Phys. Status Solidi B **249**, No. 2, 392–400 (2012) / **DOI** 10.1002/pssb.201100536

physica status solidi b
www.pss-b.com
basic solid state physics

Feature Article

Excited states of the green fluorescent protein chromophore: Performance of *ab initio* and semi-empirical methods

Marius Wanko[1], Pablo García-Risueño[2,3], and Angel Rubio*[,1]

[1] Nano-Bio Spectroscopy Group and ETSF Scientific Development Centre, Departamento de Física de Materiales, Universidad del País Vasco, Centro de Física de Materiales CSIC-UPV/EHU-MPC and DIPC Avenida de Tolosa 72, 20018 San Sebastián, Spain
[2] Instituto de Química Física Rocasolano, CSIC, C/Serrano 119, 28006 Madrid, Spain
[3] BIFI/Departamento de Física Teórica, Universidad de Zaragoza, C/Cerbuna 12, 50009 Zaragoza, Spain

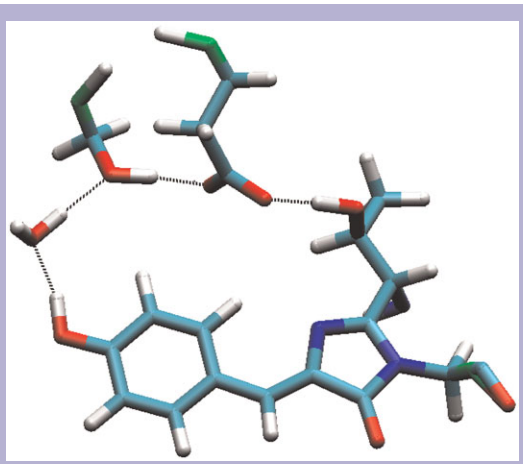
Received 5 September 2011, revised 19 October 2011, accepted 20 October 2011
Published online 23 December 2011

Dedicated to Thomas Frauenheim on the occasion of his 60th birthday

Keywords *ab initio*, chromophore, density functional theory, green fluoroscent protein, semi-empirical, time-dependent DFT

* Corresponding author: e-mail angel.rubio@ehu.es, Phone: +34 94301 8292, Fax: +34 94301 8390

The complex photophysical properties of fluorescent proteins give rise to a wide field of applications as markers in molecular biology. Understanding these properties is essential for rational genetic engineering of new fluorescent proteins. Here, theoretical models are required to support the interpretation of structural and spectroscopic experimental data. This requires the accurate and reliable prediction of excited-state features of the chromophore and its interactions with the protein matrix. Here, we compare calculations of absorption and emission energies of semi-empirical (OM2/MRCI, ZINDO/S, and TD-DFTB) and *ab initio* (SORCI, CC2, and TDDFT) approaches for the HBDI chromophore *in vacuo* and wild-type green fluorescent protein (GFP) using QM/MM models. We discuss the influence of electrostatic fields, the chromophore geometry, the size of the QM region, and methodological aspects, in particular charge-transfer states in TDDFT and the applicability of real-space TDDFT codes. We revisit previous opposing theoretical studies and benchmark gas-phase measurements.



Proton transfer wire of wild-type green fluorescent protein (wt-GFP).

1 Introduction Since the publication of the first X-ray crystal structures of the green fluorescent protein (GFP) in 1996 [1, 2], many theoretical studies have been performed to determine the protonation states and structural features of the chromophore and binding pocket that give rise to the fascinating spectroscopic features of this protein. A challenge that appears to remain until now is to accurately predict the absorption maximum of its chromophore, 4-hydroxybenzylidene-2,3-dimethyl-imidazolinone (HBDI) in different protonation states and different environments. Early INDO/S calculations [3] assigned the two absorption peaks at 477 and 397 nm to the zwitterionic and cationic form, respectively. This was later revised by OM2/MRCI calculations [4], which obtained 2.79 eV for the anionic and 3.71 eV for the neutral HBDI *in vacuo* and suggested a realistic scheme of GFP photophysics based on calculated

excited-state reaction barriers for photoisomerization. Afterwards, the first experimental absorption energy for anionic HBDI *in vacuo* was published (2.59 eV) [5]. A SAC-CI study [6] obtained considerably lower excitation energies on B3LYP geometries for anionic (2.22 eV) and neutral (3.33 eV) HBDI and also fluorescence energies based on the CIS/6-31G* geometry of the S_1 minimum (2.14 and 2.82 eV for anionic and neutral forms). A following CASPT2(12,11) study [7] reported anionic state excitation (emission) energies of the smaller HBI analog (CASSCF geometry) of 2.67 (2.44 eV), respectively, again closer to the OM2/MRCI results, but the corresponding Stokes shift is much larger than the experimental one for GFP [8, 9]. A proceeding CASSCF/CASPT2 study [10], however, emphasized the importance of solvent effects on chromophore geometry and charge distribution, in particular for anionic form. They observed an enhanced charge-transfer (CT), which may increase the potential of the protein to tune the absorption energy and suggests a combination of electrostatic and steric interactions responsible for the quenching of internal conversion in the protein.

Using a larger CAS(14,13) reference, Li et al. [11] reported MS-CASPT2 results similar to those of [7] and SAC-CI excitation energies that are lower (anionic 2.55 eV and neutral 3.67 eV). Also TD-B3LYP (2.57 eV) and ZINDO/S (2.60 eV) calculations [12] reproduced the experimental gas-phase reference for anionic HBDI. Filippi et al. [13] performed MS-CASPT2/cc-pVDZ calculation using a different zeroth-order Hamiltonian, which changes the result by up to 0.45 eV. They obtained significantly higher absorption energies for the anionic (2.92) and lower for the neutral HBDI (3.58 eV). They report a multiconfigurational character of S_1 in neutral HBDI and require a CAS(14,13) for converged CASPT2 results. They find a high anionic value also from EOM-CCSD, TDDFT, and QMC (2.93–3.04 eV) and question the experimental value of 2.59 eV. Rajput et al. [14] measured action spectra of neutral HBDI analogs that are augmented with a positive spectator charge. For their analog "E," where the spectator charge (N^+Me_3 fragment) is separated from the HBDI by methyl groups, they also presented CC2 and TD-B3LYP calculations, which are in excellent agreement with experiment. These calculations, however, were not in agreement regarding the shift between the analogs with spectator charge and the true neutral HBDI. The case of the neutral chromophore is interesting because most theoretical methods predict gas-phase absorption significantly blue-shifted compared to solution spectra (3.30–3.40 eV) [15]. The absorption maximum in GFP (3.10 eV) is even further red-shifted than in solution, whereas the anionic B form in GFP absorbs still around 2.60 eV, similar to many gas-phase calculations.

In this article, we revisit the performance of theoretical methods for application to HBDI *in vacuo* and to GFP in an QM/MM approach and discuss their dependence on basis sets, geometry, and aspects of QM/MM models, in particular equilibration via MD simulation and the extend of the QM region. We report SORCI and semi-empirical OM2/MRCI calculations, which are much less sensitive to the quality of the reference than CASPT2 calculations, and compare them with CC2 calculations to elucidate the role of static correlation. Further, we reinvestigate the shift between neutral HBDI and analog "E" of Rajput et al. to alleviate the interpretation of the experimental benchmark data. Finally, we test a real-space TDDFT code for its applicability to QM/MM problems.

2 Computational details For gas-phase calculations on HBDI, the geometry was optimized using HF, DFT (PBE and B3LYP functionals), MP2, and the CC2 method, as implemented in turbomole [16]. For geometry optimization, the def2-TZVP basis set was used. Ground-state single point calculations with the CEPA/1 and CCSD(T) methods [17] were performed with orca [18] using B3LYP-optimized geometries and the aug-cc-pVTZ basis set. Excited-state calculations were performed with TDDFT (PBE and B3LYP functionals) and CC2 (turbomole), SORCI [19] (orca [18]), and semi-empirical methods (see below). For excitation energy calculations, the employed basis sets are 6-311++G(d,p) (TDDFT) and Ahlrich's SV(P) basis augmented with diffuse functions from the aug-cc-pVDZ set (CC2 and SORCI). CC2 test calculations with larger basis sets (see Supporting Information) indicate convergence of the aug-SV(P) results within 0.05 eV. In the SORCI calculations, we use a CAS(2,2) reference in the preliminary MRDDCI2 calculations to produce approximate natural orbitals (ANO). The final MRDDCI3+Q calculations are carried out using a reference which covers 13 ANO's of π character and configurations selected ($T_{\mathrm{pre}} = 10^{-3}$) from a RAS(14: 6 2/2/5 2) space. The unselected configurations ($T_{\mathrm{sel}} = 10^{-6}$) are treated with perturbation theory.

Semi-empirical OM2/MRCISD calculations [20] were performed with version 6.1 of the MNDO99 program using active orbital windows of (26,24) and (28,28) for QM1 and QM2 regions, respectively. The active orbitals are selected from a preliminary CI run. The reference space comprised the HF and the HOMO–LUMO singly excited determinants. ZINDO/S and TDDFT calculations with the CAM-B3LYP [21] and LC-BLYP [22] functionals were performed with Gaussian03 [23]. For ZINDO/S excitation energies, all valence orbitals were included in the CIS expansion. TD-DFTB [24, 25] calculations were performed using the mio parameter set [26].

Real space TDDFT calculations were performed with octopus [27] (revision 6194) using Troullier and Martins type pseudopotentials for core electrons and a grid spacing of 0.175 Å. The simulation box consists of atom-centered spheres with a fixed radius, the influence of which is discussed in detail in Section 3.3.

GFP calculations employ the hybrid quantum mechanical molecular mechanical (QM/MM) approach using the CHARMM force field [28] for the MM region. For molecular dynamics (MD) simulations and geometry optimization, the SCC-DFTB method [26] with third-order correction [29]

was used as the QM method. For excited-state calculations, the point charges of the CHARMM force field are used in the external potential of the QM Hamiltonian (electronic embedding). The protein setup is based on the 1.85 Å X-ray crystallographic structure of Van Thor et al. [30] (PDB code 1W7S). Standard protonation states are assumed, except for the chromophore and Glu222 in the I and B state. The screening effect of bulk solvent on the surface-exposed charged amino acids was accounted for by the charge scaling scheme proposed by Dinner et al. [31] The stability of the setup was tested by extended MD simulations (1 ns for each setup). In the geometry optimizations and MD simulations, heavy atoms separated by more than 10 Å from any chromophore atom where harmonically constrained to their crystallographic positions with force constants based on their B-factors. In a range between 10 and 5 Å, the force constants were scaled down to zero, that is, within 5 Å from the chromophore no constraints were applied.

Despite high resolution, the 1W7S structure does not resolve two additional water molecules that are found in the 1EMA structure [1] of the S65T mutant (W383 and W320). These connect to the three water molecules above the chromophore that form a hydrogen bonded network (HBN) connecting Glu222, Thr203, Gln69, Ser65, and the backbone of Val68. During MD simulations, we find one of these water molecule repeatedly escaping into the cavity that is occupied by W383 and W320. As suggested by Helms et al. [32], we include these two water molecules in our setup, which yields a stable HBN and water positions in good agreement with the X-ray structure. The same cavity provides space for a third water molecule. When added, we obtain a very similar HBN that is also consistent with the X-ray structure. These two alternative setups are referred to as ''add2w'' and ''add3w'' below.

The structure of the B-form was built by manually transferring the proton from Tyr66 to Glu222 and re-orienting Thr203 to form a hydrogen bond with the phenolate group of the anionic chromophore (see Fig. 1). This structure is stable in our MD simulations.

We use two different QM regions. A minimal QM1 region (36 atoms) that contains the three chromophore residues and backbone C and O of Phe64. A larger QM2 region (73 atoms) adds the HBN that connects the chromophore phenol with the proton acceptor Glu222 (Thr203, water 223, and Ser205) and waters 80, 220, and 245 (see Fig. 1). The larger QM region allows for inter-residual charge exchange that may affect the excitation energy. It also yields a more accurate structure from the QM/MM optimization.

3 Results and discussion

3.1 Proton affinities As Table 1 shows, the proton affinities of the glutamic acid side chain (Glu) and the HBDI chromophore are calculated in very good agreement by DFT and post-HF methods. We use the CEPA/1 method as a reference, which yields results that reside within those of CCSD and CCSD(T). In case of Glu, the CCSD(T) value is

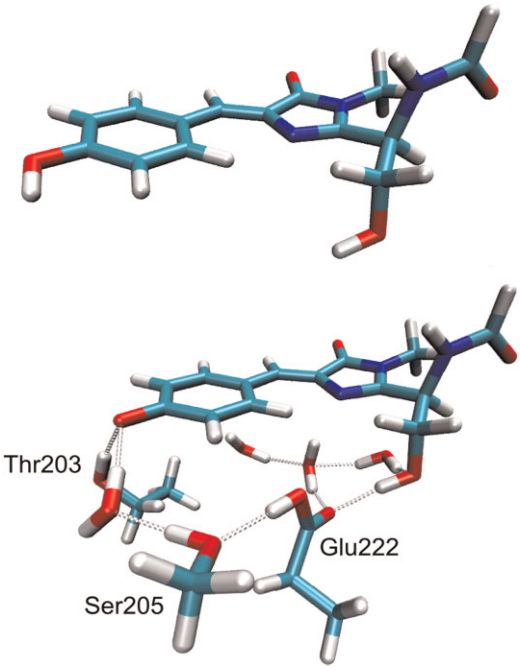


Figure 1 (online color at: www.pss-b.com) QM regions and different protonation states of HBDI in the GFP setup. Upper: QM1, A form. Lower: QM2, B form.

15.39 eV and is framed by the MP2 and the CEPA/1 results. OM2 underestimates the PA of both molecules, but the resulting difference compares well, due to error cancelation. Adding π correlation to the OM2 calculation of HBDI does not alter the PA. DFTB and the PBE functional underestimate the PA of HBDI by 0.4 eV (10 kcal/mol). In systems with an excess charge that can delocalize within a conjugated π-system, GGA's tend to exaggerate this delocalization and the charged species is described too stable with respect to the neutral one. In the case of a protonated Schiff base, for example, this leads to an overestimation of the PA by 9 kcal/mol, as the charged species is the protonated cation [33]. A natural population analysis, however, shows that the negative excess charge of the anionic form is already delocalized at the HF level. PBE, therefore, yields an almost even distribution between the phenolate and the imidazole rings, similar to HF or CASSCF (see Ref. [10]). As shown by Altoé et al., charge localization increases in solution. As we discuss below, this applies also to the protein environment,

Table 1 Proton affinities (eV).

	DFTB	OM2	PBE	B3LYP	MP2	CC2	CEPA/1
HBDI	14.49	13.43	14.44	14.63	14.64	14.43	14.92
Glu	15.42	14.00	15.24	15.33	15.27	15.17	15.47
Δ	0.93	0.57	0.80	0.70	0.63	0.74	0.55

The aug-cc-pVTZ basis was used for single point calculations and the def2-TZVP set was used for geometry optimization.

Table 2 HBDI gas-phase absorption energy (eV): Method dependency.

	TD-DFTB	ZINDO	OM2/MRCI	TD-PBE	TD-B3LYP	CAM[b]	LC[b]	CC2	SORCI	Exp[c]
anionic	2.88	2.61	2.61	2.93	3.07	3.14	3.20	2.88	2.57	2.59
neutral	3.17	3.48	3.80	3.20	3.47	3.72	3.96	3.72	3.57	
neutral$^+$ (E)[a]	3.15	3.21	3.52	3.12	3.35	3.53	3.71	3.42	3.17	3.35
$\Delta_{\text{anionic–neutral}}$	−0.29	−0.87	−1.19	−0.27	−0.40	−0.58	−0.76	−0.84	−1.00	
$\Delta_{\text{anionic–E}}$	−0.28	−0.60	−0.96	−0.19	−0.28	−0.39	−0.51	−0.54	−0.60	−0.76

The applied basis sets are 6-311++G(d,p) (TDDFT) and aug-SV(P) (CC2, SORCI). Geometries were optimized using B3LYP and the def2-TZVP basis set. [a]"Model E" analog of neutral HBDI from Ref. [14]. [b]Adiabatic TDDFT with the CAM-B3LYP or the LC-BLYP functional. [c]Refs. [5, 14].

where a larger variation of the predicted PA's should be expected.

3.2 Excitation energies If noted otherwise, our discussion of excitation energies always refers to the lowest valence $\pi - \pi^*$ transition, which dominates the low-energy absorption spectrum. In the case of TDDFT with pure LDA/GGA functionals and TD-DFTB, there are both Rydberg and CT excitations that appear at lower energies than the spectroscopic state, but mostly with low oscillator strength. The states of different methods are assigned in terms of their molecular orbitals and transition weights (TDDFT), amplitudes (CC2), or configuration coefficients (CI-expansions). The oscillator strength is used to cross-check this assignment.

As Table 2 shows, the absolute excitation energy of the HBDI chromophore *in vacuo* is reasonably well reproduced by the applied semi-empirical and *ab initio* methods, with errors in the usual range. Merely the overestimation of the anionic HBDI by TD-B3LYP (0.48 eV) and TD-CAM-B3LYP (0.65 eV) appears remarkable. More interesting is the prediction of the shift due to protonation of the chromophore. Here, we find quite some variation between the methods ranging from 0.2 to 1.0 eV. No experimental data of neutral HBDI *in vacuo* is available, as action spectroscopy requires charged species. The closest experimental benchmark system is a neutral HBDI augmented with a "spectator charge" shielded by methyl groups (model E analog in Ref. [14]). The shift between this analog and anionic HBDI is best reproduced by SORCI and ZINDO/S, which slightly underestimate the shift, and OM2/MRCI, which slightly overestimates it. The CC2 shift is smaller than the SORCI one (29% smaller than exp) while TDDFT underestimates it by a factor of 4 (PBE), 2.7 (B3LYP), 1.9 (CAM-B3LYP), or 1.5 (LC-BLYP). Obviously, HF exchange reduces the error of the adiabatic DFT functionals, but still insufficiently in the case of CAM-B3LYP, which employs 60% HF exchange in the long-range limit. Merely the LC-BLYP functional, which established 100% HF exchange in the long-range limit, yields a shift that comes close to the CC2 one, but overestimates the absolute excitation energies. TD-DFTB reproduces well the TD-PBE results, but the shift from the E analog to the anionic HBDI agrees with TD-B3LYP. The comparison between the true neutral and anionic HBDI shows the same trend but more pronounced.

Our SORCI results for the neutral HBDI are consistent with recent CASPT2/cc-pVDZ calculations using large CASSCF(13,14) references [13]. For anionic HBDI, they agree better with the earlier CASPT2 [7] and with experiment than with the recent CASPT2 study, which reported a higher value (2.92 eV) using a new zero-order Hamiltonian. On the other hand, the deviation from experiment in the latter work can also be accounted for by the lack of diffuse basis functions in the CASPT2 calculations. We obtain a blueshift of 0.33 eV (SORCI) for both protonation states when omitting the diffuse functions, the effect is less drastic for CC2 (see Supporting Information). Note that also our CC2 results indicate a rather high (−0.84 eV) shift between anionic and neutral protonation states and that we reproduce exactly the CC2 result for model E of [14], but not that for the true neutral HBDI, which appears to be a misprint in Ref. [14].

Further, we do not find any multi-reference character for the S_1 state in neutral HBDI, as indicated by the CASSCF calculations of Ref. [13]. The D1 diagnostic from the CC2 calculation is small (0.075), even smaller than that of the anionic HBDI S_1 state (0.108). The same applies to the norm of the T2 amplitude, also an empirical indicator for multi-reference character, which is 0.078 for neutral and 0.092 for anionic HBDI. In our SORCI calculations, we obtain the same high coefficient for the HOMO–LUMO singly excited configuration and the same reference weight for both protonation states. Also the S_0–S_1-averaged natural orbital occupations from the MRDDCI2 wave functions are similar.

Table 3 shows how calculated excitation energies depend on the method used for geometry optimization. PBE geometries agree remarkably well with CC2 optimized ones, whereas B3LYP geometries result in slightly higher excitation energies. HF is not recommended, resulting in significantly blue-shifted spectra. The same applies for CASSCF (see Li et al. [11]).

Depending on the treatment of electron correlation, the resonance structures are differently weighted, which is reflected in the bond lengths of conjugated systems. As these weights are very different for the ground and excited state, this variation in geometry translates directly into a variation in excitation energy. In the case of linear polyenes, a simple linear correlation between the bond length alternation

Table 3 HBDI gas-phase absorption energy (eV): Geometry dependency.

geometry	GBLA	TD-PBE	TD-B3LYP	CC2	SORCI
anionic HBDI					
DFTB	−0.820	2.83	2.96	2.77	2.44
PBE	−0.811	2.89	3.02	2.84	2.51
CC2	−0.822	2.88	3.02	2.84	2.49
B3LYP	−0.797	2.93	3.07	2.88	2.57
HF	−0.767	3.00	3.16	2.99	2.72
neutral HBDI					
DFTB	−0.212	3.06	3.32	3.50	3.39
PBE	−0.209	3.13	3.38	3.58	3.46
CC2	−0.209	3.15	3.41	3.61	3.48
B3LYP	−0.190	3.20	3.47	3.72	3.57
HF	−0.147	3.37	3.67	3.92	3.80

The applied basis sets are 6-311++G(d,p) (TDDFT), aug-SV(P) (CC2, SORCI). Geometries were optimized with the specified method and the TZVP (DFT) and def2-TZVP (HF, CC2) basis sets.

(difference between the average single and double bond length) and the excitation energy is found. In the case of HBDI, the situation is complicated by the partial aromaticity of the phenol ring. Nonetheless, a correlation between Raman C=O stretching modes and excitation energy was found [34]. Laino et al. [35] suggested a linear combination of bond lengths with fitted coefficients, which correlates with the excitation energy. We use a more simple and less empirical coordinate by projecting the cartesian gradient of the excitation energy onto bond lengths. This generalized bond length alternation (GBLA) is defined in the Supporting Information (online at: www.pss-b.com). As shown in Fig. 2, a strong linear correlation is obtained for all gas-phase CC2 calculations. The weight vector has been normalized to 1 for the plot, that is, the GBLA has formally the angstrom unit and the different slopes for anionic and neutral chromophore reflect the different sensitivity of the excitation energy to changes in the bond lengths. The same GBLA coordinates are expected to correlate also with electrostatically induced shifts in different protein environments, but we do not have enough data to investigate such correlation.

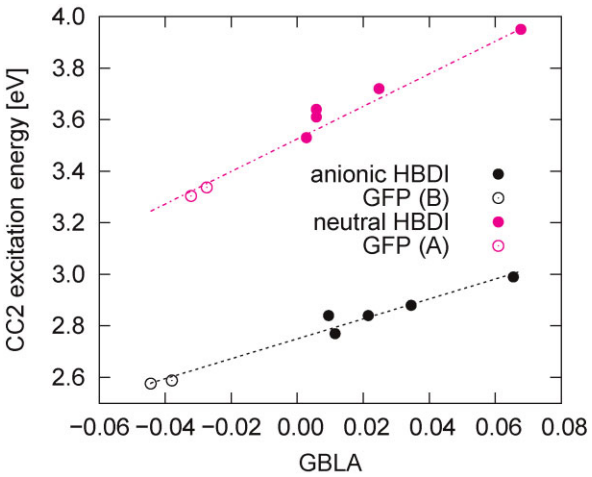


Figure 2 (online color at: www.pss-b.com) Correlation between GBLA and CC2 excitation energy. The plot contains the data from Table 3 and GFP QM/MM calculations (QM1) without point charges.

We calculated vertical excitation energies for eight different structures (Table 4): QM1 or QM2 region, add2w or add3w setup, QM/MM optimization before (opt) or after (MD) 1 ns of MD simulation and quenching. For the A form, structural variations between the setups cause only small shifts in absorption (QM1: ±0.03 eV and QM2: ±0.05 eV). Interestingly, these shifts are not found in the TDDFT results. For the B form, the resulting shifts are predicted very large (±0.08 to ±0.10 eV) by multi-reference methods, whereas they are not reproduced by CC2 and TDDFT. The extension of the QM region results in a redshift. OM2/MRCI does not reproduce this because the active orbital window is insufficient for the QM2 region. Due to the many HF levels from the added QM fragments, the active window contains fewer π orbitals than the QM1 calculation, which results in a blueshift.

The averaged absolute excitation energies agree very well with experiment, except for the B3LYP values for the B form. This is essentially due to the underestimation of the A–B shift, which leads to the same error as for HBDI *in vacuo* (compare Table 2). Using the QM1 region, SORCI and OM2/MRCI reproduce the A–B shift within the experimental accuracy, whereas the CC2 shift is slightly too small, like for HBDI in the gas phase.

For the add2w (MD) structure, we also report the "vacuo" value, which is obtained by omitting the point charges. This yields the electrostatic shift ΔES. Multireference methods and CC2 describe a bathochromic shift for the A form and a hypsochromic shift for the B form. The latter is partially preserved when omitting the point charges in the QM2 calculation and hence partially caused by residues therein. Interestingly, the ΔES shift is strongly underestimated by TDDFT methods, TD-PBE even inverts the shift for the A form.

An explanation for this behavior can be found by analyzing the difference dipole moments, that is, the CT associated with the excitation (Table 5). *In vacuo*, the $S_1 \leftarrow S_0$ excitation carries very little CT. SORCI, OM2/MRCI, and CC2 predict an increased CT character for this excitation in the protein, in agreement with CASSCF calculations with polarizable continuum models [10]. The obtained difference dipole moment for the B state agrees fairly well with the experimental estimate from Stark spectroscopy at 77 K of 6 Debye [9]. Compared with the retinal chromophore in rhodopsins, this CT is still small and explains the moderate influence of the protein electrostatic environment. The PBE functional, like other GGA's strongly underestimates the CT and even predicts an inverted direction from the imidazole to the phenolate moiety in most cases. In cases where the $\pi - \pi^*$ state mixes with CT

Table 4 Vertical transition energies (eV) of GFP.

	SORCI QM1	CC2 QM1	OM2 QM1	B3LYP QM1	PBE QM1	SORCI QM2	CC2 QM2	OM2 QM2	B3LYP QM2	PBE QM2	Exp[a]
A (neutral *cis*) form											
add2w (opt)	3.20	3.18	3.20	3.15	2.97	2.97	3.07	3.32	3.07	2.75	3.10–3.14
add2w (MD)	3.14	3.16	3.15	3.16	2.96	2.87	3.02	3.24	3.06	–[c]	
vacuo	(3.35)	(3.30)	(3.44)	(3.16)	(2.87)	(3.04)	(3.15)	(3.38)	(3.06)	(2.75)	
ΔES	−0.21	−0.14	−0.29	0.00	0.09	−0.17	−0.13	−0.14	0.00	–[c]	
add3w (opt)	3.15	3.13	3.13	3.14	2.97	2.94	3.06	3.29	3.04	2.74	
add3w (MD)	3.19	3.19	3.19	3.16	2.97	2.98	3.11	3.35	3.06	2.75	
vacuo	(3.38)	(3.34)	(3.47)	(3.18)	(2.89)	(3.15)	(3.23)	(3.43)	(3.09)	(2.75)	
ΔES	−0.19	−0.15	−0.28	−0.02	0.08	−0.16	−0.12	−0.08	−0.03	0.00	
average	3.17	3.17	3.17	3.15	2.97	2.94	3.06	3.30	3.06	2.75	
emission[b]	2.75	2.89	2.73	2.87	2.74						2.64–2.95
Stokes shift	0.40	0.27	0.42	0.29	0.22						
B (anionic *cis*) form											
add2w (opt)	2.58	2.76	2.55	2.97	2.83	2.40	2.70	2.64	2.91	2.76	2.59–2.63
add2w (MD)	2.68	2.78	2.66	2.98	2.81	2.55	2.71	2.74	2.91	2.75	
vacuo	(2.44)	(2.59)	(2.45)	(2.72)	(2.66)	(2.50)	(2.70)	(2.79)	(2.86)	(2.60)	
ΔES	0.24	0.19	0.21	0.26	0.15	0.05	0.01	−0.05	0.05	0.15	
add3w (opt)	2.62	2.76	2.57	2.96	2.80	2.48	2.71	2.67	2.92	–[c]	
add3w (MD)	2.73	2.79	2.71	2.98	2.80	2.59	2.72	2.78	2.92	–[c]	
vacuo	(2.42)	(2.58)	(2.44)	(2.73)	(2.74)	(2.48)	(2.67)	(2.77)	(2.85)	–[c]	
ΔES	0.32	0.22	0.27	0.25	0.06	0.11	0.05	0.01	0.07	–[c]	
average	2.65	2.77	2.62	2.97	2.81	2.51	2.71	2.70	2.92	2.76	
emission[b]	2.56	2.65	2.54	2.86	2.70						2.57
Stokes shift	0.11	0.13	0.12	0.12	0.11						
shift A → B	−0.52	−0.39	−0.54	−0.18	−0.16	−0.44	−0.35	−0.60	−0.14	0.01	0.52

[a]Refs. [8, 9, 36, 37]. [b]Geometry from QM/MM CC2/aug-SV(P) excited-state optimization starting from the add2w (MD) setup. [c]The spectroscopic $\pi - \pi^*$ state is strongly mixed-up with CT states.

states (not shown), GGA's predict large unphysical difference dipole moments that are not along the molecular axis.

In Table 4, we report also emission energies. The geometries of the fluorescent state were obtained from excited-state optimization at the CC2 level using a fixed MM environment. Based on this structure, all applied methods predict a rather small Stokes shift (0.1 eV) for the B form, in agreement with measurements at 77 K (0.06 eV) and RT (0.14 eV) [38]. A previous CASPT2 study on CASSCF/CHARMM QM/MM geometries obtained a larger value of 0.29 eV. For the A form, a much larger Stokes shift of 0.3–0.4 eV is predicted. This is within the range of experimental estimates (2.64–2.95 eV) [9, 8], but a reliable experimental reference is missing due to the short lifetime of the A^* species. A previous SAC-CI study obtained a slightly larger value of 0.53 eV [39].

Table 5 Difference dipole moments $|\mu(S_1) - \mu(S_2)|$ for the lowest $\pi - \pi^*$ excitation (debye).

	TD-PBE	TD-B3LYP	CC2	SORCI	OM2/MRCI
neutral HBDI (vacuo)	−1.2[b]	0.4	0.8	3.7	1.8
anionic HBDI (vacuo)	2.7	1.7	1.6	2.3	1.5
GFP A (neutral)[a]	−2.3[b]	1.0	5.0	5.4	5.8
GFP B (anionic)[a]	−0.6[b]	0.5	3.8	4.6	5.1

[a]add2w (MD) GFP setup with QM1 region. [b]The minus sign indicates transfer of electron density from the imidazole to the phenolate moiety.

3.3 QM/MM on a real-space grid In this section, we test the octopus real space code for QM/MM applications. Instead of the basis set, the numerical accuracy of real space calculations is limited by the size of the simulation box and the grid spacing. The simulation box is defined as a superposition of atom-centered spheres with a radius R. Octopus offers two ways to calculate an TDDFT absorption spectrum. For the most common Casida formalism of linear density functional response theory, the number of virtual states n_{virt} is not limited by the number of basis functions and has to be specified as input parameter. Alternatively, the spectrum can be obtained from a time propagation of the density. The ground-state density is perturbed by a potential of the form $x_i\delta(t - t_0)$, where x_i defines the polarization direction of the light. The absorption cross-section is obtained from the Fourier transformation of the time-dependent dipole moment. Figure 3 illustrates the effect of the two parameters R and n_{virt} for a protonated Schiff base (CNH_4^+) in presence of a point charge $(-1.2\,e)$ located 2.5 Å

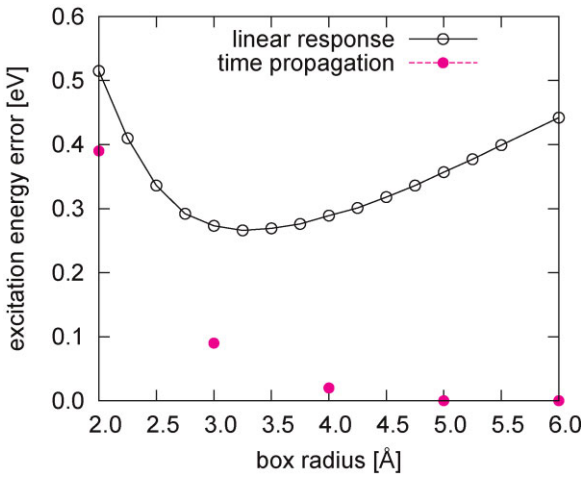


Figure 3 (online color at: www.pss-b.com) Convergence of the $\pi - \pi^*$ excitation energy in CNH_4^+ w.r.t. the simulation box radius. The error refers to an LCAO basis set limit calculation. The SVWN functional was used.

from the nitrogen. The Casida-type calculation does not converge to the LCAO limit because a constant n_{virt} is used. Both box size and n_{virt} must be increased to approach the LCAO limit. With increasing box size, the shape and order of the virtual states changes and convergence with n_{virt} can change to the worse. As excited states have in general a more extended density than the ground state, the finite-box error always produces a blue shift, which systematically reduces with the parameter R.

When point charges are introduced in a real space QM/MM calculation, the problem arises that electrons can leave the QM region and be bound in regions of low electrostatic potential. In LCAO, this is prevented by the exponential decay of the atom-centered basis functions. In order to investigate this problem in a realistic QM/MM situation, we compare LCAO with octopus results for GFP in the A and B forms using different QM regions and box sizes. As Fig. 4

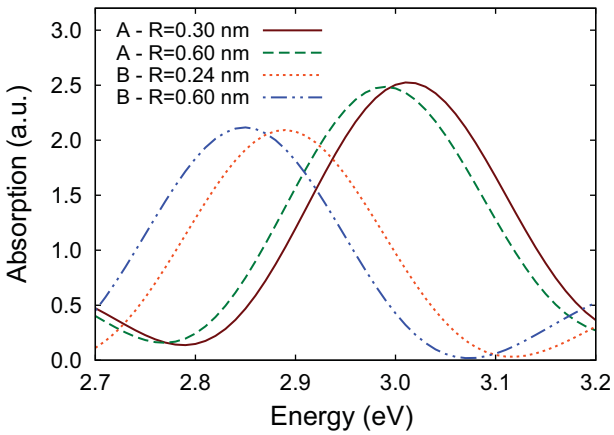


Figure 4 (online color at: www.pss-b.com) Effect of the box radius R on the $\pi - \pi^*$ peak position of GFP in A and B form.

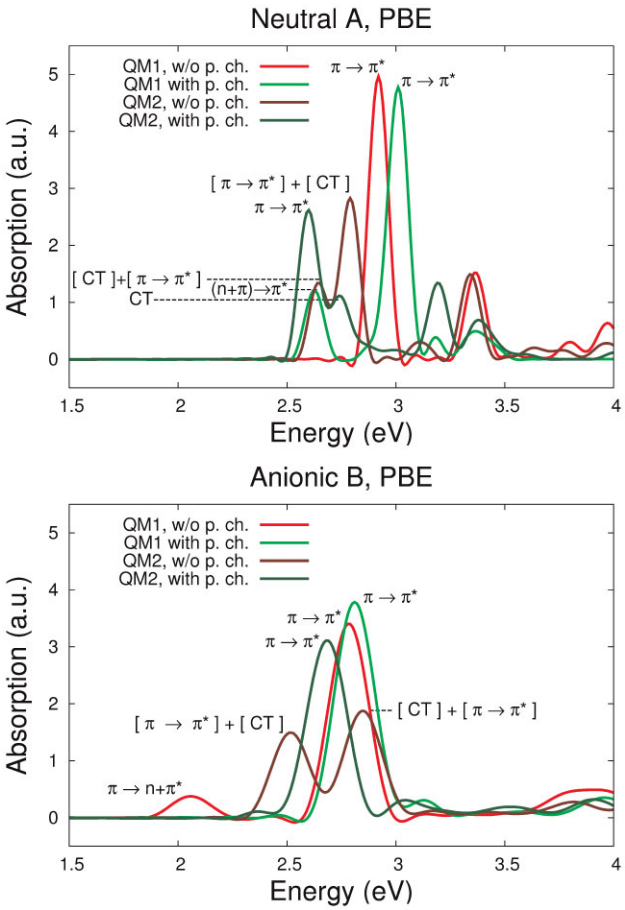


Figure 5 (online color at: www.pss-b.com) Assigned peaks in the time-propagation TD-PBE spectrum of A (upper) and B (lower). Calculations with and without point charges are shown using the QM1 and QM2 regions. The total propagation time was 39.5 fs for the A and 19.7 fs for the B form.

shows, the problem does not occur for the S_1 excitation when increasing the box radius to 6 Å and the peak positions converge as expected. An acceptable error is obtained for values between 2.4 and 3.0 Å, the latter is used in the proceeding calculations.

Figure 5 shows the time-propagation spectra of TD-PBE using either the QM1 or QM2 region and the effect of the electrostatic embedding. The peaks were assigned to states of the corresponding linear-response calculation and were characterized based on their single-particle transition weights and underlying molecular orbitals. In general, the LCAO results (Table 4) are well reproduced. Like in the LCAO case, many CT states occur in the region of the $\pi - \pi^*$ states and below. Where no mixing occurs, the peak positions agree, reproducing the small blue shift due to the point charges and the underestimated A–B shift. In many cases, the $\pi - \pi^*$ state couples to CT states, which strongly perturbs the spectrum due to redistribution of oscillator strength. For the QM2 calculation of the B form without point charges, for example, the absorption band appears splitted. The

coupling of $\pi - \pi^*$ and CT states occurs in an unpredictable fashion and can drastically change due to small structural fluctuations or the use of different GGA functionals. In the Supporting Information, we provide the spectra from BLYP and LDA functionals, which show the same trends.

4 Conclusions For both gas phase and GFP environment, our SORCI, OM2/MRCI, and CC2 calculations agree closely with experiment for anionic and neutral protonation states. Our results for the anionic HBDI *in vacuo* corroborate the gas-phase measurements of Nielsen et al. [5] and we disagree with Filippi et al. that the vertical excitation energy should be in the range 2.92–3.04 eV. We rather see the CASPT2 results critical for the large influence of the zeroth-order Hamiltonian (0.45 eV). Also the EOM-CCSD method, which yields 3.04 eV in Ref. [13] tends to overestimate excitation energies of conjugated systems [40].

Our results suggest that neutral HBDI *in vacuo* is blueshifted by 0.3–0.4 eV compared to the analog "E" investigated by Rajput et al. [14]. The neutral chromophore in GFP on the other hand is more red-shifted than in any solvent [15] and absorbs 0.4–0.5 eV lower in energy than HBDI *in vacuo*. On the other hand, the anionic form, which shows much larger solvent shifts absorbs rather similar *in vacuo* and in the protein. Nonetheless, the electrostatic influence of the protein matrix is rather limited in both cases, inline with the SAC-CI study of Hasegawa et al. [39]. Although the protein environment enhances the CT character of the $\pi - \pi^*$ excitation in both A and B forms to ca. 6 debye, the electrostatic influence of the protein is significantly smaller than in rhodopsins. Larger spectral changes are achieved rather by hydrogen bonds, structural changes, and genetic or chemical modification of the chromophore.

DFT (PBE) yields ground-state geometries in good agreement with CC2. TDDFT, however, using the adiabatic approximation with LDA/GGA functionals drastically underestimates the shifts between neutral and anionic form, both *in vacuo* and in GFP. Further, it underestimates the response to the external field of the protein matrix due to an underestimated CT upon excitation. TD-PBE even inverts the CT in some cases, leading to an increase of charge density on the phenol ring. Furthermore, a plentitude of artificial CT states appears in the TDDFT spectrum of the neutral A form in GFP, which can lead to perturbed unphysical spectra when these CT states mix with the spectroscopically allowed state. The use of hybrid functionals avoids these problems with CT states, whereas the spectral shift between neutral and anionic forms is still severely underestimated. Range-separated hybrids that employ the full nonlocal HF exchange in the long-range limit seem to minimize the error. It is, therefore, likely that the problem is caused by the spatial locality of the density functional rather than the locality in time (adiabatic approximation). Apart from the shortcomings of the density functionals, the octopus real space code is applicable to QM/MM problems and reproduces the results of the LCAO calculations.

Acknowledgements Funding by the Spanish MICINN ("Juan de la Cierva," FIS2010-65702-C02-01, PIB2010US-00652, and/CSD2010-00044/), the "Grupos Consolidados UPV/EHU del Gobierno Vasco" (IT-319-07), European Research Council Advanced Grant DYNamo (ERC-2010-AdG No. 267374), Ikerbasque and by the Barcelona Supercomputing Center, "Red Espanola de Supercomputacion," SGIker ARINA (UPV/EHU), and ACI-Promociona (ACI2009–1036).

Supporting Information Provided are data regarding basis set convergence, parameters of the GBLA, and additional TDDFT spectra. Available online at www.pss-b.com.

References

[1] M. Ormo and S. J. Remington, Science **273**, 1392–1395 (1996).
[2] F. Yang, L. G. Moss, and G. N. Phillips, Jr., Nature Biotechnol. **14**, 1246–1251 (1996).
[3] A. A. Voityuk, M. E. Michel-Beyerle, and N. Rösch, Chem. Phys. Lett. **272**, 162–167 (1997).
[4] W. Weber, V. Helms, J. A. McCammon, and P. W. Langhoff, Proc. Natl. Acad. Sci. USA **96**, 6177–6182 (1999).
[5] S. B. Nielsen, A. Lapierre, J. U. Andersen, U. V. Pedersen, S. Tomita, and L. H. Andersen, Phys. Rev. Lett. **87**(22), 228102 (2001).
[6] A. K. Das, J. Y. Hasegawa, T. Miyahara, M. Ehara, and H. Nakatsuji, J. Comput. Chem. **24**, 1421–1431 (2003).
[7] M. E. Martin, F. Negri, and M. Olivucci, J. Am. Chem. Soc. **126**(17), 5452–5464 (2004).
[8] A. F. L. Creemers, C. H. W. Klaassen, P. H. M. Bovee-Geurts, R. Kelle, U. Kragl, J. Raap, W. J. de Grip, J. Lugtenburg, and H. J. M. de Groot, Biochemistry **38**(22), 7195–7199 (1999).
[9] M. Chattoraj, B. A. King, G. U. Bublitz, and S. G. Boxer, Proc. Natl. Acad. Sci. USA **93**, 8362–8367 (1996).
[10] P. Altoé, F. Bernardi, M. Garavelli, G. Orlandi, and F. Negri, J. Am. Chem. Soc. **127**, 3952–3963 (2005).
[11] X. Li, L. W. Chung, H. Mizuno, A. Miyawaki, and K. Morokuma, J. Phys. Chem. B **114**, 1114–1126 (2010).
[12] S. B. Wan, S. S. Liu, G. J. Zhao, M. D. Chen, K. Han, and M. T. Sun, Biophys. Chem. **129**(2–3), 218–223 (2007).
[13] C. Filippi, M. Zaccheddu, and F. Buda, J. Chem. Theory Comput. **5**(8), 2074–2087 (2009).
[14] J. Rajput, D. B. Rahbek, L. H. Andersen, T. Rocha-Rinza, O. Christiansen, K. B. Bravaya, A. V. Erokhin, A. V. Bochenkova, K. M. Solntsev, J. Dong, J. Kowalik, L. M. Tolbert, MÅ. Petersen, and M. B. Nielsen, Phys. Chem. Chem. Phys. **11**(43), 9996–10002 (2009).
[15] J. Dong, K. M. Solntsev, and L. M. Tolbert, J. Am. Chem. Soc. **128**(37), 12038–12039 (2006).
[16] TURBOMOLE V6. 1 2009, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989–2007, TURBOMOLE GmbH, since 2007; available from http://www.turbomole.com.
[17] F. Neese, A. Hansen, F. Wennmohs, and S. Grimme, Acc. Chem. Res. **42**(5), 641–648 (2009).
[18] F. Neese, ORCA – An *Ab Initio*, DFT and Semiempirical electronic structure package, Version 2.7 - Revision 0, University Bonn, June 2009.
[19] F. Neese, J. Chem. Phys. **119**, 9428–9443 (2003).
[20] W. Weber and W. Thiel, Theor. Chem. Acc. **103**, 495–506 (2000).

[21] T. Yanai, D. P. Tew, and N. C. Handy, Chem. Phys. Lett. **393**, 51–57 (2004).
[22] H. Iikura, T. Tsuneda, T. Yanai, and K. Hirao, J. Chem. Phys. **115**, 3540–3544 (2001).
[23] Gaussian 03, Revision B.04, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, J. A. Pople, Gaussian, Inc., Pittsburgh PA, 2003.
[24] T. A. Niehaus, J. Mol. Struct.: Theochem. **914**(1–3), 38–49 (2009).
[25] T. A. Niehaus, S. Suhai, F. Della Sala, P. Lugli, M. Elstner, G. Seifert, and T. Frauenheim, Phys. Rev. B **63**, 085108 (2001).
[26] M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, T. Frauenheim, S. Suhai, and G. Seifert, Phys. Rev. B **58**, 7260–7268 (1998).
[27] A. Castro, H. Appel, M. Oliveira, C. Rozzi, X. Andrade, F. Lorenzen, M. Marques, E. Gross, and A. Rubio, Phys. Status Solidi B **243**, 2465–2488 (2006).
[28] A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, III, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, and M. Karplus, J. Phys. Chem. B **102**, 3586–3616 (1998).
[29] M. Gaus, Q. A. Cui, and M. Elstner, J. Chem. Theory Comput. **7**(4), 931–948 (2011).
[30] J. J. van Thor, G. Y. Georgiev, M. Towrie, and J. T. Sage, J. Biol. Chem. **280**(39), 33652–33659 (2005).
[31] A. R. Dinner, X. Lopez, and M. Karplus, Theor. Chem. Acc. **109**, 118 (2003).
[32] V. Helms, T. P. Straatsma, and J. A. McCammon, J. Phys. Chem. B **103**, 3263–3269 (1999).
[33] M. Wanko, M. Hoffmann, T. Frauenheim, and M. Elstner, J. Comput.-Aided Mol. Des. **20**, 511–518 (2006).
[34] A. F. Bell, X. He, R. M. Wachter, and P. J. Tonge, Biochemistry **39**(15), 4423–4431 (2000).
[35] T. Laino, R. Nifos, and V. Tozzini, Chem. Phys. **298**, 17–28 (2004).
[36] R. Y. Tsien, Annu. Rev. Biochem. **67**, 509–544 (1998).
[37] J. T. M. Kennis, D. S. Larsen, N. H. M. van Stokkum, M. Vengris, J. J. van Thor, and R. van Grondelle, Proc. Natl. Acad. Sci. USA **101**(52), 17988–17993 (2004).
[38] K. Brejc, T. K. Sixma, P. A. Kitts, S. R. Kain, R. Y. Tsien, M. Ormo, and S. J. Remington, Proc. Natl. Acad. Sci. USA **94**(6), 2306–2311 (1997).
[39] J. Y. Hasegawa, K. Fujimoto, B. Swerts, T. Miyahara, and H. Nakatsuji, J. Comput. Chem. **28**(15), 2443–2452 (2007).
[40] M. Schreiber, M. R. Silva-Junior, S. P. A. Sauer, and W. Thiel, J. Chem. Phys. **128**(13), 134110 (2008).

