Facultad de Ciencias Naturales y Exactas
Universidad del Valle

DFT-BASED INVESTIGATION OF THE ELECTRONIC STRUCTURE OF A DOUBLE-STRANDED AC B-DNA DIMMER AT DIFFERENT LEVELS OF THEORY

Emerson A. Rengifo
Universidad del Valle

Gustavo Murillo
Universidad del Valle

Received: april 12, 2012 Accepted: july 5, 2012

Pág. 117-122

Abstract

Calculations of the electronic structure of a stacked dimmer sequence from the D-(GCAAACGTTTGC)2 B-DNA dodecamer resolved in a PDB file 1HQ7 are performed within density functional theory. Seeking to understand the minimum level of theory that yields a reliable description for these systems, the basis sets 6-31g*, 6-31g*+BSSE, 6-311g*, 6-311g**, 6-311++g** along with the B3LYP and PBE0 exchange-correlation functionals were employed. These results are then used to implement a one-dimensional model of long stacked systems to obtain a new semi-empirical method that can be employed at low computational cost.

Keywords: DFT calculations, B3LYP functional, PBE0 functional, Stacked systems.

1 Introduction

Semi-empirical calculations have been the most successful for the description of organic compounds, from calculations of electronically excited states of polyenes, (both cyclic and linear) [1,2] up to the investigation of physical properties of phenol compounds (steric energy, oxygen charge, ionization potential, dipole moment, LUMO and bond length) [3]. Besides, semi-empirical approaches have also been applied for a phenomenological approach to the treatment of the electronic properties of single-stranded DNA [4]. Semi-empirical models should, however, be validated with accurate electronic structure results. Therefore, we undertook a study of the electronic structure based on density functional theory (DFT). Currently, an important number of publications show the efficiency of hybrid exchange-correlation (x-c) functionals like B3LYP and PBE0 to satisfactorily predict different features of organic [5,6] and biochemical [7,8]large size systems without requiring excessive amounts of computer time, memory, or disk space.

The goal of this work was to find the minimum level of theory that yields a reliable description of these systems and so, use the DFT output for each base pair as input for the one-dimensional model, which enables the description of much longer sequences. This procedure allows the description of arbitrarily low sequences by employing only the Kohn-Sham DFT energy levels (HOMO and LUMO) of individual pairs. The validation

of the model is carried out by comparing the model results for dimmers with those of DFT. These results allow calibrating our model to predict the highest occupied molecular orbital (HOMO), lowest unoccupied molecular orbital (LUMO), and other orbitals close to the HOMO-LUMO gap of long sequences of single- and double- DNA strands.

2 Methodology

From the D-(GCAAACGTTTGC)2 B-DNA dodecamer resolved in PDB file 1HQ7 (Fig. 1), the central fragment that is an AC dimmer is extracted, namely an A-T pair stacked with a C-G pair in the B-DNA conformation; the HyperChem package was used [9]. Only Hydrogen optimization was performed for the AC dimmer. All electronic structure calculations were carried out with Gaussian 03 [10] on an IBM parallel PC cluster, using 1 node and 2 processors. The B3LYP and PBE0 hybrid exchange-correlation (x-c) functionals and the most popular and successful basis sets for DNA-based stacked systems [11-13] were employed, namely 6-31g*, 6-31g*+BSSE, 6-311g*, 6-311g**, and 6-311++g**.

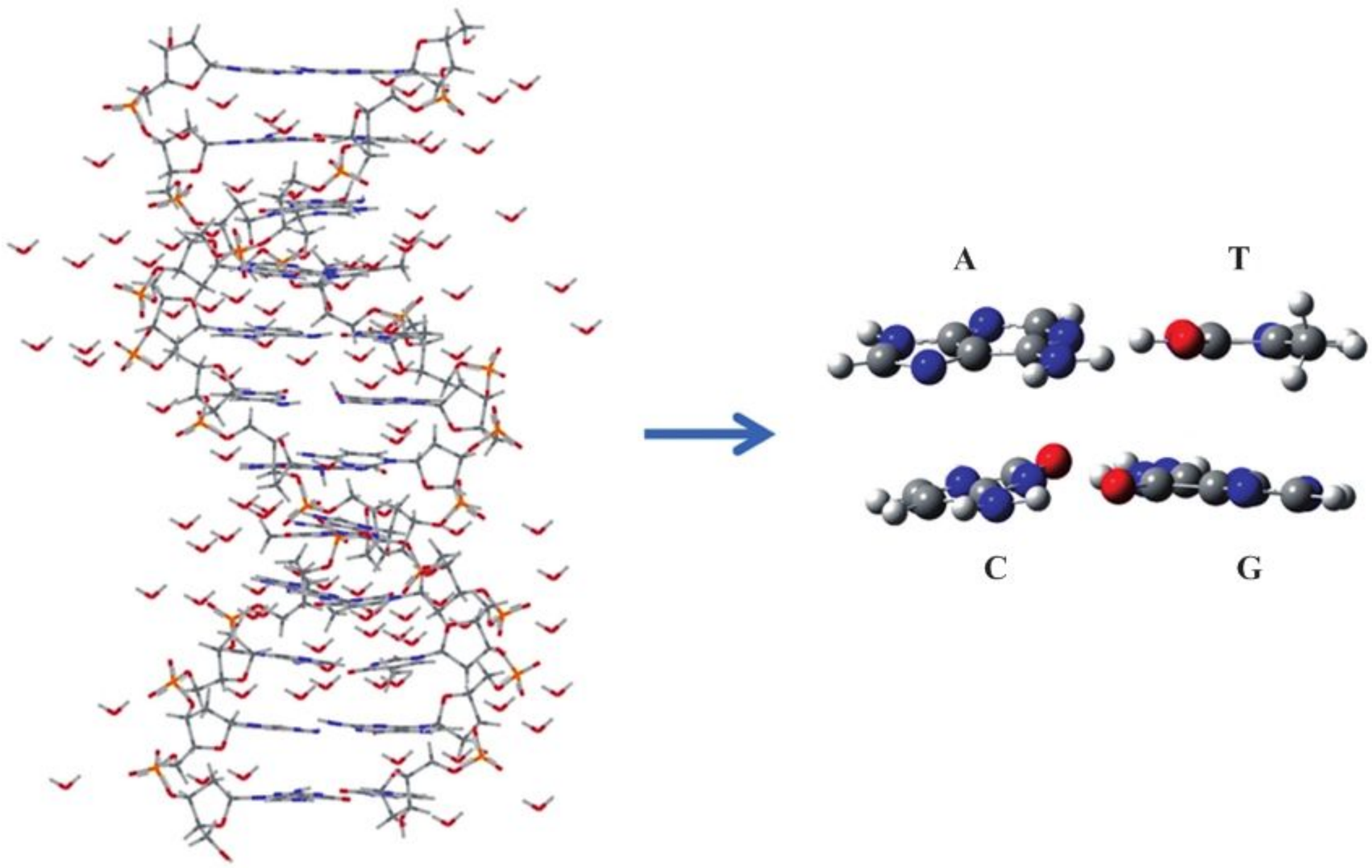

Figure 1. Structure of the dodecamer (PDB-ID 1HQ7) from which the stacked dimer AC was extracted.

3 Results and discussion

One of the most important parameters in the electronic structure is the energy gap between LUMO and the HOMO. The results for the HOMO-LUMO gap (E_{LUMO}-E_{HOMO}) with the different basis sets and x-c functionals are shown in Tables 1 and 2.

It is known that DFT underestimates the HOMO-LUMO gap in molecules and the band gap in semiconductors and insulators. Therefore, as a possible guideline in selecting the most accurate level of theory among different basis sets and exchange-correlation functional, one could select the method that produces the highest numerical value for the

gap. Nevertheless, this is not obviously the best choice and the accuracy may depend on the gap, but also for the stability of the results: if for a very poor basis set the highest gap is obtained, and for all the other larger basis sets almost the same value results, and quite different from the highest value, the larger basis sets are the good choice. Indeed, the most reasonable rule is to select the smallest basis set that yields results agreeing with those obtained with the largest affordable basis set.

From Tables 1 and 2, we found that if one discards the basis set superposition error (BSSE), all the investigated basis sets with both B3LYP and PBE0 would yield similar numerical values for the HOMO-LUMO gap. Particulary, the B3LYP gaps obtained with the most extended basis set 6-311++g** and that obtained with the smallest basis set 6-31g* differ by only 0.023 eV. The same quantities obtained with the PBE0 functional differ by 0.039 eV. Thus, if one is interested in physico-chemical properties that involve only the HOMO-LUMO gap and no other features of the electronic structure, the basis set 6-31g* would be a reasonable option. In general, the PBE0 x-c functional yields higher values of the fundamental gap and therefore it should be preferred.

Table 1. *Values of electronic-structure and total-energy parameters obtained with the B3LYP x-c functional and different basis sets for an AC dimmer.*

B3LYP A...T C...G	ΔE_{gap} (eV)	E_{homo} (eV)	E_{lumo} (eV)	E_{homo}-E_{homo-1} (eV)	E_{tot} (a.u.)	ΔE_{tot}(eV)
6-31g*	3.503	-4.782	-1.279	0.978	-1858.98	0.000
6-31g*+BSSE	5.588	-5.911	-0.323	0.760	-1858.96	-0.544
6-311g*	3.489	-4.993	-1.504	0.957	-1859.44	12.517
6-311g**	3.464	-5.016	-1.551	0.964	-1859.49	13.878
6-311++g**	3.480	-5.258	-1.778	0.916	-1859.53	14.966

Table 2. *Values of electronic-structure and total-energy parameters obtained with the PBE0 x-c functional and different basis sets for AC dimmer*

PBE0 A...T C...G	ΔE_{gap} (eV)	E_{homo} (eV)	E_{lumo} (eV)	E_{homo}-E_{homo-1} (eV)	E_{tot} (a.u.)	ΔE_{tot}(eV)
6-31g*	3.894	-5.038	-1.144	0.991	-1856.96	0.000
6-31g*+BSSE	5.990	-6.183	-0.193	0.823	-1856.93	-0.816
6-311g*	3.878	-5.201	-1.322	0.971	-1857.37	11.157
6-311g**	3.850	-5.222	-1.371	0.980	-1857.43	12.789
6-311++g**	3.855	-5.429	-1.574	0.936	-1857.46	13.606

If one looks at the absolute energy levels of the HOMO and LUMO, the conclusion would be somewhat different; in fact, the smallest basis set, 6-31g*, significantly underestimates the positions of both the HOMO and the LUMO relative to the vacuum level, both with B3LYP and PBE0. Only the largest basis set. 6-311++g**. would be a safe choice.

The energy gap between the HOMO and HOMO-1, which are both occupied states and, consequently, well estimated within DFT, is somewhat overestimated by all small basis sets. The difference between the smallest and largest basis values is only 0.062 and 0.055 eV with B3LYP and PBE0, respectively. Note that for each chosen basis set, the values of E_{HOMO}-E_{HOMO-1} obtained with B3LYP and PBE0 are quite similar, while this is not true for the fundamental HOMO-LUMO gap. This agrees with the fact that the occupied energy levels in DFT are more accurate than the unoccupied ones and, therefore, energy differences between occupied states are not expected to depend strongly on the x-c functional.

Regarding the total energy, an error exists with small basis sets, independent of the adopted functional. Hence, for applications where the total energy is important, one must employ the largest basis set available, in this case 6-311++g**; although the 6-311g* and 6-311g** sets are not dramatically off. To summarize the analysis of Tables 1 and 2, we infer that the extended basis set 6-311++g** is needed to produce reliable results on both the electronic structure and the total energy. For the fundamental gap only, even the smallest basis set 6-31g* would be satisfactory. Both explored functionals give the same trends and could, thus, be used with the same level of confidence. A slight preference for PBE0 is because it systematically yields higher values of the HOMO-LUMO gap for any basis sets.

On the other hand, the analysis of other quantities like isosurfaces that result from electronic structure DFT calculations could help us to identify the proper level of theory. Figure 2 illustrates the isosurface plots of the orbitals around the fundamental gap obtained with the B3LYP functional and the 6-311g* basis set. From these plots, we can note that the HOMO is localized on the guanine, the LUMO on cytosine, and the HOMO-1 on adenine. The locations of the HOMO and LUMO on guanine and cytosine, respectively, agree with theoretical and experimental data on the order of ionization potentials of the four nucleobases [14]. The same result was obtained with other basis sets.

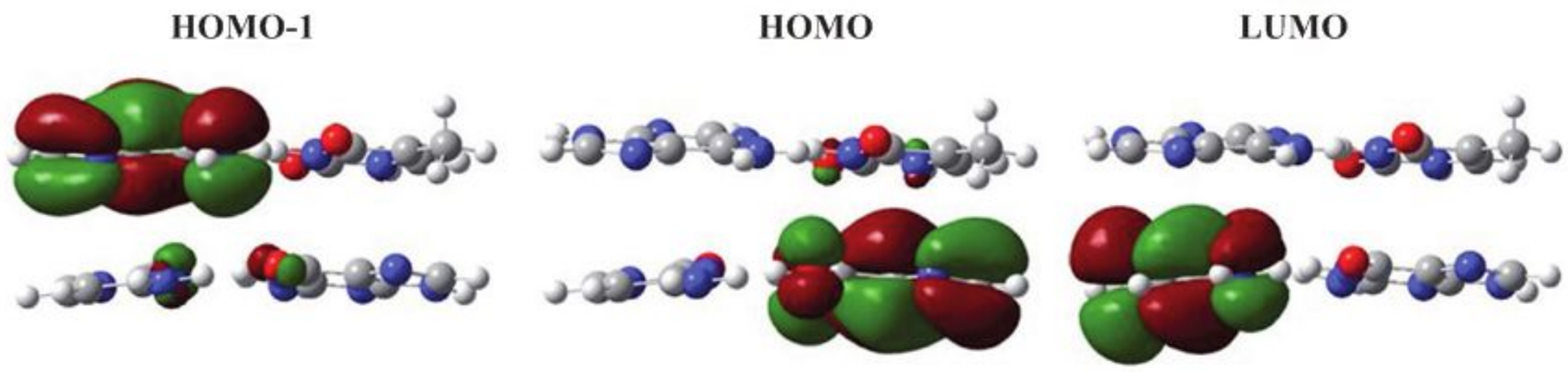

Figure 2. Isosurface plots of orbitals from Homo-1, Homo and Lumo, obtained with the B3LYP functional and the 6-311g* basis set.

4 Conclusions

The comparison of the HOMO and LUMO values extracted through different methods is not always correct. This is because the HOMOs are generated by the actual states of electronic interactions, while the LUMO's values are virtual states and the description of these is not always accurate. However, for qualitative purposes it is possible to make comparisons with virtual levels because they can offer an idea of the physical behavior of the system. Selecting a method to calculate a system is not easy and it requires clarity on what parameters we want to calculate.

For systems derived from B-DNA stacks, if we seek to describe load flow, it is possible to perform calculations with larger basis sets such as, 6-311++g**, because they have built-in polarization and diffuse functions. However, if we only were interested in the fundamental gap, even the smallest basis set, 6-31g*, would be satisfactory. Both functionals B3LYP and PBE0 give quite similar results, but a slight preference for PBE0 is because it systematically yields higher values of the HOMO-LUMO gap for any basis sets.

Acknowledgments

The authors are grateful to COLCIENCIAS for financial support and to Professor Rosa Di Felice for valuable discussions about the DFT theory and for granting us access to his cluster.

References

[1] Pariser, R.; Parr, R. *J. Chem. Phys* (1953), 21, 466.

[2] Pople, J. A. *Trans. Faraday Soc* (1953), 49, 1375.

[3] Ibrahim, A.; Abdalrazaq, E. *Am. J. Applied Sci* (2009), 6, 1385.

[4] Iguchi, K. *Int. J. Mod. Phys B*. (2004), 18, 1845.

[5] Wathelet, V.; Preat, J.; Bouhy, M.; Fontaine, M.; Perpète, E. A.; André, J-M.; Jacquemin, D. *Int. J. Quant. Chem*. (2006), 106 1853.

[6] Bohr, H. G.; Jalkanen, K. J.; Elstner, M.; Frimand, K.; Suhai, S. *Chem. Phys* (1999), 246, 13.

[7] Improta, R.; Barone, V.; Kudin, K.; Scuseria, G. E. *J. Chem. Phys* (2001), 114, 2541.

[8] Jacquemin, D.; Wathelet, V.; Preat, J.; Perpète, E. A. *Spectrochimica Acta Part A* (2007), 67, 334.

[9] *Hyper Chem Release 7 for windows:*Tools for Molecular Modeling; Hypercube Inc.: Publication HC70-00-01-00. 2002.

[10] Frisch, M. J. et al., Gaussian 03, Revision D.01G. Gaussian, Inc., Wallingford CT, (2004).

[11] Davidson, E. R.; Feller, D. *Chem. Rev*. (1986), 86, 681.

[12] González Mao, M. J.; Mandado, M.; Mosquera, R. *J. Phys. Chem. A* (2007), 111, 1998.

[13] Migliore, A.; Corni, S. Varsano, D.; Klein, M. L.; Di Felice, R. *J. Phys. Chem. B* (2009), 113, 9402.

[14] Cauët, E.; Valiev, M.; Weare, J. H. *J. Phys. Chem. B*. (2010), 114, 5886-5894.

Author's address

Emerson Rengifo
Estudiante de Doctorado en Ciencias Químicas
Departamento de Química, Universidad del Valle, Cali - Colombia
emerengi@univalle.edu.co

Gustavo Murillo
Estudiante de Doctorado en Ciencias Físicas
Departamento de Física, Universidad del Valle, Cali - Colombia
gumurillo@yahoo.com