JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

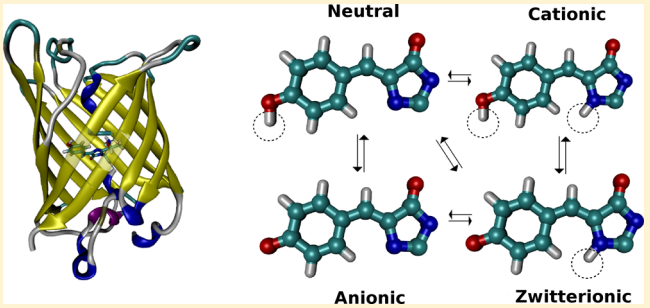
# Studies of pH-Sensitive Optical Properties of the deGFP1 Green Fluorescent Protein Using a Unique Polarizable Force Field

I. Harczuk, N. Arul Murugan, O. Vahtras, and H. Ågren*

Division of Theoretical Chemistry and Biology, School of Biotechnology, KTH Royal Institute of Technology, S-106 91 Stockholm, Sweden

Supporting Information

**ABSTRACT:** The aim of this study is to identify the responsible molecular forms for the pH dependent optical properties of the deGFP1 green fluorescent protein mutant. We have carried out static and dynamic type calculations for all four protonation states of the chromophore to unravel the contributions due to finite temperature and the flexible protein backbone on the pH dependent optical properties. In particular, we have used a combined molecular dynamics and density functional−molecular mechanics linear response approach by means of which the optical property calculations were carried out for the chromophore in the explicitly treated solvent and bioenvironment. Two different models were used to describe the environment—electronic embedding and polarizable electronic embedding—accounting for the polarization of the chromophore and the mutual polarization between the chromophore and the environment, respectively. For this purpose a polarizable force field was derived quantum mechanically for the protein environment by use of analytical response theory. While the gas-phase calculations for the chromophore predict that the induced red shift going from low to high pH is attributed to the change of molecular forms from neutral to zwitterionic, the two more advanced models that explicitly account for the protein backbone attribute the pH shift to a neutral to anionic conversion. Some ramifications of the results for the use of GFPs as pH sensors are discussed.




## 1. INTRODUCTION

Wild type and mutant green fluorescent proteins (GFPs) are widely used for *in vivo* imaging[1,2] due to their unique fluorescence properties and the ease by which they can be integrated into target biostructures such as proteins.[3,4] GFPs have been applied as reporters of gene expression and as sensors for a number of environmental variables such as the dielectric constant, environment micropolarity, pH change, metal ion concentration, temperature, and pressure.[5−8] In many cases, however, the response to the external variables is still not sufficiently strong to allow for effective real world sensing. As shown experimentally the optical properties of GFPs and their sensitivity to particular external variables can be tuned by performing mutations on the specific residues in the vicinity of the chromophores. Much research effort is thus in progress to identify the "favorable mutations" to improve the optical properties of GFPs for practical applications. GFPs have recently also been studied for imaging applications through nonlinear spectroscopic techniques such as two-photon absorption[9] and second harmonic generation,[10,11] and furthermore by dual imaging[12] where different proteins fused together exhibit different absorption/fluorescence spectra.

The sensitivity of mutant GFPs to external variables has found particular application in pH sensing—the mutant GFP referred to as deGFP1, obtained by the mutations S65T, Q80R, H148G, and T203C of the wild type GFP, has been tested as an intracellular pH indicator due to its pH sensitive optical properties. Its one-photon absorption spectrum has been studied in detail by Hanson et al.[13] for pH values ranging between low (5.0) and normal (9.5). A red shift in the absorption maximum amounting to 104.1 nm has been observed when going from low to high pH.[13] The pH dependence of the optical properties can be attributed to protonated (neutral-protonated at an oxygen site, zwitterionic-protonated at a nitrogen site and deprotonated at an oxygen site, cationic-protonated at both sites) and deprotonated (anionic) states of the GFP chromophore. However, it is not clear which of these protonation states of the chromophore are responsible for the low-pH and high-pH behavior of deGFP1.

In this study, we take interest in the analysis of the absorption spectra and the pH-induced red shift that occur for the dual emission deGFP1[13] as the pH level is raised from low to high (from pH 5.5 to 9.0). The choice of the deGFP1 mutant is motivated by its stable character in a broad pH range, making it possible to directly measure the pH in this range from its absorption/emission spectrum. The original experiments performed by Hanson et al.[13] suggest that the neutral and anionic forms are responsible for the pH-dependent optical properties of deGFP1. The analysis of the structural changes in

Received: February 14, 2014
Published: June 17, 2014

   dx.doi.org/10.1021/ct5001318 | J. Chem. Theory Comput. 2014, 10, 3492−3502

the crystals obtained at pH 5.5 and 9.0 indicated that the chain reorganization is heavily affected by even the smallest amount of anionic chromophore population, as the absorption spectrum shows a 65:35 ratio of protonated to deprotonated chromophores at higher pH, while the electron density map changes significantly around the chromophore at higher pH. Although all forms contribute to the overall absorption spectrum, there should in principle be two forms which contribute the most to the absorption. This is assumed for the simple reason that two different forms are not expected to have identical absorption peaks as they differ in net charge or, in the case of the zwitterionic and neutral forms, in heavy charge redistribution. In the current study, the one-photon absorption (OPA) spectrum is calculated for all different protonation states in order to elucidate the chromophore types responsible for the pH dependent shift. The aim is to find a systematic way of predicting absorption properties for general GFP mutants in order to synthesize molecular probes with the purpose of measuring intracellular pH.

In addition to understanding the working mechanism of a particular GFP, computational modeling studies[14] can be used to establish structure−property relationships.[15] Here the chromophore or protein backbone structures can be modified in order to tune their properties for practical purposes.[15,16] However, modeling (linear and nonlinear optical) properties of GFP proteins has been challenging because a quantum mechanical description of the whole protein is evidently not possible. Either a supermolecular approach[16] or quantum mechanics/molecular mechanics (QM/MM)[14,15] has been used to handle the size of the protein where in the former case the chromophore and the residues close to it are considered as a quantum mechanical supermolecule,[16] while in the latter case the whole system is considered as divided into two subsystems where an accurate electronic structure theory is used for the chromophore responsible for the optical response of the GFP and a less accurate molecular mechanics force-field-based description is used for its remaining part. The interaction between the two subsystems is accounted for by using an effective Hamiltonian which can be evaluated at different levels of sophistication with respect to the interaction terms included—mechanical embedding, electronic embedding, and polarizable electronic embedding. In order to study the optical properties of the chromophores with an explicitly accounted protein environment, we use in the current study a linear response time dependent density functional theory/molecular mechanics (TD-DFT/MM) approach employing the electronic and polarizable electronic embedding schemes to account for the interactions of the subsystems. In addition to this there are other issues related to sampling over solvent and protein configurations and finite temperature effects. In order to cope with such effects for the optical property of the GFP, we use an integrated type of approach where the finite temperature structure of the GFP protein in solvent is studied using molecular dynamics (MD) simulations and the optical property calculations are carried out for various configurations using the TD-DFT/MM response method[17,18] until the property converges.

Even though the machinery for modeling structures and optical properties is available, further issues must be attended referring to development of polarizable force fields for the protein environment to be used for TD-DFT in the polarizable embedding scheme involving the generation of charges, polarizabilities, and multipole moments. In the simplistic electronic embedding scheme only the charges of the protein are considered, so that the polarization of the chromophore by the environment is accounted for.[19] In the polarizable electronic embedding scheme the polarizabilities are included for the atomic sites in addition to the charges, making the protein polarizable by the chromophore so the mutual polarization of the chromophore and protein is accounted for.[17,20] Due to the difficulty associated in computing charges and polarizabilities for the whole protein, they are here computed using a methyl capping model for the individual amino acids. In this way, the presence of amino acids in the peptide-like environment is accounted for to some extent. The strategy of calculating properties in these systems has previously been investigated by Steindal et al.[20] by using the MFCC[21,22] method. The properties of the subsystems of the protein, obtained by MFCC, were calculated from the perturbed density caused by a static external field obtained by finite field calculations, employing the techniques described by Gagliardi et al.[23] In this work, LoProp is implemented in a novel approach and is based on analytical linear response theory.

## 2. FORCE-FIELD IMPLEMENTATION

Because the majority of standard force fields have been derived for the purpose of studying structure and dynamics, it is our experience that these are not sufficiently accurate for property calculations. This is the main reason why we have generated our own force field from ab initio theory. The force fields used for the QM/MM calculations in this study were generated with a so-called LoProp approach[23] which is summarized as follows together with its implementation for the Dalton program.[24] In particular we outline how linear response theory is used to obtain polarizable force fields.

**2.1. LoProp Transformation.** It has now been 10 years since Gagliardi et al.[23] introduced the LoProp method for extracting atomic and interatomic contributions to molecular properties. It was formulated as a sequence of transformations of the atomic overlap matrix, giving a localized orthonormal basis that depends only on molecular structure and the initial atomic orbital basis set. The localized properties were obtained by forming partial traces over basis functions associated with one chosen center or a pair of centers. The aim was to obtain physically sound and transferable local properties which sum up to the molecular property and which are reasonably constant for a given atom in chemically similar systems. The implications of this are that these localized properties can be used as force fields in QM/MM calculations.

The requirement on the initial basis set is that the atomic orbitals can be grouped into occupied and virtual orbitals, e.g., if they resemble the actual orbitals in atomic states. Given this division, the sequence of transformations can be briefly summarized for the combined atomic−orbital (AO) basis set of the molecule, orthonormalized first within atomic subblocks and second within the combined occupied AO space. Then any occupied components in the virtual AO space can be projected out , and finally the virtual space is orthonormalized. This leads to an orthonormal localized basis set, the LoProp basis, where each orbital is associated with an atomic center.

These steps generate a total matrix $T_l^\mu$ describing a transformation between the AO space ($\xi_\mu$) and the LoProp ($\phi_l$) basis functions

$$\phi_l = \sum_\mu \xi_\mu T_l^\mu \tag{1}$$

For integral representations of one-electron operators

$$O_{\mu\nu} = \int \mathrm{d}V\ \xi_\mu(\vec{r})\ O(\vec{r})\ \xi_\nu(\vec{r}) \quad O_{lm} = \int \mathrm{d}V\ \phi_l(\vec{r})\ O(\vec{r})\ \phi_m(\vec{r}) \tag{2}$$

we thus have the transformation rule

$$O_{lm} = \sum_{\mu\nu} O_{\mu\nu} T_l^\mu T_m^\nu \tag{3}$$

while for the density matrix we have the inverse transformation

$$D^{\mu\nu} = \sum_{lm} D^{lm} T_l^\mu T_m^\nu \tag{4}$$

Next the atom and "bond" contributions are given by grouping the summations involved in the trace operations into partial traces over the atomic subsets, denoted by $A$, $B$, ...,

$$\mathrm{tr}(OD) = \sum_{lm} O_{lm} D^{lm} = \sum_{AB} \sum_{l\in A, m\in B} O_{lm} D^{lm} \equiv \sum_{AB} O_{AB} D^{AB} \tag{5}$$

Localized charges and isotropic polarizabilities are thus the parameters which form the force fields for the QM/MM calculations in this work. The decomposition of the total electronic charge according to eq 5 gives

$$Q_{\mathrm{el}} = -\sum_{AB} \sum_{l\in A m,m\in B} S_{lm} D^{lm} \tag{6}$$

where $S$ is the overlap matrix. Since the LoProp basis is orthonormal ($S_{lm} = \delta_{lm}$) as well as localized, this becomes

$$Q_{\mathrm{el}} = -\sum_A \sum_{l\in A} D^{ll} \tag{7}$$

i.e., there are no bond contributions to the total charge, and the localized charge associated with an atomic site $A$ with nuclear charge $Z_A$ is

$$Q_A = Z_A - \sum_{l\in A} D^{ll} \tag{8}$$

**2.2. Analytical Local Polarizabilities.** We now present an alternative formulation and implementation of the LoProp approach. The original implementation was based on finite field perturbation theory applied to localized dipole moments. Here we apply analytical response theory where the change of a wave function in an external field is represented by a unitary transformation. The theory derives from the classical work by Olsen and Jørgensen[25] for MCSCF theory and later generalized for DFT by Sałek et al.[26] Following the notation of the latter, using carets ( $\hat{}$ ) to distinguish operators in second quantization from the matrix representation of the same operator, we have that a wave function $|0\rangle$ is transformed by a unitary operator as a result of the perturbation

$$|\tilde{0}\rangle \equiv \mathrm{e}^{-\hat{\kappa}}|0\rangle \tag{9}$$

The operator $\kappa$ has a real antisymmetric matrix representation which in second quantization is given by

$$\hat{\kappa} = \kappa_{pq} a_p^\dagger a_q \tag{10}$$

If we restrict ourselves to static perturbations, the Ehrenfest theorem gives that the first-order properties can be derived from

$$\delta\langle 0|[\hat{Q},\ \mathrm{e}^{\kappa}\hat{H}\mathrm{e}^{-\kappa}]|0\rangle = 0 \tag{11}$$

for an arbitrary operator $\hat{Q}$. By choosing the same excitation operators that are used to expand $\hat{\kappa}$ we obtain a system of equations for the unknowns, the matrix elements $\kappa_{pq}$. Because the details of these derivations are quite lengthy, we refer the interested reader to ref 26.

After the linear response eq 11 is solved, some matrix algebra allows us to express the first-order change of an expectation value in terms of a trace of the product of the operator matrix and the first-order density matrix

$$\delta\langle \mathrm{e}^{\hat{\kappa}}\hat{A}\mathrm{e}^{-\hat{\kappa}}\rangle|_{\kappa=0} = \langle[\delta\hat{\kappa},\ \hat{A}]\rangle = \sum_{pq} A_{pq}\delta D^{pq} \tag{12}$$

where the first-order change in the density matrix can be expressed as

$$\delta D^{pq} = [\delta\kappa^T,\ D]^{pq} \tag{13}$$

all in MO basis. It is now straightforward to transform the same quantities to AO (with MO coefficients $C_p^\mu$) or LoProp basis (with LoProp transformation $T_l^\mu$)

$$\delta D^{\mu\nu} = \sum_{pq} \delta D^{pq} C_p^\mu C_q^\nu = \sum_{lm} \delta D^{lm} T_l^\mu T_l^\nu \tag{14}$$

With the first-order densities at hand we are now in a position to proceed to evaluate the localized polarizabilities in the LoProp approach. These were introduced in ref 23. By considering the electric field derivatives of the molecular dipole moment with respect to an arbitrary gauge origin $C$

$$\langle -\hat{\vec{r}}_C\rangle = -\sum_{AB} \vec{r}_{AB} D^{AB} + \sum_A Q_A(\vec{R}_A - \vec{R}_C) \tag{15}$$

$\vec{r}_{AB}$ is the electronic coordinate with respect to the midpoint between atoms $A$ and $B$, if $A \neq B$, or with respect to an atomic site $\vec{R}_A$ if $A = B$. Considering at first a first-order variation of the dipole moment

$$\delta\langle -\hat{\vec{r}}_C\rangle = \sum_{AB} -\vec{r}_{AB}\delta D^{AB} + \sum_A \delta Q_A \vec{R}_A \tag{16}$$

it is seen that the last term in (15) containing $C$ vanishes due to charge conservation. The fact that the individual terms in (16) are origin dependent were solved by Gagliardi by introducing charge transfer between atomic sites—an antisymmetric charge transfer matrix $\Delta Q_{AB}$ satisfying

$$\delta Q_A = \sum_B \Delta Q_{AB} \tag{17}$$

where the charge shift on a site $A$ is

$$\delta Q_A = -\sum_{l\in A} \delta D^{ll} \tag{18}$$

This leads to localized polarizability of the form

$$\alpha_{AB} = \delta\langle -\hat{\vec{r}}\rangle_{AB} = -\vec{r}_{AB}\delta D^{AB} + \Delta Q_{AB}(\vec{R}_A - \vec{R}_B) \tag{19}$$

The charge transfer matrix is not uniquely defined, but it solves the problem with gauge dependence and one gets a physically reasonable interpretation of the local polarizability—as a change of the local dipole moment and, in the case of bond


dx.doi.org/10.1021/ct5001318 | J. Chem. Theory Comput. 2014, 10, 3492−3502

contributions, a charge shift between the two sites *A* and *B*. In our calculations we redistribute the bond contributions to their atomic sites, in the same way as the original implementation (equally to each atom forming the bond, respectively), and only consider local atomic polarizabilities.

To summarize, the main difference between this approach and the original implementation is that we base our calculations on analytical response theory rather than finite field perturbation theory. An advantage of our formulation is that it is straightforward to generalize the LoProp concept to local dynamical polarizabilities and to higher order polarizabilities. The analytical LoProp code is a postprocessing tool for Dalton[24] written in the Python programming language.

## 3. COMPUTATIONAL DETAILS

Our integrated approach consists of calculating the OPA spectrum as an average over configurations from the trajectory obtained from MD simulations. We develop the aforementioned polarizable force field to describe the protein as the MM part in the TD-DFT/MM calculations. The procedure of calculating and transferring force-field parameters is described in the following section (for more details see the Supporting Information.)

**3.1. Polarizable Force Field for deGFP1 Protein.** X-ray crystal structures for both the low- (referred as 1JBY.pdb) and high-pH (1JBZ.pdb) deGFP1 protein are reported in the literature[13] with resolution 1.8 and 1.5 Å, respectively. We have used the low-pH structure for the development of the polarizable force field. In the original PDB files, the coordinates were missing for some residues in both the N- and C-terminal atoms as well as for some of the heavy atoms in the side chains of the residues. The missing residues were omitted as they are far away from the chromophore. Rather residues GLY4 and ILE229 for which the coordinates were reported in the PDB file were treated as the terminal residues for GFP with the protonated and deprotonated chromophores. Since the atoms of GLY4 and ILE229 are 20.4 and 20.3 Å away from the CB atom of the CRO66 (which is the chromophore residue), respectively, these residues were treated as the N- and C-terminal amino acids. The missing heavy atoms and hydrogens were added to the crystal structure by using the xleap module of the Amber11 software.[27] The final structures were separated into individual amino acids, with each one capped by including peptide bonds from neighboring amino acids with an extra methyl group at both ends, as depicted in Figure 1. Exceptional

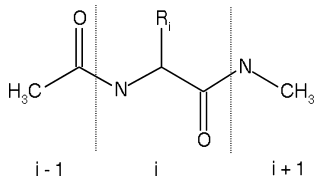



**Figure 1.** Illustration of how each amino acid is represented in the methyl capping scheme. Index *i* denotes residue id.

residues are the N-terminal and C-terminal ones, which only have the methyl cap on one side, and a protonated amide or deprotonated carboxyl representing the N- or C-terminal on the other side, respectively.

Empirical formulas[28−31] predict nonstandard protonation of the side chains in D82, E17, E222, H169, H181, and H199, which are located at a distance of 12.8, 9.0, 3.4, 8.5, 6.7, and 13.0 Å from the chromophore, respectively. Since the only one significantly close to the chromophore is GLU222, a separate calculation (for one photon absorption) is made by protonating it, and it can be seen that the results do not differ much (a few nanometers). Thus, all protonation states of the side chains are kept at their natural p$K_a$. The LoProp parameters obtained for each amino acid were transferred to the full protein MM environment atomwise, as depicted in Figure 2.

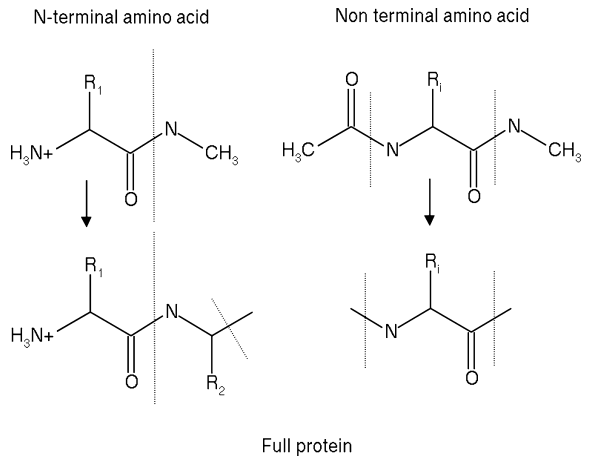



**Figure 2.** Moving properties from separate QM calculations on the amino acids to the MM region defined by the PDB X-ray structure.

All properties were atom-centered, and no multipole moments were included. An ANO basis set[32] corresponding to a 6-31+G* contraction was used together with the TD-DFT method by employing the B3LYP functional in DALTON. We cap all residues with the $CH_3CO$- and -$NHCH_3$ groups from the neighboring residues in order to mimic the environment of each residue inside the protein. The properties are then calculated for each residue by subtracting the extra contribution from the neighboring side groups, obtained from a separate calculation on the molecule $CH_3CONHCH_3$ (as shown in Figure 3), which is optimized in gas phase at the B3LYP/6-

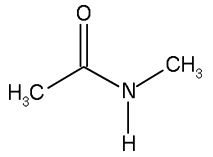



**Figure 3.** Model system used to determine the atomic properties in the side groups of each residue.

31+G* level of theory. First, the properties (charges and atomic polarizabilities) are computed for the residue depicted in Figure 3. From these results, the atomic properties for the groups $CH_3CO$- and -$NHCH_3$ (referred to as *i*−1 and *i*+1 in Figure 1), are subtracted to get the properties for the *i*th residue. The net sum remainder is transferred to the N (from the $CH_3CO$-group) and to the C atom of the *i*th residue (from the -$NHCH_3$ group).

**3.2. Excitation Energies from the Static Approach.** Different protonation states will have different optical properties, leading to pH dependent absorption spectra. In particular, the GFP chromophore can exist in four (protonation and deprotonation) states, namely, neutral, zwitterionic, anionic, and cationic. The structures of these four forms are depicted in


dx.doi.org/10.1021/ct5001318 | J. Chem. Theory Comput. 2014, 10, 3492−3502

Figure 4. We investigate all these protonation/deprotonation states in order to establish the ones responsible for the pH

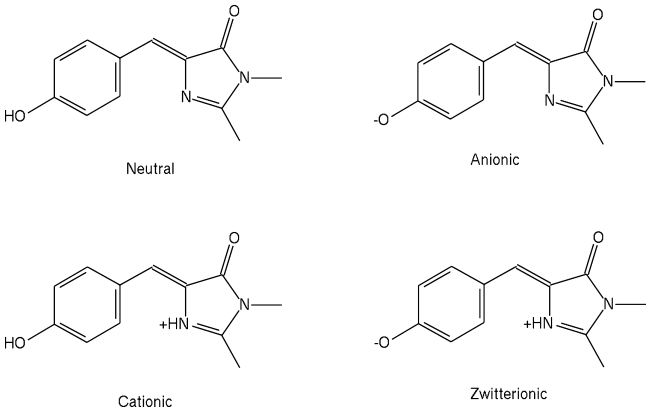



**Figure 4.** Different protonation states of the chromophore.

dependent optical properties of the deGFP1 protein. The heavy atom coordinates for the models are taken directly from the PDB structures with hydrogens added by means of the Avogadro[33] program. The chromophore, which consists of the amino acids T65, Y66, and G67, can be represented by the five different models as seen in Figure 5. These models are used in gas-phase calculations to establish which model is to be used during QM/MM calculations. Since the calculations involves sampling over hundreds of configurations, it is essential to choose a model of minimal size which at the same time gives converged results for absorption spectra. In the static calculations, we also perform PCM and QM/MM on the crystal-structure geometry of the chromophore. In the PCM description we use the CPCM[34,35] module in Gaussian 09. The radii for the atoms of the chromophore are based on the UAHF[36] model as implemented in Gaussian 09.[37] For the QM/MM calculations, we use the Ahlstrom[38] force field on all of the water molecules in the crystal structure.

Before each property calculation, the hydrogen positions are optimized by B3LYP/6-31G* in gas phase while keeping the positions of the heavy atoms fixed using the Gaussian 09 package.[37] We employ the TZVP[39] basis set with the CAM-B3LYP[40,41] functional for all one-photon absorption (OPA) calculations. The strongest excitation is the $S_0 \rightarrow S_1$ transition, unless otherwise explicitly stated. The OPA properties (absorption wavelength and oscillator strengths) were calculated using the linear response function formulation of TD-DFT as implemented in DALTON.[24] For the OPA calculations using the PCM model, however, we use the Gaussian 09 package. For the QM/MM calculations, the LoProp quantities of the N atom in the V68 residue and the CA atom in the F64 were too close to the QM region after introducing the linking hydrogen atom, and the properties on those sites were thus transferred to their neighbors.

### 3.3. Excitation Energies from a Dynamic Approach.

The static calculations described above correspond to a single configuration. However, it has often been demonstrated that sampling over different configurations of the environment[42,43] and incorporation of the vibrational motion of a structure[44,45] are important factors needed for a correct reproduction of the features observed in experimental spectra, including the line shape. So, the absorption properties can be obtained by averaging over different configurations picked up from the molecular dynamics simulations—a procedure we also adopt here. Even though the force-field MD can account for the finite temperature structure and dynamics of the chromophores and the protein backbone, the accuracy is evidently limited by the employed force field. The use of ff03 and Amber force fields for proteins and peptides has been justified through many earlier benchmarking studies, but their use for GFP chromophores has not been studied since the chromophore is not an amino acid, which is something that can limit the accurate prediction of

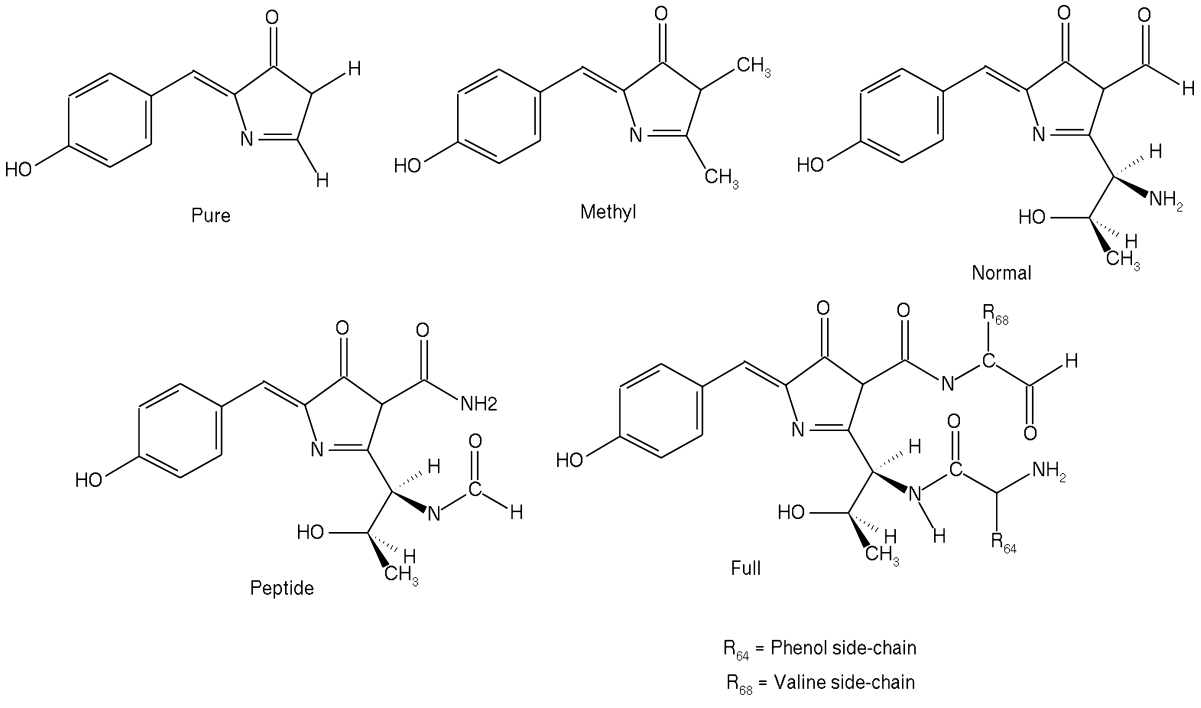



**Figure 5.** Models used for calculations representing the QM system in the gas phase.


dx.doi.org/10.1021/ct5001318 | J. Chem. Theory Comput. 2014, 10, 3492−3502

Table 1. OPA Wavelength (nm) with Oscillator Strengths in Parentheses in the Gas Phase Obtained from the Crystal Structure at Low pH[a]

| model | cationic | neutral | zwitterionic | anionic |
|---|---|---|---|---|
| pure | 393.1 (0.762) | 362.6 (0.823) | 436.4 (0.784) | 395.1 (0.993) |
| methyl | 379.6 (0.877) | 364.2 (0.867) | 432.6 (0.887) | 396.3 (1.086) |
| normal | 384.5 (0.875) | 365.9 (0.897) | 442.9 (0.867) | 408.4 (0.987) |
| peptide | 390.9 (0.903) | 370.0 (0.931) | 450.1 (0.901) | 412.2 (1.090) |
| full | 390.7 (0.938) | 371.4 (0.961) | 451.1 (0.936) | 414.8 (1.109) |

[a]Experimental $\lambda_{max}$: $\lambda_1$ = 399.9 nm; $\lambda_2$ = 504.0 nm; $\Delta\lambda_{expt}$ = 104.1 nm.

absorption properties. Since the molecular geometry of the chromophore is of primary importance to accurately reproduce its optical properties, it is judicious to use a rigid molecular model for the chromophore when a suitable force field is not available.[46,47] We follow a slightly different approach in this study where the chromophore coordinates are restrained using a harmonic potential with a force constant of 100 kcal/mol which allows small-amplitude motion of each of the atoms with respect to its equilibrium position that corresponds to the crystal structure. To some extent this accounts for the vibrational motion of the chromophore and the finite temperature effect on the structure. The most accurate protocol is to derive a suitable force field for the chromophore and to carry out fully flexible molecular dynamics for the chromophore−protein in the solvent. Because the derivation of force fields for a nonstandard residue itself is a demanding task, we have used restraint MD to describe the chromophore dynamics. However, we have kept the solvents and protein environment flexible, and let their average structure and dynamics be dictated by the employed force field. In the subsequent discussion we briefly describe the computational details related to the MD calculations.

Four different simulations are carried out for the cationic, neutral, anionic, and zwitterionic forms of the chromophore. The histidine residues can exist in different protonation states, and since they are far away from the chromophore (the closest distance is 6.7 Å), we have used the same protein structure for all four cases. We have also carried out calculations to test the effect of different protonation states of histidine on the optical properties of the anionic form using two different force fields for the proteins. The results (not shown here) indicate that the absorption maximum only shows a few nanometer shift in these cases, suggesting that the effect of pH on the protein is not significant to the optical properties of the chromophore. The different protonation states of the chromophore is the only difference between the four sets of simulations. The 1JBY crystal structure has been used as the starting structure to prepare the input configuration for the simulations. Here the crystalline water was retained, and in addition solvent molecules were added up to 10 Å beyond the edge atoms of the protein along the three directions. The simulation box approximately includes 22800 water molecules and a sufficient number of sodium ions to neutralize the system. The simulations were carried out for an isothermal−isobaric ensemble to mimic the experimental condition, with ff03 and TIP3P force fields to describe the protein and water. The time step for the integration of equation of motion was kept at 1 fs. The equilibration run is carried out until the density of the whole system is converged, within a time scale of about 20 ns. Energy minimization, temperature scaling, equilibration, and production runs were carried out in each set of the simulation.

We have picked up 100 snapshots from the trajectory (corresponding to last 4 ns) for the linear response TD-DFT and the linear response TD-DFT/MM calculations within the polarizable electronic embedding scheme. Six sets of calculations were performed using the MD trajectory, referred to as MM-0, MM-S, MM-1, MM-2A, MM2B, and MM-2. The MM-0 set uses only the chromophore coordinates from the MD trajectory, while the remaining sets use either protein or solvent coordinates or both. The MM-1 model refers to an electronic embedding where the protein and solvent environment are described as point charges. The sets MM-2A, MM2B, and MM-2 refer to the polarizable electronic embedding scheme where the charges and polarizabilities are placed on each atomic site. The charges of the ff03 force field used in the MD simulation are also used in the MM-1 set, while the polarizable force field developed for deGFP1 is used in the MM-2 set of the calculations. Similarly, for the water solvent we use the TIP3P force field in the case of the MM-1 set, and the Ahlstrom[38] polarizable force field in the case of the MM-2 set. MM-2A refers to the case where only the protein environment is included, MM-2B refers to the case where only the solvents are included while MM-2 refers to inclusion of both protein and solvent molecules. Moreover, in MM-1 and MM-2 the whole protein is included into the MM region and only those solvent molecules within a 10 Å distance from the chromophore center of mass are included. Finally, MM-S refers to the supermolecular calculations including the chromophore and water molecules up to 10 Å from the chromophore center of mass. The polarizable electronic embedding accounts for the mutual electrostatic and polarization interaction between the chromophore and bioenvironment subsystems during the response calculations. However, it does not account for the charge transfer between these subsystems, and that is the reason we employ some supermolecular calculations. The MM-S set of calculations are carried out to elucidate the solvent contribution to the optical properties of the chromophore, which accounts for the intermolecular charge transfer interactions between the chromophore and the surrounding water molecules. Since the chromophore is connected to the protein through a covalent bond, a hydrogen link atom is introduced so that the QM region has correct valency. The LoProp quantities of the MM-region atoms too close to the QM region were moved to the neighboring atoms in order to avoid overpolarization due to the additional hydrogen atom introduced in the QM region.

## 4. RESULTS AND DISCUSSION

When going from low pH to high pH, there are theoretically four different possibilities for the change in the protonation state of the chromophore to occur: (i) neutral ⇒ anionic, (ii) cationic ⇒ neutral, (iii) cationic ⇒ zwitterionic, and (iv) zwitterionic ⇒ anionic. It is, however, not clear which of these proposed conversions conforms with the experimentally


dx.doi.org/10.1021/ct5001318 | J. Chem. Theory Comput. 2014, 10, 3492−3502

**Table 2. OPA (nm) with Oscillator Strengths Given in Parentheses for the Normal Model Obtained from Crystal Structures at Low and High pH**

| | cationic | neutral | zwitterionic | anionic |
|---|---|---|---|---|
| low pH | 384.5 (0.875) | 365.9 (0.896) | 442.9 (0.867) | 408.4 (0.986) |
| high pH | 390.5 (0.801) | 373.7 (0.826) | 443.9 (0.798) | 412.3 (0.947) |
| $\Delta\lambda_{high-low}$ | 6.0 | 7.8 | 1.0 | 3.9 |

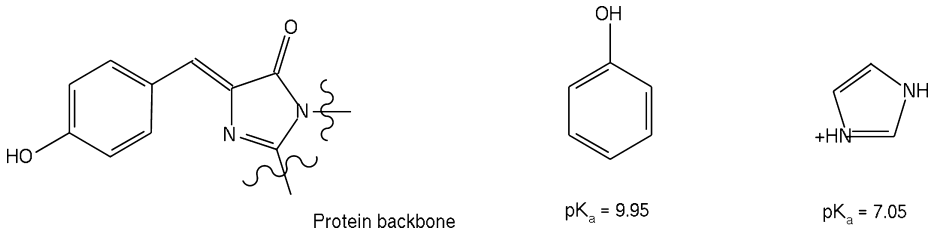



**Figure 6.** p$K_a$ of phenol and imidazole in water for comparison with chromophore protonation.

**Table 3. OPA (nm) with Oscillator Strengths Given in Parentheses for PCM Calculations Obtained for Diethyl Ether $\epsilon \approx 4.0$, Which Is Used To Approximate the Protein Environment**

| | cationic | neutral | zwitter-ionic | anionic |
|---|---|---|---|---|
| low pH | 395.9 (1.047) | 385.2 (1.037) | 444.8 (1.126) | 423.7 (1.205) |
| high pH | 402.4 (0.968) | 393.1 (0.963) | 446.6 (1.043) | 429.0 (1.118) |

**Table 4. OPA (nm) with Oscillator Strengths Given in Parentheses Obtained by QM/MM for All Chromophore States in the Crystal Environments Corresponding to Low and High pH**

| | cationic | neutral | zwitter-ionic | anionic |
|---|---|---|---|---|
| QM/MM—low pH | 363.5 (0.645) | 402.4 (0.628) | 424.4 (1.264) | 420.6 (1.210) |
| QM/MM—high pH | 381.0 (0.972) | 392.6 (0.838) | 429.4 (1.182) | 423.4 (1.177) |

reported shift in the absorption when going from low pH to high pH. The OPA spectra have been computed for all four protonation states of the deGFP1 chromophore using static and dynamic approaches as described above. In particular, the interconversion process from protonated to deprotonated form that reproduces the pH induced red shift of 104 nm has been explored.[13]

Below we present the static results in order to explore the (i) effect of the QM size on the convergence of the optical properties, (ii) the effect of the PCM model on the optical properties of different protonation states of the chromophore, and (iii) the effect of including a polarizable electronic embedding scheme in describing the optical properties of the chromophore in its various protonation states.

The OPA data for the different QM models for all chromophore protonation states in gas phase obtained from the crystal structure at low pH are presented in Table 1. A trend is seen as the oscillator strength increases for the OPA when the QM size is increased along the peptide backbone. Another observed feature is that small QM sizes that do not contain the T65 side chain for the cation and zwitterion give nonmonotonic ordering of the absolute excitation energies. Increasing the QM size for the cation and zwitterion forms gives the same trend as for the neutral and anionic forms. For the purpose of studying how the OPA differs between the different protonation states, the size of the QM region is thus not of large importance compared to the actual protonation and conformation. We will use the normal form for the remaining analysis, as it is sufficiently accurate and requires less computational resources.

A comparison of the OPA in gas phase for the normal model for both low- and high-pH crystal structures is presented in Table 2. It is evident that the OPA for the neutral form varies the most with respect to the chromophore conformation. It is highly unlikely that cationic → neutral or zwitterionic → anionic conversions are the responsible ones as they yield a blue shift going from low to high pH, which would reverse the pH dependent trend observed in the original experiments. The neutral to zwitterionic shift $\Delta\lambda_{neu,low\rightarrow zwi,high}$ = 78.0 nm is in best agreement with the experimental value of $\Delta\lambda_{exp}$ = 104.1 nm. The other important shifts are $\Delta\lambda_{cat,low\rightarrow zwi,high}$ = 59.4 nm and $\Delta\lambda_{neu,low\rightarrow ani,high}$ = 46.4 nm. However, the static gas-phase results do not agree with the view in the literature that the neutral and anionic forms are the ones responsible for the pH induced optical behavior of deGFP1. This disagreement can be attributed to the neglect of environmental effects, as will be made evident by the PCM and QM/MM calculations discussed subsequently. Furthermore, although the shift $\Delta\lambda_{cat,low\rightarrow zwi,high}$ is closer to experiment than $\Delta\lambda_{neu,low\rightarrow ani,high}$, it is unphysical since an increased pH value would more likely lead to a deprotonation of the protonated imidazole, rather than the neutral phenol, as seen by the p$K_a$ values in Figure 6.

In Table 3, OPA data are presented for the normal model of the chromophores with diethyl ether acting as an implicit solvent which has a dielectric constant of $\epsilon \approx 4$, similar to the interior of a protein. The zwitterionic state is now blue-shifted, while the other states are red-shifted. The relevant wavelength shifts are $\Delta\lambda_{neu,low\rightarrow zwi,high}$ = 61.4 nm, $\Delta\lambda_{cat,low\rightarrow zwi,high}$ = 50.7 nm, and $\Delta\lambda_{neu,low\rightarrow ani,high}$ = 43.8 nm. Including a continuous solvent environment thus drastically changes the gas-phase results—


dx.doi.org/10.1021/ct5001318 | J. Chem. Theory Comput. 2014, 10, 3492−3502

they decrease by 21.3%, 14.6%, and 5.6% for $\Delta\lambda_{neu,low\rightarrow zwi,high}$, $\Delta\lambda_{cat,low\rightarrow zwi,high}$, and $\Delta\lambda_{neu,low\rightarrow ani,high}$, respectively, which clearly indicates the significance of an environment. Using our definitions to describe the MM and QM regions, respectively, the static QM/MM calculations on the low- and high-pH crystal structures are performed on the normal chromophore by modeling the MM environment by localized charges and isotropic polarizabilities. The results are summarized in Table 4. We see that the predicted shifts now become $\Delta\lambda^{QM/MM}_{neu,low\rightarrow zwi,high}$ = 27.0 nm, $\Delta\lambda^{QM/MM}_{cat,low\rightarrow zwi,high}$ = 65.9 nm, and $\Delta\lambda^{QM/MM}_{neu,low\rightarrow ani,high}$ = 21.0 nm. Qualitative results are obtained by the static QM/MM approach similar to those with PCM, but with a blue shift of the OPA in the cationic state and an even larger blue shift of the OPA for the zwitterionic state. The inclusion of the polarizable force field in the MM region now predicts a similar shift, albeit low, for the neutral ⇒ anionic and neutral ⇒ zwitterionic interconversions, which indicate that the inclusion of environment using a polarizable force field has substantial effect. Calculations in other studies[20] as well as the ones performed here show that the OPAs for the neutral and anionic forms are very sensitive to the chosen conformation, and thus that sampling the configurational spaces for the neutral and the anionic forms becomes necessary. The static calculations of the excitation energies of the GFP chromophore in gas phase in its four different protonation states predict that the neutral to zwitterionic conversion is the responsible mechanism for the pH induced change in the optical spectrum of the GFP protein.

Turning to the results of the dynamic calculations, we find that the OPA wavelengths for the four forms, namely, neutral, zwitterionic, cationic, and anionic forms, at the MM-0 level are 352, 529, 388, and 449 nm, respectively. Moreover, the shifts corresponding to interconversion between cationic ⇒ neutral, cationic ⇒ zwitterionic, neutral ⇒ anionic, and neutral ⇒ zwitterionic forms are −36, 41, 97, and 177 nm, respectively. As we can see, the shift corresponding to neutral to anionic interconversion, $\Delta\lambda$ = 97 nm, is close to the experimental value. The temperature sampling thus suggests that neutral ⇒ anionic interconversion is the responsible mechanism behind the pH induced red shift. At the MM-1 level we have $\Delta\lambda_{neu-ani}$ = 61.7 nm, which refers to the shift in absorption maximum due to neutral ⇒ anionic interconversion and is blue-shifted when compared to MM-0. This means that the anionic form is stabilized in comparison to the neutral form when the environment is described using an electronic embedding scheme. The absolute $\lambda_{max}$ for the anionic form is blue-shifted to 424 nm. The absolute $\lambda_{max}$ for the neutral form, however, is red-shifted by 10 nm compared to MM-0, which is closer to the experimental value. Furthermore, at the MM-2 level we see that the $\Delta\lambda_{neu-ani}$ shift is reduced to 43 nm, which shows that the polarizable electronic embedding scheme stabilizes the anionic form to an even larger extent than the neutral form. It is notable that $\Delta\lambda_{cation-anion}$, as calculated by MM-2, is 68 nm, which suggests that even the cationic to anionic interconversion could be attributed to the pH induced shift. The models MM-2A and MM-2B predict $\Delta\lambda_{neu-ani}$ to 82 and 71 nm, and the corresponding values for $\Delta\lambda_{cation-anion}$ are 78 and 52 nm, respectively. Moreover, the ordering of the absorption maximum obtained by MM-2A and MM-2B for all four forms of the chromophore differs significantly from the ones obtained by the MM-2 model. This explains why it might be important to account for both solvent and protein environments when modeling the optical properties of chromophores. Finally, the similarities in the absorption spectra (i.e., absorption maximum and intensity) predicted by the MM-0 and MM-S models suggest that the intermolecular charge transfer effect between the chromophore and water solvent does not play a significant role for the optical properties of the chromophore. However, the differences seen in the second band of the absorption spectra of the chromophores still cannot be ignored.

Overall, the pH induced red shift can be attributed to the neutral to anionic interconversion based on the implicit and polarizable electronic embedding descriptions of the bioenvironment and single configuration calculations of the chromophore. The dynamic molecular models without environment and with electronic and polarizable electronic embedding predict the same. However, when going from a no-solvent description to a polarizable electronic embedding scheme, the value of the red shift reduces, which is due to the relative stabilization of the anionic form compared to the neutral form. The latter models also conclude that cationic to anionic interconversion cannot be excluded as a possible mechanism to explain the pH induced red shift. It is worthwhile here to recall that in a recent study on a methoxy naphthalene pH probe we have reported similar results where the dynamic calculations (with a polarizable electronic embedding scheme for the media) underestimates the pH induced shift when compared to the PCM model.[48] Furthermore, we also suggest that it is important to account for the temperature effects in order to reproduce the optical properties of the chromophore as it also has been shown in the case of the mkeima protein.[49]

Since the broadening parameter of the line-shape function of the one-photon spectra can be obtained from the excitation energy calculations for various configurations, it is possible to plot the full absorption spectra. We plot these spectra by convoluting the bands corresponding to the six lowest energy excitations in Figure 7. The lowest energy excitation associated with the largest intensity in the chromophore has $\pi \Rightarrow \pi^*$ character and involves the HOMO and LUMO. Detailed discussions on this transition can be found in the

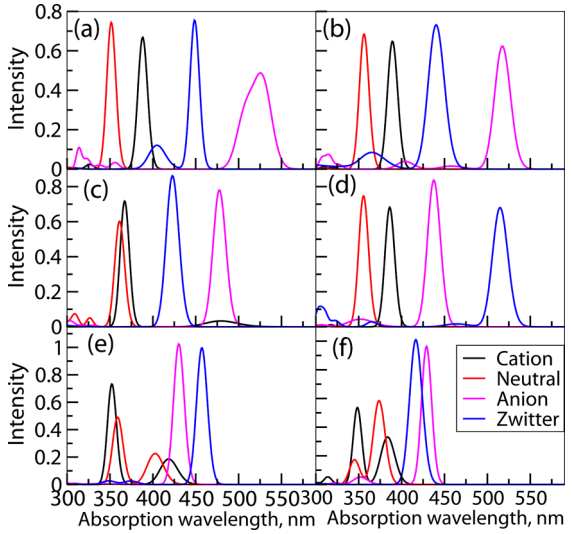



**Figure 7.** (a) Absorption spectra from dynamical sampling using no force field, (b) supermolecular calculations for the chromophore and water molecules included up to 10 Å from the chromophore center of mass, (c) using ff03 force field with only charges and (d–f) using our polarizable force field with the MM-2A, MM-2B, and MM-2 method, respectively.


dx.doi.org/10.1021/ct5001318 | J. Chem. Theory Comput. 2014, 10, 3492−3502

references.[14,50] It is interesting to notice that not only the absorption maximum but also the shape of the spectra and the relative position of the intense absorption band can be significantly changed by the different descriptions for the environments.

We have analyzed the solvation shell structure around the four different protonation states. Interestingly, as shown in Figure 8, the chromophore has access only to a limited number

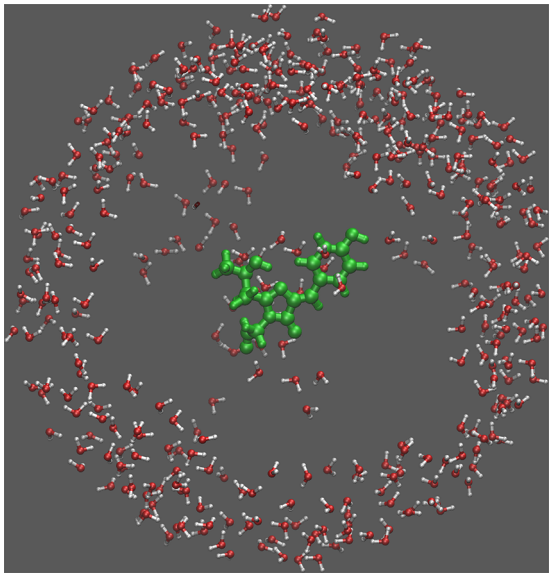


**Figure 8.** Solvation shell structure for neutral chromophore.

of solvent molecules since it is buried within the protein. The chromophore−solvent center of mass radial distribution function exhibits a dip beyond the distance ≈7.5 Å (until 15 Å) which shows that the solvent density is drastically reduced at this distance range. Most of the solvent contribution to the structure and optical properties are thus dictated by the solvent molecules located within a distance of 7.5 Å from the center of mass of the chromophore.

We have also analyzed the solvent density around the chromophores in its different forms. The results are shown in Figure 9. As expected, the water density is mostly localized around the polar groups of the chromophore. In addition, solvent molecules are localized around the imidazole ring with a density that depends on the specific (i.e., protonated/deprotonated) form of the chromophore. Striking features in the solvent density are seen in the case of the anionic form when compared to the other forms. An increased solvent density (of delocalized nature) is observed around the chromophore which reveals that there is water transport between the two polar oxygen atoms mediated by the oxygen atom in the five membered imidazole ring. The increase in solvent density can be attributed to the accumulated charge in the anionic form when compared to the neutral form of the chromophore. The relative shifts in the absorption spectra of the different forms seen in the MM-0, and MM-1 or MM-2 models have to be attributed to such differences in the solvent density (since the protein backbone structure is not altered significantly for the different forms of the chromophore). The difference in the electron density reported based on the X-ray diffraction measurements for the low-pH and high-pH crystal structures can also be attributed to such differences in the solvent density.

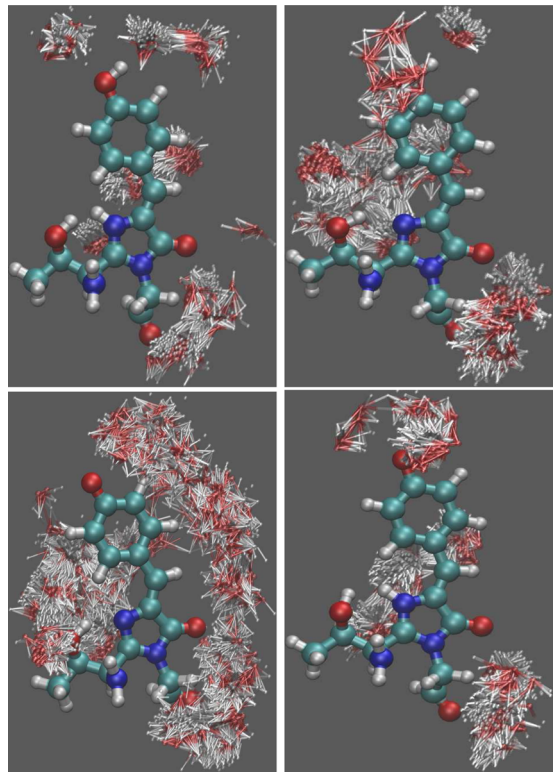


**Figure 9.** Solvent density around the cationic (top left), neutral (top right), anionic (bottom left), and zwitterionic (bottom right) forms of the chromophore.

The solvent density of the neutral, cationic, and zwitterionic forms show significant changes near the oxygen sites that undergo protonation/deprotonation. The average number of solvent molecules within a 7.5 Å cutoff radius from the center of mass of the chromophore for cationic, neutral, anionic, and zwitterionic forms is 23, 27, 23, and 31, respectively. Such differences in the protonation state dependent solvent population around the chromophore can only be accounted for through a dynamic approach as the one employed here. So, this analysis clearly demonstrates the need for explicitly incorporating the solvent to model the relative shifts of the absorption maximum for the different chromophore forms.

## 5. CONCLUSION

The deGFP1 variant of GFP mutants has attracted much interest for use in pH sensing due to its pH dependent optical properties covering a large range of pH values. It has been postulated that the different protonation states—neutral, zwitterionic, anionic, and cationic—are responsible for its pH dependent behavior; however, it has not been clearly established which of these are responsible for the low- and high-pH optical behavior. Our work is motivated by the fact that the responsible pH dependent forms can potentially be clarified by modern molecular modeling techniques as the optical properties of each of these forms then can be studied separately with sufficient precision. The electrostatic interaction, with polarization included between the chromophore and protein, was considered in linear response TD-DFT calculations by developing a unique polarizable force field for the deGFP1 protein. An integrated approach was applied where configurational sampling by means of molecular dynamics simulations was followed by electronic and polarizable


dx.doi.org/10.1021/ct5001318 | J. Chem. Theory Comput. 2014, 10, 3492−3502

electronic embedding of the chromophore using a linear response quantum mechanical−molecular mechanical method for the final optical spectra. We employed this integrated model for all four protonation states of the chromophore to model the structure and optical property relationship using both static and dynamic calculations, where the latter accounts for the explicit inclusion of the protein and solvent environments. Two different pictures emerge from these calculations: The static approach, without the description of environment, attributes the neutral and zwitterionic forms as responsible for the pH dependent behavior of the chromophore, while the static and dynamic models with solvent description propose that the neutral and anionic forms are the responsible molecular forms for the low- and high-pH behavior of the GFP chromophore, respectively. It is also found that cationic to anionic interconversion cannot be excluded as a molecular mechanism responsible for the pH induced red shift in the deGFP1 protein. The polarizable electronic embedding scheme stabilizes the anionic form to a larger extent than the neutral form, so the shift in the absorption maximum for the neutral to anionic forms is reduced when compared to the electronic embedding scheme or when using no description at all for the environment. Our study thus clearly suggests that it is important to account for the solvent and protein environments to correctly reproduce the pH dependent behavior of the deGFP1 protein. The findings are useful as modeling pH dependent properties often has relied on empirical data and extrapolations from noncomparable systems, and in that identifying the right cause for observed behavior underpins the possibilities to make practical guides for how to optimize properties by particular mutations.

## ■ ASSOCIATED CONTENT

### Supporting Information

The supporting information text gives the detailed hydrogen optimization procedure, conformational study of neutral and anionic chromophores, basis set tests, detailed gas phase and QM/MM calculations on the crystal structure of the chromophores, additional CPCM details, comparison of LoProp properties between our and the MFCC method along with the transfer of properties scheme and investigation of protonated GLU222 using QM/MM calculations, tables listing gas phase OPA, the first excitation and DFT energies with respect to the minimum between the two, the basis set dependence of the OPA, OPA for crystal structures, properties of studied residues, and OPA with oscillator strengths for PCM and QM/MM calculations, and figures showing conformations of the neutral and anionic states, the transferring LoProp quantities, and the protonated GLU222 residue. This material is available free of charge via the Internet at http://pubs.acs.org.

## ■ AUTHOR INFORMATION

### Corresponding Author

*E-mail: agren@theochem.kth.se.

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

We thank SNIC for the computer time allocation under Projects SNIC-2013-26-31 and SNIC-2013-1-160.

## ■ REFERENCES

(1) Lippincott-Schwartz, J.; Patterson, G. H. Development and Use of Fluorescent Protein Markers in Living Cells. *Science* **2003**, *300*, 87−91.

(2) Hoffman, R. M. The Multiple Uses of Fluorescent Proteins To Visualize Cancer in Vivo. *Nat. Rev. Cancer* **2005**, *5*, 796−806.

(3) Zimmer, M. Green Fluorescent Protein (GFP): Applications, Structure, and Related Photophysical Behavior. *Chem. Rev.* **2002**, *102*, 759−782.

(4) Tsien, R. Y. The Green Fluorescent Protein. *Annu. Rev. Biochem.* **1998**, *67*, 509−544.

(5) Miyawaki, A.; Llopis, J.; Heim, R.; McCaffery, J. M.; Adams, J. A.; Ikura, M.; Tsien, R. Y. Fluorescent Indicators for $Ca^{2+}$; Based On Green Fluorescent Proteins and Calmodulin. *Nature* **1997**, *388*, 882−887.

(6) Roberto, F. F.; Barnes, J. M.; Bruhn, D. F. Evaluation of A GFP Reporter Gene Construct for Environmental Arsenic Detection. *Talanta* **2002**, *58*, 181−188.

(7) Kneen, M.; Farinas, J.; Li, Y.; Verkman, A. Green Fluorescent Protein as a Noninvasive Intracellular pH Indicator. *Biophys. J.* **1998**, *74*, 1591−1599.

(8) Llopis, J.; McCaffery, J. M.; Miyawaki, A.; Farquhar, M. G.; Tsien, R. Y. Measurement of Cytosolic, Mitochondrial, and Golgi pH in Single Living Cells with Green Fluorescent Proteins. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 6803−6808.

(9) Drobizhev, M.; Makarov, N. S.; Tillo, S. E.; Hughes, T. E.; Rebane, A. Two-Photon Absorption Properties of Fluorescent Proteins. *Nat. Methods* **2011**, *8*, 393−399.

(10) Meulenaere, E. D.; Bich, N. N.; de Wergifosse, M.; Hecke, K. V.; Meervelt, L. V.; Vanderleyden, J.; Champagne, B.; Clays, K. Improving the Second-Order Nonlinear Optical Response of Fluorescent Proteins: The Symmetry Argument. *J. Am. Chem. Soc.* **2013**, *135*, 4061−4069.

(11) Asselberghs, I.; Flors, C.; Ferrighi, L.; Botek, E.; Champagne, B.; Mizuno, H.; Ando, R.; Miyawaki, A.; Hofkens, J.; der Auweraer, M. V.; Clays, K. Second-Harmonic Generation in GFP-like Proteins. *J. Am. Chem. Soc.* **2008**, *130*, 15713−15719.

(12) Hu, C.-D.; Kerppola, T. K. Simultaneous Visualization of Multiple Protein Interactions in Living Cells Using Multicolor Fluorescence Complementation Analysis. *Nat. Biotechnol.* **2003**, *21*, 539−545.

(13) Hanson, G. T.; McAnaney, T. B.; Park, E. S.; Rendell, M. E. P.; Yarbrough, D. K.; Chu, S.; Xi, L.; Boxer, S. G.; Montrose, M. H.; Remington, S. J. Green Fluorescent Protein Variants as Ratiometric Dual Emission pH Sensors. 1. Structural Characterization and Preliminary Application. *Biochemistry* **2002**, *41*, 15477−15488.

(14) Bravaya, K. B.; Grigorenko, B. L.; Nemukhin, A. V.; Krylov, A. I. Quantum Chemistry Behind Bioimaging: Insights from ab Initio Studies of Fluorescent Proteins and Their Chromophores. *Acc. Chem. Res.* **2011**, *45*, 265−275.

(15) List, N. H.; Olsen, J. M. H.; Jensen, H. J. A.; Steindal, A. H.; Kongsted, J. Molecular-Level Insight into the Spectral Tuning Mechanism of the DsRed Chromophore. *J. Phys. Chem. Lett.* **2012**, *3*, 3513−3521.

(16) Topol, I.; Collins, J.; Savitsky, A.; Nemukhin, A. Computational Strategy for Tuning Spectral Properties of Red Fluorescent Proteins. *Biophys. Chem.* **2011**, *158*, 91−95.

(17) Olsen, J. M.; Kongsted, J. Molecular Properties Through Polarizable Embedding. *Adv. Quantum Chem.* **2011**, *61*, 107−143.

(18) Olsen, J. M.; Aidas, K.; Kongsted, J. Excited States in Solution Through Polarizable Embedding. *J. Chem. Theory Comput.* **2010**, *6*, 3721−3734.

(19) Murugan, N. A.; Kongsted, J.; Rinkevicius, Z.; Ågren, H. Color Modeling of Protein Optical Probes. *Phys. Chem. Chem. Phys.* **2012**, *14*, 1107−1112.

(20) Steindal, A. H.; Olsen, J. M. H.; Ruud, K.; Frediani, L.; Kongsted, J. A Combined Quantum Mechanics/Molecular Mechanics Study of the One- and Two-Photon Absorption in the Green Fluorescent Protein. *Phys. Chem. Chem. Phys.* **2012**, *14*, 5440−5451.

(21) Zhang, D. W.; Zhang, J. Z. H. Molecular Fractionation with Conjugate Caps for Full Quantum Mechanical Calculation of Protein−Molecule Interaction Energy. *J. Chem. Phys.* **2003**, *119*, 3599.
(22) Søderhjelm, P.; Ryde, U. How Accurate Can a Force Field Become? A Polarizable Multipole Model Combined with Fragment-wise Quantum-Mechanical Calculations. *J. Phys. Chem. A* **2009**, *113*, 617−627.
(23) Gagliardi, L.; Lindh, R.; Karlström, G. Local Properties of Quantum Chemical Systems: The LoProp Approach. *J. Chem. Phys.* **2004**, *121*, 4494.
(24) Aidas, K.; et al. The Dalton Quantum Chemistry Program System. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2013**, DOI: 10.1002/wcms.1172.
(25) Olsen, J.; Jørgensen, P. Linear and Nonlinear Response Functions for an Exact State and for an MCSCF State. *J. Chem. Phys.* **1985**, *82*, 3235−3264.
(26) Sałek, P.; Vahtras, O.; Helgaker, T.; Ågren, H. Density-Functional Theory of Linear and Nonlinear Time-Dependent Molecular Properties. *J. Chem. Phys.* **2002**, *117*, 9630−9645.
(27) Case, D.; et al. *AMBER 12*; University of California: San Francisco, CA, USA, 2012.
(28) Li, H.; Robertson, A. D.; Jensen, J. H. Very Fast Empirical Prediction and Interpretation of Protein PKa Values. *Proteins: Struct., Funct., Bioinf.* **2005**, *61*, 704−721.
(29) Bas, D. C.; Rogers, D. M.; Jensen, J. H. Very Fast Prediction and Rationalization of PKa Values for Protein-Ligand Complexes. *Proteins: Struct., Funct., Bioinf.* **2008**, *73*, 765−783.
(30) Olsson, M. H.; Søndergard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical p$K_a$ Predictions. *J. Chem. Theory. Comput.* **2011**, *7*, 525−537.
(31) Søndergaard, C. R.; Olsson, M. H.; Rostkowski, M.; Jensen, J. H. Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of p$K_a$ Values. *J. Chem. Theory Comput.* **2011**, 2284−2295.
(32) Widmark, P. O.; Malmqvist, P. Å.; Roos, B. O. Density Matrix Averaged Atomic Natural Orbital (ANO) Basis Sets for Correlated Molecular Wave Functions. *Theor. Chim. Acta* **1990**, *77*, 291−306.
(33) Hanwell, M. D.; Curtis, D. E.; Lonie, D. C.; Vandermeersch, T.; Zurek, E.; Hutchison, G. R. *Avogadro: An Open-Source Molecular Builder and Visualization Tool*, Version 1.1.1; http://Avogadro.Openmolecules.Net/, 2011.
(34) Barone, V.; Cossi, M. Quantum Calculation of Molecular Energies and Energy Gradients in Solution by a Conductor Solvent Model. *J. Phys. Chem. A* **1998**, *102*, 1995−2001.
(35) Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. Energies, Structures, and Electronic Properties of Molecules in Solution with the C-PCM Solvation Model. *J. Comput. Chem.* **2003**, *24* (6), 669−681.
(36) Barone, V.; Cossi, M.; Tomasi, J. A New Definition of Cavities for the Computation of Solvation Free Energies by the Polarizable Continuum Model. *J. Chem. Phys.* **1997**, *107*, 3210−3221.
(37) Frisch, M. J.; et al. *Gaussian 09*, Revision C.1; Gaussian: Wallingford, CT, USA, 2009.
(38) Ahlström, P.; Wallqvist, A.; Engström, S.; Jönsson, B. A Molecular Dynamics Study of Polarizable Water. *Mol. Phys.* **1989**, *68*, 563−581.
(39) Schafer, A.; Huber, C.; Ahlrichs, R. Fully Optimized Contracted Gaussian Basis Sets of Triple Zeta Valence Quality for Atoms Li To Kr. *J. Chem. Phys.* **1994**, *100*, 5829−5835.
(40) Yanai, T.; Tew, D. P.; Handy, N. C. A New Hybrid Exchange-Correlation Functional Using the Coulomb-Attenuating Method (CAM-B3LYP). *Chem. Phys. Lett.* **2004**, *393*, 51−57.
(41) Peach, M. J. G.; Helgaker, T.; Sałek, P.; Keal, T. W.; Lutnæs, O. B.; Tozer, D. J.; Handy, N. C. Assessment of A Coulomb-Attenuated Exchange-Correlation Energy Functional. *Phys. Chem. Chem. Phys.* **2006**, *8*, 558−562.
(42) Aidas, K.; Møgelhøj, A.; Nilsson, E. J.; Johnson, M. S.; Mikkelsen, K. V.; Christiansen, O.; Söderhjelm, P.; Kongsted, J. On the Performance of Quantum Chemical Methods To Predict Solvatochromic Effects: The Case of Acrolein in Aqueous Solution. *J. Chem. Phys.* **2008**, *128*, No. 194503.
(43) Biczysko, M.; Bloino, J.; Brancato, G.; Cacelli, I.; Cappelli, C.; Ferretti, A.; Lami, A.; Monti, S.; Pedone, A.; Prampolini, G. Integrated Computational Approaches for Spectroscopic Studies of Molecular Systems in the Gas Phase and in Solution: Pyrimidine as a Test Case. *Theor. Chem. Acc.* **2012**, *131*, 1−19.
(44) Improta, R.; Barone, V.; Santoro, F. Ab Initio Calculations of Absorption Spectra of Large Molecules in Solution: Coumarin C153. *Angew. Chem.* **2007**, *119*, 409−412.
(45) Barone, V.; Baiardi, A.; Biczysko, M.; Bloino, J.; Cappelli, C.; Lipparini, F. Implementation and Validation of A Multi-Purpose Virtual Spectrometer for Large Systems in Complex Environments. *Phys. Chem. Chem. Phys.* **2012**, *14*, 12404−12422.
(46) Aidas, K.; Kongsted, J.; Osted, A.; Mikkelsen, K. V.; Christiansen, O. Coupled Cluster Calculation of the n $\Rightarrow$ $\pi^*$ Electronic Transition of Acetone in Aqueous Solution. *J. Phys. Chem. A* **2005**, *109*, 8001−8010.
(47) Olsen, J. M.; Aidas, K.; Mikkelsen, K. V.; Kongsted, J. Solvatochromic Shifts in Uracil: A Combined MD-QM/MM Study. *J. Chem. Theory Comput.* **2009**, *6*, 249−256.
(48) Murugan, N. A.; Kongsted, J.; Ågren, H. pH Induced Modulation of One- and Two-Photon Absorption Properties in a Naphthalene Based Molecular Probe. *J. Chem. Theory Comput.* **2013**, *9*, 3660−3669.
(49) Nadal-Ferret, M.; Gelabert, R.; Moreno, M.; Lluch, J. M. How Does the Environment Affect the Absorption Spectrum of the Fluorescent Protein mKeima? *J. Chem. Theory Comput.* **2013**, *9*, 1731−1742.
(50) Marques, M. A.; López, X.; Varsano, D.; Castro, A.; Rubio, A. Time-Dependent Density-Functional Approach for Biological Chromophores: The Case of the Green Fluorescent Protein. *Phys. Rev. Lett.* **2003**, *90*, 258101.