
α	a_0^3	Isotropic polarizability
ϵ_{HOMO}	Ha	Energy of HOMO
ϵ_{LUMO}	Ha	Energy of LUMO
ϵ_{gap}	Ha	Gap($\epsilon_LUMO - \epsilon_HOMO$)
R^2	a_0^2	Electronic spatial extent
zpve	Ha	Zero point vibrational energy
U_0	Ha	Internal energy at 0K
U	Ha	Internal energy at 298.15K
H	Ha	Enthalpy at 298.15K
G	Ha	Free energy at 298.15K
C_v	$\frac{cal}{molK}$	Heat capacity at 298.15K

ISSN: 2710-0588 DOI: 10.29556/FSE.202110_1(7).0012

The molecular properties calculated by the density functional theory (DFT) are listed in table 1. In this paper, the highest occupied molecular orbital (HOMO), lowest unoccupied molecular orbital (LUMO) and the band gaps are mainly focused due to their important role in academic application fields such as organic semi-conductors.

5. Analysis of Molecular Properties

To obtain the molecular properties at the quantum chemistry level, solving the Schrodinger equation is necessary. Because of approximations in HF method, interactions of electrons are interpreted roughly thus the wavefunction acquired from HF is not the exact solution of Schrödinger equation [11-12]. It is difficult to use HF calculation on quantitative evaluation of chemical property. The difference between the exact solution and the HF solution is called as electron correlation. The correlation energy is a measurement of how much the movement of one electron is influenced by all other electrons, which is necessary for more quantitative evaluation. In this report, density functional theory is used for analyzing the molecular properties. The Many-body Schrodinger equation expressed in atomic units (Hartree) is written as a function of a wavefunction $\Psi(\boldsymbol{r}_1, \boldsymbol{r}_2 \dots, \boldsymbol{r}_N)$,

$$\hat{H}|\Psi\rangle = \left(\hat{T} + \hat{U} + \hat{V}\right)|\Psi\rangle = \mathrm{E}|\Psi\rangle$$

where the kinetic term $\hat{T}$ is,

$$\hat{\mathrm{T}} = \sum_{i}^{N} -\frac{1}{2}\nabla_i^2$$

the electron interaction term $\hat{U}$ is,

$$U = \sum_{i<j} U(\boldsymbol{r}_i, \boldsymbol{r}_j) = \frac{1}{2}\sum_{i,j} U(\boldsymbol{r}_i, \boldsymbol{r}_j)$$

$$U(\boldsymbol{r}_i, \boldsymbol{r}_j) = U(\boldsymbol{r}_j, \boldsymbol{r}_i) = \frac{1}{|\boldsymbol{r}_i - \boldsymbol{r}_j|}$$

and the nuclei-electron interaction $\hat{V}$ is,

$$\hat{V} = \sum_{i}^{N} v(\boldsymbol{r}_i)$$

$$v(\boldsymbol{r}_i) = \sum_{k} \frac{-Z_k}{|\boldsymbol{r}_i - \boldsymbol{R}_k|}$$

Where $\boldsymbol{R_k}$ is the positions of nuclei and Z_k is the number of nucleons in an atom. $|\Psi|^2 = \Psi * \Psi$ is the probability density of existence of the first electron at $\boldsymbol{r_1}$, the second at $\boldsymbol{r_2}$, and the Nth electron at $\boldsymbol{r_N}$. The wavefunction is normalized thus can satisfy $\int |\Psi|^2 d^3\boldsymbol{r_1} d^3\boldsymbol{r_2} \dots d^3\boldsymbol{r_N} = 1$. Also, Ψ obeys Pauli's principle, therefore it is naturally antisymmetric. If we integrate $|\Psi|^2$ over the first to N-1th electron, we can obtain the probability density that the Nth electron is at the position $\boldsymbol{r_N}$.

HOMO stands for "Highest Occupied Molecular Orbital", and LUMO stands for "Lowest Unoccupied Molecular Orbital". Of the orbitals that have electrons, the HOMO is the highest in energy, and of the orbitals that don't, the LUMO is the lowest in energy. That means they are closest in energy out of all orbitals in the molecule. Due to the energies of these orbitals being the closest of any orbitals of different energy levels, the HOMO-LUMO gap is where the most likely excitations can occur. Hence, it is the most important energy gap to consider. The diagram of HOMO, LUMO and band gap is shown in Fig. 1, where each circle represents an electron in an orbital. In solid-state physics, if the valence band is completely full and the conduction band is completely empty, then electrons cannot move in the solid; however, if some electrons transfer

ISSN: 2710-0588 DOI: 10.29556/FSE.202110_1(7).0012

from the valence to the conduction band, then current can flow (see carrier generation and recombination). Therefore, the band gap is a major factor determining the electrical conductivity of a solid. Substances with large band gaps are generally insulators, those with smaller band gaps are semiconductors, while conductors either have very small band gaps or none, because the valence and conduction bands overlap. Therefore, to study the band-gap and HOMO/LUMO is important for both theoretical study and real applications [13].

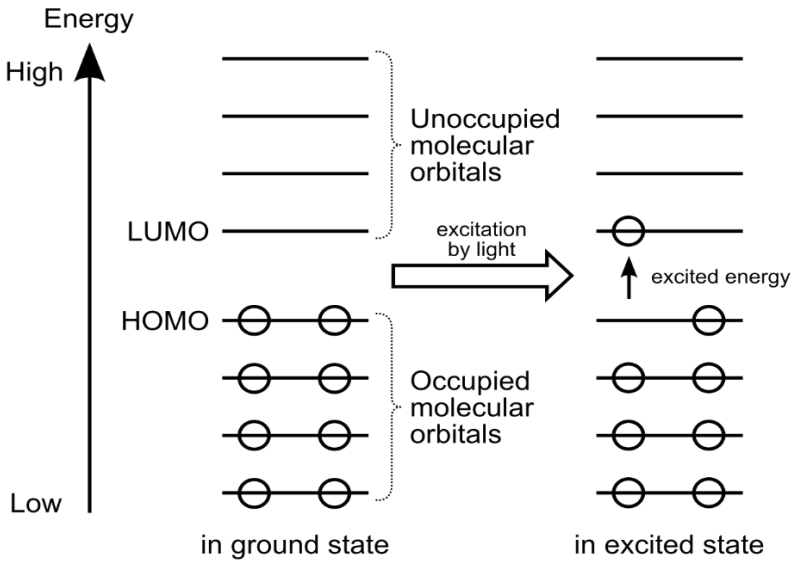

Fig. 1 Diagram of the HOMO and LUMO of a molecule

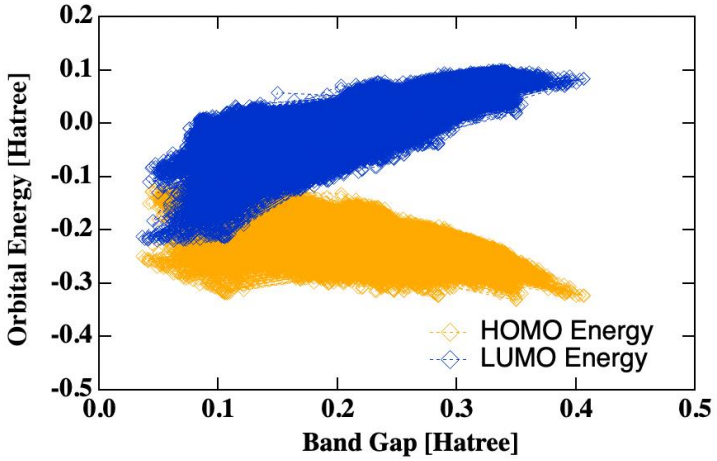

Fig. 2 Band-gap versus HOMO and LUMO orbital energies.

The molecules from QM9 dataset with HOMO, LUMO and band-gap information are extracted for analysis, as shown in Fig. 2. It is clear that the HOMO energy, which is the orange dot, is in positive correlation with the band gap value, while the LUMO energy, which is the blue dot, is in negative correlation with the band gap. In the meanwhile, HOMO energy and LUMO energy themselves don't exhibit strong correlation, which is consistent with human intuition because these two values are absolute values of the molecules, while the relative values such as band-gap are more meaningful. In the semi-conductor design, researchers use band-gap information for finding better material. Since the band-gap is positively correlated with LUMO energy, researchers can also switch to the LUMO energy for screening if the targeted molecule doesn't have the band-gap data.

ISSN: 2710-0588 DOI: 10.29556/FSE.202110_1(7).0012

References

[1] Hartree D R. The wave mechanics of an atom with a non-Coulomb central field. Part I. Theory and methods[J]. 1928.

[2] Fock V. Näherungsmethode zur Lösung des quantenmechanischen Mehrkörperproblems[J]. Zeitschrift für Physik, 1930, 61(1-2): 126-148.

[3] Sherrill C D. An introduction to Hartree-Fock molecular orbital theory[J]. School of Chemistry and Biochemistry Georgia Institute of Technology, 2000.

[4] Martin R M. Electronic structure: basic theory and practical methods[M]. Cambridge university press, 2020.

[5] Löwdin PO. Quantum theory of many-particle systems. Physical interpretations by means of density matrices, natural spin-orbitals, and convergence problems in the method of configurational interaction[J]. Physical Review, 1955, 97(6): 1474.

[6] Parr R G. Density functional theory of atoms and molecules[M]//Horizons of quantum chemistry. Springer, Dordrecht, 1980: 5-15.

[7] Kirkpatrick K, Schlein B, Staffilani G. Derivation of the two-dimensional nonlinear Schrödinger equation from many body quantum dynamics[J]. American journal of mathematics, 2011, 133(1): 91-130.

[8] L. Ruddigkeit, R. van Deursen, L. C. Blum, J.-L. Reymond, Enumeration of 166 billion organic small molecules in the chemical universe database GDB-17, J. Chem. Inf. Model. 52, 2864–2875, 2012.

[9] R. Ramakrishnan, P. O. Dral, M. Rupp, O. A. von Lilienfeld, Quantum chemistry structures and properties of 134 kilo molecules, Scientific Data 1, 140022, 2014. [bibtex].

[10] Chen G, Chen P, Hsieh C Y, et al. Alchemy: A quantum chemistry dataset for benchmarking ai models[J]. arXiv preprint arXiv:1906.09427, 2019.

[11] Møller C, Plesset M S. Note on an approximation treatment for many-electron systems[J]. Physical review, 1934, 46(7): 618.

[12] James H M, Coolidge A S. The ground state of the hydrogen molecule[J]. The Journal of Chemical Physics, 1933, 1(12): 825-835.

[13] IUPAC, Compendium of Chemical Terminology, 2nd ed. (the "Gold Book") (1997). Online corrected version: (2006–) "SOMO". doi:10.1351/goldbook.S05765.