The Topological Analysis of the ELFx Localization Function: Quantitative Prediction of Hydrogen Bonds in the Guanine–Cytosine Pair

Johanna Klein, Paul Fleurat-Lessard, Julien Pilmé

molecules

MDPI

Article

The Topological Analysis of the ELF_x Localization Function: Quantitative Prediction of Hydrogen Bonds in the Guanine–Cytosine Pair

Johanna Klein [1], Paul Fleurat-Lessard [2] and Julien Pilmé [1,*]

[1] Sorbonne Université, Laboratoire de Chimie Théorique, UMR 7616 CNRS, CC 137, 4, Place Jussieu F, CEDEX 05, 75252 Paris, France; johanna.klein@sorbonne-universite.fr
[2] Institut de Chimie Moléculaire de l'Université de Bourgogne (ICMUB), CNRS UMR 6302, 9 Avenue Alain Savary, BP 47870, CEDEX, 21078 Dijon, France; Paul.Fleurat-Lessard@u-bourgogne.fr
* Correspondence: julien.pilme@sorbonne-universite.fr

Abstract: In this contribution, we recall and test a new methodology designed to identify the favorable reaction pathway between two reactants. Applied to the formation of the DNA guanine (G)–cytosine (C) pair, we successfully predict the best orientation between the base pairs held together by hydrogen bonds and leading to the formation of the typical Watson Crick structure of the GC pair. Beyond the global minimum, some local stationary points of the targeted pair are also clearly identified.

Keywords: ELF; ELF_X; DNA; hydrogen bond; nucleophilic; electrophilic; base pair; cytosine; guanine

Citation: Klein, J.; Fleurat-Lessard, P.; Pilmé, J. The Topological Analysis of the ELF_x Localization Function: Quantitative Prediction of Hydrogen Bonds in the Guanine–Cytosine Pair. *Molecules* **2021**, *26*, 3336. https://doi.org/10.3390/molecules26113336

Academic Editors: Miroslav Kohout, Carlo Gatti, David L. Cooper and Maxim L. Kuznetsov

Received: 13 May 2021
Accepted: 29 May 2021
Published: 1 June 2021

1. Introduction

Among numerous ideas published by Linus Pauling, he proposed with Robert B. Corey in 1953 a pioneering triple DNA helix structure with the bases on the outside [1].

Although this Pauling's structure soon turned out to be false, this work has paved the way for the discovery of DNA's double-helix structure [2]. In recent decades, DFT quantum chemical studies of the Watson Crick base pairs investigated the geometry, the energy and other typical properties of the hydrogen bonds (HB) that hold together adenine–thymine (AT) and guanine–cytosine (GC) pairs [3–8]. Note that L. Pauling and R. B. Corey have already highlighted the role of hydrogen bonding in proteins [9]. The main concern assessed in this work is related to the validity of the molecular orbital point of view regarding the geometries of these base pair systems where hydrogen bonds play a crucial role. Indeed, it has been shown that the stability of the Watson Crick base pairs is related to a charge-transfer due to donor/acceptor orbital interactions (oxygen and nitrogen lone pairs, N-H σ* character) [10]. For example, consider the well-known most stable pair structure guanine (G)–cytosine (C), as depicted in Figure 1 [7,10–12].

In this article, we tackle the possibility to find the guanine (G)–cytosine (C) pair geometry simply by looking at the orientation of the donor and acceptor domains of the bases.

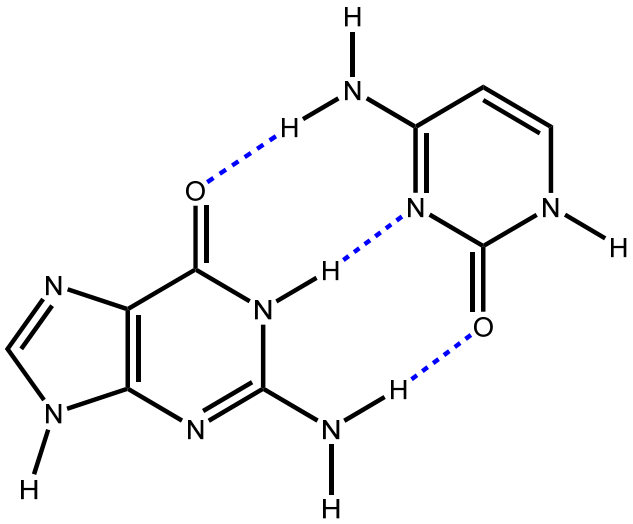

Figure 1. Guanine (G) cytosine (C) pair structure. The intermolecular hydrogen-bonds are displayed in blue.

2. Electron Localization Function for Chemical Reactivity

Nowadays, the topological analysis of the electron localization function (ELF) is a well-established tool to describe both covalent and non-covalent interactions [13–20]. However, the tricky question is to determine the most favorable relative orientations between the ELF topological domains of two reactants and, thus, to identify the preferred pathways when both molecules approach each other remains a tremendous challenge. Intuitively, it is established that favorable chemical reactions happen when electron acceptor and electron donor domains are adequately oriented. Recently, we have proposed a methodology designed to identify the favorable orientations between two reactants [21]. In this work, the topological domains are the ones of the modified ELF, termed ELF_x [22] defined from ELF as follows:

$$\chi_x(\mathbf{r}) = \frac{\chi(\mathbf{r})}{2\,x(\mathbf{r})} \text{ and } ELF_x(\mathbf{r}) = \frac{1}{1+\chi_x(\mathbf{r})^2} \quad (1)$$

The kernel $\chi(\mathbf{r})$ of ELF being defined as:

$$\chi(\mathbf{r}) = \frac{\tau_N(\mathbf{r}) - \frac{1}{8}\frac{\left|\nabla\rho(\mathbf{r})_N(\mathbf{r})\right|^2}{\rho(\mathbf{r})_N}}{c_F\rho(\mathbf{r})_N{}^{5/3}} \quad (2)$$

where $c_F = \frac{3}{10}\left(3\pi^2\right)^{\frac{2}{3}}$ is the Fermi constant, $\tau_N(\mathbf{r})$ is the positive definite kinetic energy density and $\rho(\mathbf{r})_N$ is the total electron density of a molecular system with N electrons. $x(\mathbf{r})$ is a normalized dimensionless quantity that can be expressed from the field of the frontier molecular orbitals [23].

$$x(\mathbf{r}) = \frac{\rho(\mathbf{r})_{HOMO}}{\rho(\mathbf{r})_N} \text{ or } x(\mathbf{r}) = \frac{\rho(\mathbf{r})_{LUMO}}{\rho(\mathbf{r})_{N+1}} \quad (3)$$

$\rho(\mathbf{r})_{N+1}$ is the total electron density of the molecular system with N + 1 electrons with the same geometry and the same orbitals that are obtained for the system with N electrons. In this latter case where $x(\mathbf{r}) = \frac{\rho(\mathbf{r})_{LUMO}}{\rho(\mathbf{r})_{N+1}}$, $\rho(\mathbf{r})_{N+1}$ (and $\tau_{N+1}(\mathbf{r})$) are also consistently used for the calculation of the ELF kernel.

The ELF_x localization domains are well suited for describing the chemical reactivity between donors and acceptors because they match with the electrophilic and the nucleophilic regions which are spread out over the molecular space. This is illustrated in Figure 2, which represents the ELF_x domains of guanine and cytosine.

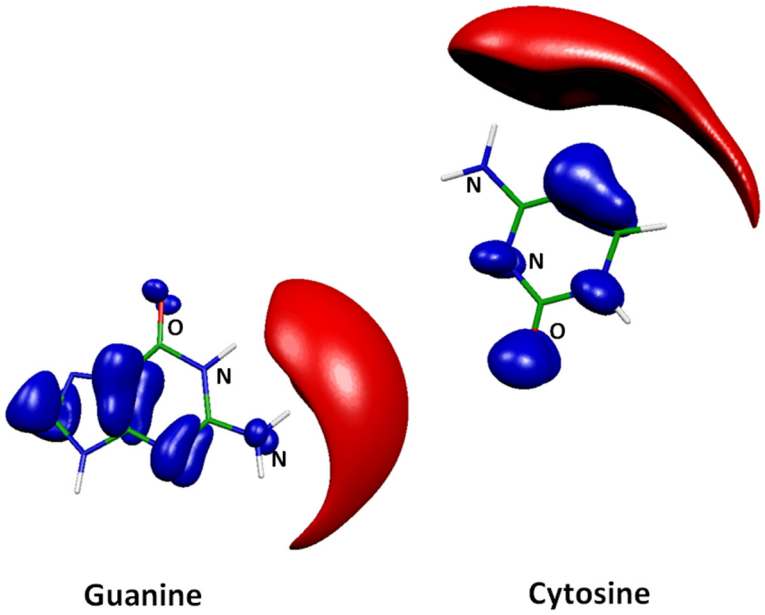

Figure 2. Main ELF_X localization domains of the guanine and cytosine molecules computed at the M06-2X/6-311++G(3df, 2pd) level of theory. Color Code: blue: nucleophilic regions and red: electrophilic regions. Carbon atoms are in green, Nitrogen atoms in blue, Oxygen atoms in red and Hydrogen atoms in white.

The ELF_X topological analysis of the guanine and cytosine molecules obtained in their isolated states yields, respectively, to valence basins accounting for electrophilic basins (red domains) and several nucleophilic basins (blue domains). The outside domains around hydrogen atoms appear as electrophilic while domains around nitrogen atoms as well as the oxygen lone-pairs clearly have a nucleophilic character.

3. Theoretical Model

3.1. Coulomb Intermolecular Interaction Energy

The total energy of a molecule or a complex can be split within the framework of the interacting quantum atoms (IQA) [24,25]. The IQA coulomb contribution between two molecules MA et MB, here termed E^{Coul}_{MA-MB}, reads:

$$E^{Coul}_{MA-MB} = \sum_{\Omega_A \in (4)MA} \sum_{\Omega_B \in MB} \left[\int_{\mathbf{r_A} \in \Omega_A} \int_{\mathbf{r_B} \in \Omega_B} \frac{[Z_A\, \delta(\mathbf{r_A} - \mathbf{R_A}) - \rho(\mathbf{r_A})]\,[Z_B\, \delta(\mathbf{r_B} - \mathbf{R_B}) - \rho(\mathbf{r_B})]}{|\mathbf{r_A} - \mathbf{r_B}|}\, d\mathbf{r_A} d\mathbf{r_B} \right] \quad (4)$$

$|\mathbf{r_A} - \mathbf{r_B}|$ being the distance between an electron in the domain Ω_A and an electron in the domain Ω_B, respectively. $\mathbf{R}_A$ and $\mathbf{R}_B$ are the nuclear locations of atoms A and B belonging to Ω_A and Ω_B domains with charges Z_A and Z_B. When MA and MB are located far from each other, we assume that E^{Coul}_{MA-MB} accounts for a large fraction of the total interaction energy [21].

3.2. Electron Transfer

The coulomb energy stabilization between an electron donor (MA) and an electron acceptor (MB) can be evaluated by the first-order variation of E^{Coul}_{MA-MB} expressed in terms of the response to changes in the number of electrons ΔN_A or ΔN_B where the external potential remains unchanged:

$$\Delta E^{Coul}_{MA-MB} = \left(\frac{\partial E^{Coul}_{MA-MB}}{\partial N_A} \right)_{N_B} \Delta N_A + \left(\frac{\partial E^{Coul}_{MA-MB}}{\partial N_B} \right)_{N_A} \Delta N_B = E^{MA/MB}_{dual}\, \Delta N_A \quad (5)$$

The total variation $\Delta N = \Delta N_A + \Delta N_B = 0$ because the total system is isolated. $E_{dual}^{MA/MB}$ is negative when the electron transfer goes spontaneously from MA (nucleophile) to MB (electrophile). After some developments previously detailed elsewhere [21], we obtain:

$$E_{dual}^{MA/MB} = \sum_{\Omega_A \in MA} \sum_{\Omega_B \in MB} \left[\int_{\mathbf{r_A} \in \Omega_A} \int_{\mathbf{r_B} \in \Omega_B} \frac{f(\mathbf{r_B}) [Z_A \delta(\mathbf{r_A} - \boldsymbol{R_A}) - \rho(\mathbf{r_A})] - f(\mathbf{r_A}) [Z_B \delta(\mathbf{r_B} - \boldsymbol{R_B}) - \rho(\mathbf{r_B})]}{|\mathbf{r_A} - \mathbf{r_B}|} d\mathbf{r_A}\, d\mathbf{r_B} \right] \quad (6)$$

where $f(\mathbf{r_A})$ and $f(\mathbf{r_B})$ are the Fukui functions [23] typically associated with reactive nucleophilic or electrophilic sites of the reactants.

The choice of the condensation scheme remains arbitrary as far as that of an electron domain or the definition of an atom in a molecule remains arbitrary. Here, we can clearly dissociate the MA and MB domains where electrophilic and the nucleophilic regions are spread out over their respective molecular space. This typically matches with the topological partition of the electron localization function ELF_x.

3.3. Practical Interactions Model

Equation (4) is exact but can be computationally expensive. In practice, it can be numerically evaluated by means of a multipole expansion (ME) [26]. We use only the first terms of the ME (that is only the monopoles). For the general case in which both molecules MA and MB exhibit some donor and acceptor sites, the monopoles' development leads to the compact equation:

$$E_{dual} = E_{dual}^{MA/MB} + E_{dual}^{MB/MA} = \sum_{\Omega_A \in MA} \sum_{\Omega_B \in MB} \left[\frac{2\, N_{\Omega_{A,Nu}}\, N_{\Omega_{B,Nu}}}{|r_{\Omega_A} - r_{\Omega_B}|} - \frac{N_{\Omega_{A,Nu}}\, N_{\Omega_{B,El}}}{|r_{\Omega_A} - r_{\Omega_B}|} - \frac{N_{\Omega_{A,El}}\, N_{\Omega_{B,Nu}}}{|r_{\Omega_A} - r_{\Omega_B}|} + \frac{Z_A \left(N_{\Omega_{B,El}} - N_{\Omega_{B,Nu}}\right)}{|r_{\Omega_B} - R_A|} + \frac{Z_B \left(N_{\Omega_{A,El}} - N_{\Omega_{A,Nu}}\right)}{|r_{\Omega_A} - R_B|} \right] \quad (7)$$

in which $N_{\Omega Nu/El}$ are the populations of nucleophile/electrophile domains, respectively, obtained from the usual condensation of the HOMO/LUMO density computed over the ELF_x basin volumes [27]. r_{Ω_A} and r_{Ω_B} are the locations of basin attractors belonging to Ω_A and Ω_B domains, respectively.

4. Results and Discussion

We explored the conformational space of the base pair GC_{WC} formation using the Equation (7) with the algorithm previously outlined elsewhere [21]. All relative rotation angles (θ, φ) of the center of mass of C around the center of mass of G have been tested, with the distance between the centers of mass of C and G being frozen to 8 Å. Note that the corresponding optimized distance between the centers of mass was found close to 6 Å at the M06-2X/6-311++G(3df, 2pd) level of theory. For a given (θ, φ) couple, the process selects the best orientation of C associated to the lowest value of E_{dual}. Figure 3 displays the obtained map $E_{dual}(\theta, \varphi)$ together with the corresponding map of the DFT intermolecular interaction energy $E_{int}^{0}(\theta, \varphi)$ computed from the relevant isolated cytosine and guanine.

It is worth noting the good mapping of E_{dual} and the DFT intermolecular interaction energy $E_{int}^{0}(\theta, \varphi)$. Indeed, the locations of critical points of E_{dual}, notably the location of the global minimum (termed (A) on Figure 3), are in agreement with the DFT intermolecular interaction energy surface. We noted that the structure associated to the global minimum corresponds to the well-known orientation between C and G leading to the natural Watson Crick base pair GC_{WC} structure where three typical HNH···O=C/NH···N/C=O···HNH intermolecular hydrogen bonds are observed. Two other stationary points (denoted by (B) and (C) on Figure 3) corresponding to already identified pair structures are also found on the E_{dual} map [7]. These latter structures highlight typical HNH···O and CH···N donor/acceptor interactions. The presence of structures (A), (B) and (C) are confirmed on the DFT interaction energy surface.

Further analysis of the E_{dual} conformational space obtained for each given (θ, φ) couple (not only for the best orientation of C in front of G) has led us to find other local stationary points. Some of them have been previously identified in the literature [7,8]. For example, the geometry of the second lowest minimum is displayed in Figure 4b: this pair appears clearly stabilized by two symmetrically NH···O=C hydrogen bonds.

Thus, in spite of numerous approximations used in this work, we show that the DFT energetic properties as well as the structural parameters of some identified pairs can be reasonably reproduced from our methodology.

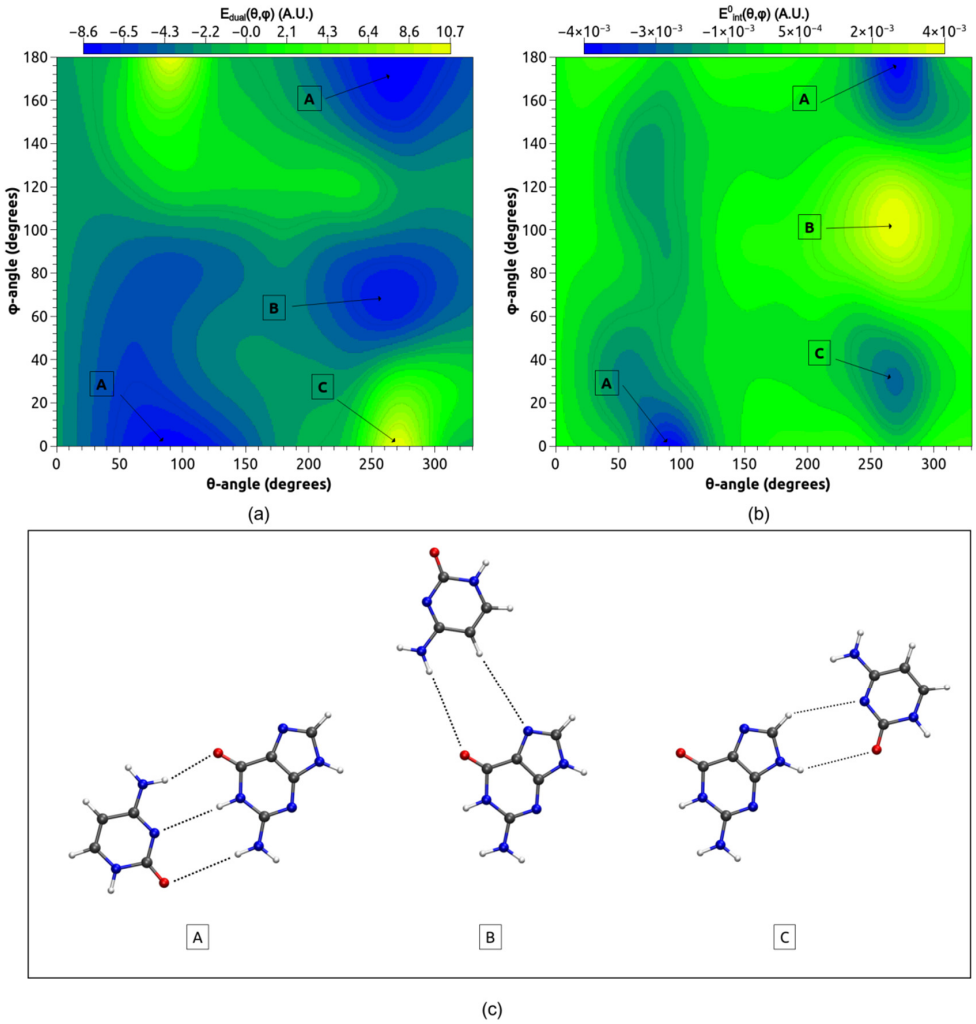

Figure 3. Two-dimensional map calculated at the M06-2X/6-311++G(3df,2pd) level of theory $E_{dual}(\theta, \varphi)$ (**a**) the DFT intermolecular interaction energy $E^0_{int}(\theta, \varphi)$ between the cytosine and the guanine vs. (**b**) $E_{dual}(\theta, \varphi)$ surfaces for base pair GC formation. (**c**) Stationary points obtained from the maps, the structure (A) is the global minimum. Carbon atoms are in grey, Nitrogen atoms in blue, Oxygen atoms in red and Hydrogen atoms in white.

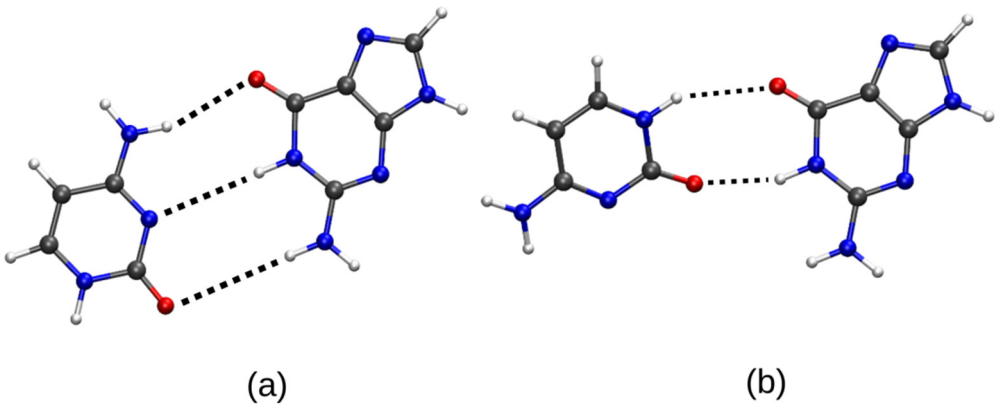

Figure 4. E_{dua}l global minima (**a**) vs. local minima (**b**) of the guanine (G)–cytosine (C) complex. Carbon atoms are in grey, Nitrogen atoms in blue, Oxygen atoms in red and Hydrogen atoms in white.

5. Conclusions

The information obtained from the domains of ELF_x function and their populations has been used to propose an empirical model coulomb stabilization energy between electron donor and electron acceptor domains. Our methodology was able to predict the best orientations between the cytosine and the guanine leading to the formation of base pair structures. We unveil a noticeable mimicking of E_{dual} onto the DFT intermolecular interaction energy E^0_{int}. In particular, we show that the global minimum, easily identified on the E_{dual} energy surface, corresponds to the well-known Watson Crick structure for the base pair GC_{WC} in which the guanine and the cytosine molecules are held together by three hydrogen bonds (see Figures 1 and 3). Some local stationary points of the GC pairs have been also identified.

Author Contributions: Conceptualization, J.P. and J.K.; methodology, J.P.; software, J.P and J.K.; validation, J.P., J.K. and P.F.-L.; formal analysis, J.P., J.K. and P.F.-L.; investigation, J.P., J.K. and P.F.-L.; resources, J.P. and J.K.; data curation, J.K. and P.F.-L.; writing—original draft preparation, J.P.; writing—review and editing, J.P., J.K. and P.F-L.; visualization, J.P. and J.K.; supervision, J.P.; project administration, J.P.; funding acquisition, J.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data that support the findings of this study are available. Additionnal information can be requested from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Pauling, L.; Corey, R.B. A Proposed Structure For The Nucleic Acids. *Proc. Natl. Acad. Sci. USA* **1953**, *39*, 84. [CrossRef]
2. Watson, J.D.; Crick, F.H.C. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* **1953**, *171*, 737–738. [CrossRef]
3. Gould, I.R.; Kollman, P.A. Theoretical Investigation of the Hydrogen Bond Strengths in Guanine-Cytosine and Adenine-Thymine Base Pairs. *J. Am. Chem. Soc.* **1994**, *116*, 2493–2499. [CrossRef]
4. Santamaria, R.; Vázquez, A. Structural and electronic property changes of the nucleic acid bases upon base pair formation. *J. Comput. Chem.* **1994**, *15*, 981–996. [CrossRef]
5. Brameld, K.; Dasgupta, S.; Goddard, W.A. Distance Dependent Hydrogen Bond Potentials for Nucleic Acid Base Pairs from ab Initio Quantum Mechanical Calculations (LMP2/cc-pVTZ). *J. ExpressionPhys. Chem. B* **1997**, *101*, 4851–4859. [CrossRef]
6. Zelený, T.; Hobza, P.; Kabeláč, M. Microhydration of guanine· · · cytosine base pairs, a theoretical Study on the role of water in stability, structure and tautomeric equilibrium. *Phys. Chem. Chem. Phys.* **2009**, *11*, 3430–3435. [CrossRef] [PubMed]
7. Thoa, T.T.; Hue, N.T.M. Theoretical study of the interaction between guanine and cytosine. *Vietnam J. Chem.* **2018**, *56*, 509–515. [CrossRef]
8. Nir, E.; Janzen, C.; Imhof, P.; Kleinermanns, K.; de Vries, M.S. Pairing of the nucleobases guanine and cytosine in the gas phase studied by IR–UV double-resonance spectroscopy and ab initio calculations. *Phys. Chem. Chem. Phys.* **2002**, *4*, 732–739. [CrossRef]
9. Pauling, L.; Corey, R.B. Two hydrogen-bonded spiral configurations of the polypeptide chain. *J. Am. Chem. Soc.* **1950**, *72*, 5349. [CrossRef]
10. Fonseca Guerra, C.; Bickelhaupt, F.M.; Snijders, J.G.; Baerends, E.J. The Nature of the Hydrogen Bond in DNA Base Pairs: The Role of Charge Transfer and Resonance Assistance. *Chem. A Eur. J.* **1999**, *5*, 3581–3594. [CrossRef]
11. Asensio, A.; Kobko, N.; Dannenberg, J.J. Cooperative Hydrogen-Bonding in Adenine−Thymine and Guanine−Cytosine Base Pairs. Density Functional Theory and Møller−Plesset Molecular Orbital Study. *J. Phys. Chem. A* **2003**, *107*, 6441–6443. [CrossRef]
12. Mo, Y. Probing the nature of hydrogen bonds in DNA base pairs. *J. Mol. Model.* **2006**, *12*, 665–672. [CrossRef] [PubMed]
13. Silvi, B.; Savin, A. Classification of chemical bonds based on topological analysis of electron localization functions. *Nature* **1994**, *371*, 683–686. [CrossRef]
14. Silvi, B.; Gillespie, R.J.; Gatti, C. 9.07—Electron Density Analysis. In *Comprehensive Inorganic Chemistry II*, 2nd ed.; Reedijk, J., Poeppelmeier, K., Eds.; Elsevier: Amsterdam, The Netherlands, 2013; pp. 187–226. [CrossRef]
15. Becke, A.D.; Edgecombe, K.E. A simple measure of electron localization in atomic and molecular systems. *J. Chem. Phys.* **1990**, *92*, 5397–5403. [CrossRef]
16. Savin, A.; Jepsen, O.; Flad, J.; Andersen, O.K.; Preuss, H.; von Schnering, H.G. Electron Localization in Solid-State Structures of the Elements: The Diamond Structure. *Angew. Chem. Int. Ed. Engl.* **1992**, *31*, 187–188. [CrossRef]
17. Fuster, F.; Silvi, B. Does the topological approach characterize the hydrogen bond? *Theor. Chem. Acc.* **2000**, *104*, 13–21. [CrossRef]

18. Pilmé, J.; Renault, E.; Ayed, T.; Montavon, G.; Galland, N. Introducing the ELF Topological Analysis in the Field of Quasirelativistic Quantum Calculations. *J. Chem. Theory Comput.* **2012**, *8*, 2985–2990. [CrossRef]
19. Polo, V.; Gonzalez-Navarrete, P.; Silvi, B.; Andres, J. An electron localization function and catastrophe theory analysis on the molecular mechanism of gas-phase identity SN2 reactions. *Theor. Chem. Acc.* **2008**, *120*, 341–349. [CrossRef]
20. Fuentealba, P.; Chamorro, E.; Santos, J.C. Chapter 5 Understanding and using the electron localization function. In *Theoretical and Computational Chemistry*; Toro-Labbé, A., Ed.; Elsevier: Amsterdam, The Netherlands, 2007; Volume 19, pp. 57–85.
21. Klein, J.; Fleurat-Lessard, P.; Pilmé, J. New insights in chemical reactivity from quantum chemical topology. *J. Comput. Chem.* **2021**, *42*, 840–854. [CrossRef]
22. Pilmé, J. Electron localization function from density components. *J. Comput. Chem.* **2017**, *38*, 204–210. [CrossRef]
23. Fukui, K.; Yonezawa, T.; Shingu, H. A Molecular Orbital Theory of Reactivity in Aromatic Hydrocarbons. *J. Chem. Phys.* **1952**, *20*, 722–725. [CrossRef]
24. Blanco, M.A.; Martín Pendás, A.; Francisco, E. Interacting Quantum Atoms: A Correlated Energy Decomposition Scheme Based on the Quantum Theory of Atoms in Molecules. *J. Chem. Theory Comput.* **2005**, *1*, 1096–1109. [CrossRef]
25. Martín Pendás, A.; Francisco, E.; Blanco, M.A. Electron–electron interactions between ELF basins. *Chem. Phys. Lett.* **2008**, *454*, 396–403. [CrossRef]
26. Stone, A.J.; Alderton, M. Distributed multipole analysis. *Mol. Phys.* **1985**, *56*, 1047–1064. [CrossRef]
27. Fuentealba, P.; Chamorro, E.; Cárdenas, C. Further exploration of the Fukui function, hardness, and other reactivity indices and its relationships within the Kohn–Sham scheme. *Int. J. Quantum Chem.* **2007**, *107*, 37–45. [CrossRef]