THE SELECTION OF AB INITIO METHOD FOR POLYETHYLENE MOLECULE

Khairun Nisa, Ponco Iswanto*, Eva Vaulina Yulistia Delsy and Ely Setiawan
Laboratory of Physical Chemistry, Department of Chemistry, Faculty of Mathematics and Natural Sciences, Universitas Jenderal Soedirman, Purwokerto 53123, Indonesia

*Email: ponco.iswanto@unsoed.ac.id

Abstract. Polyethylene is the simplest vinylic polymer in structure. Polyethylene chains consist of hydrocarbon chains. The selection of calculation methods and basis sets is important part through an ab initio approach based on quantum mechanics principles. This aims of research to select the influence of basis set variations on the simulation results of polyethylene tetramer molecules. The six basis sets used were 6-31G*, 6-31G**, 6-31++G**, 6-311+G*, 6-311++G**, and cc-pVDZ. Molecular modeling was performed in the HyperChem software into four main parameters: geometric optimize such as bond length and bond angle, infrared (IR) spectrum, and nuclear magnetic resonance (NMR) spectrum. The selection was performed by calculating the Predicted Residual of Sum Squares (PRESS) between theoretical and experimental data. The results indicated that the 6-311++G** basis set was the most suitable basis set to use based on the smallest total PRESS value obtained. The PRESS value for bond lengths and bond angles was 6.047889479, the PRESS value for infrared spectra was 76110.2513, and the PRESS value for NMR spectra was 1.546474.

Keywords: Ab Initio, polyethylene, bond length, IR spectrum, NMR spectrum

1. Introduction

Computational chemistry has emerged as one of the central pillars in modern chemical research, offering theoretical insights that are often difficult or even impossible to obtain solely through experimental approaches. By employing advanced mathematical algorithms and quantum mechanical models, computational methods enable the prediction and analysis of fundamental molecular properties such as geometry, total energy, infrared (IR) spectra, and nuclear magnetic resonance (NMR) shifts [1]. These approaches provide an efficient and cost-effective complement to experimental investigations, reducing the reliance on resource-intensive and potentially hazardous laboratory procedures [2].

Among a wide variety of materials studied in computational chemistry, polyethylene (PE) holds a particularly significant position. As one of the most widely produced and utilized synthetic polymers, polyethylene plays an essential role in daily life, ranging from packaging and household applications to industrial components. Its lightweight, durability, and chemical resistance make polyethylene indispensable across diverse sectors. Consequently, an accurate understanding of its structural and spectroscopic properties is crucial not only for advancing polymer science but also for supporting the development of more reliable computational methodologies [3].

The structural and physicochemical properties of polymers such as polyethylene are inherently determined by their molecular geometry. Experimental studies have provided important insights into the vibrational and electronic features of PE, yet computational methods offer additional advantages in simulating and predicting these properties with high precision. In particular, ab initio methods, grounded in first-principles quantum mechanics, allow the accurate resolution of the Schrödinger equation without requiring empirical parameters. This

makes ab initio approaches highly suitable for investigating spectroscopic properties such as vibrational frequencies in IR spectroscopy and chemical shifts in proton nuclear magnetic resonance (NMR) spectroscopy [4].

A central challenge in ab initio calculations is the selection of an appropriate basis set, which defines the mathematical functions used to describe atomic orbitals. The choice of basis set has a profound influence on the accuracy and efficiency of computations, affecting parameters such as bond lengths, bond angles, and predicted spectra. Although a variety of basis sets are available ranging from minimal to extended forms with polarization and diffuse functions no universal standard exists for selecting the most suitable basis set across different molecular systems. This often necessitates comparative studies to evaluate the reliability and consistency of computational results against experimental benchmarks [5].

In this context, the present study focuses on evaluating the structural and spectroscopic properties of polyethylene tetramer using ab initio methods with several basis sets, including 6-31G*, 6-31G**, 6-31++G**, 6-311+G*, 6-311++G**, and cc-pVDZ. Polyethylene tetramer was chosen as the model system due to its relatively simple structure that remains representative of the larger polymer chain. The objective of this work is to identify the basis set that provides the most accurate geometric and spectroscopic predictions compared with experimental data. By systematically analyzing both IR and ^{1}H-NMR spectra alongside geometrical parameters, this study aims to establish a computational framework for selecting optimal methods in polymer modeling, thereby bridging the gap between theory and experiment in computational polymer chemistry.

2. Materials and Methods

2.1. Instruments and Materials

The computational work was performed using hardware equipped with an Intel(R) Core(TM) i5-6500 processor, 8 GB RAM, and a Windows 10 64-bit operating system (x64-based). The software used in this study was HyperChem version 8.0, which serves as a molecular modeling and optimization tool.

The material used in this research was a three-dimensional (3D) structural model of a polyethylene tetramer, consisting of four ethylene monomer units.

2.2. Research Procedure

a) Polyethylene Molecular Modeling

The initial stage of this study involved constructing the polyethylene molecular model using HyperChem. The molecule was built in a three-dimensional (3D) and subsequently optimized using computational methods to obtain geometric data, including bond lengths and bond angles, as well as infrared (IR) and proton nuclear magnetic resonance (^{1}H-NMR) spectra.

b) Geometry Optimization

After the polyethylene tetramer molecule was successfully modeled in 3D, geometry optimization was carried out using the Ab Initio method with six different basis sets: 6-31G*, 6-31G**, 6-31++G**, 6-311+G*, 6-311++G**, and cc-pVDZ. Each optimization was performed sequentially using HyperChem to determine the optimized geometry parameters. The optimization settings for all methods are summarized in Table 1.

Table 1. Polyethylene Compound Optimization Conditions

Method (Basis Set)	Algorithm	Termination Condition	Screen Refresh Period
6-31G*, 6-31G**, 6-31++G**, 6-311+G*, 6-311++G** and cc-pVDZ	Polak-Ribiere (Conjugate gradient)	0,01 kcal/mol 10.000 maximum cycles	1

c) Molecular Geometry Analysis (Bond Length and Bond Angle)

Geometry analysis was conducted after the polyethylene tetramer structure reached convergence during the optimization process. The analyzed geometric parameters included bond lengths and bond angles. The optimized data were obtained from the geometry output of each computational method. Bond lengths were measured in Ångströms (Å) and bond angles in degrees (°). Theoretical results were then compared with experimental reference data, where the C–C bond length is 1.536 Å, the C–H bond length is 1.06 Å, and the bond angles of C–C–C and H–C–H are 112.8° and 109° respectively [6].

d) Infrared (IR) Spectrum Analysis

After the polyethylene tetramer model reached a stable energy state, infrared (IR) spectrum calculations were performed using HyperChem. This process was conducted by selecting the Compute menu and choosing the Vibrational Spectrum option within the software. The calculation provided molecular vibration data and wavenumber (frequency) values that represent the characteristic functional group vibrations of the compound.

e) ^{1}H Nuclear Magnetic Resonance (NMR) Spectrum Analysis

The next step involved calculating the ^{1}H-NMR spectrum by selecting all hydrogen (proton) atoms in the molecular structure. The hydrogen atoms were manually selected using the Select menu in HyperChem. After the selection, the ^{1}H-NMR spectrum was calculated with an instrument frequency of 75 MHz and a reference shielding value of 24.4. The resulting data included the proton chemical shift values and the corresponding NMR spectrum of the compound.

f) PRESS Test

The predicted data of molecular geometry, IR spectra, and H-NMR spectra of polyethylene were evaluated using the PRESS (Predicted Residual Error Sum of Squares) method to identify the most accurate computational model for the polyethylene molecule. The PRESS value was obtained by calculating the sum of the squared differences between the predicted and experimental data, providing a quantitative measure of model accuracy. The smaller the PRESS value, the closer the theoretical results are to the experimental observations. The PRESS calculation is shown in Equation (1).

$$PRESS = \sum_{i=1}^{n} \left(y_i - \hat{y}_{i\,(-i)}\right)^2 \ldots\ldots\ldots (1)$$

3. Results And Discussion

3.1. Modeling and Optimization of Polyethylene Molecular Geometry

The initial stage of the computational process involved constructing the polyethylene tetramer model. This fundamental structure consists of four repeating ethylene (C_2H_4) monomer units linked sequentially, forming a linear, non-aromatic hydrocarbon chain devoid of specific functional groups [7]. Modeling was performed using the HyperChem software, where the initial double bonds of the ethylene units were converted into single bonds. The tetramer size was chosen because it is relatively small, enabling efficient analysis across multiple basis sets without compromising the chemical representation necessary for accurate prediction of structural stability and spectroscopic properties.

Polyethylene tetramer is one of the simplest and most stable synthetic polymers. The structure of the polyethylene tetramer, both before and after optimization, is shown in Figure 1. The molecule was modelled using the balls and cylinders representation, depicting eight carbon atoms (light blue spheres) and eighteen hydrogen atoms (white spheres), corresponding to the molecular formula C_8H_{18}.

Geometry optimization was performed using the Ab Initio method with six different basis sets to compare the resulting energy profiles and final geometries. The basis sets used were: 6-31G*, 6-31G**, 6-31++G**, 6-311+G*, 6-311++G** and cc-pVDZ.

Each molecular structure was optimized until it reached convergence at the minimum energy point. The goal of this optimization was to find the lowest-energy conformation, which represents the most stable molecular structure [8]. The resulting low-energy polyethylene tetramer, shown in Figure 1, exhibits the characteristic zigzag configuration of a saturated carbon chain, maintaining a linear backbone with hydrogen atoms uniformly distributed around the carbons [7].

Tabel 2. Geometry Optimization Results

Calculation Method	Total Energy (kcal/mol)	Dipole Moment (Debye)
6-31G*	-196682.6162831	0.00005
6-31G**	-196699.9602448	0.00002
6-31++G**	-196701.2909371	0.00013
6-311+G*	-196718.6224671	0.00664
6-311++G**	-196732.8611352	0.00002
cc-pVDZ	-196697.1874192	0.00004

The optimization results, including total energy and dipole moment values obtained using various basis sets, are summarized in Table 2. The calculated total energy of the polyethylene tetramer molecule shows significant variation depending on the basis set applied. In principle, the lower (more negative) the total energy, the more stable the molecular structure. This indicates that the basis set effectively represents the electron distribution and molecular geometry with high accuracy [9].

The calculation results show that the 6-311++G** basis set produces the lowest total energy, -196732.8611352 kcal/mol. This is because the 6-311++G** basis set excels in capturing complex electronic interactions due to its triple-zeta configuration, which includes complete polarization and diffuse functions for both heavy and hydrogen atoms. Triple-zeta basis sets such as 6-311++G** provide lower energy because their flexible orbitals can adapt better to the variational principle, making them superior in energy accuracy compared other basis sets [10].

The dipole moment represents the uneven distribution of electrons, caused by differences in electronegativity between bonded atoms, where electrons tend to move toward the more electronegative atom. The magnitude of the dipole moment correlates with the polarity of the compound smaller dipole moments indicate lower polarity. Compounds with small dipole moments require less energy to release hydrogen atoms in the gas phase or in nonpolar solvents. The dipole moment is expressed in Debye (D) units [11]. The obtained dipole moments across all basis sets are relatively small, indicating that the polyethylene molecule is highly nonpolar, consistent with its saturated hydrocarbon structure lacking any polar functional groups.

3.2. Molecular Geometry Analysis of Polyethylene

Molecular geometry is an essential parameter for evaluating the accuracy of optimized molecular structures. This geometry includes bond lengths and bond angles. Bond length refers to the distance between the nuclei of two bonded atoms, while bond angle represents the angle formed between two bonded atoms and a third atom [12].

The infrared spectrum analysis in Table 3 shows that all six basis sets produce geometric values within the experimental tolerance range, with bond length deviations of less than ±0.01 Å and bond angle deviations of about ±2° [13]. This indicates that the six basis sets are sufficiently accurate in describing atomic interactions.

Table 3 Bond length and bond angle result

Method	Bond Length (A°)			Bond Angles (°)		
	C-C	C-H	PRESS	C-C-C	H-C-H	PRESS
Experiment	1.532	1.060	-	112.8	109	-
6-31G*	1.531	1.087	0.0007480	114.0	106.9	5.7646
6-31G**	1.531	1.088	0.0007659	114.1	106.9	5.8519
6-31++G**	1.531	1.088	0.0007803	114.1	107.0	5.7902
6-311+G*	1.531	1.088	0.0007635	114.2	107.0	6.9793
6-311++G**	1.531	1.088	0.0007861	114.1	107.0	6.0471
cc-pVDZ	1.529	1.095	0.0012007	114.1	107.0	6.2614

The optimized bond length and bond angle analysis shows that the 6-31G* basis set provides the closest agreement with experimental data. Based on the PRESS test, it yields the smallest PRESS value for bond length 0.0007480 and bond angle 5.7646. This result demonstrates that the 6-31G* basis set offers a good balance between computational complexity and accuracy. Although it is a medium-sized basis set that includes polarization functions without diffuse functions, it can produce a more rigid and less flexible structure. Consequently, the resulting molecular structure is more stable and closely resembles the experimental structure in terms of bond length and bond angle.

3.3. Infrared Spectrum Analysis

Infrared (IR) spectroscopy has become an essential analytical technique for chemical structure characterization and material research. This method provides valuable insights that enable the identification of functional groups and the determination of molecular interactions within a sample. Each type of bond and functional group in a molecule possesses unique intrinsic vibrational frequencies. The resulting spectrum acts as a molecular fingerprint, allowing for precise structural and qualitative determination.

The computational results indicate that not all basis sets produced absorption values close to the experimental data across the three spectral regions. This variation occurs because simpler basis sets are less capable of accurately predicting vibrational energies, as vibrational modeling strongly depends on the inclusion of diffuse functions [14].

Table 4 Results of polyethylene infrared spectrum analysis

Method	Spectrum 1	Spectrum 2	Spectrum 3	PRESS
Experiment	2920.23	1485.50	725.23	-
6-31G*	3261.54	1555.41	1199.99	346,776.9818
6-31G**	3242.33	1235.75	848.01	181,198.4009
6-31++G**	2564.21	1729.41	816.69	194,607.2601
6-311+G*	2517.60	1756.49	645.163	241,957.2215
6-311++G**	3183.80	1540.72	785.25	76,110.2513
cc-pVDZ	3231.31	1525.10	841.02	111,746.2505

PRESS test analysis in Table 4 shows that the 6-311++G** basis set provides the best agreement with the experimental IR spectrum, indicated by the smallest PRESS value of 76,110.2513. This superior performance is attributed to its triple-zeta configuration with added polarization and diffuse functions, which allow for a more flexible description of the electron cloud and better capture of subtle vibrational changes [15].

3.4. H-NMR (Nuclear Magnetic Resonance) Spectrum Analysis

Nuclear Magnetic Resonance (NMR) spectroscopy is a highly effective instrumental technique for determining molecular structures. This method provides valuable information about the chemical environment surrounding atomic nuclei and the connectivity patterns between atoms, particularly for hydrogen (^{1}H) and carbon (^{13}C). The principle of NMR spectroscopy is based on the absorption of radiofrequency radiation by magnetically active nuclei placed within an external magnetic field. When combined with infrared (IR)

spectroscopy, ^1H-NMR analysis offers a more comprehensive and complementary approach to molecular structure elucidation. Structural verification becomes more accurate when both datasets are integrated [16].

The results obtained using six different basis sets showed that the calculated chemical shift values were generally higher than the experimental data for both CH_2 and CH_3 groups. This discrepancy arises from theoretical simulation limitations, which often lead to an overestimation of deshielding effects, especially when larger basis sets are applied.

Table 4 H-NMR (Nuclear Magnetic Resonance) spectrum analysis

Method	Chemical shift (ppm)		
	CH_2	CH_2	CH_2
Experiment	1.27	0.88	-
6-31G*	1.53	1.09	1.404999
6-31G**	1.53	1.097	1.455573
6-31++G**	1.54	1.10	1.676490
6-311+G*	1.53	1.10	1.462384
6-311++G**	1.54	1.11	1.546474
cc-pVDZ	1.69	1.27	3.283875

Based on the PRESS test results presented in Table 5, the 6-31G* basis set achieved the lowest PRESS value, indicating the highest level of agreement between theoretical and experimental data. This can be attributed to the balanced nature of the 6-31G* basis set, which, despite lacking diffuse functions, remains sufficiently accurate for saturated hydrocarbon systems such as polyethylene. Its simplicity and moderate polarization functions provide stable results, as the polyethylene tetramer is a relatively non-polar molecule that does not require extensive electron distribution representation.

4. Conclusion

The most suitable Ab Initio computational method for polyethylene tetramer was determined using six different basis sets, evaluated through molecular geometry, IR spectrum, and ^{1}H-NMR spectrum as key parameters. The PRESS (Predicted Residual Error Sum of Squares) method was applied to quantitatively assess the accuracy of each computational model's predictions against experimental data, thereby improving the reliability of the results.

The results of the study showed that the 6-311++G** basis set exhibited the best overall performance, with the lowest total PRESS value across all basis sets. The PRESS values were 0.0007861 for bond length, 6.0471 for bond angle, 76110.2513 for infrared spectrum, and 1.546474 for H-NMR spectrum. These results confirm that the 6-311++G** basis set provides the most accurate representation of the structural and spectroscopic properties of polyethylene tetramer among the basis sets tested.

5. Acknowledgement

We express our gratitude to LPPM Universitas Jenderal Soedirman for providing financial support via contract number: 16.57/UN23.34/PT.01/V/2025.

References

[1] A. M. Mroz, V. Posligua, A. Tarzia, E. H. Wolpert, and K. E. Jelfs, “Into the Unknown: How Computation Can Help Explore Uncharted Material Space,” Oct. 19, 2022, *American Chemical Society*. doi: 10.1021/jacs.2c06833.

[2] G. M. Gutiérrez-Finol, A. Ullah, M. González-Béjar, and A. Gaita-Ariño, “A call for frugal modelling: two case studies involving molecular spin dynamics,” *Green Chemistry*, vol. 27, no. 12, pp. 3167–3177, 2025, doi: 10.1039/D4GC04900D.

[3] M. Oliveira, “Microplastic contamination in freshwater ecosystems: A global perspective,” *Science of the Total Environment*, vol. 705, 2020.

[4] Z. Wang, S. L. Lee, T. C. Huang, G. S. Lin, K. L. Tung, and T. Yoshioka, "Molecular modelling and simulation of membrane formation," *60 Years of the Loeb-Sourirajan Membrane: Principles, New Materials, Modelling, Characterization, and Applications*, pp. 463–497, Jan. 2022, doi: 10.1016/B978-0-323-89977-2.00017-8.

[5] W. Hu and M. Chen, "Editorial: Advances in Density Functional Theory and Beyond for Computational Chemistry," Jul. 12, 2021, *Frontiers Media S.A.* doi: 10.3389/fchem.2021.705762.

[6] R. Caminiti, L. Pandolfi, and P. Ballirano, "Structure of Polyethylene from X-Ray Powder Diffraction: Influence of the Amorphous Fraction on Data Analysis," *J Macromol Sci Phys*, vol. 39, no. 4, pp. 481–492, Jul. 2000, doi: 10.1081/MB-100100400.

[7] D. A. Tanjung and S. M. Munte, *Pembuatan Komposit Bioplastik*. Medan: LEMKOMINDO, 2023. [Online]. Available: https://penerbit.lemkomindo.org/

[8] K. Terayama, M. Sumita, M. Katouda, K. Tsuda, and Y. Okuno, "Efficient Search for Energetically Favorable Molecular Conformations against Metastable States via Gray-Box Optimization," *J Chem Theory Comput*, vol. 17, no. 8, pp. 5419–5427, Aug. 2021, doi: 10.1021/acs.jctc.1c00301.

[9] N. Sari Duwila and dan Muhammad Amin, "Density Functional Theory Senyawa Kompleks Ni 2+ , Zn 2+ Dan Pt 2+ Pirolidin-Ditiokarbamat," *Saintifik*, vol. 8, no. 1, pp. 23–31, 2023.

[10] A. Zaichenko and J. Autschbach, "Triplet-Singlet Emission of d-Block Metal Complexes Characterized by Spin-Orbit Natural Transition Orbitals," *ChemistryOpen*, vol. 13, no. 8, Aug. 2024, doi: 10.1002/open.202300291.

[11] Thoriq, M. Ragil, and D. Nurfadilla Chandra, "Stability Analysis of Several Glycine Conformations Using a Molecular Quantum Mechanics Computational Approach," *CHEMVIRO: Jurnal Kimia dan Ilmu Lingkungan*, vol. 3, no. 1, pp. 187–201, 2025.

[12] Y. T. Male, I. Wayan Sutapa, and O. Merion Ranglalin, "Computational Study Natural Color Essence (Dyes) As Active Material On Organic Solar Cell With Density Functional Theory (DFT)," 2015.

[13] F. Jensen, *Introduction to Computational Chemistry*. Wiley, 2017.

[14] W. J. Hehre, L. Radom, P. v. R. Schleyer, and J. A. Pople, *Ab Initio Molecular Orbital Theory*. Wiley, 1986.

[15] J. B. Foresman and Æ. Frisch, *Exploring Chemistry With Electronic Structure Methods*. Gaussian Inc, 2015.

[16] Z. Huang, M. S. Chen, C. P. Woroch, T. E. Markland, and M. W. Kanan, "A framework for automated structure elucidation from routine NMR spectra," *Chem Sci*, vol. 12, no. 46, pp. 15329–15338, Dec. 2021, doi: 10.1039/d1sc04105c.