Improved description of the structural and optoelectronic properties of DNA/RNA nucleobase anhydrous crystals: Experiment and dispersion-corrected density functional theory calculations

M. B. da Silva,[1] T. S. Francisco,[2] F. F. Maia, Jr.,[3] E. W. S. Caetano,[4] U. L. Fulco,[5] E. L. Albuquerque,[5] and V. N. Freire[1]

[1]*Departamento de Física, Universidade Federal do Ceará, Caixa Postal 6030, 60455-760 Fortaleza, CE, Brazil*

[2]*Universidade Estadual Vale do Acaraú, Av. da Universidade, 850–Campus da Betânia, 62040-370, Sobral, CE, Brazil*

[3]*Universidade Federal Rural do Semi-Árido, Av. Francisco Mota, 572, 59625-900, Mossoró, RN, Brazil*

[4]*Instituto Federal de Educação, Ciência e Tecnologia do Ceará, 60040-531 Fortaleza, CE, Brazil*

[5]*Departamento de Biofísica e Farmacologia, Universidade Federal do Rio Grande do Norte, 59072-970 Natal, RN, Brazil*

(Received 8 February 2017; published 17 August 2017)

The development of low cost and environmentally friendly organic electronic/optoelectronic devices has attracted a lot of interest. The integration of DNA and RNA nucleobases to improve the performance of organic light-emitting diodes has been proposed recently [Gomez *et al.*, Sci. Rep. **4**, 7105 (2014)], notwithstanding limited experimental and theoretical information on the optoelectronic properties of DNA/RNA thin films. As a contribution to an improved understanding of DNA/RNA-based devices in the solid state, we have performed in this paper dispersion corrected density functional theory (DFT) and time-dependent DFT (TDDFT) calculations to obtain the optimized geometries, Kohn-Sham band structures and orbitals, charge distribution, optical absorption, Frenkel exciton binding energies, and complex dielectric functions of the five DNA/RNA nucleobase anhydrous crystals, namely cytosine, guanine, adenine, thymine, and uracil. Optical absorption measurements on DNA/RNA nucleobase powders were also performed for comparison with the simulations. An improvement on the local density approximation (LDA) description of the lattice parameter estimates was achieved considering the generalized gradient approach (GGA) with a semiempirical dispersion correction scheme in comparison with structural x-ray data found in the literature. Energy gap correction using the Δ-sol methodology provided a good agreement between theory and experimental estimates from our optical absorption data, greatly surpassing the quality of previous simulations. Effective masses for the carriers were also found, indicating that the guanine crystal as well as the cytosine one (although with some drawbacks) has potential applications in optoelectronics as a direct gap semiconductor, with the other nucleobases presenting either a semiconductor or an insulator character depending on the carrier type. The complex dielectric function exhibits a high degree of anisotropy for different states of light polarization relative to the molecular stacking planes, while the Frenkel exciton binding energy estimation for the adenine crystal is very close to the optical absorption experimental data.

DOI: 10.1103/PhysRevB.96.085206

I. INTRODUCTION

Only about ten years after the publication of the landmark work of Watson and Crick [1] on the helical DNA structure, Eley and Spivey [2] have proposed a pathway for rapid, one-dimensional charge separation in double-stranded DNA through π-π interactions. The dream of using DNA as a nanoscale semiconductor was born [3–7], but after all these years, contradictory results were obtained: DNA can act as a high conduction wire [8], as a proximity induced superconductor [9], as a semiconductor [10], or even behave like an insulator [11,12]. DNA nucleobases, on the other hand, are demonstrating to be versatile materials in natural electronics (or bioelectronics) and photonics [6,13,14], a very promising field of research in which biological materials are employed to build environmentally friendly microelectronic and optoelectronic devices. In comparison with DNA, the electron affinities of the nucleobases cover a wider energy interval, giving additional room for the design of new applications.

Nevertheless, there are much fewer reports describing the use of nucleobases in bioelectronics than for DNA. A field effect transistor using a deoxyguanosine derivative was demonstrated by Maruccio *et al.* in 2003 [15], while adenine and guanine solid-state layers (SSLs) were used to produce an inorganic-organic gate dielectric in a biodegradable field effect transistor [16]. An organic field effect transistor (OFET) incorporating guanine was shown to decrease contact resistance, leading to an increase in the saturation current, improving charge carrier mobility [17]. In particular, SSLs of the DNA bases thymine and adenine in bioorganic light-emitting diodes were reported [18]. On the other hand, guanine SSLs embedded in oxide dielectrics were shown to be able to capture hydrogen and work as a charge-trapping layer for inorganic and inorganic-field-effect transistors [19]. Recently, performance improvements of organic light-emitting diodes (OLEDs) using a cellulose/epoxy substrate and employing adenine SSLs as a hole injection layer were achieved [20].

It is worth remarking that highest occupied-lowest unoccupied molecular orbital (HOMO-LUMO) energy gaps were considered in the theoretical description of all thin films of DNA nucleobases in the aforesaid devices, instead of taking into account the energy gaps of reduced dimension calculations using the corresponding nucleobase crystals as a starting point. As a matter of fact, molecular crystals such as those of the DNA/RNA nucleobases and amino acids have been extensively studied in the last few years [21]. In particular, our research group has consistently investigated the electronic, optical, and vibrational properties of several amino acid crystals using density functional theory (DFT) methods,

such as glycine, alanine, and asparagine [22–24], cysteine [25], aspartic acid [26–28], and serine [29]. In previous papers, we also proposed that anhydrous crystals of DNA bases were wide gap semiconductors according to optical absorption measurements and DFT calculations performed within the local density approximation (LDA) [30,31]. Band gaps and electron effective masses calculated from the curvature of Kohn-Sham bands at valence and conduction band suggest that these crystals can be semiconducting for electron transport along a direction perpendicular to the molecular stacking planes [30,31].

Without a good description of London dispersion forces, originated from electron correlation in π stacking and hydrogen bonds in DNA nucleobases crystals, however, there was a lack of accuracy in the structural features predicted by those LDA simulations. As a matter of fact, distinct DFT dispersion-corrected functionals have been employed in recent papers to improve the structural and vibrational properties of systems where van der Waals forces are important, such as aromatic molecules interacting with graphene [32], nucleobases and their dimers [33], nucleobases with wrinkled graphene surfaces [34], the spectroscopic signatures of buckyonions [35], aromatic and noncovalent interactions in general [36,37]. A dispersion-corrected density functional tight binding approach was also used to model accurately organic molecular crystals, opening a new avenue to be explored in order to achieve organic crystal structure prediction [38].

In this paper, we consider the crystals of the four DNA nucleobases, cytosine, guanine, adenine, and thymine, and the RNA nucleobase uracil to employ a dispersion-corrected generalized gradient approach (GGA) exchange-correlation functional. Optimized geometries, Kohn-Sham band structures and orbitals, atomic charges, and optical properties like the optical absorption and the complex dielectric function were obtained. Optical absorption measurements were also performed on the nucleobase powders to compare with the theoretical calculations and to estimate the main band gap of nucleobases in solid state. The Δ-sol scheme was also employed to correct the Kohn-Sham band gaps, sensibly improving the theoretical calculations to compare with the experimental values.

We show that lattice parameters employing the GGA dispersion corrected functional are much closer to the x-ray experimental data available in the literature in comparison with the LDA case and surpass the HOMO-LUMO energy gap estimates of recently published papers [18–20]. As a matter of fact, we have obtained energy gaps for the cytosine, guanine, adenine, and thymine anhydrous crystals, respectively, which agree with the optical absorption we have measured and are much better than the previous ones calculated using DFT-LDA. The effective masses obtained for the carriers point to the guanine anhydrous crystal as a promising candidate for the development of optoelectronic nanobiodevices in the ultraviolet/visible range, while the complex dielectric function of the nucleobase crystals is strongly anisotropic for distinct polarization states of incident light. An estimate of the binding energy of the Frenkel exciton in the solid state nucleobases was also accomplished using time-dependent DFT (TDDFT) calculations, predicting an exciton binding energy for the adenine anhydrous crystal very close to the experimental value.

II. MATERIALS AND METHODS

A. Experimental procedure

Powders of all anhydrous nucleobases, namely cytosine 99% (C3506), guanine 99% (G11950), adenine 99% (A8626), thymine 99% (T0376), and uracil 99% (U0750) were purchased from Sigma-Aldrich without further purification, and x-ray measurements were performed to check their crystalline structure (results not included in this paper). They were mixed with KBr to form pellets for each nucleobase. The ultraviolet optical absorption spectra of those pellets were measured by transmittance employing a Varian Cary 5000 UV-visible NIR spectrophotometer with solid sample holders. The wavelength range of the measurements varied from 200 to 800 nm (50 000 down to 12 500 cm^{-1}). Background removal was accomplished comparing the absorption spectrum of a pure KBr pellet, with baseline corrections being made if necessary.

B. Computational details

The initial anhydrous crystal structures of the nucleobase crystals used here were extracted from previously published experimental measurements. For the anhydrous cytosine crystal, x-ray diffraction [39] reveals an orthorhombic unit cell with the number of nucleobases $Z = 4$ and space group $P2_12_12_1$ where the amino nitrogen atom participates in two relatively long hydrogen bonds to carbonyl oxygen atoms from neighbor molecules. It has two symmetrically intercalated stacking planes: (2 0 1) and (−2 0 1) (see Fig. 1, C). The anhydrous guanine crystal (see Fig. 1, G), on the other hand,

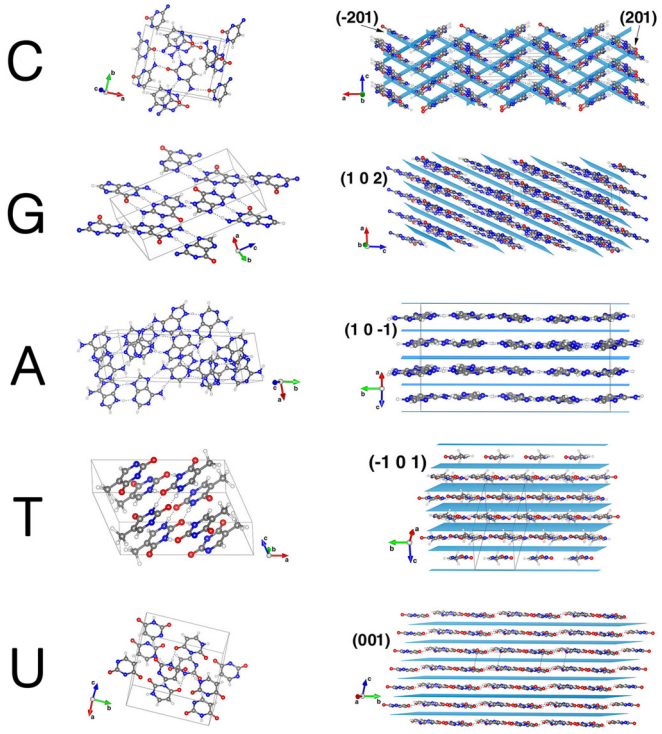

FIG. 1. Nucleobases cytosine (C), guanine (G), adenine (A), thymine (T), and uracil (U): anhydrous crystal unit cells (left) and stacking planes (right). Carbon, hydrogen, nitrogen, and oxygen are depicted in gray, white, blue, and red, respectively.

was investigated by Guille and Clegg [40] using synchrotron radiation. It has a monoclinic unit cell with $Z = 4$, space group $P2_1/c$ and is held by N-H ... N and N-H ... O hydrogen bonds, leading to a stacking of molecular sheets along (1 0 2) planes interacting through π-π effects. The crystal structure of anhydrous adenine was recently determined by Mahapatra *et al.* [41] using single crystal x-ray diffraction, giving rise to a monoclinic unit cell with $Z = 8$, space group $P2_1/c$ with the molecules connected through N-H ... N hydrogen bonds and stacking along (1 0 −1) planes (see Fig. 1, A). Thymine anhydrous crystals are monoclinic with space group $P2_1/c$ [42]. Its unit cell has four molecules and exhibits twofold screw axes with molecular units being connected through two N-H ... O=C hydrogen bonds creating infinite chains along the b direction (see Fig. 1, T) and (−1 0 1) stacking planes. Lastly, uracil anhydrous crystals are monoclinic with $Z = 4$ and space group $P2_1/a$, as determined by x-ray diffraction measurements [43]. This crystal is formed by the (0 0 1) plane stacking of molecular layers interacting through NH ... O hydrogen bonds with its pyrimidine ring being similar in size to that of adenine hydrochloride (see Fig. 1, U). Other crystal polymorphs of the nucleobase crystals in anhydrous or hydrated form have been discovered or suggested recently [44–49], but they are not considered in this paper.

In order to optimize the geometry of the unit cells, the plane-wave DFT code CASTEP was used [50,51]. Two exchange-correlation functionals were adopted in the simulation: the generalized gradient approximation of Perdew-Burke-Ernzerhof [52] (GGA-PBE) and the LDA [53]. The semiempirical dispersion correction energy term of Tkatchenko and Scheffler [54] (TS) was added to the first functional in order to estimate van der Waals noncovalent interactions. This method has presented better structural results for molecular crystals than the Grimme DFT-D dispersion methodology also available in the CASTEP code [55,56]. As a matter of fact, dispersion-correction schemes can be classified into three main groups [57], depending on the type of mathematical treatment, namely: (a) semiclassical description of the dispersion interaction, either adding the dispersion energy between atom pairs to the electronic energy, such as in the method of Tkatchenko and Scheffler [54], or the D3 proposal of Grimme *et al.* [58]; (b) nonlocal density dependent dispersion corrections, which use the electron density to evaluate the dispersion energy and employ an exchange-correlation functional kernel dependent on two nonseparable electron coordinates (van der Waals density functionals; vdW-DF) [59,60]; (c) effective one-electron potentials describing London dispersion by local properties using atom-centered external potentials [61,62], and semilocal density functionals, such as the flexible functional form developed by Mardirossian and Head-Gordon [63], which can exhibit some limitations [64]. The first two groups (a) and (b) are, in particular, very accurate for various noncovalent systems. When used in combination with the methods of the group (c), they lead to very high accuracy by taking into account long- and short-range electron correlations [57].

The core electrons were represented by ultrasoft [65] (LDA functional) and norm-conserving [66] (GGA+TS approach) pseudopotentials, while valence electrons with orbital configurations H $1s^1$, C $2s^2\ 2p^2$, N $2s^2\ 2p^3$, and O $2s^2\ 2p^4$ were taken into account explicitly during the calculations. The Kohn-Sham orbitals were expanded in two plane-wave basis sets with energy cutoffs of 500 and 830 eV for the LDA and GGA computations, respectively. These values were selected after performing a convergence study of the per-atom forces in a previous set of geometry optimizations. Self-consistency was assumed to be achieved when the total energy (electron eigenenergy) variation through three successive iterations was smaller than 10^{-6} eV/atom (0.5×10^{-6} eV). Geometry optimization thresholds were set to: (a) maximum force per atom smaller than 0.03 eV/Å; (b) total energy variation smaller than 10^{-5} eV/atom; (c) maximum atomic displacement smaller than 0.001 Å; (d) maximum stress component smaller than 0.05 GPa. The Broyden-Fletcher-Goldfarb-Shanno (BFGS) minimizer [67] was used to relax both the lattice parameters and internal atomic coordinates. Mulliken [68] and Hirshfeld [69] population analysis calculations were also performed on the electron density to evaluate the electric charge partition to each atom in the nucleobase crystals.

As it is well known, DFT band gap values tend to underestimate the real band gap of solid state samples [70]. More sophisticated approximations to improve the gaps, such as the the one-particle Green function and the screened Coulomb interaction (*GW*) method [71] and hybrid functionals [72], are computationally too expensive for molecular crystals such as the ones investigated in this paper. So in order to improve our gap predictions, we have also the less demanding Δ-sol scheme of Chan and Ceder [73], which generalizes the Δ self-consistent field (ΔSCF) theory for molecules to periodic structures taking into account the dielectric screening properties of electrons, decreasing the mean absolute error of Kohn-Sham gaps by 70% in average. The Δ-sol gaps were used to adjust the calculated optical absorption spectra for comparison with our corresponding experimental data.

The DMOL3 code [74] was employed to find optimal geometries for the individual nucleobases solvated in water to estimate the binding energy of the Frankel exciton in the respective nucleobase solid state systems. Geometry optimizations were carried out using the GGA+TS approach and a double numerical plus polarization (DNP) basis set obeying the following convergence thresholds: maximum force per atom smaller than 0.002 Ha Å^{-1}, total energy variation smaller than 10^{-5} Ha, and maximum displacement per atom less than 0.005 Å. Water solvation was eventually considered through the COSMO model [75]. The first optically active excited state of each nucleobase molecule solvated in water was optimized using TDDFT [76–78] within the GGA+TS approach. The binding energy of the Frenkel exciton E_b^{exc} can be approximated by considering a single nucleobase molecule solvated in water and taking the difference between its fundamental gap E_g^{mol} [79]

$$E_g^{\text{mol}} = E(N+1) + E(N-1) - 2E(N), \tag{1}$$

and the optical gap E_g^{opt} [80], i.e.

$$E_b^{\text{exc}} = E_g^{\text{mol}} - E_g^{\text{opt}}. \tag{2}$$

Here, $E(X)$ is the total energy of the relaxed molecule with X electrons, with N being the number of electrons in the neutral molecule. Also, E_g^{opt} can be obtained by evaluating the

TABLE I. DFT calculated at the LDA and GGA+TS levels, with cutoff energies of 500 and 830 eV, respectively, a, b, c lattice parameters (in Å), unit cell volume V (in Å^3), β angle (in degrees), and distance d between adjacent stacked molecular planes for the bases uracil (U), cytosine (C), thymine (T), adenine (A), and guanine (G). The experimental values are those of Refs. [82] for uracil, [83] for cytosine, [42] for thymine, [41] for adenine, and [40] for guanine. The deviations Δ from the experimental values are also indicated in specific columns.

LDA at 500 eV												
BASE	a	Δa	b	Δb	c	Δc	V	ΔV	β	$\Delta\beta$	d	Δd
C	12.83	− 0.22	9.19	− 0.30	3.27	− 0.54	385.71	− 86.71	—	—	2.91	− 0.38
G	3.26	− 0.29	9.38	− 0.31	15.94	− 0.41	479.82	− 80.26	100.26	4.51	2.80	− 0.33
A	7.52	− 0.37	21.60	− 0.64	6.47	− 0.98	987.12	− 214.45	109.89	3.31	2.83	− 0.37
T	11.70	− 1.17	6.62	− 0.21	5.61	− 1.09	418.99	− 149.89	105.61	0.61	2.74	− 0.49
U	11.50	− 0.44	12.05	− 0.32	3.34	− 0.31	383.28	− 80.11	124.16	3.27	2.76	− 0.38
GGA+TS at 830 eV												
BASE	a	Δa	b	Δb	C	Δc	V	ΔV	β	$\Delta\beta$	d	Δd
C	13.01	− 0.03	9.52	0.02	3.77	− 0.05	466.31	− 6.11	—	—	3.26	− 0.03
G	3.59	0.04	9.71	0.02	16.45	0.11	571.77	11.69	95.42	0.33	3.17	0.04
A	7.90	0.01	22.09	− 0.16	7.31	− 0.14	1171.21	− 30.36	113.21	0.02	3.16	− 0.04
T	12.37	− 0.50	6.89	0.06	6.95	0.25	547.79	− 21.08	112.40	7.40	3.41	0.18
U	11.93	− 0.01	12.30	− 0.08	3.64	− 0.02	455.43	− 7.96	121.58	0.68	3.10	− 0.04
EXP												
BASE	a	Δa	b	Δb	c	Δc	V	ΔV	β	$\Delta\beta$	d	Δd
C [83]	13.04	—	9.50	—	3.81	—	472.42	—	—	—	3.29	—
G [40]	3.55	—	9.69	—	16.35	—	560.08	—	95.75	—	3.13	—
A [41]	7.89	—	22.24	—	7.45	—	1201.57	—	113.19	—	3.20	—
T [42]	12.87	—	6.83	—	6.70	—	568.88	—	105.00	—	3.23	—
U [82]	11.94	—	12.38	—	3.66	—	463.39	—	120.90	—	3.14	—

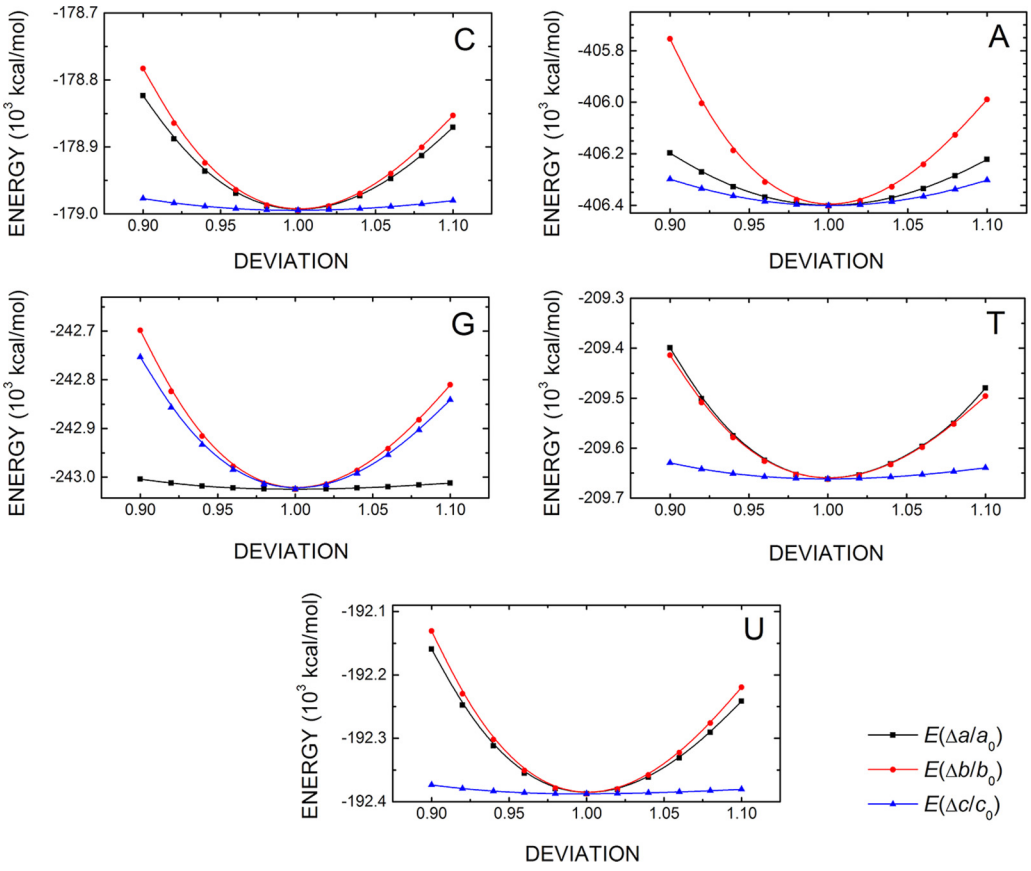

FIG. 2. Nucleobases cytosine (C), guanine (G), adenine (A), thymine (T), and uracil (U) anhydrous crystals: variation of the total unit cell energy with changes in the lattice parameters. The black (red, blue) squares (circles, triangles) depict the calculated variation of the energy as a function of the relative lattice parameter variation $\Delta a/a_0$ ($\Delta b/b_0, \Delta c/c_0$). Solid curves interpolate the calculated data as a visual aid.

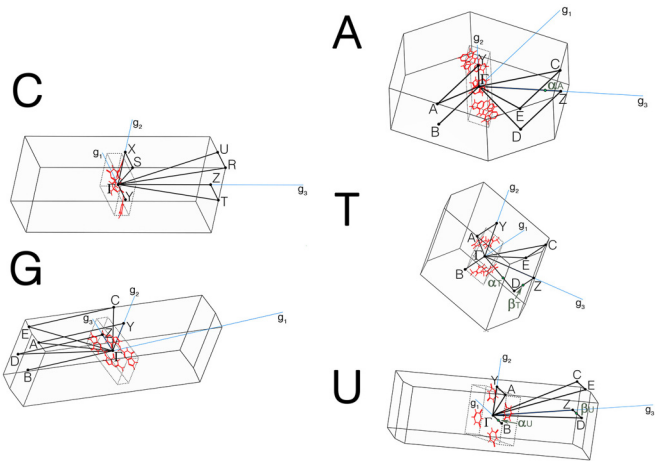

FIG. 3. Cytosine (C), guanine (G), adenine (A), thymine (T), and uracil (U): first Brillouin zones, special points in reciprocal space, and paths adopted in the band structure calculations.

difference between the first optically active excited state total energy $E^*(N)$ achieved after structural relaxation [81] and the molecular ground state total energy $E(N)$, i.e.

$$E_g^{\text{opt}} = E^*(N) - E(N). \tag{3}$$

III. STRUCTURAL PROPERTIES

Table I shows the results obtained after the geometry optimization procedure for the lattice parameters of each nucleobase anhydrous crystal in contrast with the experimental data from x-ray measurements obtained elsewhere in the literature [40–42,82,83]. As one can see, the LDA optimized unit cells are much smaller than those of the x-ray measurements, which is mainly due to the tendency of LDA functionals to overestimate interatomic forces. For uracil, the largest error occurs in the lattice parameter c, being about -8.4%, while for cytosine it reaches -14%. For thymine and adenine, the numbers are, respectively, -16 and -13%. For guanine, the

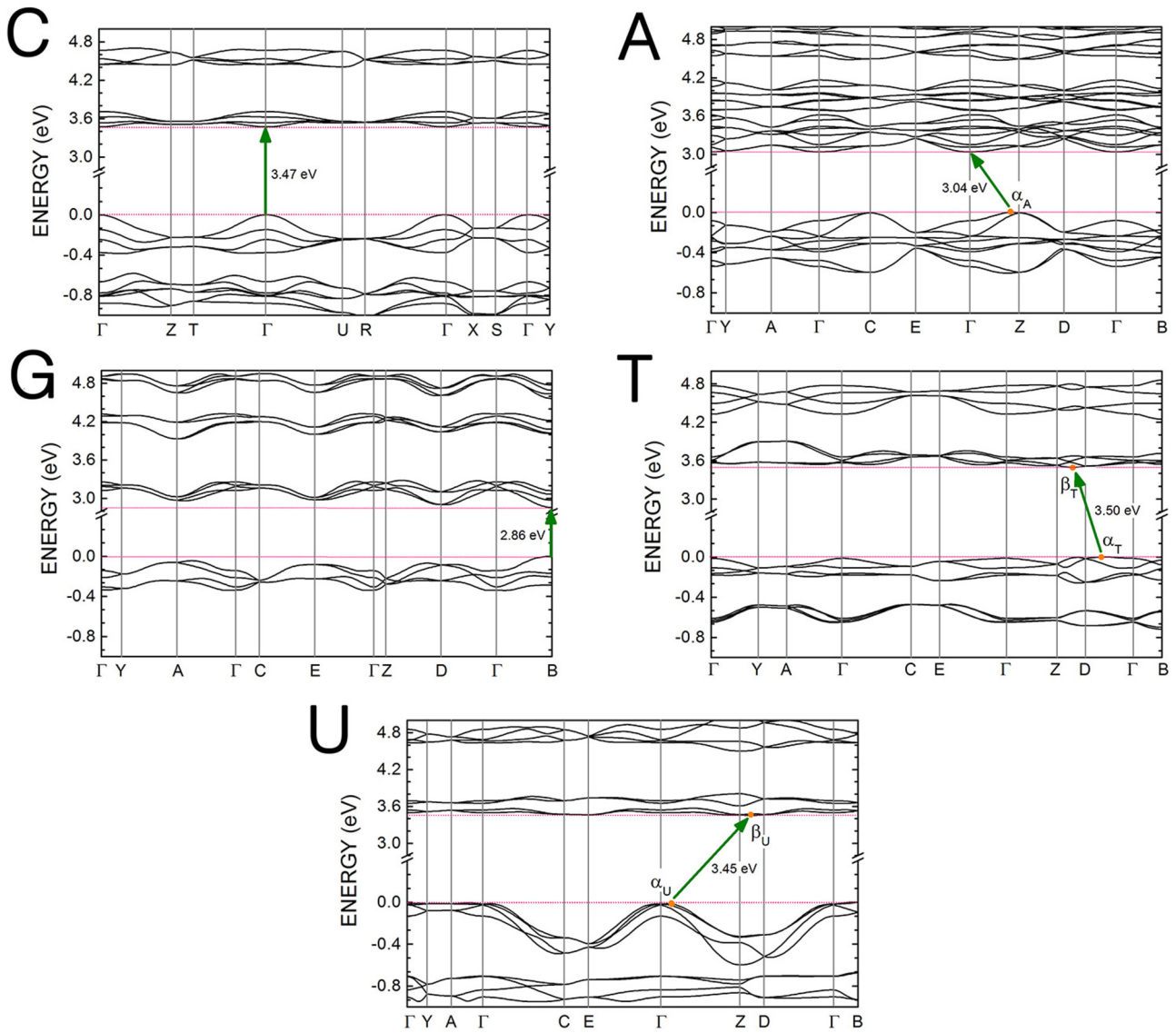

FIG. 4. Kohn-Sham band structures near the main band gap of the nucleobase anhydrous crystals: cytosine (C), guanine (G), adenine (A), thymine (T), and uracil (U). The arrows indicate the most important transitions and corresponding energy gap values.

largest error occurs for the lattice parameter a, calculated to be approximately 8% smaller than the experimental value. It is worthwhile to note that these lattice parameters correspond in general to the stacking of molecular planes of each crystal, shown at the column with distances d. The d values for cytosine, guanine, adenine, thymine, and uracil within the LDA approach are 11.5, 10.5, 11.6, 15.2, and 12.1% smaller than the measurement data. Between these planes, van der Waals interactions are significant, but not accounted for within the LDA of the exchange-correlation functional. The monoclinic angle β, on the other hand, tends to be overestimated by the LDA approach, being larger than the measurements by 0.61° for thymine, 3.27° for uracil, 3.31° for adenine, and 4.51° for guanine.

Looking at the GGA+TS results at the bottom of Table I, one can see a great improvement in comparison both to the experimental unit cell parameters and to the LDA estimated ones. In the case of cytosine, the relative difference is now only −1.3% for the worst-case scenario—the lattice parameter c—while for guanine, the corresponding figure for the lattice parameter a is +1.1%. The adenine and thymine crystals exhibit deviations of only −1.9 and +3.7% for the c lattice parameter, while for uracil, the corresponding deviation is −0.55%. Overall, the GGA+TS approach tends to slightly underestimate the unit cell sizes except for guanine. The molecular stacking planes distance d are also very close to the measured data, their relative differences to experiment being just −0.91, 1.3, −1.3, 5.6, and −1.3% for cytosine, guanine, adenine, thymine (the worst figure), and uracil, respectively. For the β angle, the GGA+TS estimates also improve on the LDA calculations except for thymine, for which the error is larger, 7.40° GGA+TS versus 0.61° LDA. However, the GGA+TS thymine c lattice parameter is much better than the LDA prediction, being about 3.7% larger than the x-ray diffraction data. Other structural parameters, such as atomic positions, bond lengths, and bond angles, are depicted in detail in the Supplemental Material (Tables S1–S10) [84].

Figure 2 depicts the variation of the unit cell total energy acquired when we modify independently each lattice parameter for the nucleobase crystals from their equilibrium values. For cytosine (top, left), the curves obtained by varying the lattice parameters a and b are very similar, while the curve for c is practically flat. This indicates that it is much easier to compress this material along the c axis in comparison with a and b, an effect related to the existence of a network of hydrogen bonds, six per molecular unit, approximately aligned to the ab plane of the crystal. Additionally, the molecular interactions along the c axis are mostly noncovalent and therefore weaker, leading to a smaller dispersion of the respective total energy curve. The same observation on the total energy behavior can be made for thymine (middle, right) and uracil (bottom), mostly because there is a clear noncovalent stacking of parallel molecular layers along the c axis and four hydrogen bonds per molecule almost aligned to the a and b directions in both phases. In the case of guanine (middle, left), the b and c curves are close to each other, and the curve for a is almost flat as a consequence of not only its eight hydrogen bonds per molecule disposed along the bc plane, but also the action of dispersive forces solely between its molecular layers, which are almost perpendicular to the a axis. Lastly, for adenine (top, right), the three curves for a, b, and c are more distinct than for the other nucleobases, the a and c curves being closer to each other. The curve for c displays the smallest dispersion about the minimum of the energy, apparently not so small as the smallest dispersion plots found for the other anhydrous nucleobase crystals, indicating stronger dispersive interactions parallel to the c direction for this system. The curvature difference between the a and b curves is probably due to the existence of more hydrogen bonds, six per molecule, more nearly aligned to the b axis than to the a axis, leading to a larger curvature for the total energy when the lattice parameter b is changed.

Under the energetic point of view, the binding energy per molecule is a measure of the strength of the interactions between the nucleobases in the solid phase. It can be evaluated as follows: let E_{cell} be the total energy of the unit cell, Z the number of molecular units contained in it, and E_{mol} the total energy of an isolated nucleobase molecule. The binding energy per molecule E_{bpm} is given by

$$E_{\text{bpm}} = \frac{E_{\text{cell}}}{Z} - E_{\text{mol}}. \tag{4}$$

The GGA+TS calculations predict the following values for E_{bpm} (in eV): −1.855 for cytosine, −2.510 for guanine, −1.837 for adenine, −1.615 for thymine, and −1.598 for uracil, with the following sequence of binding energy strengths: G > C > A > T > U. This result is in agreement with the data of our previous paper [30] using the LDA exchange-correlation functional for the DNA nucleobases only, but contrasts with the simulations of Sponer *et al.* [85,86] for guanine, adenine, cytosine, and uracil stacked dimers using the MP2 approach (G > A > C > U). However, as our calculations predict a very small difference between cytosine

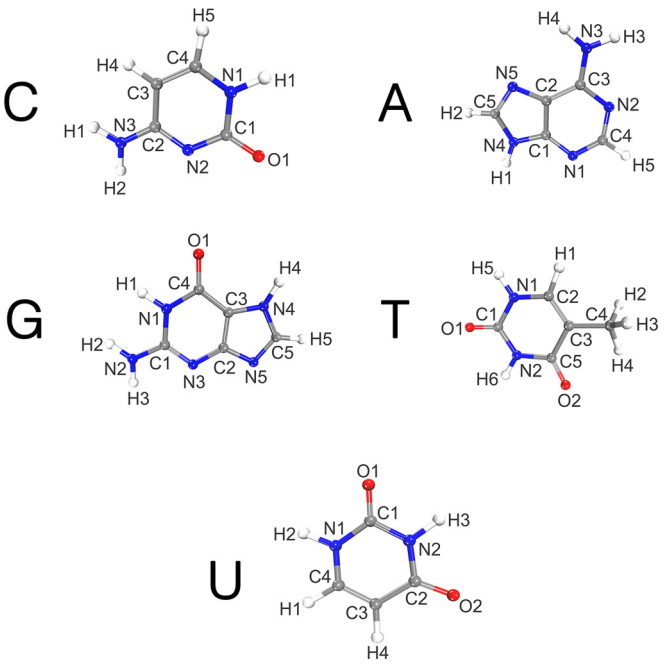

FIG. 5. Labels used to identify the atoms of cytosine (C), guanine (G), adenine (A), thymine (T), and uracil (U). Carbon, hydrogen, nitrogen, and oxygen are depicted in gray, white, blue, and red, respectively.

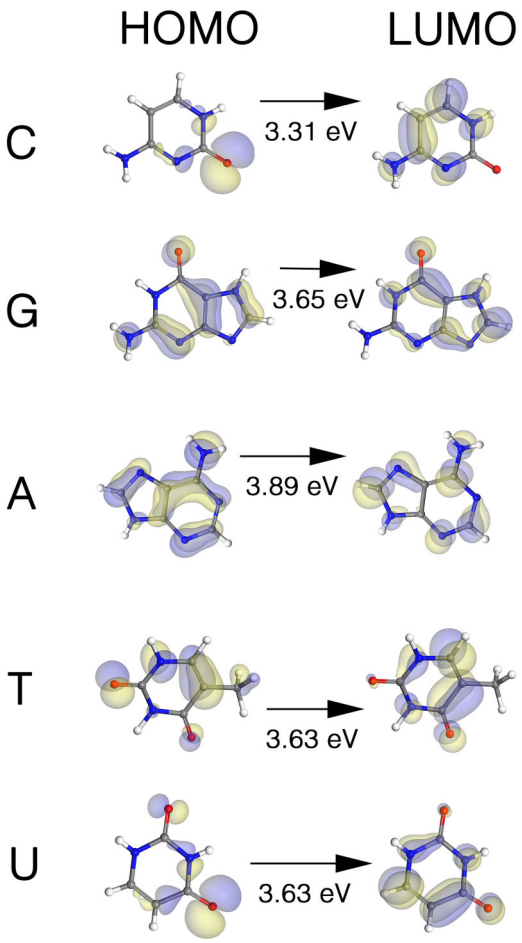

FIG. 6. HOMO and LUMO molecular orbitals of cytosine (C), guanine (G), adenine (A), thymine (T), and uracil (U). HOMO-LUMO gaps are indicated. Blue and yellow colors identify positive and negative phases of the wave function, respectively.

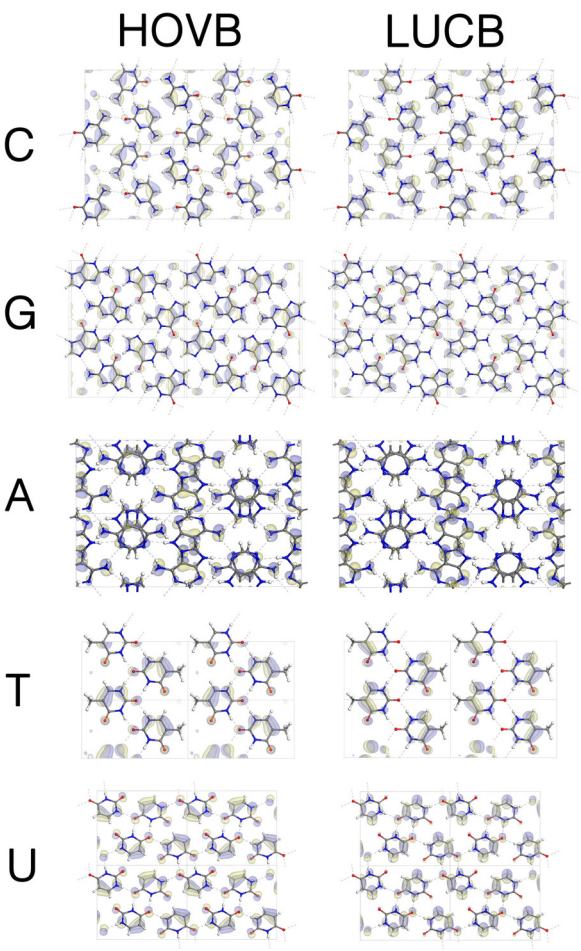

FIG. 7. HOVB and LUCB Kohn-Sham states at $\mathbf{k} = 0(\Gamma)$ of cytosine (C), guanine (G), adenine (A), thymine (T), and uracil (U) anhydrous crystals. Blue and yellow colors identify positive and negative phases of the wave function, respectively.

and adenine binding energies (18 meV or 0.41 kcal/mol), while the MP2 simulations predict a difference of the same order (0.50 kcal/mol), one can be certain that the binding interactions between cytosine and adenine units are very similar in strength. In addition, the thermal stability of the nucleobase thin films reveals the following sequence: G > C > A > T > U, in agreement with the order of binding energies we found in our LDA and GGA+TS simulations, notwithstanding the fact that they are obtained for a bulk and not a small size periodic structure [87]. As a matter of fact, even if it is preferable to carry out thin film DFT simulations for direct comparison with these systems, their thicknesses—typically of the order of 100 nm—could demand a very high computational cost.

IV. ELECTRONIC BAND STRUCTURE AND EFFECTIVE MASSES

The same Brillouin zone path was used to plot the Kohn-Sham band structures for monoclinic guanine, adenine, thymine, and uracil, following the high symmetry points $\Gamma(0,0,0) \rightarrow Y(0,1/2,0) \rightarrow A(-1/2,1/2,0) \rightarrow \Gamma \rightarrow C(0,1/2,1/2) \rightarrow E(-1/2,1/2,1/2) \rightarrow \Gamma \rightarrow Z(0,0,1/2) \rightarrow D(-1/2,0,1/2) \rightarrow \Gamma \rightarrow B(-1/2,0,0)$ (see Fig. 3). For orthorhombic cytosine, the reciprocal space path sequence of points was $\Gamma(0,0,0) \rightarrow Z(0,0,1/2) \rightarrow T(-1/2,0,1/2) \rightarrow \Gamma \rightarrow U(0,1/2,1/2) \rightarrow R(-1/2,1/2,1/2) \rightarrow \Gamma \rightarrow X(0,1/2,0) \rightarrow S(-1/2,1/2,0) \rightarrow \Gamma \rightarrow Y(-1/2,0,0)$. Figure 4 depicts the Kohn-Sham band structures near the main band gap for the nucleobase crystals. Out of the five nucleobase crystals, two have a direct gap, namely cytosine ($\Gamma \rightarrow \Gamma$, 3.47 eV) and guanine ($B \rightarrow B$, 2.86 eV), while adenine, thymine, and uracil have indirect gaps, the first being 3.04 eV ($\alpha_A \rightarrow \Gamma$), the second 3.50 eV ($\alpha_T \rightarrow \beta_T$), and the last 3.45 eV ($\alpha_U \rightarrow \beta_U$), respectively. The α and β labels are used to indicate the valence band maximum and conduction band minimum, respectively, of the subscripted nucleobase when these points do not coincide with a high symmetry point in the first Brillouin zone. The origin of the different gap types cannot be assigned to the crystal structure only, since out of the four monoclinic nucleobase crystals with same space group symmetry (adenine, guanine, thymine, and uracil), three have an indirect gap and one (guanine) has a direct main electronic transition, while the orthorhombic crystal (cytosine) has a direct main band gap. It seems also that the presence or absence of oxygen atoms in the nucleobase

TABLE II. Hirshfeld charges of cytosine (C), guanine (G), adenine (A), thymine (T), and uracil (U) anhydrous crystals after the GGA+TS calculations. The free electron charge is −1.

Cytosine		Guanine		Adenine		Thymine		Uracil	
Atom	Charge	Atom	Charge	Atom	Charge	Atom	Charge	Atom	Charge
C1	0.18	C1	0.17	C1	0.08	C1	0.20	C1	0.20
C2	0.12	C2	0.07	C2	−0.01	C2	0.03	C2	0.15
C3	−0.09	C3	−0.02	C3	0.06	C3	−0.02	C3	−0.09
C4	0.04	C4	0.14	C4	0.11	C4	−0.12	C4	0.04
O1	−0.25	C5	0.07	C5	0.06	C5	0.15	O1	−0.23
N1	−0.08	N1	−0.08	C′1	−0.01	O1	−0.21	O2	−0.20
N2	−0.16	N2	−0.17	C′2	0.08	O2	−0.23	N1	−0.07
N3	−0.15	N3	−0.16	C′3	0.11	N1	−0.07	N2	−0.09
H1	0.09	N4	−0.06	C′4	0.06	N2	−0.09	H1	0.05
H2	0.09	N5	−0.16	C′5	0.06	H1	0.06	H2	0.10
H3	0.10	O1	−0.21	N1	−0.14	H2	0.04	H3	0.09
H4	0.04	H1	0.09	N2	−0.07	H3	0.04	H4	0.04
H5	0.06	H2	0.09	N3	−0.15	H4	0.04		
		H3	0.09	N4	−0.16	H5	0.10		
		H4	0.10	N5	−0.16	H6	0.10		
		H5	0.05	N′1	−0.14				
				N′2	−0.07				
				N′3	−0.15				
				N′4	−0.16				
				N′5	−0.16				
				H1	0.10				
				H2	0.09				
				H3	0.09				
				H4	0.04				
				H5	0.06				
				H′1	0.10				
				H′2	0.09				
				H′3	0.09				
				H′4	0.06				
				H′5	0.04				

does not play a role in determining the nature of the gap: four nucleobases are oxygenated and two of their crystals have an indirect gap, while the other two have a direct gap. No correlation was found for the gap type due to the presence of pyrimidine and imidazole rings as well, which implies that the nature of the main band gap in these materials must result from a more complex interplay between molecular properties and intermolecular couplings.

If we compare the present results with those found in our previous paper on the DNA nucleobase crystals [30], the GGA+TS exchange-correlation functional predicts larger band gaps for all systems while preserving their types (direct or indirect) and the approximate location of the band maxima and minima, except for thymine (from B to a point along the ΓD line). After applying the Δ-sol correction to the GGA+TS gaps, we have obtained the following values: 3.79 eV for cytosine, 3.34 eV for guanine, 3.76 eV for adenine, 3.88 eV for thymine, and 3.85 eV for uracil. The LDA-estimated band gap energy values of our previous paper [30] were 3.30 eV for cytosine, 2.68 eV for guanine, 2.83 eV for adenine, and 3.22 eV for thymine, which are −12 to −22% smaller than the Δ-sol correction of the GGA+TS estimated energy gaps of this paper. The corresponding estimates from our optical absorption measurements are 4.05 eV (cytosine), 3.60 eV (guanine), 4.03 eV (adenine), 4.13 eV (thymine), and 4.15 eV (uracil), so the average band gap error with (without) the Δ-sol method is of about −6.7% (−18%).

In order to understand better the electronic states at the valence and conduction band extrema, we have calculated the Kohn-Sham molecular orbitals for the isolated molecules using an equivalent supercell with spacing larger than 10 Å between molecular images using the GGA+TS exchange-correlation functional. Atom labels used in our analysis follow the description shown in Fig. 5. The HOMO and LUMO orbitals thus obtained are depicted in Fig. 6, together with the respective HOMO-LUMO gaps for each nucleobase. For cytosine, the HOMO-LUMO gap is the smallest among the five nucleobases, namely 3.31 eV, with the HOMO orbital consisting mainly in the superposition of O1, N1, and N2 $2p$ states, while the LUMO involves N3 $2p$ and a C2-C3 π orbital. For guanine, we have 3.65 eV for the HOMO-LUMO gap, O1, N2 $2p$, C2-C3-C4, C1-N3, and C5-N4 π states contributing strongly to the HOMO and O1 $2p$, and C3-C4, C2-N5 π orbitals forming the LUMO. In the case of adenine, one can observe the largest HOMO-LUMO gap of all nucleobases, 3.89 eV, with the HOMO orbital consisting

mainly in the superposition of π orbitals along the hexagonal ring (C1-C2-C3-N2 and C4-N1), and a N3 $2p$ orbital. For the LUMO, the π orbitals at the hexagonal ring are replaced by $2p$ contributions perpendicular to the ring plane at all atoms, except C1 and C2. Thymine and uracil have the same HOMO-LUMO gap, 3.63 eV, and somewhat similar contributions to the frontier orbitals: O1, O2, N, and C $2p$ states contribute to the HOMO and C-C π orbitals contribute to the LUMO. Our results compare well, except for the relative magnitudes of guanine and adenine, with those presented by Gomez *et al.* [14], who predicted HOMO-LUMO gaps (in eV) of 3.6, 3.9, 3.8, 3.7, and 3.7 for cytosine, guanine, adenine, thymine, and uracil, following the sequence of decreasing values G > A > T ≈ U > C. The Kohn-Sham frontier states for the nucleobase crystals at the Γ point, shown in Fig. 7 [highest occupied valence band (HOVB), lowest unoccupied conduction band (LUCB) states at $\mathbf{k} = 0$], are very similar to the pure molecular orbitals, pointing to a strong noncovalent character in the interaction between neighbor molecular units, with small wave function overlap. One can remark here that the solid-state gaps tend to be smaller than the molecular gaps, with an exception made for the cytosine crystal with a direct gap about 0.15 eV larger than its isolated molecule.

The electric charge of each atom in a crystal unit cell can be estimated using charge partition techniques such as Mulliken population analysis (MPA [68]) and Hirshfeld population analysis (HPA [69]). Hirshfeld population analysis, in particular, is capable of predicting better Fukui function indices [88], which are used to estimate the reactivity trends of a molecule better than traditional Mulliken charge partitions. It is also able to minimize the loss of information due to the formation of chemical bonds [89], notwithstanding its tendency to underestimate charge values [90]. This last limitation, however, can be removed by using an iterative approach successfully implemented for the solid state [91]. The results of both methods are detailed in the Supplemental Material (Tables S11–S15) [84] for all nucleobase crystals using the GGA+TS and LDA exchange-correlation functionals. One must note that these partial charge methods are orbital based, not density based, which means that they work better as trend indications of charge distribution than as quantitatively accurate estimates. Here, we discuss only the GGA+TS results for the Hirshfeld charges of each nucleobase molecule in the crystal shown in Table II as well as in Fig. 8.

In all nucleobase crystals, except for adenine, there are concentrations of negative charge around the oxygen atoms, with cytosine exhibiting the largest absolute value, −0.25 in electron charge units, and uracil the smallest (−0.20). The nitrogen atoms also tend to be negatively charged, with the N2 atom of the guanine crystal reaching −0.17. Carbon and hydrogen atoms, on the other hand, have positive charge. For example, C1 has a charge of 0.18 in cytosine and 0.17 in guanine, while C4 in adenine has a charge of 0.11, and C5 (C2) in thymine (uracil) has a charge of 0.15. Hydrogen atoms have charges of about 0.10 in all nucleobase crystals. The charge distribution shown in Fig. 8 suggests that cytosine, thymine, and uracil have a higher dipole moment than guanine and adenine. As guanine and thymine are the only crystals with direct gap, the former less polarized than the latter, it seems that one cannot establish correlations between the degree of charge

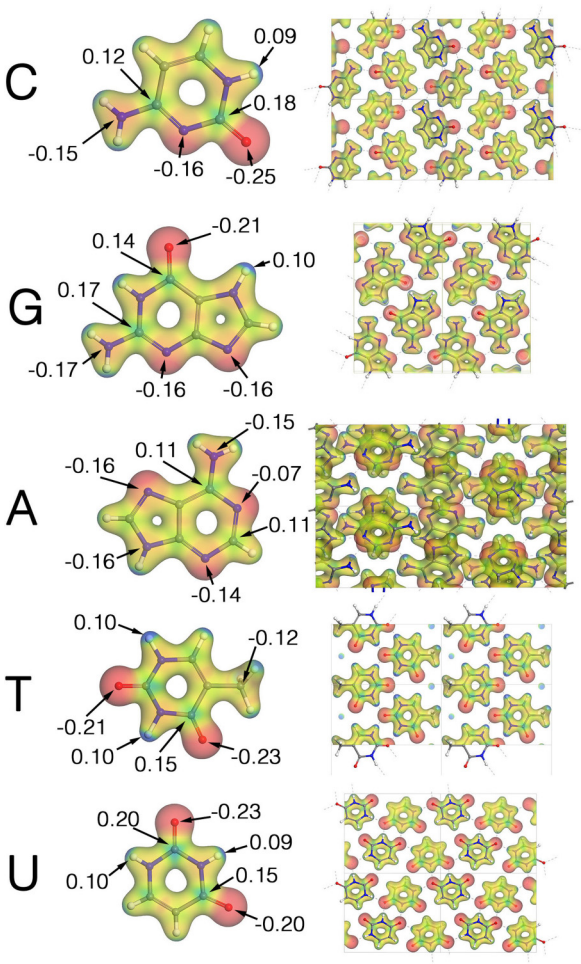

FIG. 8. Constant electron density isosurfaces for cytosine (C), guanine (G), adenine (A), thymine (T), and uracil (U) molecular units (left) in the solid state (right). The electrostatic potential is projected onto each isosurface, increasing from red to blue. Hirshfeld charges for some atoms are indicated.

separation in the molecular units of the crystalline systems and gap types. However, the DFT band gaps of cytosine, thymine, and uracil are closer, being about 0.5 eV larger than the gaps of adenine and guanine, suggesting that the degree of charge polarization is related to the magnitude of the band gap and, consequently, the onset of the optical absorption. A possible mechanism by which this could happen is that the polarization's increase produced by optical excitation in the crystal demands a larger amount of energy in a system already in a high polarization state if the excitation involves the creation of a localized electron-hole pair in a spatial configuration energetically disfavored by the polarization geometry of the crystal. Finally, the charge distribution inside a unit cell (right side of Fig. 8) reveals how electrons are spread out in such a way as to maximize the electrostatic binding interaction, thus minimizing the total energy, with the negatively charged groups of one molecule being close to the positively charged groups of its neighbors.

Figure 9 shows the per-atom electronic partial density of states for the nucleobase crystals between −1.0 and 4.9 eV, spanning the uppermost valence bands and the lowest energy

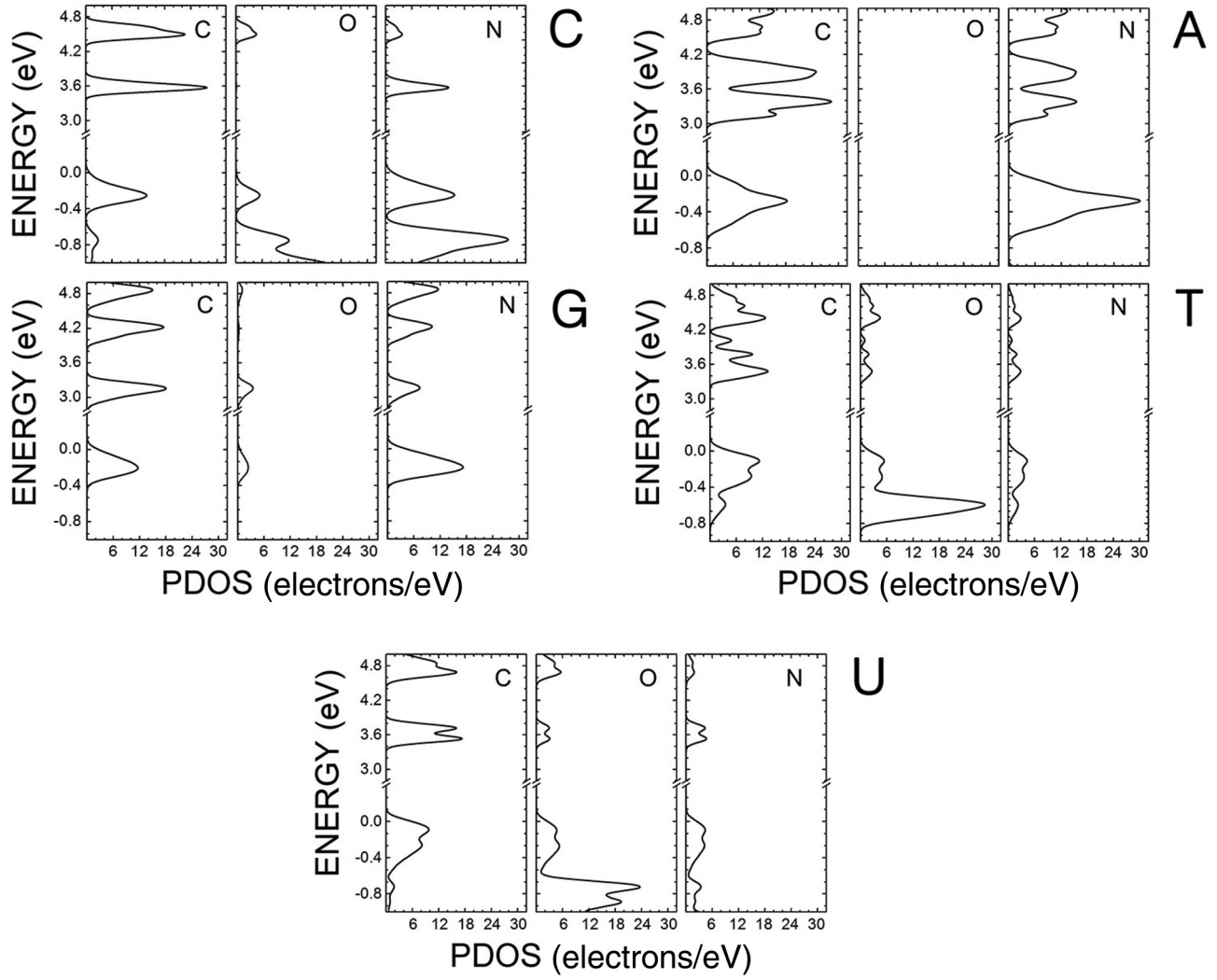

FIG. 9. Electron per-atom densities of states for the nucleobases cytosine (C), guanine (G), adenine (A), thymine (T), and uracil (U) anhydrous crystals near the main band gap. Contributions from carbon (C), oxygen (O), and nitrogen (N) are also shown.

conduction bands. One can note that, for all nucleobases, the top of the valence band has a significant contribution from the nitrogen atoms, especially adenine, originating from N $2p$ levels. Here, O $2p$ states have a similar (smaller) contribution for thymine and uracil (cytosine and guanine), while C $2p$ states have a strong role in the valence states for cytosine, adenine, thymine, and uracil. The electronic states at the bottom of the conduction band are also dominated by C $2p$ contributions for cytosine, adenine, thymine, and uracil, while O $2p$ levels are not relevant in the case of cytosine, although they contribute to the conduction band for thymine and uracil in a proportion similar to the N $2p$ orbitals. A more detailed description of the partial density of states, with simultaneous per-atom and per-orbital contributions being depicted, can be found in the Supplemental Material (Figs. S6–S10) [84].

Band structure plots at the valence and conduction extrema along directions parallel and perpendicular to the molecular stacking planes of the solid-state nucleobases are shown in Fig. 10. The electron energy dispersion curve at the valence band maximum and conduction band minimum can be fitted to a parabola in order to estimate the carrier effective masses, relevant for charge transport. In part, we evaluate them here, motivated by the lack of understanding of the conductivity behavior of DNA strands [8–12], which are formed from the stacking of distinct nucleobases. Table III shows the results obtained for each nucleobase crystal. The perpendicular mass is measured along directions perpendicular to the stacking planes of each nucleobase crystal. The parallel mass, on the other hand, was calculated along some selected in-plane hydrogen bonds for each system (see Fig. 11). For cytosine, which has a direct gap at the Γ point, effective masses parallel to its molecular planes are smaller than along the perpendicular direction, while the electron effective masses are larger and more anisotropic than the hole masses: $m_h = 4.8$ and 3.3 for the perpendicular $\perp$ and parallel $\parallel$ cases, respectively; the free electron has $m = 1.0$. In comparison with our previous paper, the perpendicular electron effective mass predicted using the GGA+TS approach (9.6) has increased significantly (the LDA value was 4.0), while the perpendicular hole effective mass increased by 20%. In the case of guanine, with its direct main transition $B \rightarrow B$, the perpendicular electron and hole effective masses are almost equal to each other

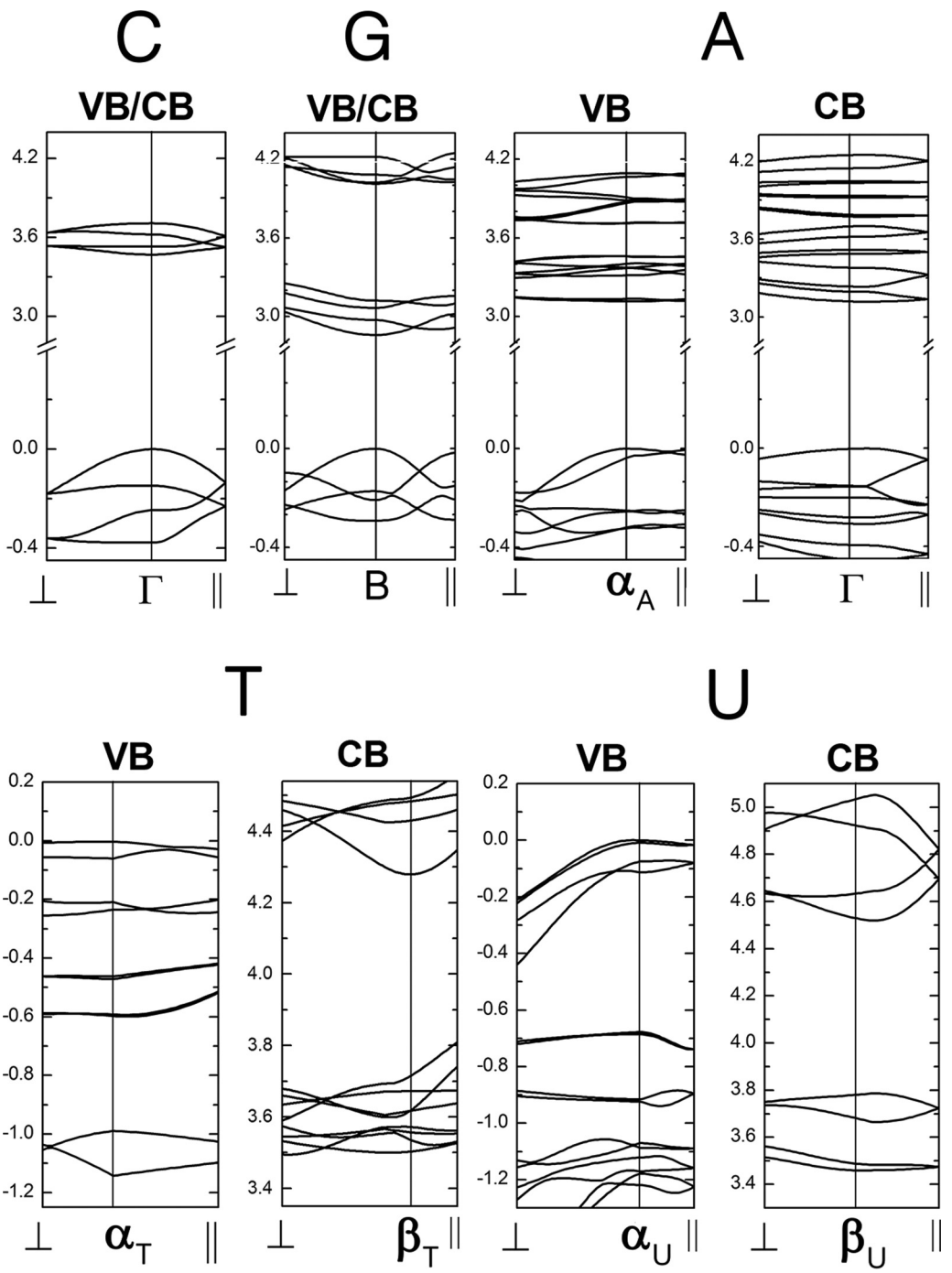

FIG. 10. Nucleobases cytosine (C), guanine (G), adenine (A), thymine (T), and uracil (U): Kohn-Sham band structures near the smallest band gaps along directions perpendicular (⊥) and parallel (||) to the stacking planes.

($m_{e,\perp} = 5.2$, $m_{h,\perp} = 5.3$), as in Ref. [30]. Along the parallel direction the same behavior can be observed, but with smaller effective mass values.

Adenine, thymine, and uracil have indirect gaps, so their band extrema occur at different points in reciprocal space. For adenine, both electron effective masses, parallel and perpendicular, are very large and anisotropic, with $m_{e,\perp} = 31.4$ and $m_{e,\parallel} = 10.6$, values, respectively, larger and smaller than the corresponding ones in Ref. [30], while its hole is relatively light, with m_h varying from 3.2 (perpendicular case) up to 5.6 (parallel case). Previous LDA calculations predicted a very large parallel hole effective mass greater than 40 [30]. In the case of thymine, the perpendicular hole effective mass is very large (30, LDA value 15 [30]) and the parallel mass is much smaller (2.0), while the electron effective masses are not very large: 6.1 and 3.4, for the perpendicular and parallel cases, respectively, very close to the LDA estimate in Ref. [30]. Finally, for uracil, the carriers moving parallel to the stacking plane are relatively heavy, with the electron (hole) mass equal to 14.4 (10.4), while those moving along the perpendicular

TABLE III. Effective masses for electrons and holes (in free electron mass units) calculated using the GGA+TS exchange correlation functional along the directions perpendicular ($\perp$) and parallel ($\|$) to the molecular stacking planes of cytosine (C), guanine (G), adenine (A), thymine (T), and uracil (U) anhydrous crystals.

C	$\Gamma \to \perp$	$\Gamma \to \|$		
m_e	9.6	4.0		
m_h	4.8	3.3		
G	$B \to \perp$	$B \to \|$		
m_e	5.2	2.9		
m_h	5.3	2.5		
A	$\alpha A \to \perp$	$\alpha A \to \|$	$\Gamma \to \perp$	$\Gamma \to \|$
m_e	—	—	31.4	10.6
m_h	3.2	5.6	—	—
T	$\alpha T \to \perp$	$\alpha T \to \|$	$\beta T \to \perp$	$\beta T \to \|$
m_e	—	—	6.1	3.4
m_h	30	2.0	—	—
U	$\alpha U \to \perp$	$\alpha U \to \|$	$\beta U \to \perp$	$\beta U \to \|$
m_e	—	—	6.1	14.4
m_h	3.8	10.4	—	—

direction have reduced effective masses in comparison (3.8 for $m_{h,\perp}$ and 6.1 for $m_{e,\perp}$). These results suggest that the guanine crystal, with an experimental direct gap of 3.60 eV, from our optical absorption measurements, and effective masses varying between 2.5 and 5.3, is more promising as a semiconductor material with potential optoelectronic applications, followed by the cytosine one, which has a direct gap of 4.05 eV. This contrasts with the conventional picture using molecular affinity trends, which points to guanine as an electron blocker [14]. Adenine, on the other hand, approaches insulator behavior with its large electron effective masses, while thymine and uracil can exhibit semiconducting characteristics due to the presence of not so heavy electrons and holes. As a matter of fact, there

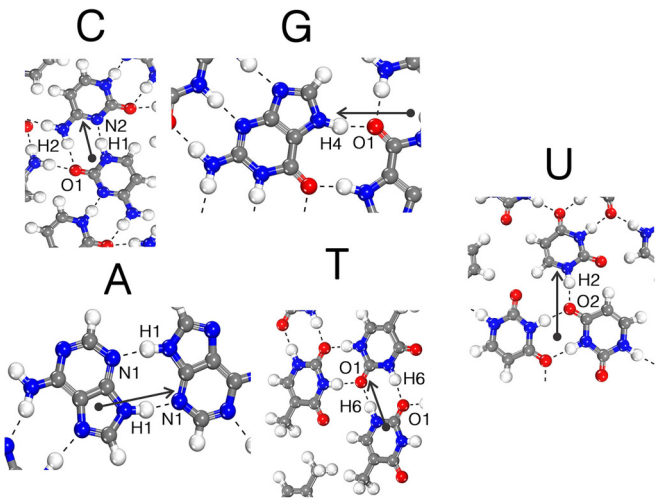

FIG. 11. Nucleobases cytosine (C), guanine (G), adenine (A), thymine (T), and uracil (U): directions parallel ($\|$) to the stacking planes used to evaluate the corresponding effective masses point along some selected in-plane hydrogen bonds.

is a report on the use of adenine and thymine in OLEDs, the former (latter) as an electron (hole) blocking (transport) [18], but no work, to our knowledge, has been published on the potential of guanine as an optically active matrix in the ultraviolet/visible range. However, recent studies on biogenic anhydrous guanine have suggested a different polymorph [92] which can be responsible for many structures used to produce brilliant colors in living beings by light scattering or reflection [93].

V. OPTICAL PROPERTIES

Figure 12 plots our experimental and theoretical optical absorption curves for the five nucleobase anhydrous crystals. In order to estimate the optical band gap from the measured data, we have taken into account the electronic gap type, whether direct or indirect, predicted by the computational simulations. For a direct (indirect) electronic gap material, the onset of the optical absorption α is proportional to the square root (square) of the photon energy [94], so in this case, we have plotted the square (square root) of α and performed a linear fit to find the main band gap (blue lines in Fig. 12). In the cases of the direct gap characteristics depicted by cytosine and guanine, the absorption onset used is barely visible, so we prefer instead to consider the theoretical curves to estimate its approximate location, obtaining 4.05 eV for the former and 3.60 eV for the latter. For adenine, thymine, and uracil, their optical gaps were estimated to be 4.03, 4.13, and 4.15 eV, respectively. The theoretical curves, on the other hand, were adjusted to the Δ-sol band gaps (shown in red) by a rigid shift. One can see that they are in reasonable agreement with our experimental data.

The electronic transitions from the valence band to the conduction band contributing to the onset of the optical absorption can be related to molecular orbital transitions [95]: for cytosine and adenine, with density of states near the band gap originating mostly from N and C $2p$ orbitals (see Figs. S6 and S8 in the Supplemental Material [84]), we can assign a $\pi \to \pi^*$ character. In the case of guanine, thymine, and uracil, a significant contribution from nonbonding O $2p$ electrons in the valence band is present (Figs. S7, S9, and S10), corresponding to a $n \to \pi^*$ molecular transition to the absorption onset. These results must be contrasted with the absorption maximum mostly due to $\pi\pi^*$ transitions observed in the molecular nucleobases solvated in water [95]. The contribution of $\pi\sigma^*$ states to the excited-state dynamics of nucleobase monomers is not a consensus in the literature [95] and is not relevant in our calculations. Our optical absorption measurements also compare well with those of Gomez *et al.* [14] for the nucleobase films, with exhibit onsets at 4.0 eV for cytosine, 3.6 eV for guanine, 3.8 eV for adenine, 4.1 eV for thymine, and 4.2 eV for uracil.

Estimates of the Frenkel excitonic binding energies, as well as fundamental gaps, optical gaps, and relative oscillator strengths for the nucleobase molecules solvated in water are shown in Table IV. The fundamental gaps range from 4.14 eV (guanine) to 4.62 eV (cytosine), increasing following the sequence G < U < T < A < C, while the optical gaps start at 3.61 eV (guanine) reaching up to 4.39 eV (adenine), in the sequence of increasing energy values G < U < T < C < A. The

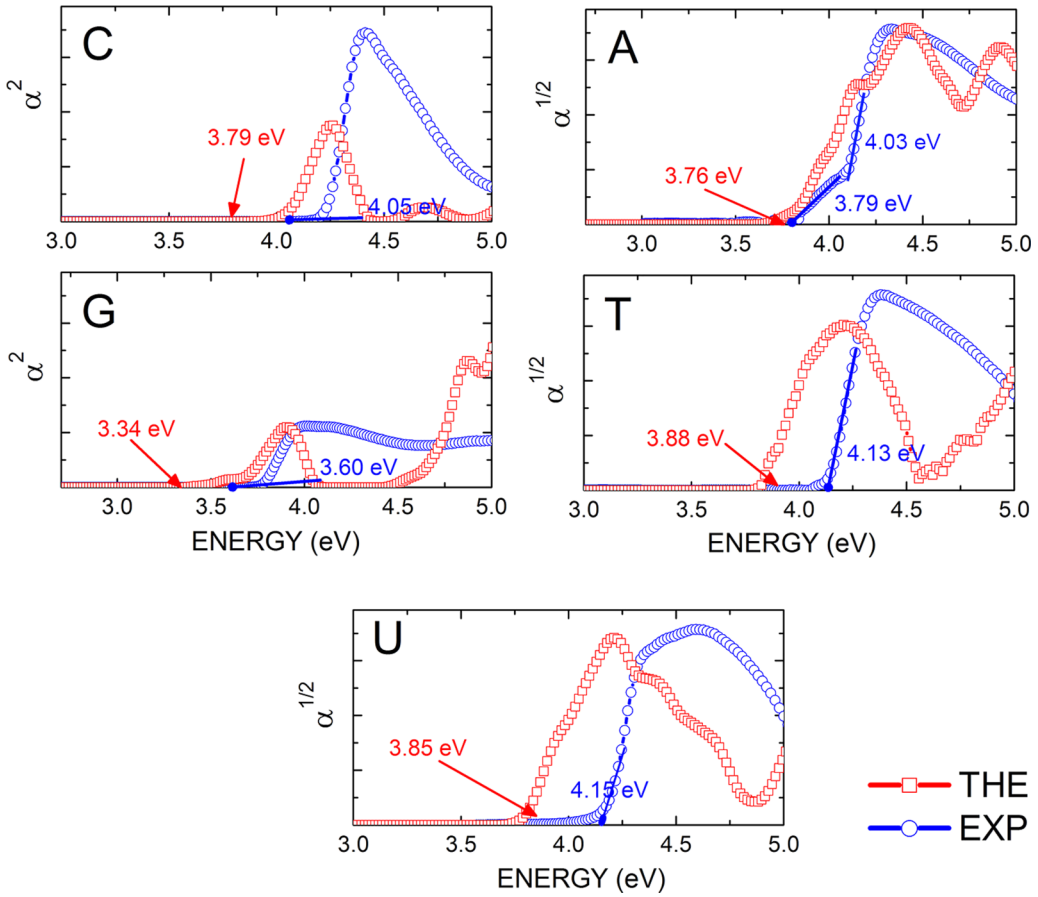

FIG. 12. Optical absorption of cytosine (C), guanine (G), adenine (A), thymine (T), and uracil (U) anhydrous crystals near the main band gap. For the direct gap materials, the square of the optical absorption is plotted, while for the indirect ones the square root of the optical absorption is shown. The red curves are the absorption calculated in this paper using the GGA+TS approach with the Δ-sol correction. The blue curves were obtained from our experimental measurements. The blue small straight lines interpolate the curves near the absorption onsets and reveal the gap estimates (numerical values in blue) for the experimental data. In the case of adenine, there are two gaps due to the Frenkel exciton. Δ-sol gaps are shown in red.

Frenkel exciton binding energy estimate, on the other hand, is the largest for thymine (0.79 eV) and the smallest for adenine (0.22 eV), increasing following the order A < G < U < C < T. Oscillator strengths are in general very small, except for adenine. This is an interesting result, as only the experimental optical absorption of the adenine crystal (Fig. 12, top, right) exhibits two absorption onsets, one starting at about 3.79 eV and the other at about 4.05 eV. Considering the first onset as due to the Frenkel exciton, one can estimate an exciton binding energy of about 0.26 eV, or even a bit smaller since, if we linearly interpolate the second absorption onset, the optical gap is found to be 4.03 eV, leading to an exciton binding energy of 0.24 eV, which is relatively close to the theoretical Frenkel exciton binding energy prediction of 0.22 eV for adenine in Table IV.

TABLE IV. Molecular gaps, Frenkel exciton binding energies, and relative oscillator strengths estimated from the nucleobases solvated in water using TDDFT.

Nucleobase	E_g^{mol} (eV)	E_g^{opt} (eV)	E_b^{exc} (eV)	f
C	4.62	3.87	0.75	10^{-2}
G	4.14	3.61	0.53	10^{-6}
A	4.61	4.39	0.22	1
T	4.55	3.76	0.79	10^{-4}
U	4.35	3.70	0.65	10^{-5}

The complex dielectric function as a function of the incident photon energy $\varepsilon = \varepsilon_1 + i\varepsilon_2$ was calculated for the nucleobase crystals in the cases of incident polarized light and light interacting with a simulated polycrystalline sample (POLY), respectively. The imaginary part ε_2 of the dielectric function is calculated using the Fermi golden rule for time-dependent perturbations

$$\varepsilon_2(\omega) = \frac{2e^2\pi}{\Omega\varepsilon_0} \sum_{\boldsymbol{k},v,c} \left|\langle \psi_{\boldsymbol{k}}^c | \boldsymbol{u}.\boldsymbol{r} | \psi_{\boldsymbol{k}}^v \rangle\right|^2 \delta\left(E_{\boldsymbol{k}}^c - E_{\boldsymbol{k}}^v - \hbar\omega\right). \quad (5)$$

Here, $\boldsymbol{k}$, v, c, and $\boldsymbol{u}$ represent, in this order, the DFT electronic wave vector in the reciprocal space, valence band, conduction band, and the vector defining the polarization of the incident electric field. Also, $E_{\boldsymbol{k}}^b$ stands for the energy of the electron with wave vector $\boldsymbol{k}$ at band b, ω is the photon angular frequency, and Ω is the unit cell volume. The real part of the

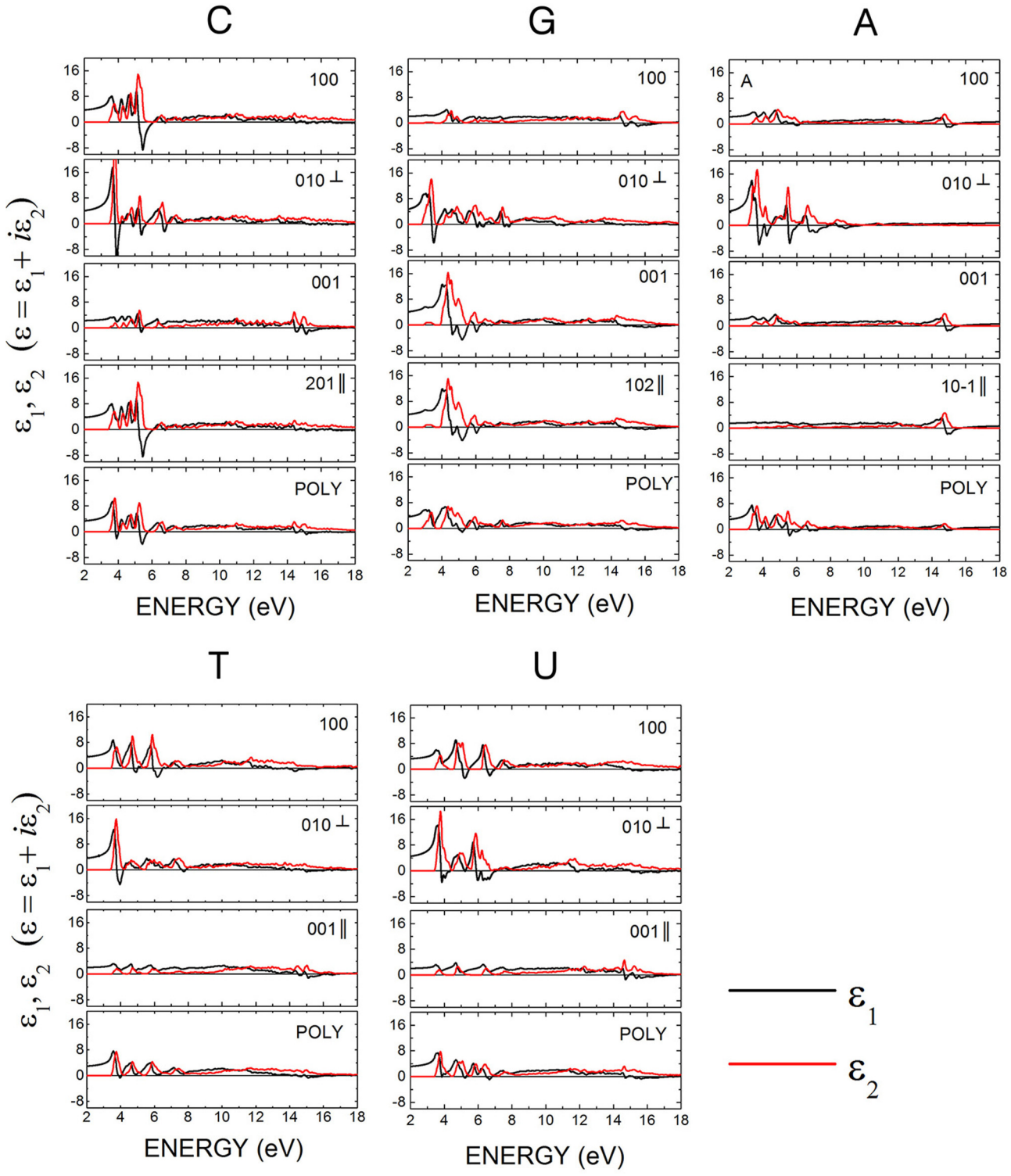

FIG. 13. Complex dielectric function $\varepsilon = \varepsilon_1 + i\varepsilon_2$ of cytosine (C), guanine (G), adenine (A), thymine (T), and uracil (U) anhydrous crystals. The real (imaginary) part $\varepsilon_1(\varepsilon_2)$ is in black (red). Dielectric function curves for light polarized along some selected crystalline planes and for a polycrystalline sample (POLY) were obtained, as well as curves for light polarized along the crystalline planes perpendicular ($\perp$) and parallel ($\|$) to the molecular stackings.

dielectric function ε_1 is obtained through the Kramers-Kronig transform due to causality [96]

$$\varepsilon_1(\omega) = \frac{1}{\pi}\mathcal{P}\int_{-\infty}^{+\infty} \frac{\varepsilon_2(\omega')}{\omega' - \omega} d\omega', \tag{6}$$

where $\mathcal{P}$ is the Cauchy principal value of the integral. Figure 13 shows the calculated complex dielectric function considering light polarized along the 100, 010, and 001 directions without the Δ-sol gap correction. For cytosine, guanine, and adenine, the directions 201, 102, and $10\bar{1}$, respectively, were also

taken into account as they define the crystalline molecular stacking planes (for thymine and uracil, the 001 polarization corresponds to the planar stacking). A quick look shows that the complex dielectric function is very anisotropic, in agreement with our previous study [30], with well-structured and intense maxima and minima for the 010 direction, which is perpendicular to the crystalline planes in all nucleobase systems, but very attenuated along other polarization states: 001 for cytosine, adenine, thymine, and uracil and 100 for guanine. For adenine, as a matter of fact, there is a very pronounced attenuation along the 100, 001, and $10\bar{1}$ directions in comparison with 010. The values of $\varepsilon_1(\omega = 0)$ for the nucleobase polycrystals are 3.03 (cytosine), 3.12 (guanine), 2.65 (adenine), 2.68 (thymine), and 2.87 (uracil). For the DNA nucleobases, these results are in good agreement with the experimental measurements of Silaghi *et al.* [97]. The maximum values of ε_1 along the 010 direction occur at 3.68 eV ($\varepsilon_1 = 17.6$), 2.94 eV ($\varepsilon_1 = 9.36$), 3.35 eV ($\varepsilon_1 = 13.9$), 3.61 eV ($\varepsilon_1 = 12.5$), and 3.61 eV ($\varepsilon_1 = 14.1$), for cytosine, guanine, adenine, thymine, and uracil, respectively. After these maxima, with increasing energy, ε_1 decreases sharply, becoming negative at 3.80 eV (cytosine), 3.40 eV (guanine), 3.67 eV (adenine), 3.78 eV (thymine), and 3.80 eV (uracil). The complex part of the dielectric function ε_2, on the other hand, is proportional to the optical absorption which was already discussed in the beginning of this section.

VI. CONCLUSIONS

In this paper, we have presented the results of optical absorption measurements and DFT GGA+TS (dispersion corrected) calculations for the anhydrous crystals of the five nucleobases of DNA/RNA in order to improve our previous paper [30], which employed a more simple DFT-LDA approach considering just the four DNA nucleobases in the solid state. The structural, electronic, and optoelectronic properties of orthorhombic cytosine and monoclinic guanine, adenine, thymine, and uracil were consistently evaluated, and their main Kohn-Sham band gaps were corrected using the Δ-sol scheme. The GGA+TS lattice parameters have shown a large improvement over the LDA data in comparison with the known experimental data, the mean error decreasing from about 12% down to less than 2% as we switch from the LDA to the GGA+TS exchange-correlation functional. The GGA+TS computations have also allowed us to estimate the binding energy per molecule in the nucleobase systems, with the following sequence of stability (decreasing binding strength): G > C > A > T > U, in agreement with the LDA data and the experimental thermal stability of nucleobase films [14], but differing with respect to MP2 results [86] (A > C). However, the GGA+TS absolute difference of binding energies for cytosine and adenine binding energies is as small (0.41 kcal/mol) as it is in the MP2 simulations (0.50 kcal/mol).

For the electronic band structure, the inclusion of dispersion effects has not changed the gap type of the nucleobase crystals in comparison to the LDA predictions, with cytosine and guanine (adenine, thymine, and uracil) exhibiting direct (indirect) band gaps of 3.79 and 3.34 (3.76, 3.88, and 3.85) eV, respectively, with Δ-sol correction. A comparison between the band gaps estimated from optical absorption and the theoretical values shows that Δ-sol figures improve the gap by almost 70%. The agreement between our calculated and experimental optical absorption curves is reasonable, especially for adenine, notwithstanding the fact that DFT is a ground state theory. The HOMO-LUMO energy gaps estimated using the GGA+TS approach are in good agreement with previous reports except for the relative ordering of guanine and adenine. We have also estimated the binding energy of the Frenkel exciton in the nucleobase crystals using TDDFT calculations performed for water solvated molecules as an approximation, obtaining the largest exciton binding energy for thymine (0.79 eV), but with a very small oscillator strength f. The nucleobase with the largest f, in contrast, was adenine, with corresponding exciton binding energy of 0.22 eV, very close to the evaluation found after analysis of its measured optical absorption spectrum (0.24 eV).

Charge population analysis, on the other hand, reveals that electrostatic effects are essential for the stabilization of the nucleobase crystals. Estimated effective masses are in general anisotropic, with large (relatively small) electron (hole) effective masses for adenine. The guanine crystal, with lighter carriers and a direct gap of 3.60 eV, is a promising semiconductor for optoelectronic devices, followed by cytosine. The other nucleobase crystals may have semiconductor features depending on the carrier type.

Finally, the complex dielectric function for polarized incident light reveals a strong optical anisotropy, with significant features being observed along the 010 direction, normal to the crystalline molecular planes for all nucleobases, and a strong attenuation for parallel polarization. The results here described can be helpful to provide a deeper understanding of potential biodevices built using nucleobase crystals in their design.

ACKNOWLEDGMENTS

V.N.F., E.L.A., U.L.F., and E.W.S.C. are researchers from the Brazilian National Research Council (CNPq) and would like to acknowledge the financial support received during the development of this paper from the Brazilian Research Agency CNPq. E.W.S.C. and M.B. da Silva received financial support from CNPq Projects No. 307843/2013-0 and No. 140898/2016-6, respectively.

[1] J. D. Watson and F. H. C. Crick, Nature **171**, 737 (1953).
[2] D. D. Eley and D. I. Spivey, Trans. Faraday Soc. **58**, 411 (1962).
[3] N. C. Seeman, Nature **421**, 427 (2003).
[4] F. L. Gervasio, A. Laio, M. Parrinello, and M. Boero, Phys. Rev. Lett. **94**, 158103 (2005).
[5] H. Wadati, K. Okazaki, Y. Niimi, A. Fujimori, H. Tabata, J. Pikus, and J. P. Lewis, Appl. Phys. Lett. **86**, 023901 (2005).
[6] A. J. Steckl, Nat. Photon. **1**, 3 (2007).
[7] W. Shi, S. Han, W. Huang, and J. Yu, Appl. Phys. Lett. **106**, 043303 (2015).

[8] R. E. Holmlin, P. J. Dandliker, and J. K. Barton, Angew. Chemie Int. Ed. English **36**, 2714 (1997).

[9] A. Y. Kasumov, M. Kociak, S. Guéron, B. Reulet, V. T. Volkov, D. V Klinov, and H. Bouchiat, Science **291**, 280 (2001).

[10] D. Porath, A. Bezryadin, S. De Vries, and C. Dekker, Nature **403**, 635 (2000).

[11] P. J. de Pablo, F. Moreno-Herrero, J. Colchero, J. Gómez Herrero, P. Herrero, A. M. Baró, P. Ordejón, J. M. Soler, and E. Artacho, Phys. Rev. Lett. **85**, 4992 (2000).

[12] C. Gómez-Navarro, F. Moreno-Herrero, P. J. de Pablo, J. Colchero, J. Gómez-Herrero, and A. M. Baró, Proc. Natl. Acad. Sci. **99**, 8484 (2002).

[13] P. Meredith, C. J. Bettinger, M. Irimia-Vladu, A. B. Mostert, and P. E. Schwenn, Rep. Prog. Phys. **76**, 034501 (2013).

[14] E. F. Gomez, V. Venkatraman, J. G. Grote, and A. J. Steckl, Adv. Mater. **27**, 7552 (2014).

[15] G. Maruccio, P. Visconti, V. Arima, S. D'Amico, A. Blasco, E. D'Amone, R. Cingolani, R. Rinaldi, S. Masiero, T. Giorgi, and G. Gottarelli, Nano Lett. **3**, 479 (2003).

[16] M. Irimia-Vladu, P. A. Troshin, M. Reisinger, L. Shmygleva, Y. Kanbur, G. Schwabegger, M. Bodea, R. Schwödiauer, A. Mumyatov, J. W. Fergus, V. F. Razumov, H. Sitter, N. S. Sariciftci, and S. Bauer, Adv. Funct. Mater. **20**, 4069 (2010).

[17] L. Shu, W. Shi, W. Huang, and J. Yu, J. Mater. Sci. Mater. Electron. **25**, 5540 (2014).

[18] E. F. Gomez, V. Venkatraman, J. G. Grote, and A. J. Steckl, Sci. Rep. **4**, 7105 (2014).

[19] J. Lee, J. H. Park, Y. T. Lee, P. J. Jeon, H. S. Lee, S. H. Nam, Y. Yi, Y. Lee, and S. Im, ACS Appl. Mater. Interfaces **6**, 4965 (2014).

[20] E. F. Gomez and A. J. Steckl, ACS Photonics **2**, 439 (2015).

[21] G. J. O. Beran, Chem. Rev. **116**, 5567 (2016).

[22] E. W. S. Caetano, J. R. Pinheiro, M. Zimmer, V. N. Freire, G. A. Farias, G. A. Bezerra, B. S. Cavada, J. R. L. Fernandez, J. R. Leite, M. C. F. De Oliveira, J. A. Pinheiro, J. L. De Lima Filho, and H. W. Leite Alves, AIP Conf. Proc. **772**, 1095 (2005).

[23] M. Z. S. Flores, V. N. Freire, R. P. Dos Santos, G. A. Farias, E. W. S. Caetano, M. C. F. De Oliveira, J. R. L. Fernandez, L. M. R. Scolfaro, M. J. B. Bezerra, T. M. Oliveira, G. A. Bezerra, B. S. Cavada, and H. W. Leite Alves, Phys. Rev. B: Condens. Matter Mater. Phys. **77**, 115104 (2008).

[24] G. Zanatta, C. Gottfried, A. M. Silva, E. W. S. Caetano, F. A. M. Sales, and V. N. Freire, J. Chem. Phys. **140**, 124511 (2014).

[25] J. R. Cândido-Júnior, F. A. M. Sales, S. N. Costa, P. de Lima-Neto, D. L. Azevedo, E. W. S. Caetano, E. L. Albuquerque, and V. N. Freire, Chem. Phys. Lett. **512**, 208 (2011).

[26] A. M. Silva, B. P. Silva, F. A. M. Sales, V. N. Freire, E. Moreira, U. L. Fulco, E. L. Albuquerque, F. F. Maia, and E. W. S. Caetano, Phys. Rev. B **86**, 195201 (2012).

[27] A. M. Silva, S. N. Costa, B. P. Silva, V. N. Freire, U. L. Fulco, E. L. Albuquerque, E. W. S. Caetano, and F. F. Maia, Cryst. Growth Des. **13**, 4844 (2013).

[28] A. M. Silva, S. N. Costa, F. A. M. Sales, V. N. Freire, E. M. Bezerra, R. P. Santos, U. L. Fulco, E. L. Albuquerque, and E. W. S. Caetano, J. Phys. Chem. A **119**, 11791 (2015).

[29] S. N. Costa, F. A. M. Sales, V. N. Freire, F. F. Maia, E. W. S. Caetano, L. O. Ladeira, E. L. Albuquerque, and U. L. Fulco, Cryst. Growth Des. **13**, 2793 (2013).

[30] F. F. Maia, V. N. Freire, E. W. S. Caetano, D. L. Azevedo, F. A. M. Sales, and E. L. Albuquerque, J. Chem. Phys. **134**, 175101 (2011).

[31] E. L. Albuquerque, U. L. Fulco, V. N. Freire, E. W. S. Caetano, M. L. Lyra, and F.A.B.F. de Moura, Phys. Rep. **535**, 139 (2014).

[32] D. Mollenhauer, C. Brieger, E. Voloshina, and B. Paulus, J. Phys. Chem. C **119**, 1898 (2015).

[33] T. Fornaro, M. Biczysko, S. Monti, and V. Barone, Phys. Chem. Chem. Phys. **16**, 10112 (2014).

[34] S. Panigrahi, A. Bhattacharya, S. Banerjee, and D. Bhattacharyya, J. Phys. Chem. C **116**, 4374 (2012).

[35] G. Casella, A. Bagno, and G. Saielli, Phys. Chem. Chem. Phys. **15**, 18030 (2013).

[36] S. Ehrlich, J. Moellmann, and S. Grimme, Acc. Chem. Res. **46**, 916 (2013).

[37] W. Hujo and S. Grimme, J. Chem. Theory Comput. **9**, 308 (2013).

[38] J. G. Brandenburg and S. Grimme, J. Phys. Chem. Lett. **5**, 1785 (2014).

[39] D. L. Barker and R. E. Marsh, Acta Crystallogr. **17**, 1581 (1964).

[40] K. Guille and W. Clegg, Acta Crystallogr. Sect. C **62**, o515 (2006).

[41] S. Mahapatra, S. K. Nayak, S. J. Prathapa, and T. N. Guru Row, Cryst. Growth Des. **8**, 1223 (2008).

[42] K. Ozeki, N. Sakabe, and J. Tanaka, Acta Crystallogr. Sect. B **25**, 1038 (1969).

[43] G. S. Parry, Acta Crystallogr. **7**, 313 (1954).

[44] H. Mineki, Y. Kaimori, T. Kawasaki, A. Matsumoto, and K. Soai, Tetrahedron: Asymmetry **24**, 1365 (2013).

[45] B. Sridhar, J. B. Nanubolu, and K. Ravikumar, Acta Crystallogr. Sect. C **71**, 128 (2015).

[46] D. E. Braun, T. Gelbrich, K. Wurst, and U. J. Griesser, Cryst. Growth Des. **16**, 3480 (2016).

[47] R. Chennuru, P. Muthudoss, S. Ramakrishnan, A. B. Mohammad, R. Ravi Chandra Babu, S. Mahapatra, and S. K. Nayak, J. Mol. Struct. **1120**, 86 (2016).

[48] T. Stolar, S. Lukin, J. Požar, M. Rubčić, G. M. Day, I. Biljan, D. Š. Jung, G. Horvat, K. Užarević, E. Meštrović, and I. Halasz, Cryst. Growth Des. **16**, 3262 (2016).

[49] D. Gur, M. Pierantoni, N. Elool Dov, A. Hirsh, Y. Feldman, S. Weiner, and L. Addadi, Cryst. Growth Des. **16**, 4975 (2016).

[50] M. D. Segall, P. J. D. Lindan, M. J. Probert, C. J. Pickard, P. J. Hasnip, S. J. Clark, and M. C. Payne, J. Phys. Condens. Matter **14**, 2717 (2002).

[51] S. J. Clark, M. D. Segall, C. J. Pickard, P. J. Hasnip, M. I. J. Probert, K. Refson, and M. C. Payne, Zeitschrift Fur Krist. **220**, 567 (2005).

[52] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).

[53] D. M. Ceperley and B. J. Alder, Phys. Rev. Lett. **45**, 566 (1980).

[54] A. Tkatchenko and M. Scheffler, Phys. Rev. Lett. **102**, 073005 (2009).

[55] S. Grimme, J. Comput. Chem. **27**, 1787 (2006).

[56] E. R. McNellis, J. Meyer, and K. Reuter, Phys. Rev. B **80**, 205414 (2009).

[57] S. Grimme, A. Hansen, J. G. Brandenburg, and C. Bannwarth, Chem. Rev. **116**, 5105 (2016).

[58] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, J. Chem. Phys. **132**, 154104 (2010).

[59] D. C. Langreth, B. I. Lundqvist, S. D. Chakarova-Käck, V. R. Cooper, M. Dion, P. Hyldgaard, A. Kelkkanen, J. Kleis, L. Kong, S. Li, P. G. Moses, E. Murray, A. Puzder, H. Rydberg, E. Schröder, and T. Thonhauser, J. Phys. Condens. Matter **21**, 084203 (2009).
[60] K. Berland, V. R. Cooper, K. Lee, E. Schröder, T. Thonhauser, P. Hyldgaard, and B. I. Lundqvist, Reports Prog. Phys. **78**, 066501 (2015).
[61] O. A. von Lilienfeld, I. Tavernelli, U. Rothlisberger, and D. Sebastiani, Phys. Rev. Lett. **93**, 153004 (2004).
[62] O. A. von Lilienfeld, I. Tavernelli, U. Rothlisberger, and D. Sebastiani, Phys. Rev. B **71**, 195119 (2005).
[63] N. Mardirossian and M. Head-Gordon, J. Chem. Theory Comput. **12**, 4303 (2016).
[64] L. Goerigk, J. Phys. Chem. Lett. **6**, 3891 (2015).
[65] C. J. Pickard, Castep Docs, 1 (2006).
[66] D. R. Hamann, M. Schlüter, and C. Chiang, Phys. Rev. Lett. **43**, 1494 (1979).
[67] B. G. Pfrommer, M. Cote, S. G. Louie, and M. L. Cohen, J. Comput. Phys. **131**, 233 (1997).
[68] R. S. Mulliken, J. Chem. Phys. **23**, 1833 (1955).
[69] F. L. Hirshfeld, Theor. Chim. Acta **44**, 129 (1977).
[70] C. S. Wang and W. E. Pickett, Phys. Rev. Lett. **51**, 597 (1983).
[71] L. Hedin, Phys. Rev. **139**, A796 (1965).
[72] J. Heyd and G. E. Scuseria, J. Chem. Phys. **121**, 1187 (2004).
[73] M. K. Y. Chan and G. Ceder, Phys. Rev. Lett. **105**, 196403 (2010).
[74] B. Delley, J. Chem. Phys. **113**, 7756 (2000).
[75] A. Klamt and G. Schuurmann, J. Chem. Soc., Perkin Trans. **2**, 799 (1993).
[76] E. Runge and E. K.U. Gross, Phys. Rev. Lett. **52**, 997 (1984).
[77] E. K. U. Gross and W. Kohn, Adv. Quantum Chem. **21**, 255 (1990).
[78] B. Delley, J. Phys. Condens. Matter **22**, 384208 (2010).
[79] J.-C. Lee, J.-D. Chai, and S.-T. Lin, RSC Adv. **5**, 101370 (2015).
[80] P. K. Nayak, Synth. Met. **174**, 42 (2013).
[81] J.-L. Bredas, Mater. Horiz. **1**, 17 (2014).
[82] R. F. Stewart and L. H. Jensen, Acta Crystallogr. **23**, 1102 (1967).
[83] R. J. McClure and B. M. Craven, Acta Crystallogr. Sect. B **29**, 1234 (1973).
[84] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevB.96.085206 for a detailed description of Kohn-Sham band structures, electron densities of states, structural properties, and charge populations calculated for the five nucleobase crystals.
[85] J. Sponer, J. Leszczynski, and P. Hobza, J. Phys. Chem. **100**, 5590 (1996).
[86] J. Sponer, K. E. Riley, and P. Hobza, Phys. Chem. Chem. Phys. **10**, 2581 (2008).
[87] F. Ouchen, E. Gomez, D. Joyce, P. Yaney, S. Kim, A. Williams, A. Steckl, N. Venkat, and J. Grote, Proc. SPIE **8817**, 88170C (2013).
[88] R. K. Roy, K. Hirao, and S. Pal, J. Chem. Phys. **110**, 8236 (1999).
[89] P. W. Ayers, R. C. Morrison, and R. K. Roy, J. Chem. Phys. **116**, 8731 (2002).
[90] P. Bultinck, C. Van Alsenoy, P. W. Ayers, and R. Carbó-Dorca, J. Chem. Phys. **126**, 144111 (2007).
[91] T. Verstraelen, S. V. Sukhomlinov, V. Van Speybroeck, M. Waroquier, and K. S. Smirnov, J. Phys. Chem. C **116**, 490 (2012).
[92] A. Hirsch, D. Gur, I. Polishchuk, D. Levy, B. Pokroy, A. J. Cruz-Cabeza, L. Addadi, L. Kronik, and L. Leiserowitz, Chem. Mater. **27**, 8289 (2015).
[93] J. Teyssier, S. V. Saenko, D. van der Marel, and M. C. Milinkovitch, Nat. Commun. **6**, 6368 (2015).
[94] A. M. Fox, *Optical Properties of Solids*, 1st ed. (Oxford University Press, New York, 2001), p. 58.
[95] M. Pollum, L. Martínez-Fernández, and C. E. Crespo-Hernández, in *Topics in Current Chemistry* (Springer International Publishing, Cham, 2014), Vol. 355, pp. 245–327.
[96] J.S. Toll, Phys. Rev. **104**, 1760 (1956).
[97] S. D. Silaghi, M. Friedrich, C. Cobet, N. Esser, W. Braun, and D. R. T. Zahn, Phys. Status Solidi **242**, 3047 (2005).