Stabilization and interaction energies of non planar DNA base pairs Guanine-Cytosine and Adenine-Thymine. A study based on the SIESTA method

C. Espejo[a,b] and R.R. Rey-González[b]

[a]*Departamento de Ciencias Básicas, Universidad de Bogotá Jorge Tadeo Lozano,*
Cra 4 # 22-61, Bogotá, Colombia.
[b]*Departamento de Física, Universidad Nacional de Colombia,*
Cra 30 calle 45, Bogotá, Colombia.

In this work we calculate interaction and stabilization energies of the the Watson-Crick H-bonded base pairs Guanine-Cytosine and Adenine-Thymine at the GGA level of DFT using the SIESTA method. Structural optimizations are performed by conjugate gradients without geometrical constraints, allowing both nonzero propeller twist angle between nucleobases and pyramidalization of amino groups. Comparison of the obtained geometries with results from the best ones reported in literature (correlated MP2 level) shows differences of the order of 1% in the intramolecular base distances. On the other hand, proportion between stabilization energies under pair formation is found to have approximately the same value as the reported results, while absolute stabilizations energies are different than the reported ones.

Keywords: DNA base pairs; DFT; SIESTA; Interaction energy.

En este trabajo calculamos las energías de interacción y estabilización de los pares de bases nitrogenadas Guanina-Citosina y Adenina-Timina en la configuración Watson-Crick, dentro del marco de la aproximación GGA de la teoría DFT, usando el método SIESTA. Las optimizaciones estructurales son realizadas con el método de los gradientes conjugados sin ligaduras geométricas, permitiendo estructuras no planas para las bases y piramidalización de los grupos amino. La comparación entre las geometrías obtenidas con los resultados provenientes de los mejores cálculos encontrados en la literatura (MP2 con correlación) indican diferecias del orden del 1% en las distancias intramoleculares de las bases. De otro lado, se encontró que la proporción entre las energías de estabilización de los pares de bases tiene aproximadamente el mismo valor que el reportado, mientras sus valores absolutos presentan diferencias.

Descriptores: Bases nitrogenadas; DFT; SIESTA; Energía de interacción.

PACS: 31.25.Qm; 31.15.Ew; 31.15.Ar

1. Introduction

Hydrogen bonds that hold together the two strands of nucleotides in DNA have been the main scope of many experimental and theoretical investigations for three decades. The importance of this molecular interaction is due to its role in DNA replication and complementarity of nucleic acid bases, which is the cornerstone of the genetic code, see Fig. 1. Bonding enthalpies of complementary bases at the gasphase [1] and bond distances and geometrical data of crystallized samples [2, 3] were obtained from the experimental research in the 70's and 80's. Nowadays, the REMPI technique allows to determine optical spectra of different types of isolated base pairs to measure the optical response of isolated bases within small clusters of water molecules [4, 5]. On the other hand, the theoretical description of base pairing has continuously evolved after improved numerical techniques and advances in computational performance arose [6]. As far as we know, the best performed calculations on these molecules are from Šponer et al. using RI-MP2 method [6] and from Fonseca Guerra et al. using a DFT approach [7, 8]. These results are considered as reference data. Recently work have been performed in order to evaluate exited states of base pairs by means of TDDFT [9].

Growing fields of nanobiology, computational chemistry and molecular biophysics have experienced a great step for-

ward since the development of a high degree of accuracy and moderated computational cost *ab initio* methods based on DFT. These methods allow one to simulate molecular systems with a large number of atoms, where quantum chemistry methods are not applicable. Such is the case of the so-called order-N methods, where linear scaling of computational work with the system's size is the main goal. The recently developed order-N *ab initio* method SIESTA [10, 11] is one of these proposals. In the present work we report some results of the SIESTA method on structural and bond properties of non planar DNA base pairs, as opposed to previous work where planar bases were studied [12].

2. The SIESTA method

The SIESTA method is one of the recently formulated *ab initio* order-N methods based on DFT. The computational implementation of the method is known as the SIESTA package [10]. Order-N scaling represents an important step forward because in traditional *ab initio* calculations the computational cost scales as N^3 [11, 13]. Order-N scaling is a consequence of the following method issues:

1. The customary diagonalization procedure is substituted by the minimization of a modified energy functional which automatically produces orthonormal states [13].
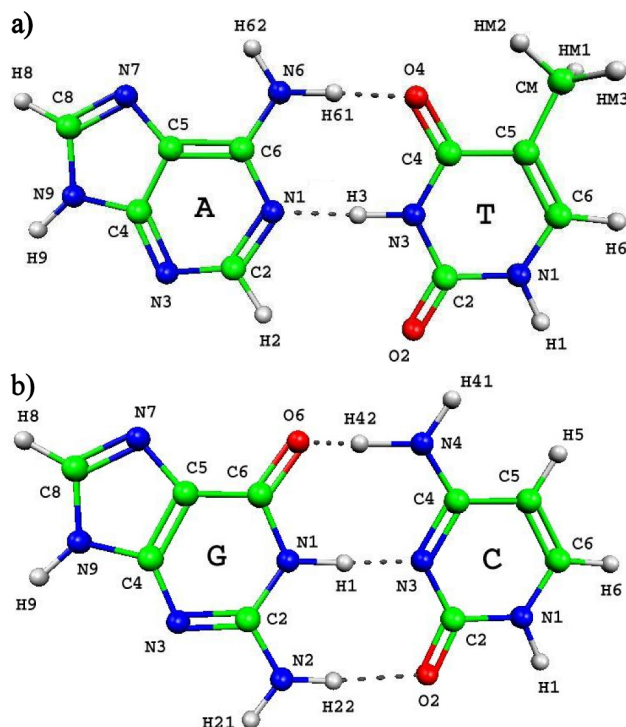
FIGURE 1. Complementary base pairs Adenine-Thymine (AT) and Guanine-Cytosine (GC) in the structure they acquire in B-DNA, the so called Watson-Crick base pair structure. Numbering of atoms is also shown.

2. Exchange and correlation interactions are treated within both local (spin) density approximation (LDA/LSD) and the generalized gradient approximation (GGA).

3. Core electrons are replaced by norm-conserving pseudopotentials [14] in their fully nonlocal form (Kleynman-Bylander form) [15].

4. Eigenfunctions are expanded on a basis set of strictly confined atomic orbitals.

5. Wavefunctions and electronic density are calculated in a three-dimensional grid yielding to low cost numerical integrals at hamiltonian evaluation.

The first issue above is used when the number of the atoms in the system is greater than 100. Hence, in the present calculations, the traditional diagonalization techniques are used to find the Kohn-Sham states of the system. Another issue of SIESTA is that it allows one to perform structural relaxations by means of several techniques such as Verlet dynamics, conjugate gradients or simulated annealing. In this work, relaxations were performed by conjugate gradients. In the software package, programs are included to generate both pseudopotentials and numerical atomic orbitals basis set.

3. Calculation details

It is well known that LDA aproximation of DFT is not suitable to calculate weak molecular interactions like hydrogen bonding; therefore, we have chosen the PBE GGA functional [16] to evaluate exchange and correlation energy.

The first step in a SIESTA calculation is the generation of the pseudopotentials for each chemical species in the molecule or solid system. In our case, norm conserving pseudopotentials are obtained for H, C, N and O atoms following the generation scheme proposed by Hamman et al. [14]. Reliability of the pseudopotentials is checked by means of a transferability test, in which the eigenenergies of the atoms in several electronic configurations are calculated using the pseudopotential, and the obtained values are compared to the results from all-electron *ab initio* calculations for the same set of configurations. If the energy difference is less than 1mRy for every configuration, then the pseudopotential is considered to be reliable.

We have used the following valence configurations to produce the pseudopotentials: H $4f^{0.5\uparrow 0.5\downarrow}$, C $2s^{1\uparrow 1\downarrow}2p^{0.5\uparrow 0.5\downarrow}3d^{0.25\uparrow 0.25\downarrow}4f^{0.25\uparrow 0.25\downarrow}$, O $2s^{1\uparrow 1\downarrow}2p^{1\uparrow 1\downarrow}3d^{0.5\uparrow 0.5\downarrow}4f^{0.5\uparrow 0.5\downarrow}$, and N $2s^{1\uparrow 1\downarrow}2p^{0.5\uparrow 0.5\downarrow}3d^{0.5\uparrow 0.5\downarrow}4f^{0.5\uparrow 0.5\downarrow}$, the superscript points out the population of spin states. Transferability tests for C, N and O GGA pseudopotentials are not so good as previosly obtained LDA pseudopotentials [17]. The average of differences in the present case are: 2mRy in carbon, 1mRy in nitrogen and 2mRy in oxygen. The latter suggests the need to adopt another pseudopotential generation scheme for higher performance simulations. For hydrogen GGA pseudopotentials, transferability tests supply much better results. Averaged energy differences are around 0.1mRy.

Basis sets in SIESTA are composed of numerical atomic orbitals confined to a finite range by means of an adequate confining potential [18]. As a result of the confinement, eigenenergies of pseudoatomic wavefunctions are increased from the isolated pseudoatomic values. The difference between them, known as *energy shift*, provides a single parameter from which basis set can be generated systematically [18]. The cut-off radius of DZP basis set calculated for an *energy shift* of 50 meV ranged from 2.055Å to 5.471Å. Atomic orbitals are constructed from pseudopotential eigenfuntions with crescent number of nodes, while polarization wave functions are obtained from the solution of the pseudoatoms inside a polarizing electric field.

Fineness of the spacial grid is measured by the energy of the shortest wavelength plain-wave that can be described with the grid [12]. In the present work, the grid cutoff used is 125 Ry. Structural relaxations are performed by a conjugate gradient technique until the force on each atom reaches a tolerance of 0.04 eV/Å. The maximum displacement of atoms under the relaxation steps are 0.2 Bohrs. No geometrical constraints are imposed to ionic coordinates, allowing optimal non planar structures to base pairs. After those optimal geometries are achieved, calculations are performed in order to obtain interaction energies as a function of the distance between bases.

TABLE I. Isolated Adenine bond distances (Å) and bond angles (degrees). Comparision with both reported calculations and experimental results [2], [3].

Distance	Here	Ref [20]	Ref [7]	Exp	Angle	Here	Ref [20]	Ref [7]	Exp
C2-N3	1.354	1.339	1.34	1.332	N3-C4-C5	126.3	127.1	127.0	126.9
N1-C2	1.364	1.353	1.34	1.338	C2-N3-C4	111.0	110.8	111.1	110.8
C6-N1	1.357	1.341	1.35	1.349	N9-C4-C5	103.8	104.3	104.4	105.7
C5-C6	1.425	1.409	1.41	1.409	C8-N9-C4	107.0	106.9	106.9	105.9
C4-C5	1.413	1.399	1.40	1.382	N7-C8-N9	113.1	113.6	113.3	113.8
N3-C4	1.349	1.343	1.34	1.342	C5-N7-C8	103.8	103.2	104.0	103.9

4. Results

4.1. Isolated bases structure

Total energy and geometrical structure of isolated bases are calculated as a prior step to base pairs calculations. Initial coordinates are taken from nucleic acid database [19], where reported data from x ray diffraction experiments on crystalized DNA samples are found. Therefore, these coordinates correspond to the structure of base pairs bonded to the DNA backbone. Results for several bond distances and bond angles in isolated Adenine after structural relaxation are shown in Table I. Comparison from higher quality, but also expensive, calculations on isolated bases [7] [20] show that both our distances and angles differ by only $\simeq 1\%$. Similar results are obtained for all other bases. This result indicates that the SIESTA method is a cheap and reliable approach to molecular structure calculations.

4.2. GC and AT structure

Once the isolated base structure and energy are calculated, relaxations for base pairs are performed. In Fig. 1, the optimal structure of both base pairs is shown. Relaxation to base pair optimal geometries took 135 steps in the AT pair and 91 steps in the GC pair. Obtained bond distances are shown in Table II. Comparison with benchmark results from [6] shows that our distances are approximately 10% shorter, except for the N2-O2 bond in GC base pair where bond distance is 3.97% longer. In Fig. 2, there is a profile view of base pairs. From this picture, it is clear that the bases in each pair are in different planes. The angle between the planes, the so-called propeller twist angle is 3.2° in the AT base pair, and 6.67° in the GC base pair. Another important feature of the molecular structure is the pyramidalization of the amino group in cytosine and adenine due to the partial sp^3 hybridization of nitrogen.

4.3. Interaction and stabilization energies

Under base pair formation, each base suffers a geometrical change from the optimal isolated base geometry acquired. Let E_{def} be the deformation energy spent. On the other hand, the interaction energy ΔE_{int} is defined as the change in the
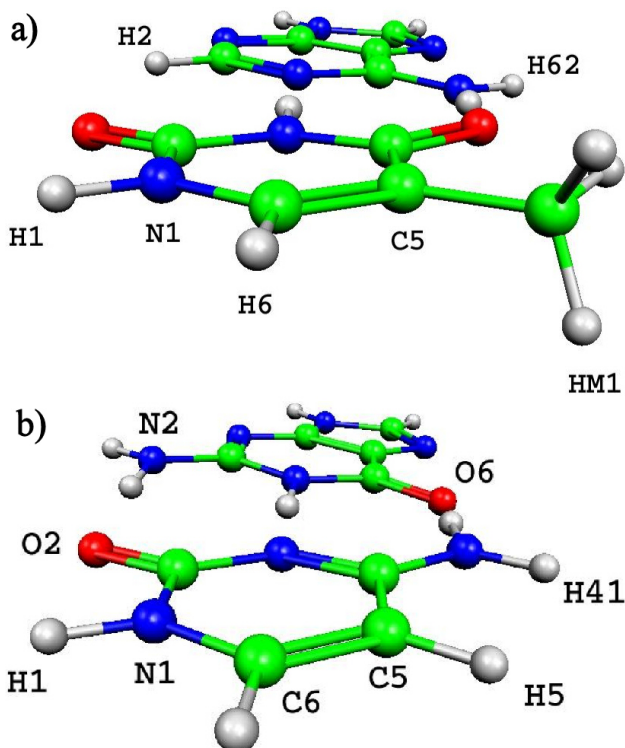

FIGURE 2. Propeller twist angle between the bases in AT and GC base pairs. Note the pyramidalization of the guanine amino group.

TABLE II. Bond distances between donor and acceptor atoms of the hydrogen bonds. Angles are measured between acceptor, hydrogen and donor atoms.

	A-T		G-C		
Bond	Distance (Å)	Angle	Bond	Distance (Å)	Angle
N6-O4	2.677	177.3	O6-N4	2.437	178.5
N1-N3	2.792	177.3	N1-N3	2.715	176.1
			N2-O2	3.005	174.0

base pair total energy when molecules are moved from infinite to bond distance, keeping the optimal base pair geometries unchanged. Therefore, base pair formation energy, stabilization energy, is $\Delta E_{stab} = \Delta E_{int} + \Delta E_{def}$. When the latter calculations are performed, basis set superposition error (BSSE) [21] must be taken into account. In order to estimate
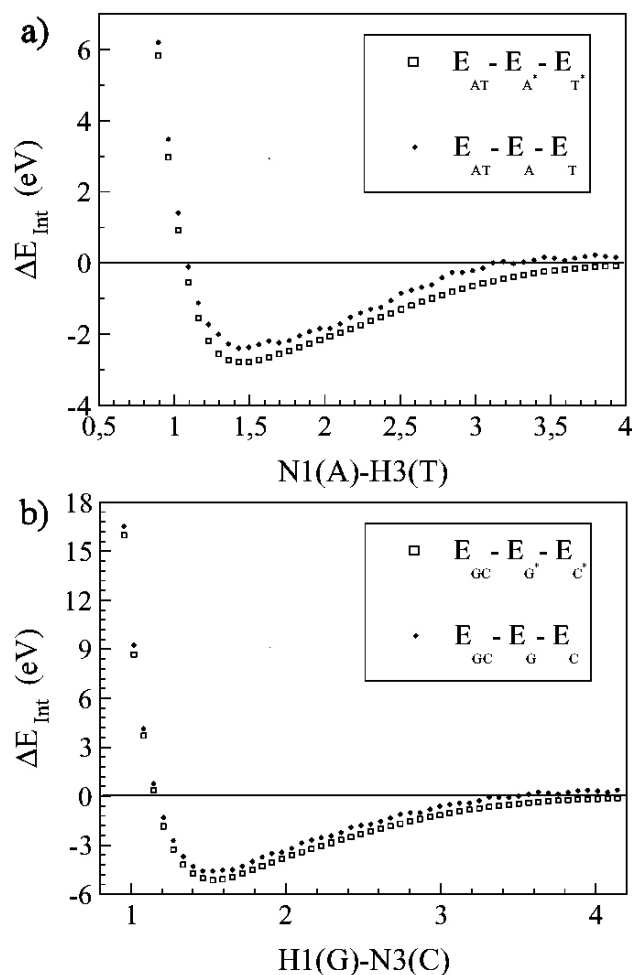
FIGURE 3. Base pair interaction energy as a function of the distance between hydrogen and acceptor atoms. Circles: BSSE not corrected energies. Squares: counterpoise function corrected interaction energies.

BSSE, we calculate the total energy of each isolated molecule adding the atomic orbitals of the other, following the counterpoise function correction scheme [21]. In Fig. 3 the interaction energy for AT and GC base pairs are displayed as a function of the distance between hydrogen and acceptor atoms. From the figure, it can be seen that the BSSE correction has approximately the same value for all distances; therefore, the correction for the stabilization energy can be taken as the average of the correction of the interaction energy (eV for A-T and eV for G-C base pair). Although our net stabilization energy values are different from benchmark results, the proportion between them ($E_{stab}^{GC}/E_{stab}^{AT}$) only differs by 2.37%.

5. Conclusions

The present results show us that SIESTA is a reliable method to study DNA base pairs; however, it is recomended to use higher efficiency pseudopotentials and a bigger basis set in order to reproduce both the bond distances and stabilization energies of base pairs from benchmark results. Intrinsically, non planar geometries for base pairs are found, GC being the base pair with a greater propeller twist angle. BSSE, as calculated by the counterpoise function, is an almost constant function of intermolecular distance for both base pairs.

Acknowledgments

1. I.K. Yanson, A.B. Teplitsky, and L.F. Sukhodub, *Biopolymers.* **18** (1979) 1149.
2. R. Taylor and O. Kennard, *J. Mol. Struct.* **78** (1982) 1.
3. K. Ozeki, N. Sakabe, and J. Tanaka, *Acta Cryst. B* **25** (1969) 1038.
4. E. Nir, K. Kleinermanns, and M.S. de Vries, *Nature* **408** (2000) 949.
5. E. Nir, Ch. Plützer, K. Kleinermanns, and M. de Vries, *Eur. Phys. J. D* **20** (2002) 317.
6. J. Šponer, P. Jurečka, and P. Hobza, *J. Am. Chem. Soc.* **126** (2004) 10142, and references therein.
7. C. Fonseca Guerra, F. Matthias Bickelhaupt, J.G. Snijders, and E.J. Baerends, *J. Am. Chem. Soc.* **122** (2000) 4117.
8. T. van der Wijst, C. Fonseca Guerra, M. Swart, and F. Matthias Bickelhaupt, *Chem. Phys. Lett.* **426** (2006) 415.
9. D. Varsano, R. Di Felice, M.A.L. Marques, and A. Rubio, *J. Phys. Chem. B* **110** (2006) 7129.
10. D. Sánchez-Portal, P. Ordejón, E. Artacho, and J. Soler, *Int. J. Quantum Chem.* **65** (1997) 453.
11. J.M. Soler *et al.*, *J. Phys.: Condens. Matter*, **14** (2002) 2745.
12. M. Machado, P. Ordejón, E. Artacho, D. Sánchez-Portal, and J.M. Soler, *http://xxx.arXiv.org/ps/cond-mat/9908022* (1999).
13. P. Ordejón, E. Artacho, and J.M. Soler, *Phys. Rev. B* **53** (1996) R10441.
14. D.R. Hamann, M. Schlüter, and C. Chiang, *Phys. Rev. Lett.* **43** (1979) 1494.
15. L. Kleinman and D.M. Bylander, *Phys. Rev. Lett.* **48** (1982) 1425.
16. J. P. Perdew, K. Burke, and M. Ernzerhof, *Phys. Rev. Lett.* **77** (1996) 3865.
17. C. Espejo and R.R. Rey-González, *Rev. Col. Física.* **38** (2006) 407.
18. E. Anglada, J.M. Soler, J. Junquera, and E. Artacho, *Phys. Rev. B* **66** (2002) 205101.

19. H.M. Berman *et al.*, *Biophys. J.* **63** (1992) 751.

20. J. Sponer and P. Hobza, *J. Phys. Chem.* **98** (1994) 3161.

21. S. F. Boys and F. Bernardi, *Mol. Phys.* **19** (1970) 553.