JCTC Journal of Chemical Theory and Computation

ARTICLE

pubs.acs.org/JCTC

Bathochromic Shift in Green Fluorescent Protein: A Puzzle for QM/MM Approaches

Claudia Filippi,*[,†] Francesco Buda,*[,‡] Leonardo Guidoni,[§] and Adalgisa Sinicropi*[,∥]

[†]Faculty of Science and Technology and MESA+ Institute of Nanotechnology, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands

[‡]Leiden Institute of Chemistry, Leiden University, P.O. Box 9502, 2300 RA Leiden, The Netherlands

[§]Dipartimento di Chimica, Ingegneria Chimica e Materiali, Universitá dell'Aquila, Via Campo di Pile, 67100 L'Aquila, Italy

[∥]Dipartimento di Chimica, Universitá di Siena, Via A. De Gasperi 2, 53100 Siena, Italy

Supporting Information

ABSTRACT: We present an extensive investigation of the vertical excitations of the anionic and neutral forms of wild-type green fluorescent protein using time-dependent density functional theory (TDDFT), multiconfigurational perturbation theory (CASPT2), and quantum Monte Carlo (QMC) methods within a quantum mechanics/molecular mechanics (QM/MM) scheme. The protein models are constructed via room-temperature QM/MM molecular dynamics simulations based on DFT and are representative of an average configuration of the chromophore−protein complex. We thoroughly verify the reliability of our structures through simulations with an extended QM region, different nonpolarizable force fields, as well as partial reoptimization with the CASPT2 approach. When computing the excitations, we find that wave function as well as density functional theory methods with long-range corrected functionals agree in the gas phase with the extrapolation of solution experiments but fail in reproducing the bathochromic shift in the protein, which should be particularly significant in the neutral case. In particular, while all methods correctly predict a shift in the absorption between the anionic and neutral forms of the protein, the location of the theoretical absorption maxima is significantly blue-shifted and too close to the gas-phase values. These results point to either an intrinsic limitation of nonpolarizable force-field embedding in the computation of the excitations or to the need to explore alternative protonation states of amino acids in the close vicinity of the chomophore.

1. INTRODUCTION

In previous decades, the development and application of intrinsically fluorescent proteins[1] has launched a revolution in molecular cell biology. Fluorescent proteins are routinely employed to dynamically visualize cellular processes in living organisms and are responsible for significant advances in fluorescence spectroscopy, enabling for instance bioimaging techniques with subdiffraction resolution.[2] Green fluorescent protein (GFP) is the prototype of this class of proteins and, together with its mutants, one of the most widely used fluorescent markers in cell biology.[3] Because of its technological relevance, GFP is very well characterized experimentally and represents therefore a perfect playground to investigate the effectiveness of commonly used as well as novel quantum mechanics/molecular mechanics (QM/MM) approaches.

In this paper, we employ a variety of electronic structure approaches to compute the vertical excitations of the neutral A and anionic B forms of wild-type GFP. These are the protonation states of the chromophore responsible for the two room-temperature absorption peaks at 398 nm (3.12 eV) and 478 nm (2.59 eV), respectively.[4] For these two forms, we construct the protein models via room-temperature QM/MM molecular dynamics simulations based on density functional theory (DFT) and thoroughly test their reliability by performing simulations with extended QM regions, different nonpolarizable force fields, and even partially relaxing the QM component with the multiconfigurational perturbation theory (CASPT2) approach. The absorption spectra of the neutral and anionic forms are then computed with the use of time-dependent density functional theory (TDDFT), CASPT2, and quantum Monte Carlo (QMC) methods within a QM/MM scheme.

We find that, in the gas phase, wave function and density functional methods with long-range corrected functionals agree well with the extrapolation of solution experiments[5] but, when introducing the protein environment, do not yield the desired bathochromic shift, which should be particularly significant in the neutral case. In particular, we obtain theoretical excitations between 2.8 and 3.2 eV for the B form and in the range of 3.4−3.6 eV for the A form. Therefore, while all methods predict a shift in the absorption between the anionic and neutral protein forms, the location of the theoretical absorption maxima is significantly blue-shifted and too close to the gas-phase values. Our extensive tests on the structures indicate that the source of the problem does not lie in the computational details of the construction within the standard QM/MM prescription. Moreover, since the theoretical techniques to compute the excitation spectrum appear to be reliable in the gas phase and display an overall agreement among each other in the protein, we also rule out that the origin of the problem lies in our choice of QM method to compute the excitation. Therefore, our results point at two possible sources of error still largely unexplored, namely, that

Received: October 4, 2011
Published: November 21, 2011

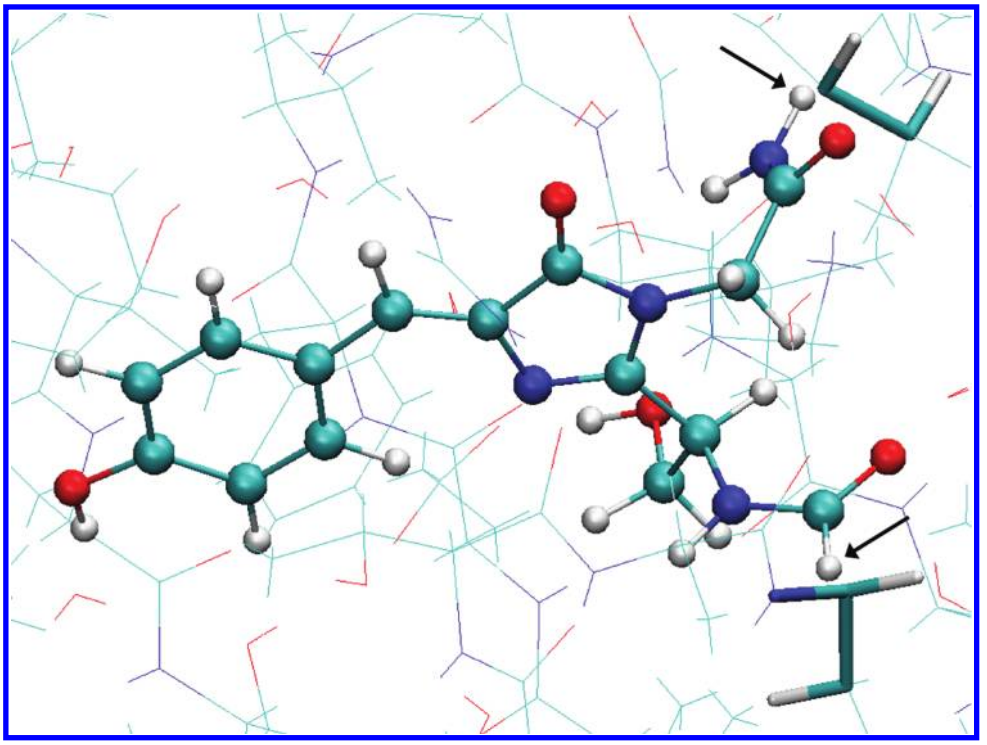

Figure 1. Partitioning between the QM and MM regions shown for the A form of wild-type GFP. The QM/MM cuts are at the $C_{OOH}-C_{\alpha}$ bond of Phe64 and the $N-C_A$ bond of Val68. The arrows indicate the hydrogen-link atoms between the QM and the MM. The MM residues closest to the QM atoms are represented with cylinders. This image is produced with VMD.[37]

some amino acids in the protein have a different protonation than commonly accepted or that the protein environment must be described beyond nonpolarizable force fields in the computation of the excitations.

In section 2, we describe the computational details and focus on the construction of the protein models for the neutral A and anionic B as well as alternative nonstandard forms. The results for the vertical excitations are presented in section 3. Finally, we discuss the structural models and the theoretical excitations in section 4 and conclude in section 5.

2. METHODS AND MODELS

2.1. Computational Details. The Amber suite of programs[6] in combination with the Amber 03 force field[7] and the TIP3P water model[8] is used in the setup of the protein models and for the subsequent MM equilibrations.

The CPMD code[9] is used for all ab initio molecular dynamics simulations described by quantum mechanics. The Gromos96 code[10] with the Amber 03 force field is employed to describe MM atoms within the QM/MM approach.[11,12] The Kohn−Sham orbitals are expanded in a plane-wave basis set with a kinetic energy cutoff of 70 Ry. We employ Martins−Troullier norm-conserving pseudopotentials[13] and the Perdew, Becke, and Ernzerhof (PBE) generalized gradient approximation for the exchange-correlation functional.[14,15] The size of the QM box is such that the distance between periodic replicas is about 8 Å in every direction. The annealing and the room-temperature Car−Parrinello molecular dynamics simulations are performed with a time step of 0.075 fs and a value of 400 au for the fictitious electronic mass in the Car−Parrinello Lagrangian.

To compute the excitation energies, we employ TDDFT, the complete active space self-consistent field (CASSCF) method with its perturbative extension (CASPT2), and QMC methods.

For the TDDFT and TDDFT/MM calculations, we use the Gaussian 09 code[16] with default convergence parameters. The BLYP,[17,18] B3LYP,[19] CAM-B3LYP,[20] LC-BLYP,[21] and the LC-ωPBE[22] functionals are employed together with the aug-cc-pVDZ basis set. The effect of using the cc-pVDZ, cc-pVTZ, and aug-cc-pVTZ basis sets is also tested. In the TDDFT/MM calculations, we use the Amber 99 point charges for consistency with the CASPT2 calculations.

The complete active space calculations are performed using MOLCAS 7.2.[23] In the CASPT2 calculations, we employ the default IPEA zero-order Hamiltonian[24] unless otherwise stated and indicate if an additional imaginary constant level shift[25] is added to the Hamiltonian. In the CASPT2 calculations, we do not correlate as many of the lowest σ orbitals as there are heavy atoms in the molecule. For all models, we use the Cholesky decomposition of the two-electron integrals[26,27] with the threshold of 10^{-8}. Default convergence criteria are used for all calculations. The CASPT2/MM calculations are performed with the MOLCAS 7.2 package coupled with a modified version of the MM package Tinker 4.2.[28] The Amber 99 force field[29] and TIP3P water model are employed.

The program package CHAMP[30] is used for the QMC calculations. We employ scalar-relativistic energy-consistent Hartree−Fock pseudopotentials[31] where the carbon and nitrogen 1s electrons are replaced by a nonsingular s-nonlocal pseudopotential and the hydrogen potential is softened by removing the Coulomb divergence. We use the Gaussian basis sets[31] specifically constructed for our pseudopotentials. In particular, we employ the cc-pVDZ basis augmented with diffuse s and p functions[32] on the heavy atoms and denoted as D+. Different Jastrow factors are

used to describe the correlation with different atom types, and for each atom type, the Jastrow factor consists of an exponential of the sum of two fifth-order polynomials of the electron−nucleus and the electron−electron distances, respectively.[33] The determinant components of the ground and excited states are obtained in state-average CASSCF calculations performed with the program GAMESS(US).[34] The final CAS expansions are expressed on the CASSCF natural orbitals and may be truncated with an appropriate threshold on the configuration state functions (CSF) coefficients for use in the QMC calculations. The union of the surviving CSFs for the states of interest is kept in the final Jastrow−Slater wave functions. The Jastrow correlation factor and the CI coefficients are optimized by energy minimization in a state-averaged fashion within variational Monte Carlo (VMC).[35] The pseudopotentials are treated beyond the locality approximation,[36] and an imaginary time step of 0.04 or 0.06 au is used in the diffusion Monte Carlo (DMC) calculations. The QMC/MM calculations are performed using the electrostatic coupling scheme as in the QM/MM approach employed in the CPMD code. For the starting trial wave function, CASSCF/MM calculations are performed within GAMESS(US), and the divergence of point charges at the origin is removed to resemble the embedding scheme employed within CPMD.

2.2. Protein Models. For all forms of GFP, the protein models are initially equilibrated via MM simulations at room temperature. The hydrogens and water molecules are first equilibrated while constraining the positions of the heavy atoms. Subsequently, a MM isothermal and isobaric simulation is performed at 300 K and 1 atm, keeping only the chromophore coordinates fixed. Finally, the structure is refined in a simulated annealing run within QM/MM, where the chromophore is treated quantum mechanically and allowed to relax. The boundary between the QM and the MM regions is set through the single $C_{OOH}-C_{\alpha}$ bond of Phe64 and the single $N-C_A$ bond of Val68, as shown in Figure 1. The possible protonation states of the QM chromophore model are shown in Figure 2.

2.2.1. The Neutral and Anionic Forms. The starting structure for the construction of the neutral A form is the X-ray structure[38] at 1.90 Å resolution (entry 1GFL in the Protein Data Bank[39]). On the basis of the most likely hydrogen-bonding configuration, the histidine residues numbered 25, 148, 181, 199, and 217 are protonated at their δ nitrogen while the remaining histidine residues are protonated at their ε nitrogen. For the protonation of the glutamic acids, particular attention is given to Glu222, which is deprotonated in the A form and will be the proton acceptor in the proton shuffle between the neutral and the anionic forms.

To simulate solution conditions, the protein is placed at the center of a cubic MM simulation box surrounded by 12 Å of water molecules in each direction. Moreover, counterions are added to achieve physiologic conditions with a saline concentration of 0.15 M and to ensure that the cell is neutral. The total simulation box contains around 70 000 atoms. A short isothermic and isobaric MM equilibration at a fixed chromophore is performed with a time step of 0.5 fs until the temperature fluctuations are less than 5% and the density is constant within 2%. A quenched structure is then obtained via an energy minimization procedure. Starting from the structure obtained in the MM equilibration, we perform a short QM/MM run of about 1 ps at 300 K, followed by an annealing period of about 0.5 ps, where the velocities are rescaled at each step by a factor in the range of 0.99−0.999. The temperature of the final cooled system is less than 0.5 K.

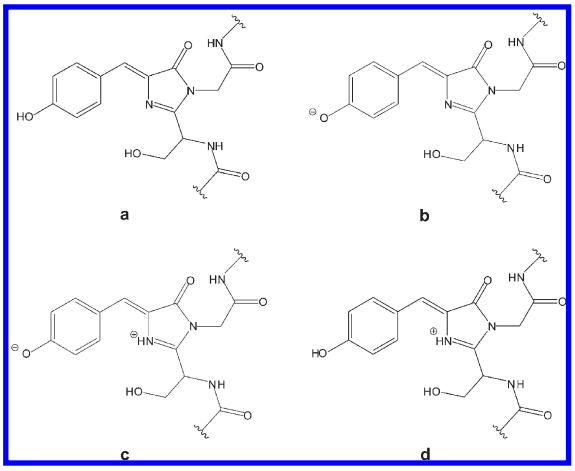

Figure 2. Possible protonation states of the QM chromophore model of GFP: (a) neutral, (b) anionic, (c) zwitterionic, and (d) cationic.

The anionic I form is obtained by deprotonating the chromophore of our equilibrated A form and adding the proton to Glu222. A QM/MM simulated annealing procedure is then performed via an isothermic molecular dynamics simulation followed by a velocity rescaling as described for the A form.

Since no crystal structure of the wild-type anionic B form is available, the model for this form is obtained starting from the crystallographic structure of a mutant stabilizing the anionic chromophore with an environment believed to be close to that of wild-type GFP in the B form. Following the theoretical work of Nifosì and Tozzini,[40] we start from the crystallographic structure[41] of mutant S65T (code 1EMG in the Protein Data Bank). This mutant differs from the wild-type structure (1GFL) for the substitution of Ser65 with a threonine (Thr65). To restore the wild-type structure, we simply undo this mutation by reintroducing Ser65 and adjusting its orientation to form a hydrogen bond with Glu222, which is now protonated. Starting from this structure, we add 10 Å of water molecules in each direction and the appropriate number of counterions, leading to a system of about 31 000 atoms. Note that the number of atoms is approximately half that of the A form, as the starting crystal structures of the A and B forms are a dimer and a monomer, respectively. We then proceed with the classical MM equilibration and the subsequent QM/MM dynamics and annealing as described for the A form.

2.2.2. Nonstandard Forms. We also investigate the stability of alternative protonation states of the chromophore and surrounding environment via additional QM/MM simulations. We test the possible stability of a solvated hydronium in proximity of the chromophore following the experimental suggestions of ref 42 as well as the existence of the zwitterionic and cationic forms in the presence of a protonated and deprotonated Glu222 residue. In these simulations, the QM region is extended to include the anionic chromophore; the side chains of Thr203, Ser205, and Glu222; and the five water molecules closest to the chromophore. One water molecule is protonated when investigating the stability of a hydronium.

Our simulations at room temperature indicate that a hydronium is not stable. In particular, in the presence of a protonated Glu222, the additional proton migrates to the nitrogen of the imidazolinone ring, leading to the formation of a zwitterionic chromophore and a neutral water. If we hypothesize that the

Table 1. CASPT2 Vertical Excitation Energies (eV) of the Minimal Neutral Chromophore Computed with the ANO-S-PVDZ Basis on the Ground-State DFT/BLYP Geometry[a]

CAS(n,m)	N	state	osc. str.	MS-PT2	SS-PT2	CASSCF	$\|\Delta\mu\|$
2,2	2	2	0.96	3.56	3.55	4.79	2.8
4,4	2	2	0.94	3.82	3.79	4.72	4.2
6,6[b]	3	3	0.75	3.87	4.24	5.37	5.3
8,8[b]	3	3	0.80	3.99	4.24	5.46	5.5
10,10[b]	3	3	0.86	3.99	4.06	5.44	5.3
12,11[b]	3	3	0.68	3.69	3.87	5.24	3.3
	4	3, 4	0.72, 0.23	3.38	3.80	5.24	3.1
14,13[b]	4	4	0.82	3.61	3.73	5.35	3.6
	5	4, 5	0.80, 0.16	3.49	3.75	5.32	3.6
16,14[b]	4	4	0.89	3.60	3.64	5.31	4.6
	5	4 ,5	0.85, 0.15	3.49	3.67	5.27	4.5
				IPEA0			
16,14[b]	4	4	0.89	3.17	3.23	5.31	4.6

[a] Different CAS expansions of n electrons in m active π orbitals are used, and the total number of π electrons in the reference is 16. The SA calculations are on the lowest N roots, and the brighter states with the corresponding CASSCF oscillator strength (osc. str.) are listed. The standard IPEA shift is used in the multistate (MS) and single-state (SS) CASPT2 calculations, and the CASSCF and single-state CASPT2 excitations are relative to the state with the highest oscillator strength. We also report the CASSCF modulus of the dipole difference ($|\Delta\mu|$) in Debye. In the last line, we also list the CASPT2 excitations for the CAS(16,14) expansion computed without IPEA shift (IPEA0). [b] Imaginary shift of 0.1 au.

proton on the hydronium is originating from Glu222 and repeat the simulation with an anionic Glu222, the proton initially forming the hydronium migrates back to Glu222 during the dynamics, yielding a neutral Glu222 and a water molecule.

Finally, we find that the zwitterionic form is only stable in the presence of a protonated Glu222, while, in a doubly protonated GFP chromophore (on the phenol and imidazolinone rings) with a deprotonated Glu222, the proton jumps during the dynamics from the nitrogen of the imidazolinone to Glu222, neutralizing the carboxyl group. Therefore, the proton shuffling between a zwitterionic and a cationic form is not possible, while the existence of a zwitterionic-neutral equilibrium can be postulated only in the presence of a neutral Glu222 in both forms of the chromophore.

3. VERTICAL EXCITATION ENERGIES

The vertical excitation energies of the A and B forms of GFP are computed within CASPT2/MM, TDDFT/MM, and QMC/MM on the ground-state DFT/PBE geometry equilibrated within CPMD/Amber03. The polarization effect of the protein environment on the excitation is accounted for by point charges at the positions obtained in the QM/MM equilibration.

3.1. CASPT2 Results. We compute here the multistate (MS) and single-state (SS) CASPT2 vertical excitation energies of the neutral A and anionic B forms using the ANO-S-PVDZ basis. For comparison, we also calculate the excitation energies of the anionic and neutral minimal (p-HBI) and methyl-terminated (p-HBDI) models in the gas phase using the same basis and the DFT/BLYP geometries obtained in ref 35. All excitations are computed using the standard IPEA Hamiltonian, and for the

Table 2. CASPT2 Vertical Excitation Energies (eV) of the Methyl-Terminated Neutral Chromophore Computed with the ANO-S-PVDZ Basis on the Ground-State DFT/BLYP Geometry[a]

CAS(n,m)	N	state	osc. str.	MS-PT2	SS-PT2	CASSCF	$\|\Delta\mu\|$
2,2	2	2	0.94	3.46	3.46	4.82	2.2
4,4	2	2	0.95	3.72	3.70	4.78	3.4
6,6[b]	2	2	0.93	3.84	3.82	5.08	4.8
8,8[b]	3	3	0.83	4.12	4.18	5.54	4.9
10,10[b]	3	3	0.87	4.01	3.99	5.50	4.4
12,11[b]	4	3, 4	0.41, 0.26	3.06	4.13	5.18	2.5
	5	3, 4	0.41, 0.25	2.92	4.15	5.20	2.8
14,13[b]	4	4	0.57	3.34	3.92	5.24	1.1
	5	4, 5	0.57, 0.34	3.14	3.92	5.23	1.2
16,14[b]	4	4	0.64	3.33	3.82	5.21	1.7
	5	4, 5	0.63, 0.33	3.15	3.85	5.20	2.0
				IPEA0			
16,14[b]	4	4	0.64	2.75	3.35	5.21	1.7

[a] The total number of π electrons on the chromophore is 16. For the notation, see the caption of Table 1. [b] Imaginary shift of 0.1 au.

CAS(16,14) expansion, we also report the excitations computed without the IPEA shift to allow for a comparison with previous CASPT2 calculations.

The computation of the vertical excitations of the neutral models in the gas phase and in the protein is considerably more problematic than the CASPT2 calculations for the anionic counterparts. In Table 1, we begin with the minimal neutral model, where we can observe signs of the complications that will become evident in the neutral methyl-terminated model. In particular, the brightest state is no longer the second one at the CASSCF level, and whenever the oscillator strength is non-negligible on more than one state, we observe a marked difference between the single- and multistate CASPT2 excitations. In fact, when increasing the expansion from CAS(10,10) to CAS(12,11), the oscillator strength decreases on the bright state and becomes non-negligible on an additional state while the multi- and single-state excitations differ by as much as 0.4 eV. In CAS(12,11), we cannot stabilize the lone-pair orbital of the phenolic oxygen but must include the nitrogen lone pair on the imidazolinone while omitting the bonding and antibonding π orbitals on the benzene. Further increasing the expansion to correlate all 16 π electrons in the reference in a minimal active space appears to cure the problem. The oscillator strength in the bright state increases, and the difference between the multi- and single-state CASPT2 excitations is only 0.2 eV when the additional state with non-negligible oscillator strength is included in the multistate calculation. Overall, we note that the single-state CASPT2 excitation is well behaved and displays a smooth convergence as a function of the size of the expansion. On the contrary, the multistate excitation is particularly sensitive to the inclusion of states other than the bright one, which leads to a significant divergence of the single- and multistate values for small CAS expansions.

The CASPT2 excitations of the methyl-terminated chromophore are listed in Table 2. Since the minimal and methyl-terminated models only differ in the termination, one would expect rather similar excitations in the two cases. This expectation appears fulfilled when inspecting the single-state CASPT2 excitations, which display a similar convergence as a function of

Table 3. CASPT2/Amber99 Vertical Excitation Energies (eV) of the A Form Computed with the ANO-S-PVDZ Basis[a]

CAS(n,m)	N	state	osc. str.	MS-PT2	SS-PT2	CASSCF	$\|\Delta\mu\|$
2,2	2	2	1.11	3.29	3.26	4.50	3.7
4,4	2	2	0.89	3.54	3.49	4.36	5.1
6,6	2	2	0.83	3.72	3.69	4.59	8.8
8,8	2	2	0.83	3.72	3.69	4.60	8.9
10,10[b]	2	2	1.03	3.56	3.53	4.52	7.5
12,11[b]	2	2	1.10	3.56	3.56	4.21	8.3
14,13[b]	2	2	0.99	3.57	3.56	4.16	9.8
16,14[b]	2	2	1.00	3.53	3.53	4.16	9.9
				IPEA0			
16,14[b]	2	2	1.00	3.14	3.13	4.16	9.9

[a] The ground-state DFT/PBE geometry equilibrated within CPMD/Amber03 is used. For the notation, see the caption of Table 1. [b] Imaginary shift of 0.1 au.

Table 4. CASPT2 Vertical Excitation Energies (eV) of the Minimal Anionic Chromophore Computed with the ANO-S-PVDZ Basis on the Ground-State DFT/BLYP Geometry[a]

CAS(n,m)	osc. str.	MS-PT2	SS-PT2	CASSCF	$\|\Delta\mu\|$
2,2	1.29	2.68	2.67	3.99	0.3
4,4	1.36	2.96	2.96	3.45	1.4
6,6	1.19	2.93	2.90	3.30	2.7
8,8	1.15	2.96	2.90	3.33	3.4
10,10	1.17	2.96	2.85	3.26	4.2
12,11	1.39	2.82	2.82	3.42	1.5
14,13	1.35	2.80	2.79	3.16	1.0
16,14	1.34	2.78	2.77	3.13	0.9
		IPEA0			
16,14	1.35	2.57	2.56	3.13	0.9

[a] All SA calculations are on the two lowest roots. For the notation, see the caption of Table 1.

Table 5. CASPT2 Vertical Excitation Energies (eV) of the Methyl-Terminated Anionic Chromophore Computed with the ANO-S-PVDZ Basis on the Ground-State DFT/BLYP Geometry[a]

CAS(n,m)	osc. str.	MS-PT2	SS-PT2	CASSCF	$\|\Delta\mu\|$
2,2	1.30	2.65	2.64	4.01	0.1
4,4	1.37	2.92	2.92	3.48	1.7
6,6	1.19	2.92	2.88	3.33	3.4
8,8	1.15	2.95	2.89	3.37	3.8
10,10	1.18	2.94	2.83	3.31	4.6
12,11	1.18	2.91	2.82	3.28	4.4
14,13[b]	1.35	2.79	2.78	3.18	1.4
16,14[b]	1.35	2.77	2.76	3.15	1.7
		IPEA0			
16, 14[b]	1.35	2.34	2.33	3.15	1.7

[a] All SA calculations are on the two lowest roots. For the notation, see the caption of Table 1. [b] Imaginary shift of 0.1 au.

the CAS size and are only blue-shifted by 0.2 eV with respect to the CAS(16,14) values of the minimal model. On the other hand, the multistate excitations differ from the single-state values even more dramatically than for the minimal model. The CAS(12,11) expansion is over a set of orbitals with the same character as in the minimal models and marks the appearance of a second state with non-negligible oscillator strength while the oscillator strength of the bright state drops to half the value of the CAS(10,10) state. The difference between multi- and single-state CASPT2 excitations is now more than 1 eV. Further increasing the size of the CAS expansion ameliorates the situation, but correlating all π electrons on the chromophore in a minimal CAS(16,14) is not sufficient to reduce the difference between single- and multistate CASPT2 results, which remains larger than 0.6 eV. Since the presence of the methyl groups might be responsible for iper-conjugation effects, we also correlate the electrons of the methyl carbons by performing CAS(18,15) and CAS(20,16) calculations over four and five states (data not shown). However, we find that iper-conjugation plays no significant role, as the excitations of these larger active spaces are identical to the CAS(16,14) results.

We believe that the large difference between multi- and single-state results is due to the inability of the minimal CAS(16,14) expansion to provide a satisfactory description of all states included in the multistate calculation. While increasing the size of the expansion to a minimal CAS(16,14) was sufficient to adjust the multistate value in the minimal model, this active space does not appear to be sufficient in the presence of methyl-termination. Unfortunately, due to the high computational cost, we were not able to add a sufficiently high number of virtual orbitals in the CAS(16,14) expansion to further stabilize the multistate results. Since the chemical behavior of the methyl-terminated model should be similar to the minimal one where the multistate excitation approaches the single-state value for larger expansions, we consider the single-state CASPT2 excitations to be more reliable estimates, and we will take them as our reference data. Moreover, it has been previously observed that large deviations of the multistate CASPT2 excitations from their single-state counterparts should be taken with caution and most likely are an indication of the failure of the multistate approach.[43,44]

The CASPT2 excitations of the A form in the presence of the MM charges are collected in Table 3. Differently from the gas-phase models, it is always possible to obtain a bright state with a large oscillator strength and with excitations that are similar in the single- and multistate calculations even though the orbitals in the various CAS expansions closely resemble the ones in vacuo. We note however that also for the A form we find energetically close CASSCF minima characterized by significantly different multi- and single-state excitations and oscillator strength spread over many states. For instance, even for the largest CAS(16,14) expansion, one can obtain a minimum where the CASSCF excitation is higher by 0.5 eV and the multistate CASPT2 value as low as 3.12 eV.

In Tables 4, 5, and 6, we report the CASPT2 excitations of the anionic mimimal, methyl-terminated, and B-form models, respectively. The behavior of the excitations is rather similar for the three anionic models, which are characterized by a bright state being clearly dominant and a close agreement between the single- and multistate results. For all anionic models, the CAS(12,11) expansion is constructed by omitting the lone-pair orbital on the imidazolinone nitrogen and the π bonding and antibonding orbitals on benzene.

For all models and the chemically relevant CAS(16,14) expansion correlating all π electrons on the chromophore, we

Table 6. CASPT2/Amber99 Vertical Excitation Energies (eV) of the B Form Computed with the ANO-S-PVDZ Basis[a]

CAS(n,m)	osc. str.	MS-PT2	SS-PT2	CASSCF	$\|\Delta\mu\|$
2,2	1.30	2.73	2.69	4.06	1.6
4,4	1.34	2.96	2.96	3.57	4.4
6,6	1.35	2.98	2.98	3.69	4.6
8,8	1.21	3.02	2.95	3.60	6.4
10,10	1.17	2.91	2.87	3.45	6.4
12,11	1.36	2.84	2.84	3.52	5.0
14,13[b]	1.28	2.84	2.84	3.26	5.9
16,14[b]	1.30	2.82	2.82	3.20	5.8
		IPEA0			
16,14[b]	1.30	2.38	2.38	3.20	5.8

[a] All SA calculations are on the two lowest roots. The ground-state DFT/PBE geometry equilibrated within CPMD/Amber03 is used. For the notation, see the caption of Table 1. [b] Imaginary shift of 0.1 au.

compute the CASPT2 excitations also without the IPEA shift and therefore with the old definition of zero-order Hamiltonian. The resulting excitations are significantly red-shifted by as much as 0.4−0.5 eV for all models. The only system where the deviation is smaller and equal to 0.2 eV is the minimal anionic model, which indicates that the minimal CAS on all π electrons captures the most relevant correlation effects for this model.

Finally, we also list the modulus of the difference between the CASSCF dipole moments of the ground and the bright excited state for all models and CAS expansions. The magnitude of the dipole moment difference is closely related to the charge transfer character of the excitation, which we estimate as the change in the Mulliken charges induced by the excitation (see Table 1 of Supporting Information). For the anions, the modulus of the dipole difference of the CAS(16,14) states increases from 0.9 to 1.7 to 5.8 D when going from the minimal to the methyl-terminated model to the B form. Correspondingly, the degree of charge transfer from the phenolic ring to the carbon bridge increases from the minimal to the methyl model, and the transfer becomes then predominantly to the imidazolinone ring in the B form. For the neutral moieties, the dipole difference decreases from 4.6 to 1.7 D from the minimal to the methyl model and increases to as much as 9.9 D in the A form. Differently from the gas-phase models, the charge transfer of the A form is from the phenolic ring mainly toward the central bridge. We note that, in all cases, the CASSCF magnitude of the dipole moment difference is very close to the value computed within CASPT2.

3.2. TDDFT Results. We compute the TDDFT excitations of the anionic and neutral methyl-terminated models and of the A and B forms using the aug-cc-pVDZ basis and several exchange-correlation functionals, namely, BLYP, B3LYP, CAM-B3LYP, LC-BLYP, and LC-ωPBE. As shown in Table 2 of the Supporting Information, the inclusion of augmentation yields a significantly faster basis-set convergence for the excitations and dipole moments, and the use of the aug-cc-pVDZ gives results converged within 0.01 eV for both the anionic and neutral states.

We collect the excitations of the neutral and anionic methyl-terminated models in Table 7. For the anion, all functionals yield excitation energies in the range 3.0−3.1 eV with the exception of the generalized gradient approximation BLYP, which gives a red-shifted value of 2.8 eV. For the neutral methyl-terminated model, the spread of results is larger, with BLYP giving a very low

Table 7. TDDFT Vertical Excitation Energies (eV) of the Methyl-Terminated Neutral and Anionic Chromophore Models Computed with Different Functionals and the aug-cc-pVDZ Basis on the Ground-State DFT/BLYP Geometry[a]

	neutral		anion	
functional	E_{exc}	$\|\Delta\mu\|$	E_{exc}	$\|\Delta\mu\|$
BLYP	3.06 (0.51)	1.3	2.79 (0.73)	2.8
B3LYP	3.33 (0.64)	0.4	2.96 (0.90)	1.7
CAM-B3LYP	3.56 (0.71)	0.6	3.05 (1.02)	1.2
LC-BLYP	3.79 (0.78)	1.0	3.10 (1.11)	1.2
LC-ωPBE	3.74 (0.75)		3.08 (1.09)	

[a] We also report the oscillator strength in brackets and the modulus of the dipole difference ($|\Delta\mu|$) in Debye.

Table 8. TDDFT/Amber99 Vertical Excitation Energies (eV) of the Neutral A and Anionic B Forms Computed with Different Functionals and the aug-cc-pVDZ Basis[a]

	A form		B form	
functional	E_{exc}	$\|\Delta\mu\|$	E_{exc}	$\|\Delta\mu\|$
BLYP	3.07 (0.47)	3.8	2.82 (0.79)	2.6
B3LYP	3.22 (0.79)	1.5	3.00 (0.97)	1.5
CAM-B3LYP	3.42 (0.86)	2.5	3.10 (1.05)	2.0
LC-BLYP	3.61 (0.91)	3.3	3.17 (1.12)	2.8
LC-ωPBE	3.57 (0.90)		3.15 (1.11)	

[a] We also report the oscillator strength in brackets and the modulus of the dipole difference ($|\Delta\mu|$) in Debye. The ground-state DFT/PBE geometry equilibrated within CPMD/Amber03 is used.

excitation energy and LC-BLYP the highest one. The use of long-range corrected functionals yields excitation energies in the range of 3.6−3.8 eV characterized by a larger oscillator strength and by dipole moments rather similar to the anionic values.

As shown in Table 8, the TDDFT excitations in the protein for the neutral A and anionic B forms follow a pattern similar to the one in the gas-phase counterparts. The BLYP excitations are red-shifted with respect to the values obtained with the other functionals, and the difference is more marked for the neutral case where the oscillator strength within BLYP is smaller and spurious excitations appear at lower energies (not shown in the table). Long-range corrected functionals yield excitations in the range of 3.4−3.6 eV and of 3.1−3.2 eV for the neutral and anionic forms, respectively. Again, the difference between the dipole moments of the neutral and anionic forms is not very large when employing hybrid or long-range corrected functionals. The change in the dipole moment induced by excitation is however generally larger in the protein environment than in the gas-phase for both neutral and anionic moieties.

3.3. QMC Results. We compute the variational (VMC) and diffusion Monte Carlo (DMC) excitations of the A and B forms in the presence of the MM environment, using the D+ basis set. In the Jastrow−Slater wave functions, we employ CASSCF expansions expressed over natural orbitals and truncated with appropriate thresholds on the CSF coefficients. Since the optimization of the orbitals does not significantly affect the QMC excitations of the anionic models in the gas phase,[35] we only optimize here the Jastrow and linear coefficients in energy

Table 9. QMC/Amber03 Vertical Excitation Energies (eV) of the Neutral A and Anionic B Forms Computed with the D+ Basis[a]

	CAS(n,m)	Thr.	Det/CSF	VMC	DMC
B form	2,2	0.00	4/3	3.35(7)	3.02(8)
	12,11	0.10	7/4	3.25(7)	3.1(1)
		0.05	80/28	3.4(1)	3.1(1)
A form	2,2	0.00	4/3	3.99(7)	3.78(9)

[a] The statistical error is indicated in brackets. We use a CAS(2,2) and a CAS(12,11) expansion in the determinantal component truncated with different thresholds (Thr.) on the CSF coefficients and report the number of determinants and CSFs retained in the expansion. The ground-state DFT/PBE geometry equilibrated within CPMD/Amber03 is used.

minimization within variational Monte Carlo in a state-averaged manner.

In Table 9, we report the VMC and DMC excitations of the B form computed with a CAS(2,2) and a larger CAS(12,11) expansion shown to be adequate in the convergence of the CASPT2 excitation. We find that the excitations computed with different CAS expansions and thresholds are equivalent within statistical error at both the VMC and the DMC levels, and even a simple CAS(2,2) wave function appears to be sufficient. The more reliable DMC excitation is systematically lower than the VMC value and close to the DMC estimate of 3.04(4) eV for the methyl-terminated model in the gas phase.[35] For the A form, we only compute the QMC excitation with a CAS(2,2) wave function due to difficulties in converging the CAS(12,11) calculation within GAMESS to the same CASSCF minimum reported in Table 3. Nevertheless, the DMC excitation obtained with a CAS(2,2) is an indication that the DMC estimate is similar to CASPT2 and TDDFT with long-range corrected functionals.

4. DISCUSSION

In this section, we first discuss the features of our structural models and compare them with other simulations available in the literature. We describe the tests we performed to investigate the reliability of our structures with respect to the choice of QM theoretical approach and MM force field. We also discuss our findings on the stabilization of alternative nonstandard protonations and of a solvated hydronium in proximity of the chromophore.

We then focus on the relative performance of the theoretical approaches employed to compute the vertical excitation energies of the neutral A and anionic B forms and then discuss their comparison with the experimental absorption spectra.

4.1. Structural Analysis of GFP Models. We discuss here the structural features of the chromophore for the neutral and anionic forms of wild-type GFP obtained in our QM/MM calculations. For the labeling of the atoms of the chromophore, we refer the reader to Figure 3.

The structure of the chromophore is expected to play a very important role in tuning its excited state properties. In particular, the degree of bond-length alternation in the conjugate chain running through the chromophore will be correlated to the size of the excitation, which is here of π to π^* character. In addition, the local features of the binding site of the chromophore can affect the spectral response of the chromophore by either tuning the internal geometrical structure of the chromophore or as a polarizing environment.

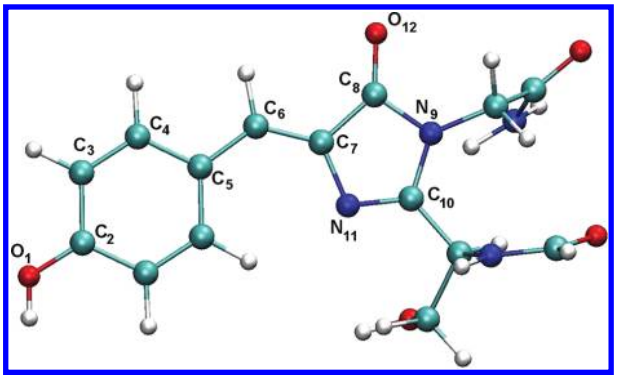

Figure 3. Atom numbering used for the chromophore of GFP. This image is produced with VMD.

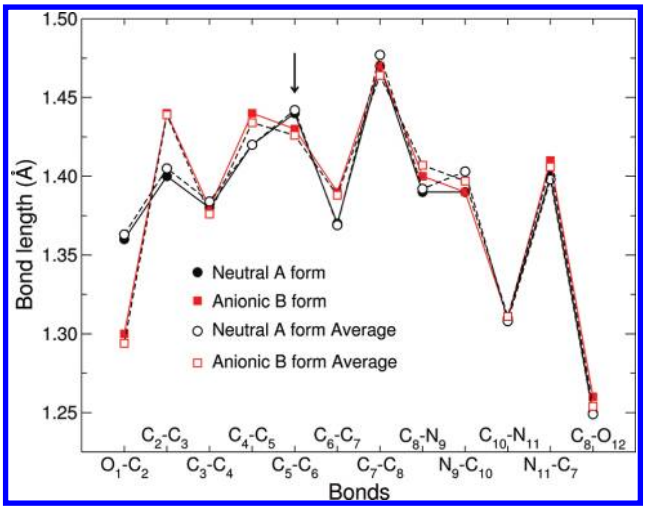

Figure 4. Bond lengths of the chromophore of the neutral A and anionic B forms as obtained in our PBE/Amber CPMD simulations. The bond lengths of the chromophore resulting from our annealing procedure are compared with the average values obtained along a room-temperature QM/MM molecular dynamics simulation. The root-mean-square fluctuations on the bond lengths are on the order of 0.02−0.03 Å, and the error bars on the averages are smaller than the size of the symbols. The C_5-C_6 bond in the central carbon bridge is indicated with an arrow.

Before analyzing the structural differences between the various forms, it is important to verify that the models obtained in our annealing procedure are indeed representative of an average configuration of the chromophore within the protein. To this end, we compare our annealed models of the A and B forms to the bond lengths obtained by averaging over approximately 1.5 ps of a molecular dynamics QM/MM simulation at room temperature. As shown in Figure 4, the annealed and average bond lengths are remarkably similar for both forms, with differences smaller than 0.01 Å, demonstrating the validity of employing our annealed structures as representative models in further analysis and in the calculations of the excitation properties of both forms.

To understand the geometrical effects of the protein environment on the chromophore, the bond lengths of the chromophore in the neutral and anionic forms are compared with the values optimized in vacuo in Figure 5. In the protein, we observe a slight shortening of about 0.01 Å in the bond lengths of the neutral A form, with larger deviations in the imidazolinone ring close to the

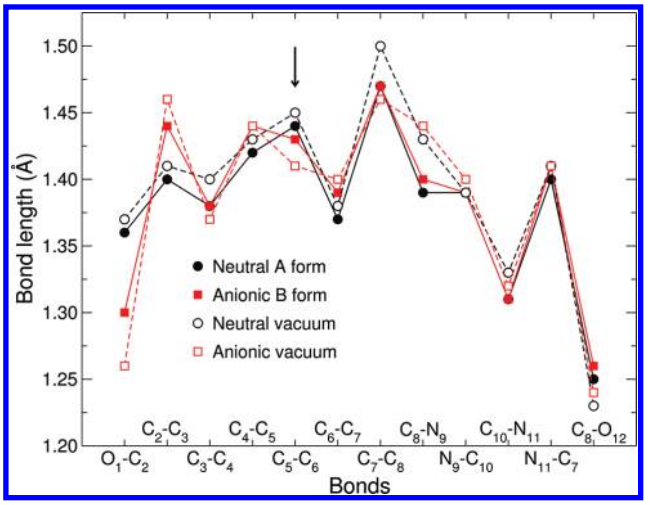

Figure 5. Bond lengths of the chromophore of the neutral A and anionic B forms as obtained in our PBE/Amber CPMD simulations. The results for the chromophore models optimized in vacuo with BLYP/cc-pVTZ are also shown. The C_5-C_6 bond in the central carbon bridge is indicated with an arrow.

Arg96 counterion. Similarly, the bond lengths of the anionic B form do not dramatically differ from the values we obtain in vacuo. The most significant difference is observed in the degree of bond alternation, which increases close to the central bridge of the anionic form. Moreover, the O_1-C_2 bond is lengthened in the protein due to hydrogen bonding to close-by residues.

When comparing the neutral and the anionic forms, we see that, in vacuo, the largest difference occurs in proximity of the O_1-C_2 bond. In fact, as expected, after deprotonation, it loses its single-bond character and significantly shortens by about 0.1 Å. As a consequence, the aromaticity (similar bond lengths) of the phenolic ring is reduced in the anionic form with respect to the neutral case. Finally, the degree of bond alternation in proximity of the central bond is smaller in the anion than in the neutral chromophore. Interestingly, we note that the difference between the neutral and the anionic bond lengths is overall smaller in the protein. In particular, the degree of bond alternation close to the bridge is more similar between anion and neutral form in the presence of the protein. Overall, it appears that the protein environment acts to partially compensate the change in protonation state and keeps the chromophore in a more similar structural conformation in the two protonation states with respect to a vacuum.

In Figure 6, we compare the geometrical properties of the chromophore of the three forms of GFP with the results of other simulations available in the literature. We first focus on the AM1/CHARMM calculations by Marques et al.,[45] as the structural features of their models most significantly differ from ours as well as from other calculations reported in the literature. We first note that Marques et al. construct the I form by deprotonation of the neutral form although they refer incorrectly to this structure as the B form. Their neutral A and anionic I forms are rather similar with the exception of the O_1-C_2 bond, which is significantly shorter in the I form, in agreement with our calculations. Both structures display however a more marked bond-length alternation along the chromophore than our models. This is particularly evident for the I form where the bond alternation in the central carbon bridge is 0.09 Å compared to 0.05 Å in our calculations.

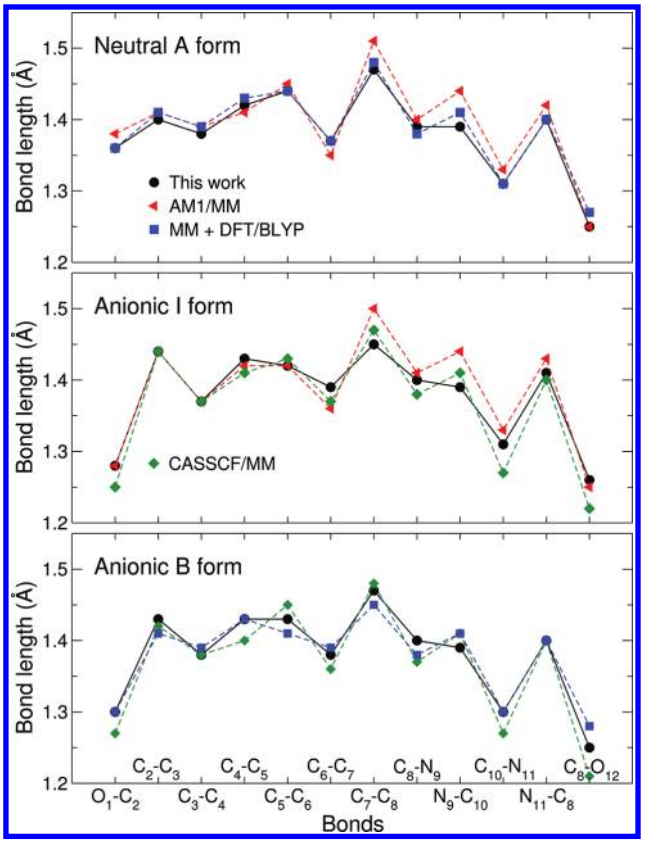

Figure 6. Bond lengths (Å) of the chromophore of the A, I, and B forms of wild-type GFP obtained with PBE/Amber in CPMD. We also show the results of the AM1/MM,[45] CASSCF/MM,[46] and MM simulations followed by partial QM relaxation[47] (denoted as MM+DFT/BLYP). The bonds of the central bridge are C_5-C_6 and C_6-C_7.

Another clear difference is the bond length of the subsequent single carbon bond of the imidazolinone ring, C_7-C_8, which is 0.05 Å longer than our value for both the neutral and anionic I forms.

Our structures for the neutral A and anionic B forms are instead in close agreement with the simulations of Laino et al.[47] denoted here as "MM + DFT/BLYP". The authors employ the Amber98 force field and parametrize the chromophore accordingly, to perform a first classical MD relaxation of the protein. They then refine the structural parameters with a DFT/BLYP optimization of an isolated cluster comprising the chromophore and its close residues with fixed heavy atoms at the boundary.

Finally, we compare our anionic I and B forms with the results of the CASSCF/CHARMM calculations by Sinicropi et al.,[46] who relax the chromophore structure within CASSCF together with three closeby classical waters. The rest of the protein is kept to the original crystallographic coordinates, and the I form is obtained starting from the A form and manually reorienting the residues Ser205 and Glu222 to form the expected hydrogen bond network. The B form is derived from the I form by rotating and relaxing Thr203 to create a hydrogen bond with the phenolic oxygen of the chromophore. Despite the lack of complete relaxation, their hydrogen-bond distances are in reasonable agreement with our values. As generally found when comparing CASSCF and DFT structures, their chromophore displays a somewhat larger degree of bond-length alternation as compared to our DFT calculations.

To understand the origin of some of the differences with previous studies, we investigate how particular choices in the description of

the QM and MM parts affect the structural features of the chromophore. We first consider the effect of the choice of the force field since the AM1[45] and CASSCF[46] calculations use CHARMM while we employ the Amber03 parametrization. Starting from our PBE/Amber03 annealed model of the B form, we optimize the chromophore alone with BLYP/CHARMM for fixed positions of the MM charges. We note here that BLYP and PBE lead to equivalent structures in the gas phase and the use of BLYP/CHARMM instead of PBE/CHARMM is simply due to technical reasons.[48] As shown in Table 3 of the Supporting Information, we find that the chromophores optimized with BLYP/CHARMM and PBE/Amber03 are equivalent with bond lengths deviating at most by 0.02 Å. Therefore, the choice of force field does not lead to significant differences in the main structural features and cannot explain the large deviations observed between the AM1/CHARMM model of ref 45 and our structure.

We then explore whether the use of the CASSCF method in optimizing the QM part and/or the lack of full relaxation in their protein model are responsible for the differences between the CASSCF/CHARMM structure of ref 46 and our model of the B form. Starting from the coordinates of ref 46, we optimize the chromophore within DFT/BLYP at fixed MM positions and, as shown in Table 3 of the Supporting Information, recover a chromophore which closely resembles the structure obtained in our very different DFT-based annealing procedure. Therefore, the observed differences must be ascribed to the use of the CASSCF approach to optimize the QM part and not to the details of the construction of the protein model. Then, we investigate how the choice of the exchange-correlation functional affects the final QM geometry. In particular, we consider the use of hybrid functionals and, starting from our model of the B form, we perform a B3LYP/Amber99 optimization of the chromophore at fixed positions of the MM charges. Even in this case, the resulting chromophore has similar bond lengths to the PBE/Amber03 values. We already note here and discuss further in the next section that the observed geometrical differences between the CASSCF, DFT/B3LYP, and our DFT/BLYP structures do not result in substantial differences in the TDDFT/MM excitations of the B form.

To further probe the reliability of our structures, we perform few steps of geometrical optimization using the CASPT2/MM method starting from our model of the B form. We employ the Amber99 force field, a CAS(12,11) expansion in the CASSCF wave function, and keep the MM charges at the fixed positions of our CPMD/MM annealed structures. We converge the CASPT2 root-mean-square fluctuations of the Cartesian coordinates and forces to less than 0.01 Å and 0.01 au, respectively, and the chromophore remains very similar to the DFT starting structure with deviations smaller than 0.01 Å. Therefore, this test gives us confidence in our choice of DFT/BLYP being able to capture the correct structural properties of the chromophore embedded in the protein.

Finally, we perform extensive QM/MM simulations of the anionic I form increasing the QM part to include residues surrounding the chromophore. The structure is first equilibrated at room temperature with CPMD/Amber03 within DFT/PBE, and the QM part consists of Arg96, Gln94, and the chromophore. The QM component is then enlarged to also include Glu222, His148, and the water close to the phenolic oxygen, and the system is slowly annealed to zero temperature. The resulting chromophore deviates by less than 0.01 Å from the original geometry obtained with only the chromophore in the QM part. These simulations are also repeated including dispersion-corrected atom-centered (DCACP) pseudopotentials.[49,50] The use of DCACP does not lead to significant changes in the structure.

Table 10. Vertical Excitation Energies (eV) of the Neutral and Anionic Methyl-Terminated Models in the Gas Phase, and the Neutral A and Anionic B Forms in the Protein[a]

	Gas phase		Protein	
	Neutral	Anion	Neutral	Anion
CAM-B3LP	3.56	3.05	3.42	3.10
LC-BLYP	3.79	3.10	3.61	3.17
CASPT2	3.82	2.76	3.53	2.82
DMC	–	3.04(4)[b]	–	3.1(1)
Expt.	3.51	2.59[c], 2.75[d], >2.60[e], 2.84[f]	3.05[g]	2.63[g]

[a] The TDDFT excitations are reported for the CAM-B3LYP and LC-BLYP functionals and the aug-cc-pVDZ basis set. The best available single-state CASPT2 and DMC excitations, obtained with the ANO-S-PVDZ and D+ bases, respectively, are shown. The ground-state DFT/BPE geometry equilibrated within CPMD/Amber03 is used. [b] DMC excitation from ref 35. [c] Photodestruction experiments of ref 52. [d] Photodestruction experiments of ref 53. [e] Photodestruction experiments of ref 54. [f] Extrapolation of Kamlet−Taft fit of solution experiments.[5] [g] Absorption maxima in the protein at 1.7 K.[4]

In summary, our structures constructed within CPMD/MM via a room-temperature MD procedure followed by annealing appear to be reliable. They are representative of an average configuration of the chromophore−protein complex as demonstrated by a comparison with the average bond lengths over room-temperature MD trajectories. Treating only the chromophore in the QM part is sufficient, and the geometry is not sensitive to the particular choice of nonpolarizable force field in the MM part, namely, CHARMM or Amber. The choice of DFT to treat the QM component is appropriate, as a partial optimization within CASPT2/MM approach leaves the structure unchanged. We can rationalize the small difference with the CASSCF structures of ref 46 as due to the use of a different QM method in the optimization. Finally, our structures compare well to the ones obtained in the MM simulations refined by further QM relaxation of ref 47. The difference with the AM1/MM structure of ref 45 remains the most significant and is possibly due to the positive protonation state of most His residues in their model.[51]

4.2. Comparison of Theoretical and Experimental Excitations. In Table 10, we summarize the most representative theoretical results for the vertical excitations of the neutral and anionic methyl-terminated models in the gas phase and of the neutral A and anionic B forms in the protein. We report the single-state CASPT2 excitations computed with the largest CAS(16,14) expansion since the single-state values are either close to their multistate counterpart or appear to be a more reliable estimate of the excitation energy as shown in the case of the neutral gas-phase models. We list the TDDFT excitations computed with the CAM-B3LYP and LC-BLYP functionals, as calculations with BLYP are in some cases plagued by the appearance of spurious low-lying excitations. Finally, we consider the best DMC values available. Our theoretical estimates are compared with the available experimental results, namely, three different photodestruction experiments for the anion in the gas phase,[52−54] the extrapolation to vacuum conditions of solution experiments for the neutral and anionic species,[5] and the low-temperature absorption spectra in the protein.[4]

We first focus on the results for the anionic models, which are remarkably consistent among the wave function and DFT methods both in the gas phase and in the protein. In the gas phase, the excitations are between 2.76 and 3.10 eV, in agreement with our previous calculations,[35] which are here extended to include a more complete set of DFT functionals. The CASPT2 excitations are computed with the ANO-S-VDZP basis for consistency with the calculations in the protein, so small differences with the CASPT2 estimates of ref 35 must be attributed to the use of a different basis set. We also list the DMC excitation in the gas phase computed without augmentation and a CAS(8,8) expansion and recall here that tests on the inclusion of augmentation and larger CAS spaces indicate that the DMC excitation of the anionic model is rather robust and consistently at about 3 eV.[35]

The computation of the anionic excitation in the gas phase is in principle complicated by the fact that the excitation lies above the ionization threshold, as demonstrated in recent photodestruction experiments.[53] Nevertheless, the DFT value is rather insensitive to the use of very large basis sets,[55] and even artificially raising the ionization threshold well above the excitation via asymptotically corrected potentials (e.g., the statistically average orbital potential approach) does not affect its value.[35] Similarly, the DMC estimate appears to be largely unaffected by the choice of basis and CAS expansions.[35] These findings indicate that this metastable state in the continuum is dominated by a well localized π to π^* transition, which renders possible the computation of this excitation in the continuum. Experimentally, early photodestruction experiments assign the vertical transition to the only observed peak in the absorption spectrum at 2.59 eV.[52] A second photodestruction experiment resolves however a rather different shape of the absorption spectrum and reconsiders its interpretation assigning the adiabatic transition to the lowest peak at 2.59 eV and the vertical transition to the second, newly resolved peak at 2.75 eV.[53] The most recent photodestruction experiment reveals a strong dependence of the shape of the spectrum on the excitation laser power and leads the authors to conclude that such a spectrum cannot reliably represent the optical absorption spectrum of the anionic GFP model in the gas phase.[54] Finally, the multivariant Kamlet−Taft fit of the absorption maxima in solution in terms of the acidic, basic, and polar solvating parameters extrapolates to 2.84 eV for the conditions of a vacuum.[5] In view of the apparent uncertainty on the shape of the spectra in photodestruction experiments, our theoretical estimates appear to be in the right range as they fall between 2.76 and 3.10 eV and display a blue shift which can be easily explained with vibronic effects and with the intrinsic theoretical limitations of predicting a metastable excitation in the continuum.

In the protein, the theoretical excitations of the anionic B form range between 2.82 and 3.17 eV and therefore display a negligible bathochromic shift with respect to the gas phase. Our theoretical estimates appear to be rather robust and even more insensitive to the internal parameters of the theory than the calculations in the gas phase. In particular, the excited state is not strongly multiconfigurational as the CASPT2 excitations converge rapidly with the size of the expansion and, for the DMC values, even a small CAS(2,2) expansion is sufficient. The DFT estimates are clustered around 3.1 eV, and the presence of the protein raises the ionization threshold (estimated as minus the HOMO eigenvalue) well above the excitation, in agreement with the findings of ref 56. Therefore, with respect to the anion in the gas phase, we no longer have the complications of describing an excitation in the continuum. The theoretical excitations computed within our QM/MM model are however significantly higher than the experimental absorption maximum of 2.63 eV measured at 1.7 K. Therefore, while experiments indicate that there should be a red shift on the order of 0.2 eV when going to the protein, theory obtains a shift less than 0.1 eV in the opposite direction, which brings the vertical excitation farther away from the location of the absorption maximum in the protein. To explain our findings within the protein, we cannot invoke vibronic or temperature effects, as the chromophore is tightly held in a rigid position within the protein pocket and the experimental absorption band of the B form is rather narrow, with the location of the maximum moving only to 2.59 eV when the measurement is performed at room temperature.

As already mentioned, the computation of the excitations of the neutral moieties is even more challenging. For instance, the CASPT2 calculations in the protein present difficulties due to the existence of multiple close minima at the CASSCF level which are characterized by different ordering of the states, oscillator strengths, and CASSCF excitations, even when all π electrons on the chromophore are correlated in a minimal active space. Both in the gas phase and in the protein, whenever the oscillator strengths are spread over multiple states, the single- and multistate excitations differ significantly, and this behavior is particularly evident in the gas-phase calculations of the methyl-terminated model. Finally, the DFT excitations in the gas phase and in the protein computed using the generalized gradient approximation BLYP and hybrid B3LYP are significantly red-shifted with respect to all other values. Focusing on our best theoretical estimates of Table 10, the vertical excitations in the gas phase range between 3.56 and 3.82 eV and are in reasonably good agreement with the Kamlet−Taft extrapolation of solution experiments to vacuum conditions at 3.51 eV.[5] Again, we need to recall that vibronic effects are expected to be very strong for the chromophore in the gas phase (and in solution) and will lead to an absorption maximum red-shifted with respect to the location of the vertical excitation. Nevertheless, both experiments and theoretical estimates in the gas phase unequivocally indicate that one must expect a significant bathochromic shift of at least 0.5 eV since the absorption maximum in the protein is located at 3.05 eV in measurements at 1.7 K. Our theoretical excitations of the A form lie instead in the range 3.42−3.53 eV, therefore significantly blue-shifted with respect to experiments, and display a too small bathocromic shift of 0.2−0.3 eV. Similarly to the B form, we do not expect vibronic or temperature effects to be of significant magnitude to explain the discrepancy between theoretical and experimental data in the protein. It therefore appears that, while theory can correctly predict the absorption in the gas phase of the neutral moiety, we cannot correctly reproduce the location of the vertical excitation in the protein and fail to see the large red shift due to the protein environment.

To explore possible origins of the discrepancy between the theoretical excitations and the experimental data, we follow two routes. On the one hand, we consider extended QM regions including the chromophore and the amino acids involved in the hydrogen-bond network surrounding the chromophore, namely, Ser205, Glu22, Arg96, Gln94, and the water close to the phenolic oxygen. On these extended models, we perform TDDFT/MM calculations with the CAM-B3LYP and LC-BLYP functionals. We find that both functionals give a red shift for the A form of about 0.15 eV. However, the corresponding excited states show non-negligible contributions of electronic transition with strong

charge transfer character also to the boundary of the QM region. Consequently, we expect that these spurious transitions will depend on the specifics of the QM/MM border and are therefore an artifact of approximate TDDFT even with the use of long-range corrected functionals. Unfortunately, it is computationally very demanding to perform such large calculations with the highly correlated approaches we consider in this paper. Therefore, as a second route, we only attempt a similar test on a QM model including only the chromophore and the counterion, Arg96, at the CASPT2/MM level and choose to explore the B form as it displays a faster and more robust convergence with CAS expansion. We find that the inclusion of a quantum counterion does not shift the excitation. In addition, attempts to stabilize a charge transfer excitation by including acceptor orbitals on the counterion in an extended CASSCF expansion are not successful as the excitation remains confined on the chromophore.

If we compare with previous calculations of the A and B form in the literature, we find the TDDFT/LDA results by Marques et al.,[45] which report an excellent agreement with the experimental absorption maxima in the protein. It is however rather difficult to assess the quality of these calculations. As already mentioned in the previous section, we need to bear in mind that they compute the I and not the B form and that the structures of their anionic and neutral chromophores are the only ones in the literature which depart by a significant extent from our models. These differences might be due to their choice of a positive protonation of most His residues in their models. Moreover, they extract the chromophores from the protein and compute the excitation without including the environment. Finally, the use of the LDA functional will naturally lead to red-shifted values as even generalized gradient approximations are affected by this problem (see section 3.2).

For the B form, previous CASPT2 calculations[46] report an excitation of 2.81 eV, blue-shifted with respect to the experimental value of 2.63 eV and in apparent agreement with our calculations. The agreement is however in part fortuitous since the authors of ref 46 employ the CASSCF method to optimize the structure, and the older definition of the zero-order Hamiltonian without IPEA shift and a smaller 6-31G* basis in their CASPT2/MM calculations. As shown in section 3.3, setting the IPEA shift to zero lowers our CASPT2 excitation of the B form from 2.82 to 2.38 eV, and the lower value of 2.38 eV should be compared to the result of ref 46. While we cannot ascribe this discrepancy to the setup of the protein model (see section 4.1), the difference between the two CASPT2 results when using the same zero-order Hamiltonian must be due to the combined use in ref 46 of a poorer basis set and the CASSCF geometry for the chomophore since both factors lead separately to a blue shift.

For the A form, Hagasawa et al.[57] report an excitation energy of 3.21 eV obtained with SAC−CI and QM/MM, in reasonable agreement with experiments. In their protein model, the QM chromophore is optimized in B3LYP/Amber, and the MM atoms are fixed at the X-ray structure. For the methyl-terminated chromophore in the gas phase, they obtain however an excitation of 3.23 eV, which is on top of the value in the protein and significantly red-shifted with respect to our estimates of 3.6−3.8 eV and to the extrapolation of solution experiments to vacuum conditions at 3.51 eV.[5] Consequently, their SAC−CI calculations do not yield the desired bathochromic shift of at least 0.5 eV. Given the failure of their approach to reproduce the gas phase absorption and the correct bathochromic shift, the apparent agreement in the protein might be fortuitous.

Finally, in a recent study by Krylov et al.,[56] the B form is investigated using the SOS-CIS(D) method within QM/MM. The excitation energy is estimated as 2.70 eV for a protein model optimized by PBE0/Amber. The bond lengths of the chromophore are rather similar to the ones of our model, and the discrepancy between the SOS-CIS(D) and our estimates in the higher-energy range of 2.82−3.17 eV is most likely due to the use of a different technique to compute the excitation energy.

5. CONCLUSIONS

In the gas phase, DFT and wave function methods show an overall good agreement with the extrapolation of solution experiments to vacuum conditions both for the anionic and neutral species. For the anion, the vertical excitation in the gas phase ranges between 2.8 and 3.1 eV, and our theoretical estimates appear to be rather robust despite the inherent difficulty of computing an excitation in the continuum. We also note that the comparison with photodestruction experiments is difficult given the uncertainty in the shape of the spectra.[52−54] For the neutral moiety, our calculations using either long-range corrected DFT functionals or wave function methods lead to vertical excitations in the range of 3.6−3.8 eV.

When introducing the protein environment, we do not obtain the proper bathochromic shift to the experimental absorption maxima of the anionic B and neutral A forms, which are located at 2.63 and 3.05 eV, respectively. For the anionic B form, we obtain theoretical excitations between 2.8 and 3.2 eV, therefore slightly blue-shifted with respect to the gas phase. For the neutral A form, we find excitations in the range of 3.4−3.6 eV, which are not sufficiently red-shifted. Similarly to the gas phase, wave function and DFT methods agree when long-range corrected functionals are used. Therefore, while the various theoretical approaches predict a shift between the excitations of the A and the B form (in the range 0.32−0.71 eV as compared to the experimental value of 0.4 eV), they fail rather dramatically in estimating the location of the absorption maxima.

Then, why does our model GFP not yield the desired bathochromic shift? Here, we have adopted the generally accepted protonation of the amino acids in wild-type GFP and a standard QM/MM prescription in the construction of the protein models. Within this paradigm, we can exclude the source of the problem lying in the computational details of the construction procedure. In particular, our extensive tests indicate that our structures are rather robust. Temperature effects do not play a role. The use of DFT as QM method is sufficiently accurate, and extending the QM region beyond the chromophore is not necessary when relaxing the protein. Finally, the structures are insensitive to the particular choice of nonpolarizable force field. Since the theoretical techniques to compute the excitation spectrum appear to be reliable in the gas phase and display an overall agreement among each other in the protein, we also rule out that the origin of the problem lies in our choice of QM method to compute the excitation.

Consequently, few possible sources of error require more extensive investigations. One possibility is that some amino acids in the protein are differently protonated than what is commonly accepted in the literature. Here, following the experimental suggestion of ref 42, we have investigated the stability of a hydronium in proximity to the chromophore, as the presence of this charged moiety could lead to significant structural and electronic changes in the model. However, the hydronium is not stable and

always donates the additional proton either to the chromophore or to Glu222 during a room-temperature QM/MM molecular dynamics simulation. Further investigations of other residues in the second shell surrounding the chromophore are needed. An alternative origin of the disagreement between theory and experiments is possibly the theoretical description of the polarization field of the protein on the excited state of the chromophore. While we find that the excitation is insensitive to the choice of the particular nonpolarizable force field used to describe the MM region, we cannot exclude that a more accurate description of the protein would lead to significantly different excitations. Ideally, one would extend the QM region well beyond the chromophore and the surrounding amino acids. Unfortunately, this is currently prohibitive with the highly correlated methods employed here, and our tests indicate that TDDFT with current approximations is affected by spurious charge-transfer effects when using extended QM regions in GFP. An alternative is to keep a partition of the system in an active site and an external region but to improve upon the MM treatment. Addressing the potential limitations of a MM description in the computation of excitations is currently an active field of investigation,[58−61] and the exploration of better embedding schemes together with alternative protonations within wild-type GFP will be the subject of future research.

■ ASSOCIATED CONTENT

Supporting Information. Mulliken charge analysis, TDDFT basis set convergence, and chromophore bond lengths of the B form computed with different QM/MM approaches. PDB files of the B form and the A chain of the A form including residues and waters within 5 Å of the protein. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: c.filippi@utwente.nl, f.buda@chem.leidenuniv.nl, adalgisa.sinicropi@unisi.it.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENT

We acknowledge the support from the Stichting Nationale Computerfaciliteiten (NCF-NWO) for the use of the SARA supercomputer facilities. We also acknowledge computational resources provided by the CASPUR and CINECA computer centres. We thank Riccardo Nifosì for useful discussions.

■ REFERENCES

(1) Day, R. N.; Davidson, M. W. *Chem. Soc. Rev.* **2009**, *38*, 2887–2921.

(2) Lippincott-Schwartz, J.; Patterson, G. H. *Trends Cell Biol.* **2009**, *19*, 555–565.

(3) Tsien, R. Y. *Annu. Rev. Biochem.* **1998**, *67*, 509–544.

(4) Creemers, T. M. H.; Lock, A. J.; Subramaniam, V.; Jovin, T. M.; Völker, S. *Nat. Struct. Biol.* **1999**, *6*, 557.

(5) Dong, J.; Solntsev, K. M.; Tolbert, L. M. *J. Am. Chem. Soc.* **2006**, *128*, 12038.

(6) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. *J. Comput. Chem.* **2005**, *26*, 1668–1688.

(7) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. *J. Comput. Chem.* **2003**, *24*, 1999–2012.

(8) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.

(9) *CPMD*, v3.11.1, C. (revision a11); IBM Corp.: Endicott, NY, 2008; MPI für Festkörperforschung Stuttgart: Stuttgart, Germany, 2001. http://www.cpmd.org/.

(10) van Gunsteren, W. F.; Billeter, S. R.; Eising, A. A.; Huenenberger, P. H.; Krueger, P.; Mark, A. E.; Scott, W. R. P.; Tironi, I. G. *Biomolecular Simulation: The Gromos96 Manual and User Guide*; Hochschulverlag an der ETH Zurich: Zurich, Switzerland, 1996.

(11) Laio, A.; VandeVondele, J.; Röthlisberger, U. *J. Chem. Phys.* **2002**, *116*, 6941–6947.

(12) Laio, A.; VandeVondele, J.; Röthlisberger, U. *J. Phys. Chem. B* **2002**, *106*, 7300–7307.

(13) Troullier, N.; Martins, J. L. *Phys. Rev. B* **1991**, *43*, 1993–2006.

(14) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865.

(15) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1997**, *78*, 1396.

(16) Frisch, M. J.; et al. *Gaussian 09*, Revision A.1; *Gaussian 09*, Revision A.1; Gaussian Inc.: Wallingford, CT, 2009.

(17) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3096.

(18) Lee, C. T.; Yang, W. T.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.

(19) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.

(20) Yanai, T.; Tew, D. P.; Handy, N. C. *Chem. Phys. Lett.* **2004**, *393*, 51–57.

(21) Iikura, H.; Tsuneda, T.; Yanai, T.; Hirao, K. *J. Chem. Phys.* **2001**, *115*, 3540–3544.

(22) Vydrov, O. A.; Scuseria, G. E. *J. Chem. Phys.* **2006**, *125*, 234109.

(23) Aquilante, F.; Vico, L. D.; Ferré, N.; Ghigo, G.; Malmqvist, P.-Å.; Neogrády, P.; Pedersen, T. B.; Pitonák, M.; Reiher, M.; Roos, B. O.; Serrano-Andrés, L.; Urban, M.; Veryazov, V.; Lindh, R. *J. Comput. Chem.* **2010**, *31*, 224–247.

(24) Ghigo, G.; Roos, B. O.; Malmqvist, P.-Å. *Chem. Phys. Lett.* **2004**, *396*, 142–149.

(25) Forsberg, N.; Malmqvist, P.-A. *Chem. Phys. Lett.* **1997**, *274*, 196–204.

(26) Aquilante, F.; Pedersen, T. B.; Lindh, R. *Theor. Chem. Acc.* **2009**, *124*, 1–10.

(27) Boström, J.; Delcey, M. G.; Aquilante, F.; Serrano-Andrés, L.; Pedersen, T. B.; Lindh, R. *J. Chem. Theory Comput.* **2010**, *6*, 747–754.

(28) Ponder, J. W.; Richards, F. M. *J. Comput. Chem.* **1987**, *8*, 1016.

(29) Wang, J.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049.

(30) CHAMP is a quantum Monte Carlo program package written by C. J. Umrigar, C. Filippi, and collaborators.

(31) Burkatzki, M.; Filippi, C.; Dolg, M. *J. Chem. Phys.* **2007**, *126*, 234105.

(32) We add one s and one p diffuse function on the carbon and the nitrogen using exponents from the aug-cc-pVDZ basis set, taken from EMSL Basis Set Library (http://bse.pnl.gov).

(33) Filippi, C.; Umrigar, C. J. *J. Chem. Phys.* **1996**, *105*, 213–226. As the Jastrow correlation factor, we use the exponential of the sum of three fifth-order polynomials of the electron−nuclear (e−n), the electron−electron (e−e), and the pure three-body mixed e−e and e−n distances, respectively. The Jastrow factor is adapted to deal with pseudo-atoms, and the scaling factor κ is set to 0.60 a.u.

(34) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; Windus, T. L.; Dupuis, M., Jr. *J. Comput. Chem.* **1993**, *14*, 1347–1363.

(35) Filippi, C.; Zaccheddu, M.; Buda, F. *J. Chem. Theory Comput.* **2009**, *5*, 2074–2087.

(36) Casula, M. *Phys. Rev. B* **2006**, *74*, 161102.

(37) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33–38.

(38) Yang, F.; Moss, L. G.; Phillips, G. N. *Nat. Biotechnol.* **1996**, *14*, 1246.
(39) Berman, H.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.; Weissig, H.; Shindyalov, I.; Bourne, P. *Nucleic Acids Res.* **2000**, *28*, 235–242.
(40) Nifosí, R.; Tozzini, V. *Proteins: Struct., Funct., Genet.* **2003**, *51*, 378–389.
(41) Elsliger, M.; Wachter, R.; Hanson, G.; Kallio, K.; Remington, S. *Biochemistry* **1999**, *38*, 5296–5301.
(42) Shinobu, A.; Palm, G. J.; Schierbeek, A. J.; Agmon, N. *J. Am. Chem. Soc.* **2010**, *132*, 11093.
(43) Zuev, D.; Bravaya, K. B.; Crawford, T. D.; Lindh, R.; Krylov, A. I. *J. Chem. Phys.* **2011**, *134*, 034310.
(44) Serrano-Andrés, L.; Merchán, M.; Lindh, R. *J. Chem. Phys.* **2005**, *122*, 104107.
(45) Marques, M. A. L.; López, X.; Varsano, D.; Castro, A.; Rubio, A. *Phys. Rev. Lett.* **2003**, *90*, 257101.
(46) Sinicropi, A.; Andruniow, T.; Ferré, N.; Basosi, R.; Olivucci, M. *J. Am. Chem. Soc.* **2005**, *127*, 11534–11535.
(47) Laino, T.; Nifosì, R.; Tozzini, V. *Chem. Phys.* **2004**, *298*, 17.
(48) The Molcas code has convergence problems with the use of DFT/PBE while no such difficulties are encountered when DFT/BLYP is employed.
(49) von Lilienfeld, O. A.; Tavernelli, I.; Röthlisberger, U.; Sebastiani, D. *Phys. Rev. Lett.* **2004**, *93*, 153004.
(50) von Lilienfeld, O. A.; Tavernelli, I.; Röthlisberger, U.; Sebastiani, D. *Phys. Rev. B* **2005**, *71*, 195119.
(51) López, X. Private communication.
(52) Nielsen, S. B.; Lapierre, A.; Andersen, J. U.; Pedersen, U. V.; Tomita, S.; Andersen, L. H. *Phys. Rev. Lett.* **2001**, *87*, 228102.
(53) Forbes, M. W.; Jockusch, R. A. *J. Am. Chem. Soc.* **2009**, *131*, 17038–17039.
(54) Chingin, K.; Balaboin, R. M.; Frankevich, V.; Barylyuk, K.; Nieckarz, R.; Sagulenko, P.; Zenobi, R. *Int. J. Mass Spectrom.* **2011**, *306*, 241–245.
(55) Epifanovsky, E.; Polyakov, I.; Grigorenko, B.; Nemukhin, A.; Krylov, A. I. *J. Chem. Theory Comput.* **2009**, *5*, 1895–1906.
(56) Bravaya, K. B.; Khrenova, M. G.; Grigorenko, B. L.; Nemukhin, A. V.; Krylov, A. I. *J. Phys. Chem. B* **2011**, *115*, 8296.
(57) Hasegawa, J.-Y.; Fujimoto, K.; Swerts, B.; Miyahara, T.; Nakatsuji, H. *J. Comput. Chem.* **2007**, *28*, 2443.
(58) Wesolowski, T. A. *Phys. Rev. A* **2008**, *77*, 012504.
(59) Pereira Gomes, A. S.; Jacob, C. R.; Visscher, L. *Phys. Chem. Chem. Phys.* **2008**, *10*, 5353.
(60) Söderhjelm, P.; Husberg, C.; Strambi, A.; Olivucci, M.; Ryde, U. *J. Chem. Theory Comput.* **2009**, *5*, 649.
(61) Olsen, J.; Aidas, K.; Kongsted, J. *J. Chem. Theory Comput.* **2010**, *6*, 3721.