JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

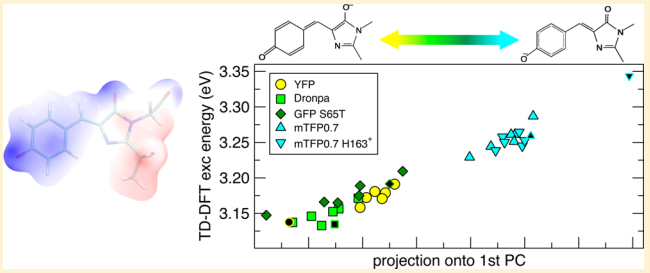
Spectral "Fine" Tuning in Fluorescent Proteins: The Case of the GFP-Like Chromophore in the Anionic Protonation State

Pietro Amat and Riccardo Nifosì*

NEST, Scuola Normale Superiore and Istituto Nanoscienze-CNR, Piazza San Silvestro 12 - 56127 Pisa, Italy

Supporting Information

ABSTRACT: Fluorescent proteins (FPs), featuring the same chromophore but different chromophore−protein interactions, display remarkable spectral variations even when the same chromophore protonation state, i.e. the anionic state, is involved. We examine the mechanisms behind this tuning by means of structural analysis, molecular dynamics simulations, and vertical excitation energy calculations using QM/MM Time-Dependent Density Functional Theory (TD-DFT), CASPT2/CASSCF, and SAC-CI. The proteins under investigation include the structurally similar, though spectrally distinct, Dronpa and mTFP0.7, with absorption peaks at 453 and 503 nm, respectively. We extend our analysis to two Green Fluorescent Protein variants, GFP-S65T (absorption peak at 484 nm), for comparison with previous computational studies, and GFP-S65G/V68L/S72A/T203Y, a yellow fluorescent protein (514 nm), in order to include one of the most red-shifted FPs containing a GFP-like chromophore. We compare different choices of the QM system, and we discuss how molecular dynamics simulations affect the calculation of excitation energies, with respect to X-ray structures. We are able to partially reproduce the spectral tuning of the FPs and correlate it to the chromophore bond-length variations, as determined by specific interactions with the chromophore environment.



INTRODUCTION

Since the cloning of the Green Fluorescent Protein (GFP) in 1992[1] and the discovery of a whole family of homologues, interest in the photophysical properties of fluorescent proteins (FPs) has constantly increased, revolutionizing the experimental approaches of molecular biology and biotechnologies through a series of new applications, featuring FPs as *in vivo* markers of cell processes.

FPs all share a common β-barrel fold embedding a chromophore moiety, derived from autocatalytic backbone cyclization at an internal tripeptide sequence (residues Ser65, Tyr66, and Gly67 in the original GFP sequence). The multiplicity of optical properties of FPs is surely one of the factors contributing to their usefulness. Indeed, the set of FPs discovered and/or engineered so far covers the whole visible spectrum from blue to far red. This variability primarily arises from the different chemical structures of the chromophore. A finer, yet considerable, tuning originates from the noncovalent interactions of the chromophore with residues and water molecules located in its vicinity. The mechanisms governing this "fine" tuning are the focus of the present computational work. We examine the case of FPs with the GFP *anionic* chromophore structure, deriving from a Tyr at the second position of the chromogenic tripeptide (residue 66 in GFP sequence).

FPs in this family can be roughly divided into three groups according to their excitation and fluorescence peak wavelength: cyan/teal, green, and yellow. Table 1 reports several such FPs, with their excitation/absorption and emission fluorescence and the residues in the chromophore environment (for the position of these residues with respect to the chromophore, see Figure 1).

FPs in the cyan−teal group are characterized by a broader absorption/excitation spectrum with respect to those in the green and yellow groups, presumably resulting from a higher degree of structural inhomogeneity at room temperature.[14] These broad spectra also prompted the hypothesis of a mixed population of protonation states of the chromophore, i.e., neutral and anionic.[14] Throughout the article, we will assume that mTFP0.7 contains an anionic chromophore, though the issue has not been investigated in the same depth as for the *Aequorea victoria* GFP family. Nonetheless, the overall shifts in the spectra of these FPs are remarkable, pointing to a considerable tuning, up to 80 nm, of the chromophore by the protein matrix. Analogously, high sensitivity to the surrounding environment is reported for the isolated chromophore model in the anionic state, displaying a much larger solvatocromic shift with respect to the neutral protonation state.[15] Several X-ray and mutagenesis studies on GFPs and mTFPs have been performed to identify the key residues responsible for the tuning (for reviews, see refs 16 and 17).

Received: August 27, 2012
Published: November 29, 2012

   dx.doi.org/10.1021/ct3007452 | J. Chem. Theory Comput. 2013, 9, 497−508

Table 1. List of Proteins Containing the GFP-Like Chromophore, with Specification of the Amino Acids Occupying Homologous Key Positions in the Chromophore Environment[a]

	environment residues (*av*GFP #)									
name	62	65	68	69	148	150	167	203	abs. max (nm)	em. max (nm)
cyan/teal									440−465	480−495
dsFP483	Pro	Gln	Asn	Lys	Ser	Glu	His	Thr	443	483[2]
cFP484	Asn	Gln	Asn	Arg	Ser	Glu	His	His	456	484[2]
amFP486	Thr	Lys	Asn	Arg	Ser	Glu	Ala	His	454	486[3]
mTFP0.7	Asn	Ala	Asn	Arg	Ser	Glu	His	His	453	488[4]
mTFP1	Thr	Ala	Asn	Arg	Ser	Glu	His	His	463	492[5]
green									490−505	500−520
mTFP1 H163M	Thr	Ala	Asn	Arg	Ser	Glu	Met	His	487	503[6]
GFP-S65T	Thr	Thr	Val	Gln	His	Val	Ile	Thr	489	507[7]
cmFP512	Met	Gln	Phe	Arg	Ala	Glu	Lys	His	502	511[8]
Dronpa	Thr	Cys	Asn	Arg	Ser	Glu	Met	His	503	518[9]
yellow									515−525	520−540
YFP	Thr	Gly	Leu	Gln	His	Val	Ile	Tyr	514	528[7]
phiYFP	Thr	Thr	Ala	Gln	His	Leu	Ile	Tyr	525	537[10]

[a]Experimental absorption and emission peak wavelengths for the anionic protonation state of the chromophore are also reported.

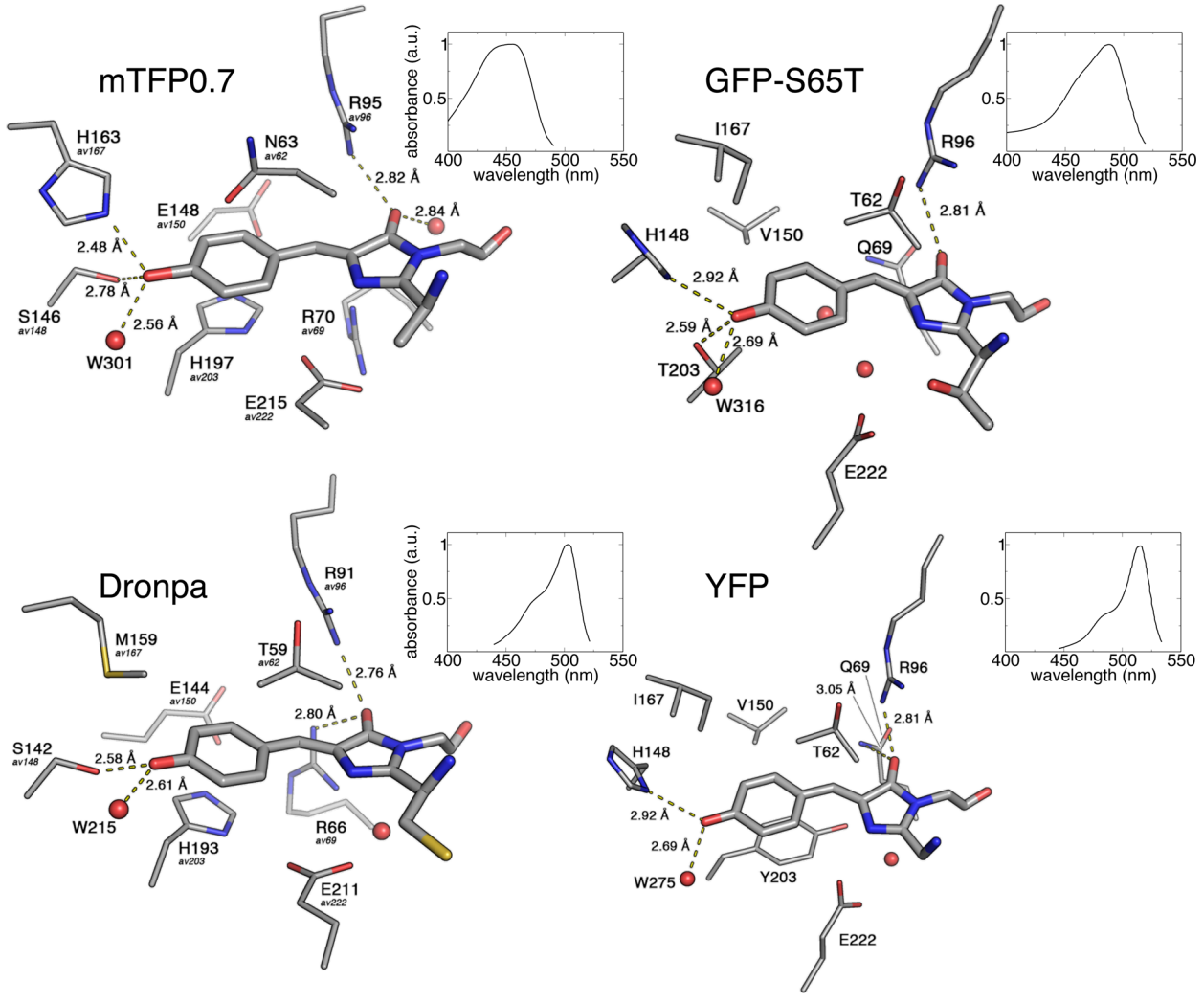


Figure 1. Chromophore environments in the four FPs examined from X-ray structures. The coordinates were taken from the Protein Data Bank, with codes 2OTB[4] (mTFP0.7), 1EMG[7] (GFP-S65T), 2IOV[11] (Dronpa), and 1YFP.[12] The *av* prefix for the residue numbers in italics refers to the corresponding position in *av*GFP after sequence alignment. Figure produced using PyMOL.[13]

The side-chain of the residue at the first position in the chromogenic tripeptide (residue 65 in GFP sequence) differs in these various FPs. However, it does not participate in the π-conjugated system and has only a slight influence on the

dx.doi.org/10.1021/ct3007452 | J. Chem. Theory Comput. 2013, 9, 497−508

excitation peak wavelength. For instance, Dronpa and cmFP512 contain respectively a Cys and a Gln at this location and have remarkably similar excitation peaks (502 and 503 nm). Analogous examples are mTFP0.7 and amFP486 with, respectively, an Ala and a Lys (454 and 453 nm). In contrast, the Ser65Thr mutation in GFP seems to have a somewhat more sizable influence, in that it shifts the anionic state absorption peak from 475 nm (in wild type GFP, state B) to 489 nm (in GFP-S65T). This influence, however, might be an indirect one, rising from the repositioning of Glu222 in the S65T mutant.[7]

Being a rather large set of similar structures, well characterized in terms of spectral and structural properties, FPs are becoming a benchmark for quantum chemistry methods. In particular, the "fine" tuning problem is an excellent test for QM/MM approaches, like the analogous case of retinal in rhodopsins.[18,19] However, computational work aimed at reproducing these spectral shifts is only limited, partly because such an approach requires a completely homogeneous treatment of the models of the different FPs. Prediction of their optical properties, starting from the structure, involves selection of the appropriate QM method and, since *ab initio* simulations of whole FPs are still hardly feasible,[20] careful choice of the models. Hybrid QM/MM approaches, treating the chromophore and possibly some surrounding residues by QM methods and describing the rest of the protein and solvent with MM force fields, are a feasible compromise if one wants to preserve the discrete nature of the surrounding molecular matrix. QM/MM approaches have been applied to FPs, including wild-type GFP[21−23] and Dronpa.[24]

Previous studies have addressed tuning in FPs with different chromophore moieties.[20,25] One of us performed a DFT study of the chromophore and its neighboring residues in various GFP mutants, aimed at understanding the relationship between structure and optical properties.[26] Hasegawa et al. compared the excitation energies in GFP, mKO, and DsRed (a green, orange, and red FP, respectively)[25] at the SAC-CI QM/MM level, while Taguchi et al. studied a YFP and a Blue FP (with His66 in the chromophore) using CIS(D) (configuration interaction singles and perturbative doubles) in conjunction with a fragment molecular orbital scheme.[20] Recently, Filippi and co-workers published extensive results from QM/MM calculations on wild-type GFP,[23] pointing out that the difficulty to reproduce experimental data for FPs may arise from incorrect building of the MM subsystems (specifically, regarding commonly assessed protonation of some amino acids) and/or failure to take into account all relevant interactions at stake (namely, a purely electrostatic, non-polarizable MM subsystem may not prove sufficient).

For our investigation, we selected four proteins within the group of FPs containing the GFP anionic chromophore: one of the cyan−teal group, mTFP0.7;[4] two of the green group, Dronpa[9] and GFP-S65T;[7] and one of the yellow group, YFP (GFP-S65G/V68L/S72A/T203Y).[12] The chromophore and its surroundings in these proteins are shown in Figure 1. In spite of displaying different spectra, Dronpa and mTFP0.7 have similar chromophore surroundings and share a higher homology than with GFP.

We examine different choices of starting structures, with or without equilibration by molecular dynamics simulations, and different choices of the QM system in the QM/MM calculation, either the chromophore alone or a cluster containing also the surrounding residues. We perform DFT QM/MM optimization and calculate the excitation energy with three methods: TD-DFT, CASPT2, and SAC-CI. Variations in the chromophore structure due to interaction with the protein matrix are discussed and correlated with variation in excitation energy.

METHODS

Preparation of the Models. Models of YFP, Dronpa, mTFP0.7, and the S65T mutant of GFP were obtained by starting from crystal structures collected from the PDB databank with codes 2YFP,[12] 2IOV,[11] 2OTB,[4] and 1EMG,[7] respectively. Since the 2IOV and 2OTB structures are a tetramer and a dimer, respectively, the first monomer (chain A) was selected in both cases, the difference between the monomers being negligible in the conformation of the amino acids which surround the chromophore (i.e., they display only minor displacements and no side-chain rotations).

Hydrogen atoms were added to the PDB structures through the Leap routine from the Amber suite of programs.[27] Na^+ or Cl^- ions were added as necessary, in order to keep the overall neutral electric charge of the models, and a truncated octahedral box of TIP3P water molecules was added to surround the proteins, with a minimum 8 Å distance from the protein surface to the box surface.

The protonation state of Glu222 in GFP-S65T and YFP was set to neutral, as suggested from X-ray studies.[7,12] The His residue π-stacked with the chromophore in Dronpa and mTFP0.7 (His193 and His197, respectively) participates in a quadrupole salt bridge common in FPs of the Anthozoa family.[28] As a consequence, this His was set to the cationic state, and Glu215 of mTFP0.7 and Glu211 of Dronpa, which correspond to Glu222 in the GFPs, were considered anionic. The other His residues were considered neutral, and their protonation state (δ or ε) was decided according to the surrounding H-bond network, as in ref 29. In the case of mTFP0.7, we also discuss the possibility of cationic His163. For the protonation state of residues in the chromophore surroundings, see Figure 3.

MM Calculations. The force field used for all MM calculations is Amber94,[30] and appropriate force field parameters for the chromophore were obtained as reported by Nifosì and Tozzini.[29] Hydrogens and bulk water molecules were first equilibrated for 100−150 ps, with 200 kcal/mol/Å^2 restraints on heavy atoms (which kept them within 0.1 Å from their original PDB positions), resulting in a first set of structures (labeled PDB_H), to be later refined by QM/MM geometry optimization.

Then, 2 ns of unrestrained MD simulation were carried out for each model. From 1 to 2 ns of such trajectories, six snapshots were extracted at 200 ps intervals, and for each of these snapshots the structure was optimized by simulated annealing (labeled MD-Snapshots).

The coordinates of the protein and a 4-Å-thick shell of water molecules were extracted from each structure, prior to further QM/MM geometry optimization.

QM/MM Geometry Optimization. We performed a QM/MM geometry optimization on all structures for each protein model, where the QM system (see Figure 3 for reference) comprises either only the chromophore ("c1" label) or the chromophore (with a slightly different cut, see Figure 3) and the surrounding residues within a distance of 4 Å ("c2+env" label).

dx.doi.org/10.1021/ct3007452 | J. Chem. Theory Comput. 2013, 9, 497−508

During geometry optimization, the coupling between the MM and QM subsystems was accounted for within the ONIOM scheme with electrostatic embedding.[31] We employed DFT with PBE0[32] as the exchange and correlation functional and a 6-31G* Gaussian basis set for the QM part and the Amber94 force field for the MM system. We stress that chromophore QM optimization is a necessary step, because the resolution of the crystal structures may not be sufficiently high nor the force field accurate enough to allow the direct use of the structures for excitation energy calculations (see Table 1 and Figure 2 in the Supporting Information, SI). QM/MM optimized geometries are available upon request.

Excitation Energies. Excitation energies were calculated on these structures with *ab initio* methods, namely, time-dependent density functional theory (TD-DFT), complete active space self-consistent field (CASSCF) with its CASPT2 correction,[33] and SAC-CI. After considering the performance of different exchange and correlation functionals in TD-DFT[34] for organic molecules,[35] we chose the CAM-B3LYP functional,[36] with the 6-31+G* basis set. SAC-CI[37] calculations were performed at level 2 with the 6-31G* basis set. Both TD-DFT and SAC-CI calculations were performed with Gaussian 09.[38] CASSCF calculations were carried out with the MOLCAS 7.6[39] package using the ANO-S-VDZP basis set[40] and a CAS(12,11). Cholesky decomposition of the two-electron integrals[41] was used for all models, and the default IPEA[42] zeroth-order Hamiltonian was employed in the CASPT2 calculations. Default criteria for convergence were used for all calculations.

For the calculation of excitation energies, QM/MM coupling was treated within a scheme slightly different than ONIOM, whereby all MM partial charges are retained, except for those belonging to MM link atoms which are substituted by capping QM atoms. Moreover, in the present scheme, when the backbone of an amino acid is split between the MM and the QM subsystems, the sum of the MM partial charges of the atoms now included in the QM subsystem is evenly redistributed over the remaining backbone MM atoms of the same amino acid, in order to maintain an overall charge neutrality of the model.

Figure 2 summarizes the various models used in the calculations, together with the procedures followed to obtain them.

RESULTS

PDB_H(c1) Models. As described above, the first set of structures was obtained from the X-ray coordinates of the four FPs under scrutiny, optimizing the geometry of the (c1) QM subsystem (PDB_H(c1) models in Figure 2). The excitation energies, calculated with different methods, are reported in Table 2. We also performed QM optimization and TD-DFT calculations on the four isolated (gas phase) chromophore structures (c1), obtaining values between 3.11 and 3.12 eV (data not shown), supporting the conclusion that the side chain at the first position in the chromophore tripeptide does not significantly influence the excitation properties.

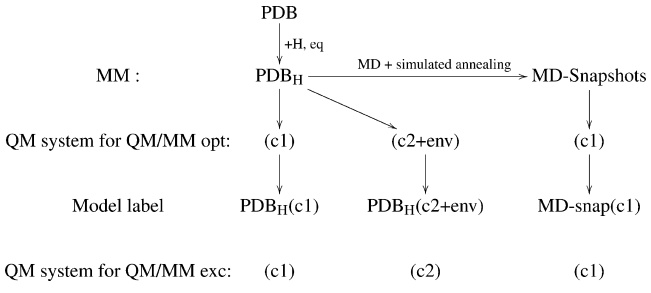


Figure 2. Scheme of the structures used in the calculations. In the QM/MM optimization, the ONIOM scheme with electrostatic embedding was used, keeping the MM part frozen, unless specifically stated.

Given that different choices were available to build the model of mTFP0.7, we prepared three different structures, the first two (mTFP0.7/$N63_{Xray}$ and mTFP0.7/$N63_{flip}$) differ for the orientation of Asn63, while the third (mTFP0.7/$N63_{flip}$/$H163^+$) contains a protonated His163 and a flipped Asn63. In the 2OTB PDB structure of mTFP0.7, the Asn63 carbonyl oxygen points toward the chromophore phenolate (see Figure 1). However, the rotamer which features the amide group pointing toward the phenolate is compatible with the X-ray structure, as well, because electronic density maps may not distinguish between nitrogen and oxygen, and as a consequence, many structures in the PDB database contain wrong Asn and Gln rotamers.[43] Furthermore, the possibility of His163 protonation was discussed in a recent paper for mTFP1,[44] a close homologue of mTFP0.7.

Choice of the QM System: PDB_H(c2+env) Models. By choosing a bigger QM system, it should be possible to evaluate the accuracy of the QM/MM coupling, both in the structural optimization and in the excitation energy calculation. Starting from the X-ray structures described above, we chose a QM system comprised of all the amino acid side chains in contact with the chromophore, i.e., the (c2+env) in Figure 3. Optimization was performed for the whole QM system, while the MM system was kept frozen. For GFP-S65T, we also tried to optimize a larger part including all MM atoms within a distance of 7 Å from the chromophore; with such an approach, we only obtained negligible variations in the structure and in the final calculated excitation energies (<0.005 eV).

The inclusion of all groups interacting with the chromophore (>150 atoms) is feasible within a TD-DFT treatment, though too heavy for CASPT2 and SAC-CI. Hence, for the latter methods we performed the excitation energy calculations on the chromophore only, (c2) QM system. Table 3 reports the excitation energy on these structures.

The QM/MM TD-DFT calculation on the (c2+env) QM system (second column in Table 3) gives two spectroscopic states for YFP. Those are due to MOs delocalized both on the chromophore and on the phenolic ring of Tyr203. Since no such two states are detected in spectroscopic measurements, we regard this feature as an artifact of the TD-DFT calculation. No such issue arises instead in the other three FPs, where the first excited state is the only one among the low-energy excitations to have sizable oscillator strength. With respect to the values obtained when considering the smaller (c2) QM system in the excitation energy calculation (third column), these TD-DFT results for the three FPs are rather uniformly red-shifted by ~0.2 eV.

MD-snap(c1) Structures. To account for internal flexibility and to correlate variations in excitation energy to conformational changes, we performed 2-ns-long MD simulations on each of the FPs, without any restraints. MD also allows us to test the plausibility of Asn63 flipped side chain and His163 protonation in mTFP0.7. Once again, for mTFP0.7 we simulated both the neutral and protonated His163 cases. The calculated excitation energies are reported in Tables 4−8.

Table 2. Comparison between Experimental Excitation Maxima and QM/MM Excitation Energies (in eV) of the PDB_H(c1) Models[a]

	TD-DFT (OS)	Δ	CASSCF (OS)	Δ	CASPT2 (OS)	Δ	SAC−CI (OS)	Δ	exptl	Δ
YFP	3.138 (1.13)	0.00	3.336 (1.36)	0.00	2.857 (1.16)	0.00	2.339 (0.87)	0.00	2.41[b]	0.00
Dronpa	3.150 (1.10)	0.01	3.276 (1.33)	−0.06	2.884 (1.17)	0.03	2.504 (0.92)	0.16	2.46[c]	0.05
GFP-S65T	3.191 (1.01)	0.05	3.392 (1.22)	0.06	3.006 (1.02)	0.15	2.493 (0.79)	0.15	2.54[d]	0.13
mTFP0.7 $N63_{Xray}$	3.192 (1.04)	0.05	3.468 (1.22)	0.13	2.923 (1.03)	0.07			2.74[e]	0.33
mTFP0.7 $N63_{flip}$	3.259 (1.02)	0.12	3.696 (1.15)	0.36	3.059 (0.99)	0.20	2.562 (0.81)	0.22	2.74[e]	0.33
mTFP0.7 $H163^+$	3.344 (0.99)	0.21	3.985 (1.10)	0.65	3.266 (0.96)	0.41	2.743 (0.80)	0.40	2.74[e]	0.33

[a]The Δ columns report the difference from the YFP value within each set. [b]From ref 7. [c]From ref 9. [d]From ref 7. [e]From ref 4.

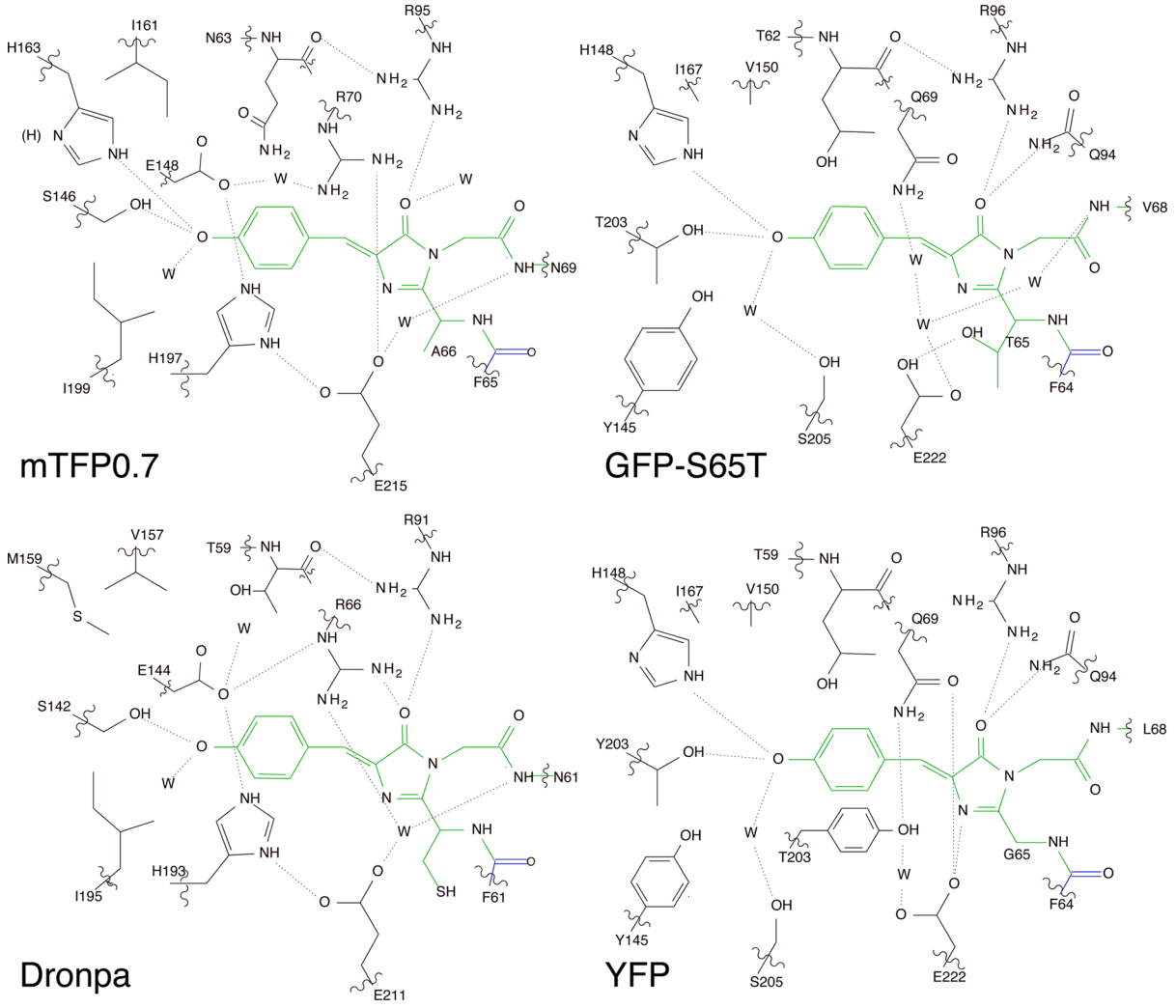


Figure 3. The QM systems used in QM/MM geometry optimization and excitation energy calculations. The smaller (c1) QM system is shown in green. The (c2) QM system is shown in green+blue, while the largest (c2+env) QM system is green+blue+black. The cut bonds are capped with hydrogen link atoms.

Table 3. QM/MM Excitation Energies of the PDB_H(2+env) Models

	TD-DFT[a] (OS)	TD-DFT[b] (OS)	Δ	CASPT2[b] (OS)	Δ	SAC-CI[b] (OS)	Δ	exptl	Δ
YFP	2.642 (0.13) 3.064 (0.82)	3.160 (1.19)	0.00	2.922 (1.19)	0.00	2.388 (0.91)	0.00	2.41	0.00
Dronpa	2.964 (0.81)	3.149 (1.13)	−0.01	2.914 (1.18)	−0.01	2.414 (0.89)	0.03	2.46	0.05
GFP-S65T	2.947 (1.00)	3.153 (1.07)	−0.01	2.864 (1.08)	−0.06	2.410 (0.83)	0.02	2.54	0.13
mTFP0.7	3.066 (0.82)	3.229 (1.06)	0.07	2.993 (1.00)	0.07	2.516 (0.82)	0.13	2.74	0.33
mTFP0.7 $H163^+$	3.186 (0.79)	3.348 (1.02)	0.19	3.220 (0.94)	0.30	2.785 (0.81)	0.40	2.74	0.33

[a]Excitation energies calculated on the (c2+env) QM subsystem. [b]Excitation energies calculated on the (c2) QM subsystem.

dx.doi.org/10.1021/ct3007452 | J. Chem. Theory Comput. 2013, 9, 497−508

Table 4. YFP MD-snap(c1)

	TD-DFT	CASPT2
1.0 ns	3.179	2.929
1.2 ns	3.181	2.915
1.4 ns	3.171	2.917
1.6 ns	3.158	2.881
1.8 ns	3.191	2.927
2.0 ns	3.172	2.897
ave	3.175 (0.011)	2.911 (0.018)

Table 5. Dronpa MD-snap(c1)

	TD-DFT	CASPT2
1.0 ns	3.146	2.905
1.2 ns	3.133	2.892
1.4 ns	3.156	2.909
1.6 ns	3.137	2.810
1.8 ns	3.152	2.877
2.0 ns	3.171	2.885
ave	3.149 (0.014)	2.880 (0.036)

Table 6. GFP-S65T MD-snap(c1)

	TD-DFT	CASPT2	SAC−CI
1.0 ns	3.166	2.911	2.394
1.2 ns	3.189	2.996	
1.4 ns	3.147	2.879	
1.6 ns	3.175	2.945	2.479
1.8 ns	3.209	3.019	2.495
2.0 ns	3.165	2.938	
ave	3.175 (0.021)	2.948 (0.052)	2.456 (0.054)

Table 7. mTFP0.7 MD-snap(c1)

	TD-DFT	CASPT2
1.0 ns	3.261	2.998
1.2 ns	3.261	2.982
1.4 ns	3.287	3.027
1.6 ns	3.251	2.942
1.8 ns	3.229	2.956
2.0 ns	3.244	2.975
ave	3.256 (0.019)	2.980 (0.030)

Table 8. mTFP0.7 H163$^+$ MD-snap(c1)

	TD-DFT	CASPT2	SAC−CI
1.0 ns	3.239	3.000	2.528
1.2 ns	3.258	2.982	2.544
1.4 ns	3.250	2.994	2.548
1.6 ns	3.245	2.971	2.535
1.8 ns	3.263	3.021	2.585
2.0 ns	3.253	2.968	2.531
ave	3.251 (0.009)	2.985 (0.013)	2.545 (0.021)

The molecular dynamics simulations on mTFP0.7 reveal a flip of the Asn63 side chain, with respect to the PDB_H structure, taking place after few tens of picoseconds both in the neutral and in the protonated His163 case. This supports the conclusion that the correct conformer of Asn63 has the amide pointing toward the chromophore, and we shall henceforth rely on this attribution.

In mTFP0.7 H163$^+$, we observe a distortion with respect to the X-ray structure, in the backbone of 163 and 164 (see Figure 3 in the SI). In addition, the hydrogen bond between the chromophore and Ser146 is not stable, Ser146 being preferentially H-bonded to W301, rather than to the chromophore (see Figure 4 in the SI). The values for the

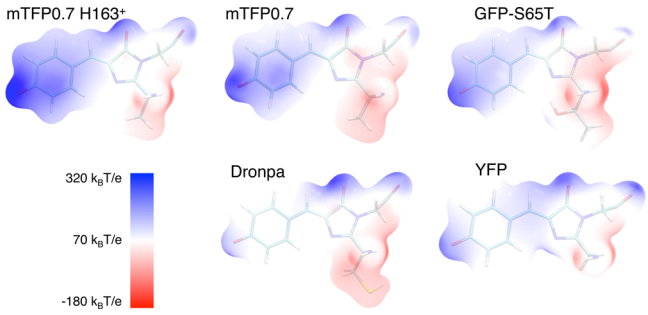


Figure 4. Electrostatic potential due to the MM charges is color-mapped around the chromophore in the PDB_H(c1) models. Figure produced using VMD.[47]

MD-snap(c1) models of mTFP0.7 H163$^+$ are considerably lower (0.2−0.3 eV) than those obtained from the PDB_H(c1) structure, resulting in the same excitation energy for the unprotonated His163 case (Table 7). Overall, the effect of the His163$^+$−chromophore salt bridge is reduced by the loss of the hydrogen bond with Ser146.

Further analysis of the mTFP0.7 MD trajectory reveals a conformational change taking place after ∼1.5 ns and involving the exposure of Ser146 side chain to the solvent from the original configuration pointing toward the protein interior and H-bonded to the chromophore phenolate. As a consequence of this conformational change, the excitation energy in the last three snapshots is reduced, on average, by around 0.02 eV in the TD-DFT and 0.04 eV in the CASPT2 results. However, this configurational change does not seem to be compatible with the X-ray structure and might be either an artifact of the MD simulation or a sign of the rather high structural inhomogeneity of cyan FPs mentioned in the Introduction. We mention that in the X-ray structure of the mTFP0.7 *trans* chromophore state Ser146 exists in two configurations, one pointing inward and the other exposed to the solvent.

In YFP, the MD structures show some differences with respect to the X-ray structures in the position of Tyr145 phenolate, with Tyr145-OH closer to the chromophore phenolate. According to the bond-length analysis provided at the end of the Results section, this change mostly accounts for the slightly blue-shifted excitation energy values of the YFP MD-snap(c1) with respect to the PDB_H(c1) (3.175 vs 3.138 eV in TD-DFT and 2.911 vs 2.857 eV in CASPT2).

The results we obtained are quite robust with respect to different MD procedures. If no MD simulated annealing is performed, the chromophore (c1) being QM/MM-optimized as before, the excitation-energy values differ for the single snapshots, as a consequence of small structure rearrangements, but the average is stable. In Dronpa, for example, an average of 3.138 (0.017) eV for the TD-DFT values is obtained, to be compared with 3.146 eV of the annealed structures, the difference being within the standard deviation.

Origin of Excitation Energy Shift between Dronpa and mTFP0.7. The chromophore surroundings of Dronpa and mTFP0.7 are remarkably similar. Despite this similarity, the measured excitation energy peak of mTFP0.7 is 0.27 eV higher than that of Dronpa. Considering the neutral His163 case, our

dx.doi.org/10.1021/ct3007452 | J. Chem. Theory Comput. 2013, 9, 497−508

calculations reproduce part of this shift, about 0.12 eV in TD-DFT and 0.15 eV in CASPT2. Three relevant differences emerge from the structural comparison (see Figure.1): Asn63 in mTFP0.7 is replaced by Thr59 in Dronpa. His163 in mTFP0.7, participating in a H bond with the chromophore phenolate, is replaced in Dronpa by apolar Met159. The arginine (Arg70 in mTFP0.7 and Arg66 in Dronpa) in the quadrupole "below" the chromophore adopts different conformations in the two proteins, as a consequence of a different amino acid at position +3 with respect to arginine, namely Thr73 in mTFP0.7 and Ala69 in Dronpa. Mutating Thr73 into Ala indeed results in a red shift of the experimental excitation peak.[45]

In order to determine if these three factors are indeed responsible for the calculated shift, as well as assess their relative importance, we performed computational "mutagenesis" trials. We took the configurations (with respect to the chromophore) of the side chain of Asn63, His163, and Arg70 from the PDB_H(c1) structure of mTFP0.7 and substituted them in the corresponding sites of Dronpa, namely, Thr59, Met159, and Arg66. We also selectively set to zero the MM charges on some relevant residues (Table 10). In all the "mutant" structures the chromophore (c1 system) was optimized with the usual ONIOM scheme, allowing the substituted residues to move during optimization. Rather than predicting the structure and optical properties of the mutants, our aim here is to evaluate the contribution of the relevant residues in the chromophore neighboring.

The results reported in Table 9 show that the biggest blue shift in TD-DFT is generated by the different configuration of Arg66. Indeed when the charge of Arg66 in Dronpa is put to 0, a 0.06 eV blue shift is also observed (Table 10). Smaller shifts are due to mutations Thr59Asn and Met159His. When all three changes are incorporated, a good part (0.05 eV vs 0.11 eV) of the blue shift from Dronpa to mTFP0.7 is reproduced in the TD-DFT calculations. The CASPT2 results are less well correlated, in particular the change in conformation of Arg66 produces a lower rather than a higher excitation energy. However, when including all three modifications, a blue shift of 0.06 eV is observed also in CASPT2. The missing part of the shift presumably comes from other differences in the environment, not accounted for with our substitutions.

Table 9. Mutagenesis from Dronpa to mTFP0.7, Starting from the PDB_H(c1) Models and Using the (c1) QM Subsystem

	TD-DFT	Δ	CASPT2	Δ	exptl	Δ
Dronpa	3.150	0.00	2.884	0.00	2.47	0.00
Dronpa Thr59Asn	3.150	0.00	2.879	0.00		
Dronpa Thr59Asn/ Met159His	3.155	0.01	2.889	0.01		
Dronpa Arg66 conf	3.172	0.02	2.818	−0.06		
Dronpa Thr59Asn/ Met159His/ Arg66 conf	3.201	0.05	2.944	0.06		
mTFP0.7	3.259	0.11	3.036	0.15	2.74	0.27

Table 10. Excitation Energies Calculated by Switching off the Charge of Selected Residues, Starting from the PDB_H(c1) Models and Using the (c1) QM Subsystem

	TD-DFT	Δ
Dronpa	3.150	0.00
Dronpa Met159→0	3.154	0.00
Dronpa His193→0	3.172	0.02
Dronpa Thr59→0	3.153	0.00
Dronpa Arg66→0	3.208	0.06
mTFP0.7	3.259	0.00
mTFP0.7 His163→0	3.204	−0.06
mTFP0.7 His197→0	3.221	−0.04
mTFP0.7 Asn63→0	3.208	−0.06

Electrostatic Potential. Figure 4 shows the electrostatic potential (ESP) around the chromophore due to the MM charges in the PDB_H(c1) models. The ESP variation highlights the strength of chromophore−protein matrix interactions at different chromophore sites. In particular, the ESP on the phenolate carbonyl region is higher in mTFP0.7 $H163^+$, mTFP0.7, and GFP-S65T (in this order) and lower in Dronpa and YFP, as a consequence of the three vs two chromophore−protein matrix H bonds. At the imidazolidinone carbonyl site, the ESP is higher in Dronpa, because of the two H-bonded arginines (Arg66 and Arg91). The slightly higher ESP around YFP imidazolidinone N2 atom is due to the presence of protonated Glu222. In the two mTFP0.7's the higher ESP around the center of the phenolate is due to the partial positive charge on the N63 amide group pointing toward the chromophore. Interestingly, the cationic His197 stacked "below" the chromophore in mTFP0.7 and Dronpa does not lead to increased ESP in the phenolate region, with respect to GFP-S65T and YFP, possibly because the positive charge is balanced by the nearby negatively charged Glu side chains (215 and 148 in mTFP0.7, 211, and 144 in Dronpa).

Chromophore Geometry. The chromophore geometries in all optimized structures are rather planar. The bridge dihedrals τ (N2−CA2−CB2−CG2, see Figure 5 for chromophore atom names) and ϕ (CA2−CB2−CG2−CD1) show maximum deviations in mTFP0.7, with values around −10° for ϕ. The dihedral values in all models and in the original X-ray structures are available in the SI and show no particular correlation with the excitation energy. Though the deviations of these dihedrals from the PDB structures are sizable, for Dronpa and mTFP0.7 in particular, the root-mean-square distance of atom coordinates of the chromophore π system are below 0.15 Å.

In Figure 5, we compare the variations in chromophore bond lengths across the structures. The plots report the differences between bond lengths in the structures and those of the gas-phase (anionic) model chromophore optimized within the same DFT PBE0 method (reported in the SI). With respect to the gas phase chromophore, CZ−OH and C2−O2 bonds become longer in all structures, mostly because of the H bonds/electrostatic interactions with the protein matrix. The amount of this variation nicely reflects the number and the strength of such interactions, as revealed by the electrostatic maps. The OH−CZ bond is longer in mTFP0.7 $H163^+$, mTFP0.7, and GFP-S65T, with three H bonds between OH and the protein matrix (see Figure 1) than in Dronpa and YFP, with only two such bonds. A characteristic of Dronpa is a longer C2−O2 bond length, resulting from the H-bond interaction with two arginines (Arg91 and Arg66).

Bond lengths in the MD-snap(c1) do not differ sensibly from their PDB_H(c1) counterparts. In YFP, a somewhat larger difference is observed in the bridge region of the chromophore, presumably a consequence of the repositioning of Tyr145. This

dx.doi.org/10.1021/ct3007452 | J. Chem. Theory Comput. 2013, 9, 497−508

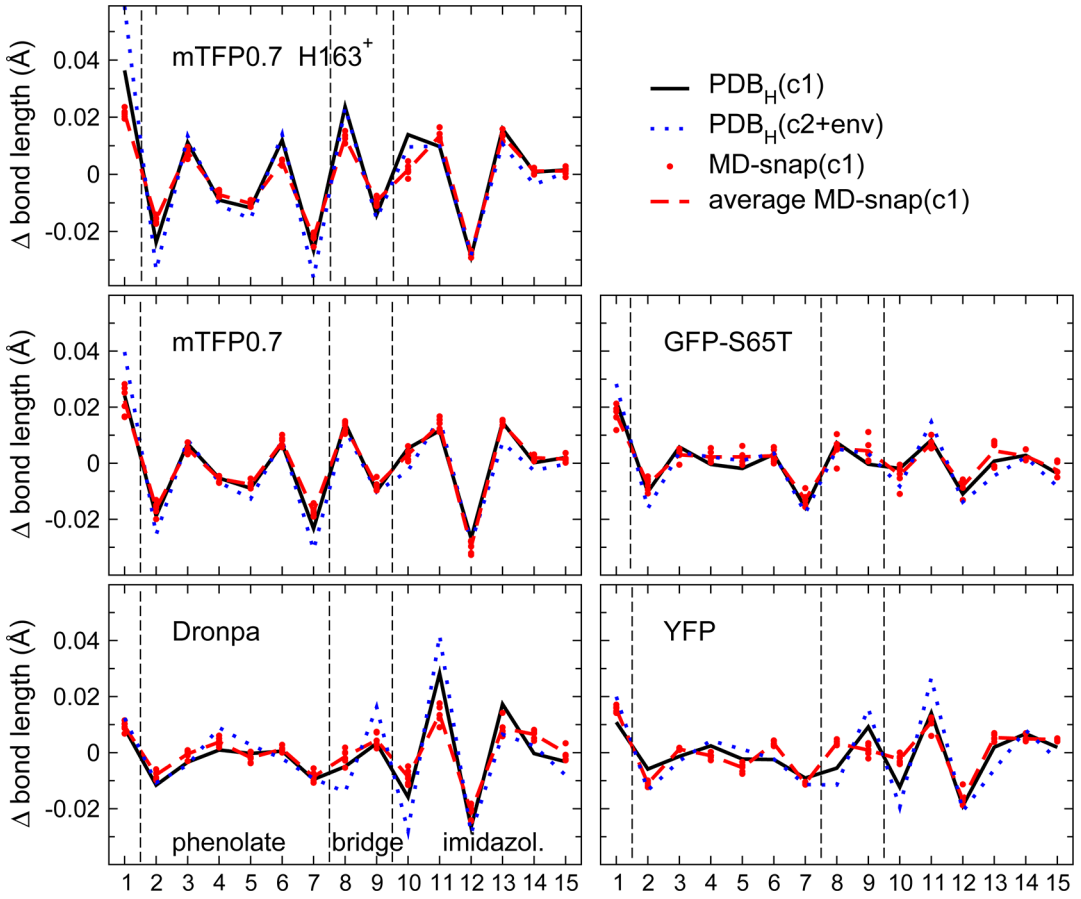


Figure 5. Comparison between Δ chromophore bond lengths (obtained by subtracting the corresponding value in the optimized gas-phase model chromophore, available in the SI) in the PDB_H(c1), PDB_H(c2+env), and MD-snap(c1) models. The numbering of bond lengths and the PDB names of the chromophore atoms are shown in the top part.

repositioning brings Tyr145 OH closer to the phenolate, and though no new H-bond is formed, this electrostatic interaction influences the chromophore bond lengths. In Dronpa, the C2−O2 distance is reduced in the MD-snap(c1) models. This can be attributed to a less ideal H-bond geometry of Arg91 and Arg66 with the chromophore O2.

Chromophore bond lengths of the PDB_H(c2+env) models differ mostly for the atoms participating in H bonds with the environment, i.e., OH−CZ and C2−O2. These differences are due to the QM description of the environment rather than to the fact that the environment is no longer constrained during optimization. Indeed, optimization of the QM-chromophore leaving the same atoms free to move leads to irrelevant differences.

To search for correlation between chromophore bond lengths across the various structures and the excitation energies, we performed a principal component (PC) analysis on the matrix of the bond-length sets. Let $\mathbf{b}^a = (b_1^a, b_2^a, ..., b_{15}^a)$ be the vector containing the 15 chromophore (heavy-atoms) bond lengths of structure a and $\tilde{\mathbf{b}}^a = (b_1^a - \bar{b}_1, b_2^a - \bar{b}_{15}, ..., b_{15}^a - \bar{b}_{15})$ be the vector of variations with respect to the average bond length in the set of N structures ($\bar{b}_i = \sum_a^N (b_i^a/N)$). The first principal component vector $\mathbf{v^1} = v_1^1, v_2^1, ..., v_{15}^1$ specifies the direction of the line in the Euclidean 15-dimension space along which the sum of squares of the distances of the points from the line is minimal, or, equivalently maximizes the sum $\sum_a (\tilde{\mathbf{b}}^a \cdot \mathbf{v}^1)^2$. Basically, the first vector contains the coefficients of the linear combination ($\tilde{\mathbf{b}}^a \cdot \mathbf{v}^1$) of bond lengths accounting for as much variability in the set of bond lengths as possible.

The sets are well described by the first two PC vectors, the components of which are plotted in Figure 6, for the collections of sets used in the analysis. The first PC vector can be associated with the bond-length alternation in the quinonoid structure (Figure 6, top). Indeed, the vector components change sign according to the changing nature (single or double) of the corresponding bond. The second-vector components mostly involve the C2−O2, C2−N3, and lengths.

We found a remarkable correlation between the projection onto the first PCA vector and the calculated excitation energies (Figure 7). The correlation is higher when considering TD-DFT rather than CASPT2 values, because the former are calculated in the same DFT scheme of the structural optimizations. Clearly, the degree of quinonoid structure is the main determinant of the excitation energy; i.e., interactions driving the chromophore toward a more quinonoid form (with lower PC component) decrease its excitation energy. The neutral gas-phase chromophore, which has an exclusively benzenoid character, features a much larger PC component and fits perfectly well in the correlation (upper left panel of Figure 7). With an increasing degree of quinonoid structure, the CZ−OH bond in the phenolate and the CG2−CB2 single bond in the bridge become shorter. Conversely, the C2−O2 bond on the imidazolidinone and the double bond in the bridge gets longer. Interestingly, however, the C2−O2 distance contributes only marginally to the first PCA vector (Figure

dx.doi.org/10.1021/ct3007452 | J. Chem. Theory Comput. 2013, 9, 497−508

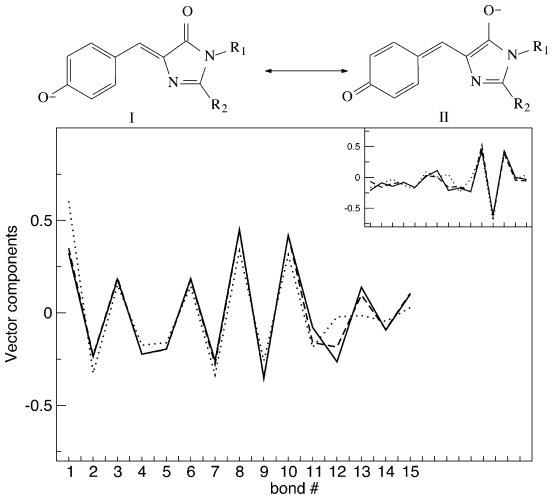


Figure 6. Top: Chromophore main resonance structures, benzenoid (I) and quinonid (II). Bottom: Components of the first PC vector (second in the inset). The continuous line refers to the analysis performed on $PDB_H(c1)$ and MD-snap(c1) sets. The dotted lines refer to the same sets as before with the addition of the gas-phase chromophores.

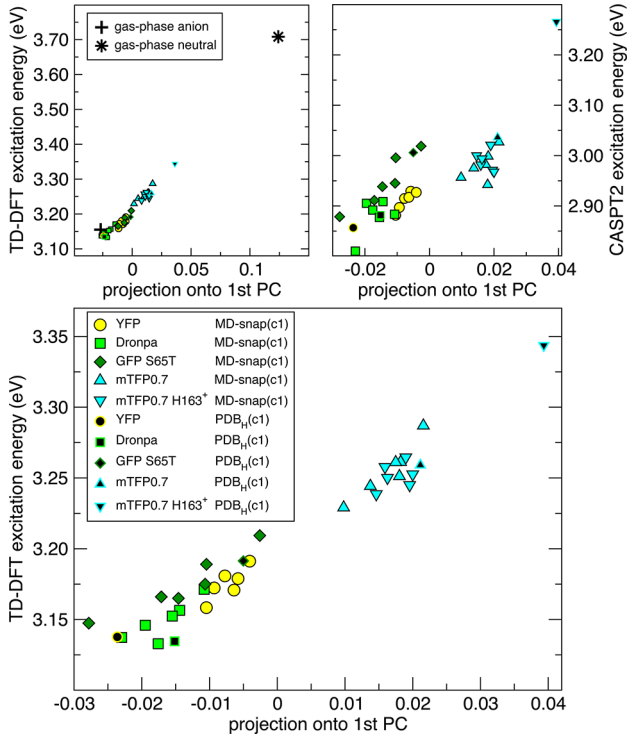


Figure 7. Calculated excitation energies (TD-DFT lower panel and left upper panel, CASPT2 left right panel) vs projections of the corresponding structure on the first principal component (see text).

6), despite changing character in going from the benzenoid to the imidazolidinone resonance structure. Presumably the "information" on the varying length of C2−O2 is already accounted for in the CA2−C2 length. The C2−O2 bond length is instead accounted for in the second PCA vector, also including C2−N3 and N3−C1, whose projection shows no correlation with the excitation energies (data not shown).

As shown in Figure 8, the projections onto the first PC of the $PDB_H(c1)$ models display a strict correlation with the

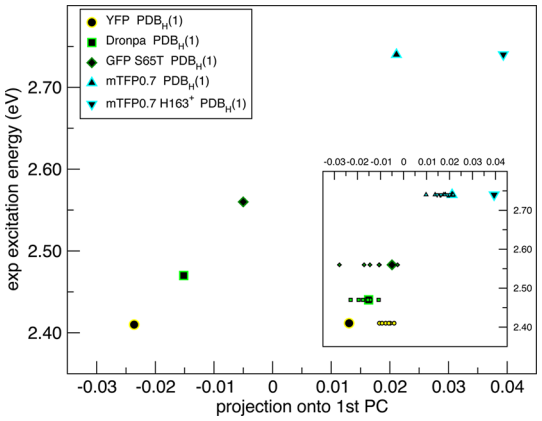


Figure 8. Experimental excitation peaks vs projections on the first principal component of the corresponding chromophore structure in the $PDB_H(c1)$ models. The inset shows also the projections of MD-snap(c1) models.

experimental peaks. Though this data set is too small for drawing strong conclusions, it is interesting to note that the mTFP0.7 $H163^+$ projection deviates from the alignment of the other four structures, suggesting that the benzenoid character of this structure is more than what is needed to explain mTFP0.7 blue-shifted excitation energy. In the case of YFP, the projections of MD-snap(c1) models shift to higher values with respect to the $PDB_H(c1)$ models, as a consequence of the already described conformational changes observed during MD.

DISCUSSION

The focus of this paper is understanding the origin and assessing the reproducibility by computational methods of spectral fine-tuning in FPs containing the same π-conjugated chromophore structure, namely the GFP-like anionic chromophore. We compare three different strategies for producing the starting models on which vertical excitation energies are calculated. All geometry optimizations use an ONIOM-(PBE0/6-31G*:AMBER) scheme. The first two strategies differ in the choice of the QM system, namely, either only the chromophore (c1) or the chromophore and neighboring residues (c2+env) in the ONIOM optimization, both starting from X-ray structures with MD equilibrated hydrogen atoms ($PDB_H(c1)$ and $PDB_H(c2+env)$ models, respectively). In the third set, we performed unrestrained MD simulations and used simulated-annealed MD snapshots as starting structures for ONIOM optimization using the small (c1) QM system (MD-snap(c1) models). We wish to stress the importance of treating the various proteins in a completely homogeneous way, as different choices of methods and models lead to sizable deviations, which become particularly relevant when the focus is on the tuning by few tenths of an electronvolt. The calculated excitation energies are summarized in Figure 9 and compared with the experimental values.

Overall, the CAM-B3LYP TD-DFT excitation energies are considerably blue-shifted (~0.8 eV) with respect to the experimental values, and the shift within the different FPs is only partially reproduced. CASPT2 values are also blue-shifted with respect to experimental values, though by a smaller amount (~0.4 eV). In the SAC-CI calculations the excitation energies are generally closer to experimental values. In $PDB_H(c1)$ calculations, the blue shift of mTFP0.7 with respect to the other FPs is better reproduced (albeit overestimated in

dx.doi.org/10.1021/ct3007452 | J. Chem. Theory Comput. 2013, 9, 497−508

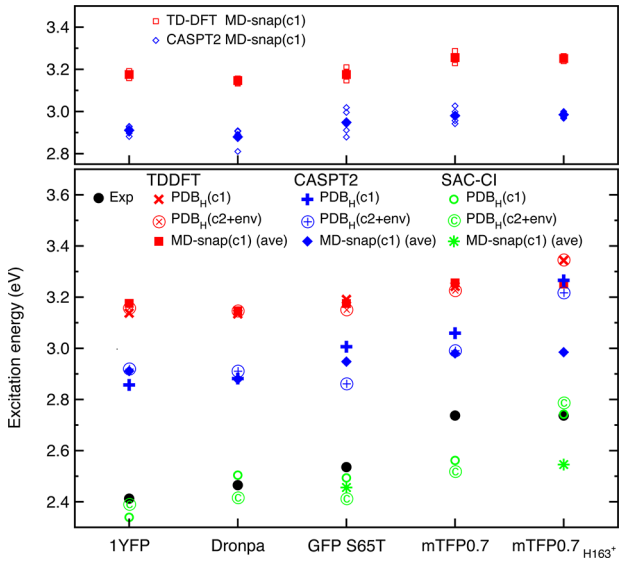


Figure 9. Excitation energies calculated with different combination of methods and structure optimization strategies compared with the experimental values.

CASPT2 and SAC-CI calculations) when considering protonated His163 (mTFP0.7 H163$^+$). However, the results on mTFP0.7 H163$^+$ are challenged by the molecular dynamics simulations.

Enlarging the QM system (PDB_H(c2+env), second set of structures) should provide a better description of the electrostatic interactions of the chromophore with the neighboring residues, despite the known deficiencies of DFT for the description of dispersive interactions. The chromophore bond lengths calculated in the two schemes (PDB_H(c1) and PDB_H(c2+env)) show a good agreement, with maximum differences of 0.02 Å, and overall RMS deviations of ∼0.0015 Å. Maximum deviations are obtained for bond lengths, such as OH−CZ in mTFP0.7 and C2−O2 in Dronpa and YFP, involving atoms that participate in H bonds with the environment. The excitation energies calculated using the (c2) QM system are similar (within 0.05 eV) to those in the PDB_H(c1) set, with the exception of a 0.09 eV red-shift of Dronpa in the SAC-CI value and a 0.07 eV red shift in the mTFP0.7 CASPT2 value. It must be stressed, however, that whereas in the PDB_H(c1) calculations the chromophore "feels" the same interactions in the excitation energy calculation and in the geometry optimization, in the PDB_H(c2+env) models the interactions of the chromophore with the near environment are replaced by the unpolarizable partial-charge model, because the QM/MM excitation energies are calculated with a (c2) QM system. This mismatch might be at the basis of the observed deviations in the excitation energy of mTFP0.7 in CASPT2 and of Dronpa in SAC-CI. Within TD-DFT we were also able to calculate the QM/MM excitation energy considering the bigger (c2+env) QM system. Apart from the spurious case of YFP, these excitation energies are rather homogeneously red-shifted by about 0.2 eV with respect to the TD-DFT values on the (c2) QM system (data not shown in Figure 9, see Table 3). So whereas the QM system enlargement has a relatively low impact on the geometry optimization, it has a sizable (though uniform in the set of structures) effect in the TD-DFT calculation of the excitation energy, improving the overall agreement with the experimental values.

The MD-snap(c1) models generally give results in line with those on the PDB_H(c1) models, both regarding the excitation energies and the chromophore structures, with the notable exception of mTFP0.7 H163$^+$. In the latter case MD simulations predict the loss of the H bond between Ser146 and other rearrangements that significantly lower the excitation energy, shifting it back to the neutral His163 case (see below and Figures 3 and 4 of the SI). Hence, the MD-snap(c1) results do not validate the quite large blue-shifted excitation energies of mTFP0.7 H163$^+$.

Regarding the excitation energy trend, the PDB_H(c1) set of calculations best reproduces the ordering of excitation energies, while the extension of the QM system in the ONIOM optimization procedure somewhat worsens this trend. When one excludes the dubious case of mTFP0.7 H163$^+$, the shifts are smaller (0.1 eV in TD-DFT and 0.2 eV in CASPT2/CAS-CI) than the experimental values (0.3 eV) with all QM methods and models.

CASPT2 and SAC-CI have very similar shifts, with an offset of 0.5 eV. Subtracting the 0.5 eV offset from CASPT2 results, the two methods agree within 0.05 eV, apart from the case of Dronpa PDB_H(c1), the results of which blue-shifted about 0.1 eV than expected from the trend (excitation energy is 2.88 eV in CASPT2 and 2.50 eV in SAC−CI).

Regarding the comparison with results from other computational studies, our SAC-CI, CASPT2, and TD-DFT excitation energies for Dronpa agree with the results of Morokuma and co-workers.[24] They find a value of 2.71 eV for the Dronpa anionic state using an ONIOM(CASSCF:AMBER) geometry and CASPT2 with the original zeroth-order Hamiltonian (no IPEA shift). Taking into account the blue shift found when using the IPEA Hamiltonian in the GFP chromophore,[23] our values around 2.90 eV are in line with their findings. By using SAC-CI, they obtain a value of 2.42 eV, which, considering the different optimization strategy (ONIOM/CASSCF), is in line with our values (2.40−2.50 eV).

The source for mTFP0.7 blue-shifted excitation energy is confirmed to be the combination of three hydrogen bonds to the phenolate and the absence of the second arginine H-bond donor to the imidazolidinone carbonyl. Indeed, we showed that a good part of the blue shift in going from Dronpa to mTFP0.7 can be reproduced substituting Dronpa's Met159 with a His (H-bonded to the chromophore phenolate) and repositioning Arg63 away from the imidazolidinone carbonyl in the Dronpa structure.

Concerning the protonation state of His163 in mTFP0.7, we cannot be conclusive. On one side, the excitation energy is blue-shifted using PDB_H(c1) models with protonated His163, improving the agreement with the experimental shift. On the other, fully equilibrated MD structures with protonated His163 show a slight but significant rearrangement of the 162−164 backbone and the loss of the chromophore−Ser146 H bond. As a result, excitation energies calculated on the His163$^+$ MD-snap(c1) models become similar to the unprotonated His163 case, with a smaller blue shift with respect to the other FPs. Until new spectral and structural studies at different pH's are available to clarify this matter, the His163 protonation state should be treated as an open problem; i.e., additional computational work should consider both charge states. The broadness of the mTFP0.7 absorption spectrum may also result from subpopulations containing different protonation states of the surrounding residues.

dx.doi.org/10.1021/ct3007452 | J. Chem. Theory Comput. 2013, 9, 497−508

The MM description of the environment may be particularly inadequate for YFP. For this protein, the π-stacking interaction of Tyr203 with the chromophore was indicated as one possible source of the red-shifted spectral properties, by a mechanism of increased local polarizability.[12] Clearly, a MM description of Tyr203 with fixed partial charges completely neglects this effect.

Within the QM/MM approach adopted in this work, the main determinant of the excitation energy is the ESP felt by the QM system, due to the MM charges. The ESP acts in modifying the chromophore structure and in selectively stabilizing the ground or excited state electronic density. The other force-field terms, i.e., the covalent terms and the van der Waals interactions, also play a role in structure optimization, but they are less variable than the ESP with a changing chromophore environment in the various FPs. Hence, the behavior of ESP on the chromophore surface (Figure 4) helps to rationalize the calculated shifts in excitation energy. Indeed, upon excitation, a small but sizable charge transfer of 0.1e takes place from the phenolate to the imidazolidinone ring.[46] A higher ESP on the phenolate ring, like in mTFP0.7 and GFP-S65T, will thus increase the energy of the excited state, i.e, blue shift the excitation energy. An opposite, red-shifting, effect originates from a higher ESP on the imidazolidinone.

The analysis of chromophore bond-length variations across the various structures highlights a correlation between excitation energy and the projection on the first vector within a principal component analysis (PCA). The first PCA vector describes the amount of quinonoid structure and is connected to the Bond Length Alternation (BLA). Whereas the three yellow/green FPs examined show overlapping PCA projection, all mTFP0.7 structure projections are rather well separated from the others, indicating a sizable difference in the chromophore structure (Figure 7). The projections of PDB_H(c1) structures correlate remarkably well with the experimental peaks (Figure 8).

This observation confirms and extends a previous study by our group,[26] where the excitation energy was correlated to a suitable linear combination of bond lengths. Similarly to other systems such as the retinal in rhodopsins,[18] also in FPs the BLA correlates with the excitation energy. Such correlation has an experimental counterpart in the observed Raman spectrum of the chromophore in different GFP mutants. By measuring the frequency of a fingerprint Raman mode (mostly involving the CB2−CA2 bond vibration), Tonge and co-workers[48] found that it displays an accurate linear correlation with the excitation energy.

The explanation of spectral tuning in terms of ESP and of chromophore structure are clearly related, because the two sources are hardly separable. Indeed, H-bonds to the phenolate both increase the degree of benzenoid resonance structure and contribute to a higher ESP in the region. Conversely, higher ESP in the imidazolidinone is due to H bonds that also enhance the quinonoid resonance character. A study of the excitation-energy tuning with different, even unrealistic, ESP around the chromophore may help to separate these effects. We preferred not to vary the ESP while maintaining the same chromophore structure, because within this approach one is simulating conditions away from the Franck−Condon region. As a final remark, the rather strong dependency of excitation energies on the fine details of chromophore structure highlights the importance of carefully optimizing the models, or, in a nonzero temperature treatment, considering deviations from the equilibrium structure that are compatible with thermal fluctuations.

CONCLUSIONS

We performed a thorough investigation on the excitation energy variation due to interactions between the chromophore and the protein matrix in fluorescent proteins containing the anionic GFP chromophore. Our results highlight the importance of carefully validating the starting structures by molecular dynamics simulations and QM/MM geometry optimization. We are able to reproduce with certain robustness regarding the choice of the method (CAM-B3LYP TD-DFT, CASPT2, and SAC-CI) and of the structure (X-ray vs MD-equilibrated) part of the blue-shift of mTFP0.7 with respect to the other analyzed FPs, i.e., the structurally similar Dronpa, and two GFP variants, YFP and GFP-S65T. The blue-shifted excitation energy of mTFP0.7 is associated with a pronouncedly different pattern of bond length variations in the chromophore phenolate and bridge, leading to the conclusion that the chromophore in mTFP0.7 has a more conspicuous benzenoid character with respect to the other examined cases. Reproducing the finer excitation-energy ordering among the latter three green/yellow FPs has proven more challenging. This can be attributed to the inadequacy of the electrostatic coupling between QM and MM system, which in our calculations is based on the fixed partial-charges approximation. Such approximation neglects polarization effects and the response of the "external" electronic charge density to the chromophore excitation. Moreover, our modeling does not include any vibronic coupling, since only the vertical excitation energies are calculated, nor finite temperature effects, because the optimization generates zero temperature structures. Taking into account these effects may improve the agreement with experimental measurements.

ASSOCIATED CONTENT

Supporting Information

QM/MM excitation energies calculated on the nonoptimized (apart from hydrogen atoms) X-ray structures, bond lengths in the gas-phase chromophore model optimized at the DFT PBE0/6-31G* level, chromophore τ and ϕ dihedrals in the various models, comparison chromophore bond lengths in the X-ray structures and in the PDB_H(c1) models, backbone around His163 in the various MD simulations, and a comparison between PDB_H and MD-Snapshots in mTFP0.7 $H163^+$. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: r.nifosi@sns.it.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This research has benefited from the allocation of computer time at the CINECA supercomputer center under the IIT Platform "Integrated Multiscale Computational Technology." We gratefully acknowledge useful discussions with Claudia Filippi, Yi Luo, and Adalgisa Sinicropi.

REFERENCES

(1) Chalfie, M.; Tu, Y.; Euskirchen, G.; Ward, W. W.; Prasher, D. C. *Science* **1994**, *263*, 802−805.
(2) Matz, M. V.; Fradkov, A. F.; Labas, Y. A.; Savitsky, A. P.; Zaraisky, A. G.; Markelov, M. L.; Lukyanov, S. A. *Nat. Biotechnol.* **1999**, *17*, 969−973.
(3) Henderson, J. N.; Remington, S. J. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 12712−12717.
(4) Henderson, J. N.; Ai, H.; Campell, R. E.; Remington, S. J. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 6672−6677.
(5) Ai, H.; Henderson, J. N.; Remington, S. J.; Campbell, R. E. *Biochem. J.* **2006**, *400*, 531−540.
(6) Ai, H.-w.; Olenych, S.; Wong, P.; Davidson, M.; Campbell, R. *BMC Biol.* **2008**, *6*, 13.
(7) Elsliger, M.-A.; Wachter, R. M.; Hanson, G. T.; Kallio, K.; Remington, S. J. *Biochemistry* **1999**, *38*, 5296−5301.
(8) Nienhaus, K.; Renzi, F.; Vallone, B.; Wiedenmann, J.; Nienhaus, G. U. *Biochemistry* **2006**, *45*, 12942−12953.
(9) Ando, R.; Mizuno, H.; Miyawaki, A. *Science* **2004**, *306*, 1370−1373.
(10) Shagin, D. A.; Barsova, E. V.; Yahuchevich, Y. H. G.; Fradkov, A. F.; Lukyanov, K. A.; Labas, Y. A.; Semenova, T. N.; Ugalde, J. A.; Meyer, A.; Nunez, J. M.; Widder, E. A.; Lukyanov, S. A.; Matz, M. V. *Mol. Biol. Evol.* **2004**, *21*, 841−850.
(11) Stiel, A. C.; Trowitzsch, S.; Weber, G.; Andresen, M.; Eggeling, C.; Hell, S. W.; Jakobs, S.; Wahl, M. C. *Biochem. J.* **2007**, *402*, 35−42.
(12) Wachter, R. M.; Elsliger, M.-A.; Kallio, K.; Hanson, G. T.; Remington, S. J. *Structure* **1998**, *6*, 1267−1277.
(13) *The PyMOL Molecular Graphics System*, Version 1.3r1; Schrödinger, LLC: New York.
(14) Malo, G. D.; Wang, M.; Wu, D.; Stelling, A. L.; Tonge, P. J.; Wachter, R. M. *J. Mol. Biol.* **2008**, *378*, 871−886.
(15) Dong, J.; Solntsev, K. M.; Tolbert, L. M. *J. Am. Chem. Soc.* **2006**, *128*, 12038−12039.
(16) Pakhomov, A. A.; Martynov, V. I. *Chem. Biol.* **2008**, *15*, 755−764.
(17) Nifosì, R.; Tozzini, V. In *Fluorescent Proteins I*; Jung, G., Ed.; Springer: Berlin, 2012; Springer Series on Fluorescence, Vol. *11*; pp 3−40.
(18) Hoffmann, M.; Wanko, M.; Strodel, P.; König, P. H.; Frauenheim, T.; Schulten, K.; Thiel, W.; Tajkhorshid, E.; Elstner, M. *J. Am. Chem. Soc.* **2006**, *128*, 10808−10818.
(19) Altun, A.; Yokoyama, S.; Morokuma, K. *Photochem. Photobiol.* **2008**, *84*, 845−854.
(20) Taguchi, N.; Mochizuki, Y.; Nakano, T. *Chem. Phys. Lett.* **2011**, *504*, 76−82.
(21) Sinicropi, A.; Andruniow, T.; Ferré, N.; Basosi, R.; Olivucci, M. *J. Am. Chem. Soc.* **2005**, *127*, 11534−11535.
(22) Hasegawa, J.-Y.; Fujimoto, K.; Swerts, B.; Miyahara, T.; Nakatsuji, H. *J. Comput. Chem.* **2007**, *28*, 2443−2452.
(23) Filippi, C.; Buda, F.; Guidoni, L.; Sinicropi, A. *J. Chem. Theory Comput.* **2012**, *8*, 112−124.
(24) Li, X.; Chung, L. W.; Mizuno, H.; Miyawaki, A.; Morokuma, K. *J. Phys. Chem. B* **2010**, *114*, 1114−1126.
(25) Hasegawa, J.-y.; Ise, T.; Fujimoto, K. J.; Kikuchi, A.; Fukumura, E.; Miyawaki, A.; Shiro, Y. *J. Phys. Chem. B* **2010**, *114*, 2971−2979.
(26) Laino, T.; Nifosì, R.; Tozzini, V. *Chem. Phys.* **2004**, *298*, 17−28.
(27) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.;Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S. R.;Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M. J.; Cui, G.; Roe, D. R.;Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A. *AMBER 10*; University of California: San Francisco, CA, 2008.
(28) Remington, S. J.; Wachter, R. M.; Yarbrough, D. K.; Branchaud, B.; Anderson, D. C.; Kallio, K.; Lukyanov, K. A. *Biochemistry* **2005**, *44*, 202−212.
(29) Nifosì, R.; Tozzini, V. *Proteins* **2003**, *51*, 378−389.
(30) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(31) (a) Dapprich, S.; Komáromi, I.; Byun, K.; Morokuma, K.; Frisch, M. J. *J. Mol. Struct.: THEOCHEM* **1999**, *461−462*, 1−21. (b) Vreven, T.; Morokuma, K.; Farkas, Ö.; Schlegel, H. B.; Frisch, M. J. *J. Comput. Chem.* **2003**, *24*, 760−769.
(32) Adamo, C.; Barone, V. *J. Chem. Phys.* **1999**, *110*, 6158−6170.
(33) (a) Andersson, K.; Malmqvist, P.-Å.; Roos, B. O.; Sadlej, A. J.; Wolinski, K. *J. Phys. Chem.* **1990**, *94*, 5483−5488. (b) Andersson, K.; Malmqvist, P.-Å.; Roos, B. O. *J. Chem. Phys.* **1992**, *96*, 1218−1226.
(34) Casida, M. E. In *Recent Advances in Density Functional Methods*; Chong, D. P., Ed.; World Scientific: Singapore, 1995; Vol. *127*, p 155.
(35) Jacquemin, D.; Wathelet, V.; Perpète, E. A.; Adamo, C. *J. Chem. Theory Comput.* **2009**, *5*, 2420−2435.
(36) Yanai, T.; Tew, D. P.; Handy, N. C. *Chem. Phys. Lett.* **2004**, *393*, 51−57.
(37) (a) Nakatsuji, H.; Hirao, K. *J. Chem. Phys.* **1978**, *68*, 2053−2065. (b) Nakatsuji, N. *Chem. Phys. Lett.* **1978**, *59*, 362−364. (c) Nakatsuji, H. *Chem. Phys. Lett.* **1979**, *67*, 329−333.
(38) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 09*, Revision B.01; Gaussian Inc.: Wallingford, CT, 2009.
(39) (a) Aquilante, F.; De Vico, L.; Ferré, N.; Ghigo, G.; Malmqvist, P.-Å.; Neogrády, P.; Pedersen, T. B.; Pitoňák, M.; Reiher, M.; Roos, B. O.; Serrano-Andrés, L.; Urban, M.; Veryazov, V.; Lindh, R. *J. Comput. Chem.* **2010**, *31*, 224−247. (b) Veryazov, V.; Widmark, P.-O.; Serrano-Andrés, L.; Lindh, R.; Roos, B. O. *Int. J. Quantum Chem.* **2004**, *100*, 626−635. (c) Karlström, G.; Lindh, R.; Malmqvist, P.-Å.; Roos, B. O.; Ryde, U.; Veryazov, V.; Widmark, P.-O.; Cossi, M.; Schimmelpfennig, B.; Neogrády, P.; Seijo, L. *Comput. Mater. Sci.* **2003**, *28*, 222−239.
(40) (a) Almlof, J.; Taylor, P. R. *J. Chem. Phys.* **1987**, *86*, 4070−4077. (b) Widmark, P.-O.; Malmqvist, P.-Å.; Roos, B. O. *Theor. Chem. Acc.* **1990**, *77*, 291−306.
(41) Pedersen, T.; Aquilante, F.; Lindh, R. *Theor. Chem. Acc.* **2009**, *124*, 1−10.
(42) Ghigo, G.; Roos, B. O.; Malmqvist, P.-Å. *Chem. Phys. Lett.* **2004**, *396*, 142−149.
(43) Weichenberger, C. X.; Sippl, M. J. *Structure* **2006**, *14*, 967−972.
(44) Topol, I.; Collins, J.; Nemukhin, A. *Biophys. Chem.* **2010**, *149*, 78−82.
(45) Campbell, R. E.; Ai, H.-W. TEAL FLUORESCENT PROTEINS (mTFP1). U.S. Patent US7790868, Sep. 7, 2007.
(46) Martin, M. E.; Negri, F.; Olivucci, M. *J. Am. Chem. Soc.* **2004**, *126*, 5452−5464.
(47) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33−38.
(48) Bell, A. F.; He, X.; Wachter, R. M.; Tonge, P. J. *Biochemistry* **2000**, *39*, 4423−4431.