First-principles *GW* calculations for DNA and RNA nucleobases

Carina Faber,[1,2] Claudio Attaccalite,[1] V. Olevano,[1] E. Runge,[2] and X. Blase[1]

[1]*Institut Néel, CNRS and Université Joseph Fourier, B.P. 166, F-38042 Grenoble Cedex 09, France*
[2]*Institut für Physik, Technische Universität Ilmenau, D-98693 Ilmenau, Germany*

On the basis of first-principles GW calculations, we study the quasiparticle properties of the guanine, adenine, cytosine, thymine, and uracil DNA and RNA nucleobases. Beyond standard G_0W_0 calculations, starting from Kohn-Sham eigenstates obtained with (semi)local functionals, a simple self-consistency on the eigenvalues allows us to obtain vertical ionization energies and electron affinities within an average 0.11 and 0.18 eV error, respectively, as compared to state-of-the-art coupled-cluster and multiconfigurational perturbative quantum chemistry approaches. Further, GW calculations predict the correct π-character of the highest occupied state, due to several level crossings between density functional and GW calculations. Our study is based on a recent Gaussian-basis implementation of GW calculations with explicit treatment of dynamical screening through contour deformation techniques.

I. INTRODUCTION

The determination of the ionization energies, electronic affinities, and character of the frontier orbitals of DNA and RNA nucleobases is an important step toward a better understanding of the electronic properties and reactivity of nucleotides and nucleosides along the DNA/RNA chains. Important phenomena such as nucleobases/protein interactions, defining the DNA functions,[1] or damage of the genetic material through oxidation or ionizing radiations[2] are strongly related to these fundamental spectroscopic quantities. Even though nucleobases in DNA/RNA strands are connected within the nucleotides to phosphate groups through a five-carbon sugar, several studies show that the highest-occupied orbital (the HOMO level) in nucleotides, which is responsible, e.g., for the sensitivity of the molecule to oxidation processes, remains localized on the nucleobases.[3] Figure 1 shows the structures of the DNA and RNA nucleobases, i.e., the purines—adenine (A) and guanine (G)—and the pyrimidines—cytosine (C) as well as thymine (T) in DNA and uracil (U) in RNA.

Besides the overarching fundamental interest in understanding complex biological processes at the microscopic level, *ab initio* calculations of isolated nucleobases are interesting since recent high-level quantum chemistry calculations[4–6] allow us to rationalize the rather large spread of experimental results concerning the electronic properties of the nucleobases in the gas phase,[7–13] in particular as due to the existence of several isomers for guanine and cytosine.[6] Thus, these molecules offer a valuable mean to explore the merits of the so-called GW formalism[14–18] for isolated organic molecules, along the line of recent systematic studies of small molecules[19] or molecules such as fullerenes or porphyrins of interest for electronic or photovoltaic applications.[20–25]

In the present work, we study by means of first-principles GW calculations the quasiparticle properties of the DNA and RNA nucleobases, namely, guanine, adenine, cytosine, thymine, and uracil. We show in particular that the GW correction to the Kohn-Sham eigenvalues brings the ionization energies in much better agreement with experiment and high-level quantum chemistry calculations. These results demonstrate the importance of self-consistency on the eigenvalues when performing GW calculations in molecular systems starting from (semi)local DFT functionals, and the merits of a simple scheme based on a G_0W_0 calculation starting from Hartree-Fock-like eigenvalues.

The GW approach is a Green's function formalism usually derived within a functional derivative treatment[14,26] allowing us to prove that the two-body Green's function (G_2), involved in the equation of motion of the one-body time-ordered Green's function G, can be recast into a nonlocal and energy-dependent self-energy operator $\Sigma(\mathbf{r},\mathbf{r}'|\omega)$. This self-energy Σ accounts for exchange and correlation in the present formalism. Since it is energy dependent, it must be evaluated at the $E = \varepsilon_i^{QP}$ quasiparticle energies, where (i) indexes the molecular energy levels. This self-energy involves $G(\mathbf{r},\mathbf{r}'|\omega)$, the dynamically screened Coulomb potential $W(\mathbf{r},\mathbf{r}'|\omega)$, and the so-called vertex correction Γ. A set of exact self-consistent (closed) equations connects G, W, Γ, and the independent-electron/full polarizabilities $\chi_0(\mathbf{r},\mathbf{r}'|\omega)$ and $\chi(\mathbf{r},\mathbf{r}'|\omega)$, respectively. In the GW approximation (GWA), the three-body vertex operator Γ is set to unity, yielding the following expression for the self-energy:

$$\Sigma(\mathbf{r},\mathbf{r}'|E) = \frac{i}{2\pi}\int d\omega e^{i\omega 0^+} G(\mathbf{r},\mathbf{r}'|E+\omega)W(\mathbf{r},\mathbf{r}'|\omega),$$

$$\tilde{W}(\mathbf{r},\mathbf{r}'|\omega) = \int d\mathbf{r_1}d\mathbf{r_2}\, v(\mathbf{r},\mathbf{r_1})\, \chi_0(\mathbf{r_1},\mathbf{r_2}|\omega)\, W(\mathbf{r_2},\mathbf{r}'|\omega),$$

$$\chi_0(\mathbf{r},\mathbf{r}'|\omega) = \sum_{i,j}(f_i - f_j)\frac{\phi_i^*(\mathbf{r})\phi_j(\mathbf{r})\phi_j^*(\mathbf{r}')\phi_i(\mathbf{r}')}{\varepsilon_i - \varepsilon_j + \omega \pm i\delta},$$

where $v(\mathbf{r},\mathbf{r}')$ is the bare (unscreened) Coulomb potential and $\tilde{W} = W - v$. The (ε_i, ϕ_i) are "zeroth-order" one-body eigenstates. Following the large bulk of work[18] devoted to GW calculations in solids, surfaces, graphene, nanotubes, or nanowires, we use here Kohn-Sham DFT-LDA eigenstates. It is shown below, and in Refs. 19,25,27,28, that Hartree-Fock (or hybrid) solutions may constitute better starting points for molecular systems. (f_i, f_j) are Fermi-Dirac occupation numbers, and δ an infinitesimal such that the poles of W fall in the second and fourth quadrants of the complex plane.
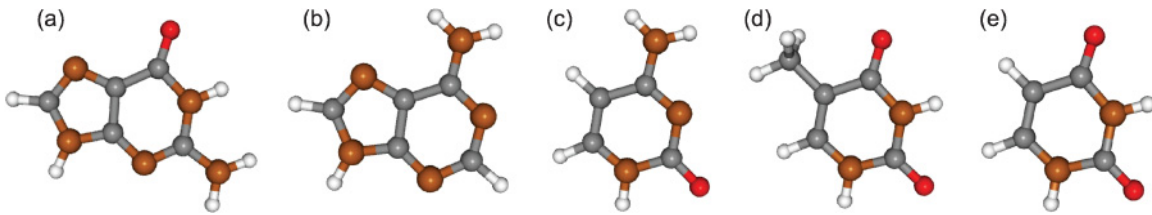
FIG. 1. (Color online) Schematic representation of the molecular structure of (a) guanine (G9K), (b) adenine, (c) cytosine (C1), (d) thymine, and (e) uracil. Black, brown, red, and white atoms are carbon, nitrogen, oxygen, and hydrogen, respectively. The G9K and C1 notations for the guanine and cytosine tautomers are consistent with Ref. 6.

In the GW approximation, the self-energy operator can be loosely interpreted as a generalization of the Hartree-Fock method by replacing the bare Coulomb potential with a dynamically screened Coulomb interaction accounting both for exchange and (dynamical) correlations. An important feature of the GW approach is that not only ionization energies and electronic affinities can be calculated, but also the full quasiparticle spectrum. Further, both localized and infinite systems can be treated on the same footing with long- and short-range screening automatically accounted for in the construction of the screened Coulomb potential W. More details about the present implementation can be found in Ref. 25.

Our calculations are based on a recently developed implementation of the GW formalism (the FIESTA code) using a Gaussian auxiliary basis to expend the two-point operators such as the Coulomb potential, the susceptibilities, or the self-energy.[25] Dynamical correlations are included explicitly through contour deformation techniques. We start with a ground-state DFT calculation using the SIESTA package[29] and a large triple-zeta with double polarization (TZDP) basis.[30] We fit the radial part of the numerical basis generated by the SIESTA code by up to five contracted Gaussians in order to facilitate the calculation of the Coulomb matrix elements and of the matrix elements $\langle \phi_i | \beta | \phi_j \rangle$ of the auxiliary basis (β) between Kohn-Sham states. Such a scheme allows us to exploit the analytic relations for the products of Gaussian orbitals centered on different atoms or for their Fourier transform.[25] Our auxiliary basis for first row elements is the tempered basis[31] developed by Kaczmarski and coworkers.[32] Such a basis was tested recently in a systematic study of several molecules of interest for photovoltaic applications.[25] Four Gaussians for each 1-channel with localization coefficients $\alpha = (0.2, 0.5, 1.25, 3.2)$ a.u. are used for the (s,p,d) channels of C, O, and N atoms, while three Gaussians with $\alpha = (0.1, 0.4, 1.5)$ a.u. describe hydrogen.[33]

II. IONIZATION ENERGIES

We now comment on the values of the calculated first ionization energy (IE) as compiled in Table I and Fig. 2. The comparison to the experimental data is complicated by the 0.2–0.3 eV range spanned by the various experimental reports (vertical arrows in Fig. 2). An additional complication in the case of cytosine and guanine, beyond the intrinsic difficulties in accurately measuring ionization energies in the gas phase, is that several gas phase tautomers exist[6] that differ from the

so-called C1-cytosine and G9K-guanine isomers commonly found in DNA (see Fig. 1). State-of-the-art *ab initio* quantum chemistry calculations, namely, coupled-cluster CCSD(T) and multiconfigurational perturbation (CASPT2) methods,[4,5] studied the nucleobase tautomers that can be found along the DNA/RNA strands. More recently, equation of motion coupled-cluster techniques (EOM-IP-CCSD) were performed on several isomers.[6] All methods agree to within 0.04 eV for the average IE of the A, G, C, T tautomers we consider here, with a maximum discrepancy of 0.09 eV in the case of thymine. The CASPT2 and CCSD(T) calculations agree to within 0.03 eV for all molecules. These theoretical IE are commonly considered as the most reliable references and land within the experimental error bars, except for the cytosine (C1) case, where the calculated IEs are slightly smaller than the experimental lower bound[34] (see Table I and Fig. 2).

Clearly, the ionization energy within DFT-LDA, as given by the negative HOMO Kohn-Sham level energy, significantly underestimates the IE by an average of ~2.5 eV (29%).[35] The self-energy correction at the G_0W_0(LDA) level improves very significantly the situation by bringing the error to an average 0.5 eV (5.7%) as compared to state-of-the-art quantum chemistry results. However, as emphasized in recent papers,[19,25,27,28] the overscreening induced by starting with LDA eigenvalues, which dramatically underestimate the band gap, tends to produce too small ionization energies. This problem can be solved at least partly by performing a simple self-consistency on the eigenvalues. We shall refer to this approach as GW henceforth. Such a self-consistency on the eigenvalues leads to a much reduced average error of 0.11 eV (~1.3%) as compared to the quantum chemistry reference. This good agreement certainly indicates the reliability of the present GW scheme for such systems. As shown in Fig. 2, the largest discrepancies are observed for guanine and adenine (the purines), while the agreement is excellent for the three remaining bases.

In recent work, it was shown that for small molecules a non-self-consistent G_0W_0 calculation starting from Hartree-Fock eigenstates leads for the ionization energy to better results than a full self-consistent GW calculation where the wavefunctions are updated as well.[19,27] Consistent with this observation, a simple scheme relying on an Hartree-Fock-like approach was successfully tested on silane, disilane, and water[28] and larger molecules such as fullerenes or porphyrins.[25] In this "G_0W_0 on Hartree-Fock (HF)" *ansatz*, the input eigenvalues ($\tilde{\epsilon}_n$) are computed within a diagonal first-order perturbation theory,

TABLE I. Vertical ionization energies and electronic affinities in eV as obtained from the negative Kohn-Sham eigenvalues (LDA-KS), from non-self-consistent G_0W_0(LDA) calculations, from a GW calculation with self-consistency on the eigenvalues (GW), and from a non-self-consistent G_0W_0(HF$_{diag}$) calculation starting from Hartree-Fock-like eigenvalues. The σ or π character of the wavefunctions is indicated when the GW correction changes the level ordering as compared to DFT/LDA (see text). The acronyms CAS, CC, and EOM stand for CASPT2, CCSD(T), and equation-of-motion coupled-cluster high-level many-body quantum chemistry calculations, respectively. Theoretical values are reported for the C1-cytosine and G9K-guanine, while the experimental values average over several tautomers. The MAE is the mean absolute error in eV as compared to the quantum chemistry reference calculations in columns 6 and 7.

	LDA-KS	G_0W_0(LDA)	GW	G_0W_0(HF$_{diag}$)	CAS[a,b]/CC[a,b]	EOM[c]	Experiment[d,e,f,g]
G-LUMO	1.80	−1.04	−1.58	−1.77	−1.14[a]/		
G-HOMO	5.69	7.49	7.81	7.76	8.09[b]/8.09[b]	8.15	8.0–8.3[d]
G-HOMO-1	6.34	8.78	9.82	9.78	9.56[b]/	9.86	9.90[g]
A-LUMO	2.22	−0.64	−1.14	−1.30	−0.91[a]/		−0.56 to −0.45[e]
A-HOMO	6.02	7.90	8.22	8.23	8.37[b]/8.40[b]	8.37	8.3–8.5,[d] 8.47[f]
A-HOMO-1	6.28	8.75	9.47	9.51	9.05[b]/	9.37	9.45[f]
C-LUMO	2.57	−0.45	−0.91	−1.05	−0.69[a]/−0.79[a]		−0.55 to −0.32[e]
C-HOMO	6.167 (σ_O)	8.21 (π)	8.73 (π)	9.05 (π)	8.73[b] (π)/8.76[b]	8.78 (π)	8.8–9.0,[d] 8.89[f]
C-HOMO-1	6.172 (π)	8.80 (σ_O)	9.52 (π')	9.87 (π')	9.42[b] (σ_O)/	9.54 (π')	9.45,[g] 9.55[f]
C-HOMO-2	6.806 (σ)	8.92 (π')	9.89 (σ_O)	10.36 (σ_O)	9.49[b] (π')/	9.65 (σ_O)	9.89[f]
C-HOMO-3	6.809 (π')	9.38 (σ)	10.22 (σ)	10.64 (σ)	9.88[b](σ)/	10.06 (σ)	11.20[f]
T-LUMO	2.83	−0.14	−0.67	−0.77	−0.60[a]/−0.65[a]		−0.53 to −0.29[e]
T-HOMO	6.54	8.64	9.05	9.05	9.07[b]/9.04[b]	9.13	9.0–9.2,[d] 9.19[f]
T-HOMO-1	6.68	9.34	10.41	10.40	9.81[b]/	10.13	9.95–10.05,[d]10.14[f]
U-LUMO	3.01	−0.11	−0.64	−0.71	−0.61[a]/−0.64[a]		−0.30 to −0.22[e]
U-HOMO	6.72 (σ_O)	9.03 (π)	9.47 (π)	9.73 (π)	9.42[b] (π)/9.43[b]		9.4–9.6[d]
U-HOMO-1	6.88 (π)	9.45 (σ_O)	10.54 (σ_O)	10.96 (σ_O)	9.83[b] (σ_O)/		10.02–10.13[d]
U-HOMO-2	7.55 (σ)	9.88 (π')	10.66 (π')	11.06 (π')	10.41[b] (π')/		10.51–10.56[d]
U-HOMO-3	7.66 (π')	10.33 (σ)	11.48 (σ)	11.90 (σ)	10.86[b] (σ)/		10.90–11.16[d]
MAE LUMO	3.29	0.33	0.18	0.31			
MAE HOMO	2.5	0.5	0.11	0.22			

[a]Refrence 5.
[b]Refrence 4.
[c]Refrence 6.
[d]Compiled in Ref. 4.
[e]Compiled in Ref. 5.
[f]Refrence 10.
[g]Refrence 8.

where the DFT exchange-correlation contribution to the eigenvalues is replaced by the Fock exchange integral, namely:

$$\tilde{\epsilon}_n = \epsilon_n^{LDA} + \langle \psi_n^{LDA} | \Sigma_x - V_{xc}^{LDA} | \psi_n^{LDA} \rangle,$$

where Σ_x is the Fock operator. This approach, labeled G_0W_0(HF$_{diag}$) in Table I, produces an average error of 0.22 eV ($\sim$2.6%). This good agreement with both the GW and quantum chemistry calculations clearly speaks in favor of this simple scheme for molecular systems, or the full G_0W_0(HF) calculations tested in Ref. 19, which also avoids seeking self-consistency. A difficult issue lying ahead concerns, e.g., hybrid systems, such as semiconducting surfaces grafted by organic molecules, for which it is not quite clear what should be the best starting point.

Next, we address the character of the HOMO level of cytosine and uracil. It changes from DFT-LDA to GW calculations. We plot in Fig. 3(a)–3(d) the C1-cytosine DFT/LDA Kohn-Sham HOMO to (HOMO-3) eigenstates. The LDA HOMO level is an in-plane σ state with a strong component on the (p_x, p_y) oxygen orbitals. Such a state is labeled σ_O in Table I

and in the following. The (HOMO-1) level is a more standard π state with weight on the oxygen (p_z) orbital and a delocalized benzene ring π molecular orbital. Within the G_0W_0(LDA), GW, and G_0W_0(HF$_{diag}$) approaches, the LDA HOMO σ_O state is pushed to a significantly lower energy, and the π state becomes the HOMO level. This level crossing brings the GW calculations in agreement with many-body quantum chemistry calculations, which all predict the π state to be the HOMO level. The same level crossing is observed in the case of uracil with the LDA HOMO and (HOMO-1) levels being σ_O and π states, respectively, while all GW results and quantum chemistry calculations predict a reverse ordering. Our interpretation is that the very localized σ_O state suffers much more from the spurious LDA self-interaction than the rather delocalized π state. Even though it would be wrong to reduce the dynamical GW self-energy operator to a self-interaction free functional, the GW correction certainly cures in part this well-known problem. The other bases, namely, guanine, adenine, and thymine, all show the correct π-character for the HOMO level.
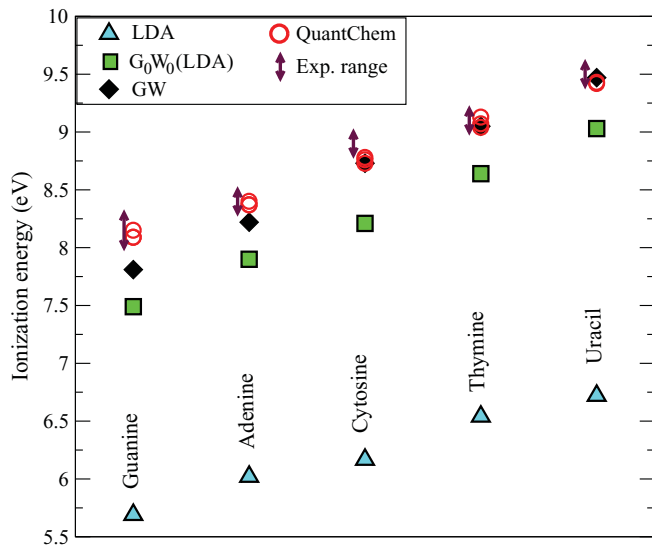
FIG. 2. (Color online) Ionization energies in eV. The vertical (maroon) error bars indicate the experimental range. Triangles (light blue): LDA values; (green) squares: G_0W_0(LDA) values; solid black diamond: GW values; (red) empty circles (QuantChem abbreviation): quantum chemistry, namely, CCSD(T), CASPT2, and EOM-IP-CCSD, values (see text).

FIG. 3. (Color online) Isodensity surface plot of the HOMO (σ_O), HOMO-1 (π), HOMO-2 (σ), and HOMO-3 (π') LDA Kohn-Sham eigenstates of cytosine. Within GW, the ordering of states becomes $\pi, \pi', \sigma_O, \sigma$ for HOMO to HOMO-3 (see text).

The HOMO to (HOMO-1) energy difference averages to 0.80 and 1.12 eV within CASPT2 and EOM-IP-CCSD, respectively. Clearly, the average LDA energy spacing of 0.22 eV is significantly too small. We find that the 0.77 eV G_0W_0(LDA) average value is close to the CASPT2 results, while the larger 1.29 eV GW result falls closer to the EOM-IP-CCSD energy difference. Averaging over all isomers, the experimental HOMO to (HOMO-1) energy spacing comes to 0.97 eV, in between the G_0W_0(LDA) or CASPT2 results and the GW or EOM-IP-CCSD values. Even though it is too early for final conclusions about the merits of the various approaches, it seems fair to state that the LDA value is significantly too small, and that the situation is improved significantly by the GW correction.

III. ELECTRONIC AFFINITIES

We conclude this study by exploring the electronic affinity (EA) of the nucleobases. They are provided in Table I as the negative sign of the LUMO Kohn-Sham energies. Experimental data for guanine are missing. Further, the CASPT2 and CCSD(T) results[5] are clearly larger (in absolute value) than the highest experimental estimates. While again part of the discrepancy may come from the presence of several tautomers in the gas phase, it certainly results as well from the fact that the electronic affinity is negative. A detailed discussion on the experimental difficulties in probing unbound states is presented in Ref. 6. Taking again the CCSD(T) and CASPT2 calculations[5] as a reference, the GW electronic affinities are quite satisfying, with an MAE of 0.18 eV. Such an agreement is rather impressive since the LDA electronic affinities show the wrong sign, with a discrepancy as compared to CASPT2
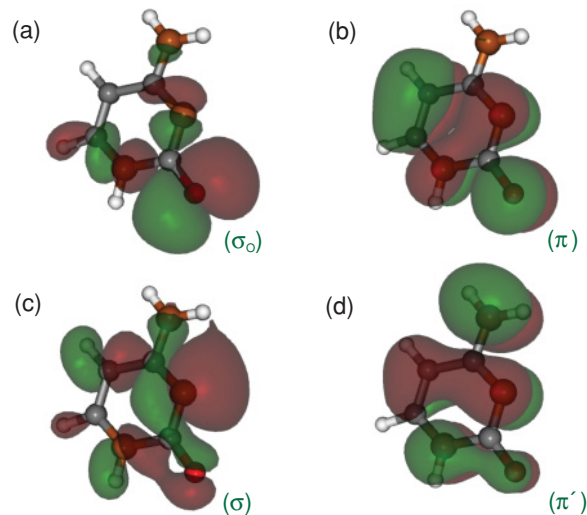
ranging from 2.9 eV to 3.6 eV. We observe that while the G_0W_0(LDA) EAs are smaller (in absolute value) than the quantum chemistry ones, the GW EAs are larger. This contrasts with the IE case where both G_0W_0(LDA) and GW values were smaller except for uracil (see Fig. 2). Similar to the quantum chemistry case, the GW values are found to systematically overestimate the experimental results. Further study is needed to understand such a discrepancy between theoretical and available experimental results.

IV. CONCLUSION

In conclusion, we have studied on the basis of *ab initio* GW calculations the ionization energies and electronic affinities of the DNA and RNA nucleobases, guanine, adenine, cytosine, thymine, and uracil. While a standard G_0W_0(LDA) calculation yields ionization energies that are 0.5 eV away from CCSD(T)/CASPT2 reference quantum chemistry calculations, self-consistency on the eigenvalues brings the agreement to an excellent 0.11 eV average absolute error. A simple G_0W_0 calculation starting from Hartree-Fock-like eigenvalues, avoiding the need for self-consistency, shifts the agreement to 0.22 eV. The possibility of bringing the calculated values to within 0.1–0.2 eV from state-of-the-art reference calculations with a scheme, the GW formalism, which allows us to treat both finite-size and extended systems with a N^4 scaling and permits us to obtain the full quasiparticle spectrum, paves the way to further studies of larger DNA strands and biological systems in general.

[1] A well-known example is that of the "replication" proteins that copy the nucleobases in DNA transcription and replication. See, e.g., A. Travers, in *DNA-Protein Interactions* (Chapman and Hall, London, 1993); P. B. Dervan, Science **232**, 464 (1983).

[2] H. Lodish *et al.*, in *Molecular Biology of the Cell*, 5th ed. (W. H. Freeman, New York, 2004).

[3] D. M. Close and K. T. Øhman, J. Phys. Chem. A **112**, 11207 (2008), and references therein.

[4] D. Roca-Sanjuan *et al.*, J. Chem. Phys. **125**, 084302 (2006).

[5] D. Roca-Sanjuan, M. Merchan, L. Serrano-Andres, and M. Rubio, J. Chem. Phys. **129**, 095104 (2008).

[6] K. B. Bravaya *et al.*, J. Phys. Chem. A **114**, 12305 (2010), and references therein.

[7] N. S. Hush and A. S. Cheung, Chem. Phys. Lett. **34**, 11 (1975).

[8] D. Dougherty *et al.*, J. Electron Spectrosc. Relat. Phenom. **13**, 379 (1978).

[9] K. W. Choi, J. H. Lee, and S. K. Kim, J. Am. Chem. Soc. **127**, 15674 (2005).

[10] A. B. Trofimov *et al.*, J. Phys. B **39**, 305 (2006).

[11] M. Schwell *et al.*, Chem. Phys. **353**, 145 (2008).

[12] I. L. Zaytseva *et al.*, J. Phys. Chem. A **113**, 15142 (2009).

[13] O. Kostko *et al.*, Phys. Chem. Chem. Phys. **12**, 2860 (2010), and references therein.

[14] L. Hedin, Phys. Rev. A **139**, 796 (1965).

[15] G. Strinati, H. J. Mattausch, and W. Hanke, Phys. Rev. Lett. **45**, 290 (1980); Phys. Rev. B **25**, 2867 (1982).

[16] M. S. Hybertsen and S. G. Louie, Phys. Rev. B **34**, 5390 (1986).

[17] R. W. Godby, M. Schlüter, and L. J. Sham, Phys. Rev. B **37**, 10159 (1988).

[18] G. Onida, L. Reining, and A. Rubio, Rev. Mod. Phys. **74**, 601 (2002).

[19] C. Rostgaard, K. W. Jacobsen, and K. S. Thygesen, Phys. Rev. B **81**, 085103 (2010).

[20] N. Dori, M. Menon, L. Kilian, M. Sokolowski, L. Kronik, and E. Umbach, Phys. Rev. B **73**, 195208 (2006).

[21] M. L. Tiago, P. R. C. Kent, R. Q. Hood, and F. A. Reboredo, J. Chem. Phys. **129**, 084311 (2008).

[22] P. Umari, G. Stenuit, and S. Baroni, Phys. Rev. B **79**, 201104(R) (2009).

[23] M. Palummo *et al.*, J. Chem. Phys. **131**, 084102 (2009).

[24] G. Stenuit *et al.*, Phys. Chem. Chem. Phys. **12**, 10817 (2010).

[25] X. Blase, C. Attaccalite, and V. Olevano, Phys. Rev. B **83**, 115103 (2011).

[26] P. C. Martin and J. Schwinger, Phys. Rev. **115**, 1342 (1959).

[27] K. Kaasbjerg and K. S. Thygesen, Phys. Rev. B **81**, 085102 (2010).

[28] P. H. Hahn, W. G. Schmidt, and F. Bechstedt, Phys. Rev. B **72**, 24545 (2005).

[29] J. M. Soler *et al.*, J. Phys. Condens. Matter **14**, 2745 (2002).

[30] It was shown in Refs. 19,25 for a large set of molecules that a standard double-zeta plus polarization (DZP) basis to expand the Kohn-Sham eigenstates already yields IE within ∼0.1 eV as compared to the much larger TZDP basis. This certainly indicates that the present TZDP basis is an extremely good basis for such calculations.

[31] For a recent analysis, see I. Cherkes, S. Klaiman, and N. Miseyev, Int. J. Quantum Chem. **109**, 2996 (2009), and references therein.

[32] M. S. Kaczmarski, Y. C. Ma, and M. Rohlfing, Phys. Rev. B **81**, 115433 (2010).

[33] We have tested the use of a larger and more diffuse even-tempered basis with decay coefficients $\alpha = (0.15, 0.32, 0.69, 1.48, 3.2)$ a.u. for C, O, and N atoms. We find that the GW ionization energies and electronic affinities change by 0.08 and 0.015 eV (MAE), respectively, with no systematic trend.

[34] The EOM-IE-CCSD calculations predict that the IP of the most stable guanine (G7K) and cytosine (C2b) isomers (the most abundant in the gas phase) are 0.14 and 0.08 eV larger than the IP of the G9K and C1 isomers we study. Our calculations yields differences of 0.08 and 0.04 eV respectively (GW value). For cytosine, our GW calculations predict that the largest IE is that of the (C3a) tautomer which lands 0.19 eV above that of (C1). This is consistent with the 0.2 eV experimental range, even though larger than the 0.12 eV difference reported in Ref. 6 (EOM-IE-CCSD calculations). Clearly, averaging over all isomers would bring our results in better agreement with experiment. We note, however, that such an average would require to know the abundance of each tautomer at the experimental temperature. This stands beyond the purpose of the present paper.

[35] Very similar results are obtained with the PBE functional. See J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).