Electronic Circular Dichroism of Fluorescent Proteins: a Computational Study

Anna Pikulska,[†] Arnfinn Hykkerud Steindal,[‡] Maarten T. P. Beerepoot,[‡] and Magdalena Pecul[*,†]

Faculty of Chemistry, University of Warsaw, Pasteura 1, 02-093 Warszawa, Poland, and Centre for Theoretical and Computational Chemistry, Department of Chemistry, University of Tromsø—The Arctic University of Norway, N-9037 Tromsø, Norway

E-mail: mpecul@chem.uw.edu.pl

[*]To whom correspondence should be addressed
[†]University of Warsaw
[‡]University of Tromsø

Abstract

The ECD properties of the green fluorescent protein and other fluorescent proteins have been calculated with density functional theory. The influence of different embedding models on the ECD signal of the chromophore has been investigated by modelling the protein environment by the polarizable continuum model (QM/PCM), by the polarizable embedding model (PE-QM/MM), by treating the minimal environment quantum mechanically at the same footing as the chromophore (QM/QM) and by adding the remaining part of the protein by means of PCM (QM/QM/PCM). The rotatory strength is found to be more sensitive than the oscillatory strength to changes in the geometry of the chromophore and its surroundings and to the type of embedding model used. In general, explicit embedding of the surrounding protein (PE-QM/MM or QM/QM) induces an increase in the rotatory strength of the chromophore. Explicit inclusion of the whole protein through polarizable embedding is found to be an affordable embedding model that gives the correct sign of the rotatory strength for all fluorescent proteins. PCM is useful as a first approximation to protein environment effects, but as a rule seems to underestimate the rotatory strength.

Keywords: Green Fluorescent Protein, Fluorescent Proteins, Electronic Circular Dichroism, Multiscale Modeling, Polarizable Embedding, Polarizable Continuum Model

Introduction

Electronic circular dichroism (ECD) spectroscopy, based on the different absorption of left- and right-circularly polarized light by chiral species, is an established method for structural investigation in biochemistry, especially for proteins.[1–5] Since α-helices and β-sheets have very distinct patterns of n to π^* and π to π^* transitions in the UV region (between 190 nm and 250 nm), this is the region most often used to investigate the secondary structure of proteins.[1,2] Moreover, if a protein has a prosthetic group absorbing in the visible region or near UV, the Cotton effect of this transition can be utilized to investigate the confor-

mation of this group and the nearby protein fragments. Such chromophores can also be introduced artificially in the protein region of interest.[6–8] Since the ECD signal is very sensitive to conformational changes and the molecular environment, it is a sensitive probe of protein structure in the vicinity of the prosthetic group. However, this sensitivity also makes computational modelling of the Cotton effect of this type a difficult task, because of many factors (such as solvation and conformational equilibria) that need to be accounted for.

Among proteins absorbing in the visible light region, the green fluorescent protein (GFP) and other fluorescent proteins have attracted a lot of attention. The GFP chromophore is a product of a cyclization reaction of three amino acids: serine, tyrosine and glycine, residues 65 to 67.[9] The chromophore of GFP exists in nature in two stable forms: neutral and anionic. The two forms of the protein differ not only in the protonation state of the chromophore, but also in its environment. Indeed, both the hydrogen bonding network and the orientation of Thr203 are known to differ between the neutral and anionic chromophore.[9] In addition, the existence of an 'intermediate' state—an anionic chromophore with an environment similar to that of the neutral chromophore—has been described.[10] The GFP chromophore with its closest environment in the anionic, neutral and intermediate states are shown in Figure 1. Wild type GFP (wtGFP) from *Aequorea victoria* has a major excitation peak at a wavelength of 395 nm (ascribed to the neutral form), a minor one at 475 nm (ascribed to the anionic form), and an additional at 495 nm (ascribed to the intermediate state).[10]

Several mutants of GFP have been engineered by mutating residues both inside and outside the chromophore. For example, blue fluorescent protein (BFP) and enhanced cyan fluorescent protein (eCFP) contain different aromatic residues in the place of residue 66, leading to chromophores with an imidazole (BFP) or indole (eCFP) ring in the place of the tyrosine ring in GFP.[11] Enhanced GFP (eGFP) contains the S65T mutation just outside the conjugated system of the chromophore, which is responsible for repression of the neutral peak in the absorption spectrum of GFP and enhancement of the anionic peak and fluorescence intensity.[12] An example of a mutation outside the chromophore is the mutation

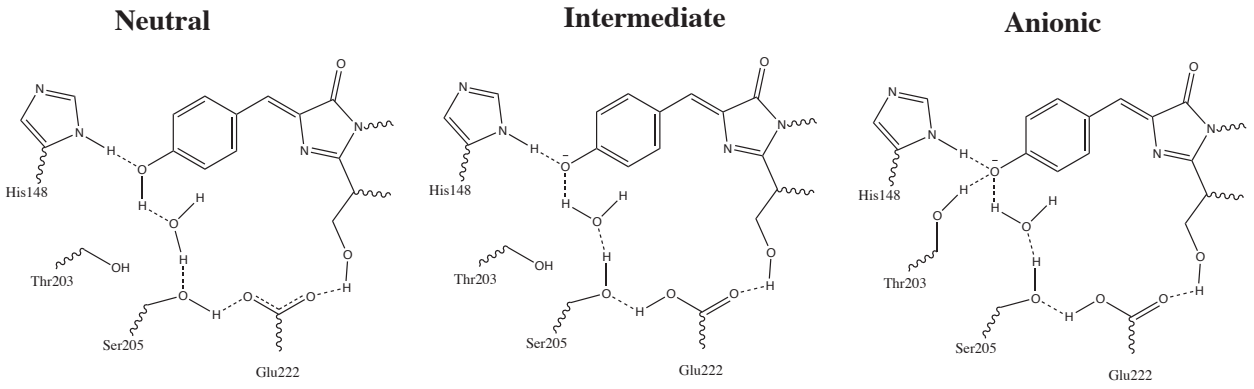

Figure 1: A schematic representation of the GFP chromophore and part of the hydrogen-bonding network around it for the protein in the neutral, anionic and intermediate states.[10]

of threonine to tyrosine in position 203 in yellow fluorescent protein (YFP), leading to π-stacking of the phenol rings of the chromophore and tyrosine 203.[13] A similar protein with a chromophore with an extended conjugated system (DsRed) and therefore a red-shifted absorption spectrum has been isolated from reef coral.[14] The structures of the chromophores of the fluorescent proteins investigated in this work are shown in Figure 2.

There have been numerous experimental[11–19] and computational[20–24] studies of absorption and fluorescence spectra of GFP and related proteins. However, the chiroptical aspects of fluorescent proteins are far less well investigated, with only few experimental studies on the topic.[25–27] Since the fluorescent chromophores are both inherently chiral and surrounded by a chiral protein environment, they offer an interesting object to study the subtle interplay between the inherent chirality and environment-induced chirality and their expression in the ECD spectra. The ECD properties of these systems are also expected to be sensitive to the effects of protonation/deprotonation processes and related conformational changes.

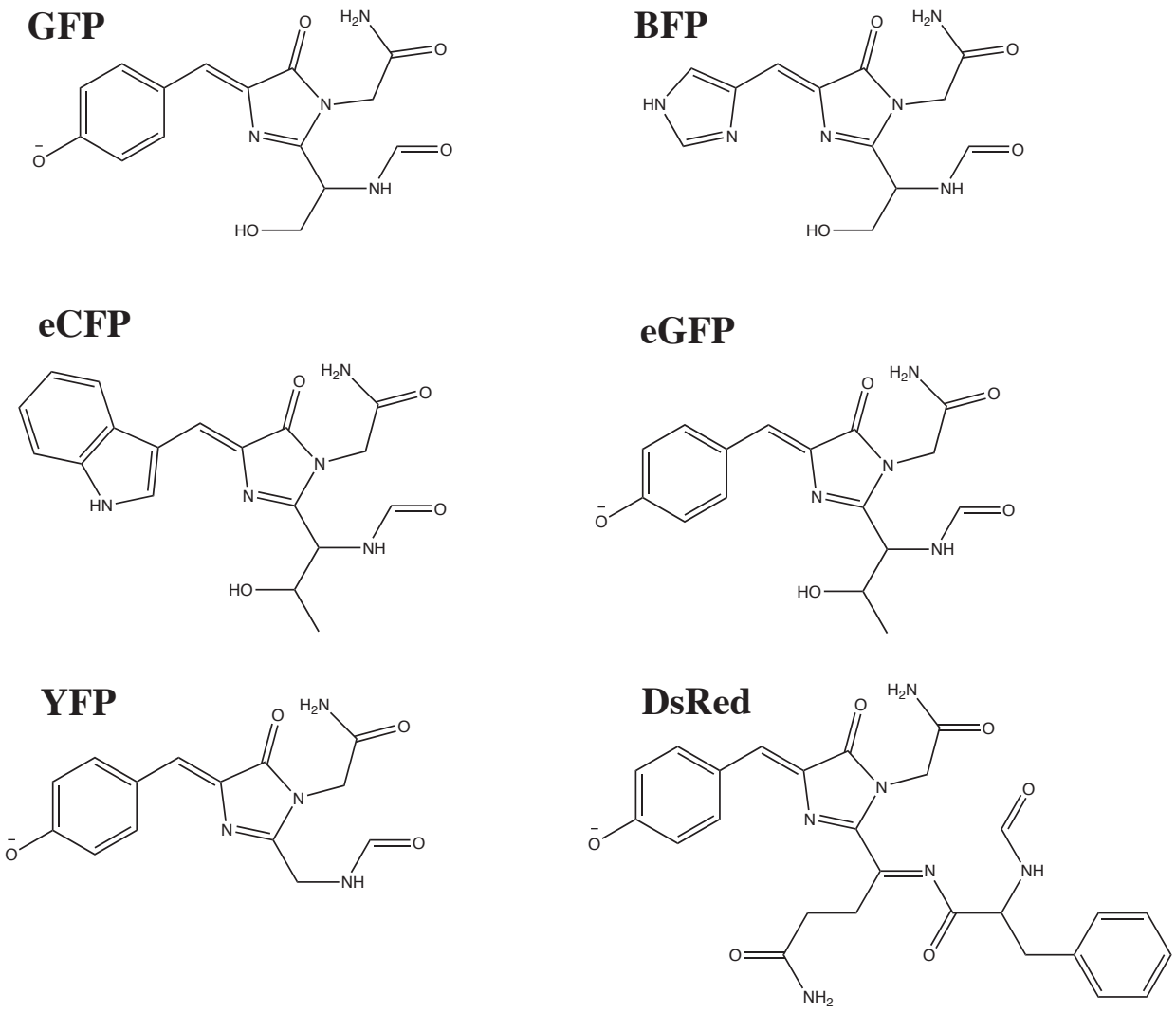

Figure 2: The structures of the chromophores of GFP, BFP, eCFP, eGFP, YFP and DsRed. GFP and YFP are here shown in their anionic forms, but have also been modelled in the neutral forms.

The aim of this study is twofold. The main aim is to check which methods of modelling the molecular environment are suitable for calculations of ECD properties in biological systems and in particular to investigate the performance of the polarizable embedding method, which takes into account the mutual polarization interaction between a central subsystem (here: the chromophore) and its surroundings (here: the protein).[28] The second aim is to investigate the influence of various structural factors on the ECD signal of the fluorescent proteins. We investigate the effect of the protein surrounding and crystal water molecules on the ECD signal of the chromophores. In particular, we attempt to verify whether calculations can confirm the hypothesis of Visser *et al.* that the so-called intermediate form of GFP has a larger rotatory strength than the anionic form, which could explain the red-shift in the ECD spectrum of eGFP in comparison with the absorption spectrum.[25]

The size of the GFP chromophore or a cluster of the chromophore and its surroundings requires the use of an efficient computational method. Even though cluster calculations on GFP have been reported using RI-CC2 with a reduction of the virtual space,[21] we here use the more favorably scaling time-dependent density functional theory (TD-DFT).[29] We note that a simultaneous good description of the excitation wavelengths and the rotatory strengths is a challenge for TD-DFT.[30]

The paper is organized as follows. After the introduction and description of the computational details, we discuss the excitation wavelength, oscillator and rotatory strength obtained for GFP using different methods of modelling the protein environment: the polarizable continuum model (QM/PCM), polarizable embedding model (PE-QM/MM), and treating the minimal environment quantum mechanically at the same footing as the chromophore (QM/QM), and the remaining part by means of PCM (QM/QM/PCM). We also consider the influence of water molecules in the vicinity of the chromophore on the ECD properties of GFP. Next, the QM/PCM and PE-QM/MM methods are employed to model the ECD properties of other fluorescent proteins. Finally, a summary and concluding remarks are given.

Computational details

The ECD calculations on GFP have been carried out using different methods to simulate the protein environment of the chromophore: using Integral Equation Formalism of the Polarizable Continuum Model (IEF-PCM),[31] the polarizable embedding quantum mechanics/molecular mechanics (PE-QM/MM) model[28] and explicitly including the nearest environment of the chromophore at the quantum mechanical level (QM/QM). In addition, the PCM model has also been added on top of the QM/QM model (QM/QM/PCM).

The computations have been carried out using time-dependent density functional theory (TD-DFT)[29] using the CAM-B3LYP[32] exchange–correlation functional. The CAM-B3LYP functional has been found to give a correct qualitative description of electronic transitions also for chromophores with extended π-systems[33] and good quantitative rotatory strengths[34] compared to RI-CC2, despite a rather large but systematic overestimation of the electronic excitation energies.[22,24,33] The more widely used B3LYP functional tends to give smaller deviations compared to experimental excitation energies, despite a more variable qualitative description for different chromophores.[33] TD-DFT results using the B3LYP functional are given in the Supplementary Information. The orbital transitions giving rise to the electronic excitations described in this work are π to π^* transitions with both orbitals delocalized over the conjugated system of the chromophore. Only in the case of DsRed, there is a significant degree of charge transfer.[35]

The 6-31+G* basis set has been employed in all ECD calculations. A basis set test on the electronic excitation properties of the neutral and anionic chromophore of GFP using the CAM-B3LYP functional has shown that a set of diffuse functions is important for this system.[36] Moreover, 6-31+G* results for one- and two-photon absorption were found to be similar to the slightly larger 6-31++G* and 6-311+G* basis sets with only small quantitative differences with the much larger aug-cc-pVDZ and aug-cc-pVTZ basis sets.[36] A study of the ECD spectrum of (neutral) flexible peptides revealed a high degree of similarity between the 6-31G* and 6-311G** spectra.[34] The calculated band positions here are vertical

excitation energies, so vibronic couplings effects are not accounted for. Rotatory strengths are calculated in the velocity gauge to ensure gauge-origin independent results.

The QM/MM model used in the ECD calculations is an explicitly polarizable embedding scheme coupled with a DFT description of the quantum system (PE-DFT). This model takes into account the mutual polarization between the chromophore and its surroundings in the ground state and upon excitation of the chromophore.[28] The chromophore is treated quantum mechanically, while the protein surrounding is described by atomic dipole–dipole (anisotropic) polarizabilities and by atomic charges, dipoles and quadrupoles from an electric multipole expansion. This approach has been used before to study fluorescent proteins.[24,35–37] The computations of the embedding potential parameters have been done using B3LYP/6-31+G* with the help of the polarizable embedding assistant script (PEAS).[38] Details are given in Ref. 35 for DsRed and in Ref. 24 for the other fluorescent proteins.

In the calculations on the bare chromophores, the chromophore structures have been optimized *in vacuo* and using PCM with DFT (B3LYP) and the 6-31+G* basis set. Two different dielectric constants (ε) have been used in the PCM calculations, corresponding to the two limiting situations the chromophore may encounter. $\varepsilon = 4$ is used in the literature to mimic the hydrophobic protein environment,[39] while $\varepsilon = 15$ corresponds to the proteins in water solution and simulates the phase boundary.[40,41]

The preparation of the crystal structures for blue fluorescent protein (BFP),[16] enhanced cyan fluorescent protein (eCFP),[15] enhanced green fluorescent protein (eGFP)[12] and neutral and anionic yellow fluorescent protein (YFP)[19] used here is described in detail in Ref. 24; the preparation of the DsRed crystal structure[42] structure is described in Ref. 35. In all cases, the chromophore of the fluorescent proteins was optimized with QM/MM with the rest of the protein described by the OPLS[43] point-charge model. For DsRed—which has an extended conjugated system compared to the other fluorescent proteins—the chromophore cut out as the QM region consists of the residues Phe65, Gln66, Tyr67 and Gly68 and atoms N and H from Gln64, appropriately capped with hydrogen link atoms.[35] For the other fluorescent

proteins, the chromophore cut out as the QM region consists of the residues 65 to 67 (Ser65, Tyr66 and Gly67 in GFP), atoms N and HN from residue 64 and atoms C and O from residue 68, appropriately capped with hydrogen atoms.[24] The chromophore structures constituting the QM region in the calculations are shown in Figure 2.

Additional steps were necessary to prepare the neutral, anionic and intermediate structures for GFP, which differ mainly in the hydrogen-bonding network around the chromophore and the orientation of Thr203 (see Figure 1). The starting structures were the neutral and anionic structures from Ref. 24, where the position of Thr203 is correct for the neutral and intermediate forms. The preparation of the anionic chromophore was described in Ref. 37, and a similar procedure is followed for the neutral and intermediate structures described here. The orientation of the hydroxyl group of Thr203 in the anionic form was changed so that it points towards the anionic chromophore.[37] The hydrogen bonding network around the chromophore (OH in Ser65, CH2OH in Ser205, COO(H) in Glu222 and H_2O between Ser205 and phenol(ate) from chromophore) was classically relaxed (conjugate gradient with convergence to a gradient of 0.05) with the rest of the protein present as a frozen constraint. Thr203 and its closest water molecule were also relaxed in the otherwise frozen protein to relieve steric stress between Thr203 and the water molecule next to it. In all three cases (neutral, anionic and intermediate), the chromophore and parts of its surroundings were optimized with QM/QM/MM. The QM region consisted of the chromophore, six amino acid residues (Gln94, Arg96, Gln148, Thr203, Ser205 and Glu222) and three water molecules (1, 3 and 4 in Figure 3). The MM region consisted of the rest of the protein described with OPLS[43] point charges. Since both the chromophore and part of the surroundings were described by QM, we will denote this optimization method QM/QM/MM. ECD calculations on these structures, however, were done with only the chromophore in the QM region and will be denoted PE-QM/MM.

The neutral, anionic and intermediate forms of the GFP chromophore and parts of the surrounding have also been optimized in a dielectric continuum with PCM. Since the chro-

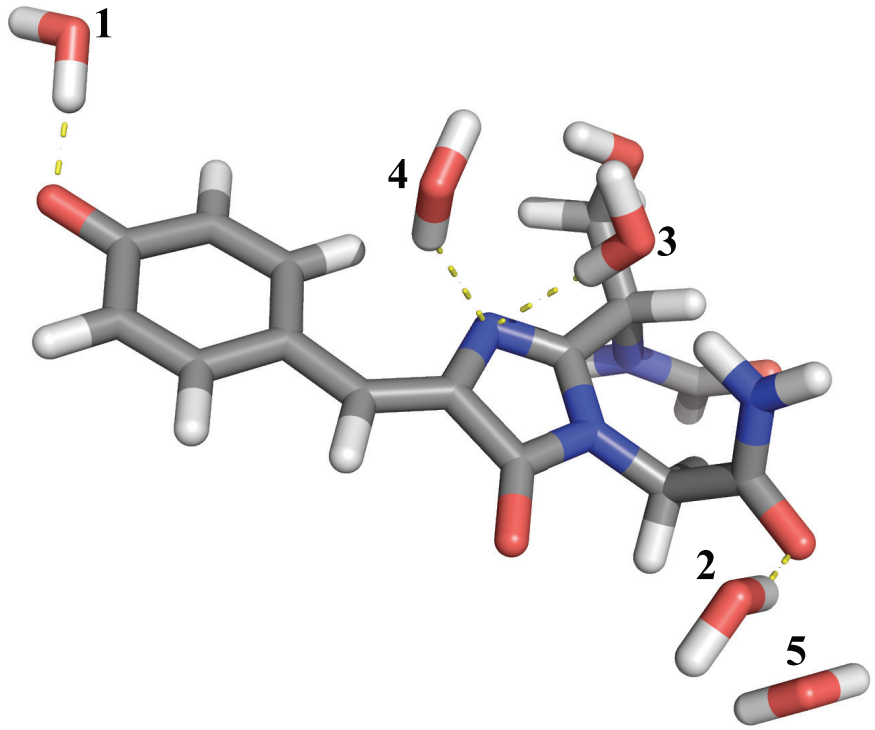

Figure 3: The anionic chromophore of GFP with the closest five crystal water molecules. Water molecule 1 is a hydrogen bond donor to the anionic chromophore and a hydrogen bond acceptor from the neutral chromophore.

mophore and its close surrounding are here described at the same level of theory, this optimization method will be denoted QM/QM/PCM. The cluster of atoms chosen in this calculation has been chosen following the work of Kaila *et al.*.[21] The cluster model includes the chromophore, 9 amino acid residues (Thr62, Gln69, Gln94, Arg96, His148, Val150, Thr203, Ser205 and Glu222) and five water molecules forming the H-bond network with the chromophore and the residues. The number of water molecules is different from the work of Kaila *et al.*,[21] who used the 1EMB crystal structure[44] where we use the 1GFL crystal structure.[45] The chromophore in these calculations includes residues Ser65, Tyr66 and Gly67 and parts of Phe64 (atoms C, O and C_α capped with three hydrogens) and Val68 (atoms N, H and C_α capped with three hydrogens). The geometry of this system has been optimized using B3LYP/6-31+G* with frozen positions of C_α (chromophore Phe64 and Val68) or C_β (other residues) linking the chromophore and other residues to the protein. The starting structures were the QM/QM/MM-optimized structures described above.

The ECD calculations *in vacuo* and with polarizable embedding (PE-QM/MM) were been performed using the Dalton program,[46,47] linked to the polarizable embedding library (PElib),[48] which makes use of the integral package GEN1INT.[49,50] The geometry optimization and ECD calculations using the PCM model were been performed using the Gaussian 09 program.[51] The QM/MM geometry optimizations (including the QM/QM/MM geometry optimizations with the extended QM region) with the whole protein described as point charges were performed using the QSite program.[52,53]

Results and Discussion

The discussion of the computational results is divided into two parts. First, the results for GFP are discussed, with emphasis on the comparison of different methods of describing the protein environment and on the interplay between contributions from inherent chromophore chirality and the chiral environment. Second, the results for other fluorescent proteins are

treated, with the emphasis on reproducing experimental data using implicit and explicit solvent models.

GFP

In the following subsections, we describe our attempts to simulate the ECD properties of GFP in the three forms (neutral, anionic and intermediate), starting from calculations for the isolated chromophore (denoted QM/vacuum) following different geometry optimization procedures, then modelling the protein environment by means of PCM (denoted QM/PCM), the polarizable embedding model (denoted PE-QM/MM), and fully quantum mechanical calculations for the chromophore with minimal protein environment with PCM (QM/QM/PCM) and without it (QM/QM). The PE-QM/MM calculations on GFP include the whole protein around the chromophore (dimer A of the 1GFL crystal structure), corresponding to around 4000 atoms. The QM/QM calculations include the chromophore and 115 atoms around it.

QM/vacuum results

The results of the calculations of the excitation wavelengths, rotatory and oscillator strengths carried out for an isolated GFP chromophore after different geometry optimization procedures are presented in Table 1. This set of calculations has been carried out to extract the steric effect of the environment on the ECD signal of the GFP chromophore and to check how employing different methods of modelling the geometry of the isolated chromophore in different protonation states influences the results. The structure of the GFP chromophore is shown in Figure 2 and is the same for all the ECD calculations reported in Table 1, regardless of the optimization method used.

First of all, we note the blue-shift of all excitation wavelengths compared to experimental peak maxima, partly related to the characteristic blue-shift of excitation wavelengths with the CAM-B3LYP exchange–correlation functional[22,24,33] and partly to the lack of environment in the calculations as indicated by the calculations with protein environment below.

Table 1: Absorption wavelengths (λ, in nm), rotatory (R, in 10^{-40} esu $\cdot$ cm $\cdot$ erg $\cdot$ G^{-1}) and oscillator strengths (*f*, dimensionless) calculated for the neutral and anionic isolated chromophores with the geometry optimized in different ways. For the calculations in which the chromophore is optimized with QM/QM/PCM and QM/QM/MM, the results for the intermediate form are shown, since the different surrounding here influences the geometry of the anionic chromophore.

Structures	Optimization method	λ	R	f
neutral	QM/vacuum	343	-23.45	0.85
	QM/PCM, ε=4	345	-23.70	0.86
	QM/QM/MM	344	-69.64	0.82
	QM/QM/PCM	346	-25.15	0.82
anionic	QM/vacuum	405	-28.20	1.12
	QM/PCM, ε=4	407	-26.66	1.11
	QM/QM/MM	412	-55.26	1.08
	QM/QM/PCM	415	-27.89	1.06
intermediate	QM/QM/MM	413	-60.64	1.09
	QM/QM/PCM	413	-41.32	1.08

The observed changes in the rotatory strength are larger than the changes in the oscillator strength—both when comparing different optimization methods for one ionization state and for the anionic and intermediate states optimized using the same method. An investigation of the structural differences in the chromophore obtained with different optimization procedures reveals that the main differences are found in the backbone parts of residues 65 and 67, *i.e.*, outside the conjugated system. This explains why the changes in the rotatory strength are more pronounced than those in the oscillator strength: the backbone atoms are not directly involved in the electronic excitation, but do affect the chirality of the chromophore. In particular, we observe differences between the anionic and intermediate forms when optimized with QM/QM/PCM, while the structural differences in the backbone are much smaller for the structures optimized with QM/QM/MM. Indeed, the larger number of atoms present in the QM/QM/MM geometry optimization leads to both a more constrained and a more realistic inclusion of the rest of the protein.

The rotatory strength of the chromophore is larger for the intermediate than for the anionic form, especially following the QM/QM/PCM optimization using a minimal sur-

rounding of the chromophore (-41 *vs.* -28 10^{-40} esu $\cdot$ cm $\cdot$ erg $\cdot$ G^{-1}). This is the result of the different environment in the geometry optimization—in particular the orientation of Thr203—and the effect it has on the structure of the chromophore. When the whole protein is included in the geometry optimization (QM/QM/MM), this effect is much less pronounced (-61 *vs.* -55 10^{-40} esu $\cdot$ cm $\cdot$ erg $\cdot$ G^{-1}). This thus can not be taken as evidence that the proposed larger rotatory strength of the intermediate state[25] is caused by induced structural differences in the chromophore.

QM/PCM results

The results of the calculations with the protein environment modelled by means of PCM with two different dielectric constants are shown in Table 2. Taking into account the influence

Table 2: Absorption wavelengths (λ, in nm), rotatory (R, in 10^{-40} esu $\cdot$ cm $\cdot$ erg $\cdot$ G^{-1}) and oscillator strengths (f, dimensionless) calculated with PCM (with two different values of the dielectric constant, ε) in both the geometry optimization and ECD calculation.

Structure	ε	λ	R	f
neutral	4	359	-22.31	1.00
	15	359	-20.39	1.00
anionic	4	423	-20.49	1.25
	15	418	-16.98	1.24

of the environment on the electron density in the calculations of the absorption properties causes a red shift (14 to 16 nm) in the calculated excitation wavelengths, a decrease in the rotatory strength (6% neutral; 23% anionic) and an increase in the oscillator strength (13 to 16%) compared to the values in Table 1. The differences between the values obtained using the dielectric constants of 4 and 15 are very small for the neutral structure, but more pronounced for the anionic chromophore. The larger effect of the dielectric medium on the anionic than on the neutral species is explained by the fact that the cavity size of the neutral and anionic chromophores is similar, while the electron density of the anionic chromophore is more diffuse and thus more affected by the dielectric medium. In general, we can conclude

that the nature of the dielectric environment (hydrophobic: $\varepsilon = 4$; hydrophilic: $\varepsilon = 15$) *does* influence the ECD signal of the chromophore, but the effect is rather small.

PE-QM/MM results

The results of polarizable embedding QM/MM calculations (PE-QM/MM) are shown in Table 3. Compared to the results obtained for the chromophore in vacuum (Table 1), we observe

Table 3: Absorption wavelengths (λ, in nm), rotatory (R, in 10^{-40} esu $\cdot$ cm $\cdot$ erg $\cdot$ G^{-1}) and oscillator strengths (f, dimensionless) for the GFP chromophores optimized with the QM/QM/MM scheme and ECD calculations performed with the rest of the protein included through polarizable embedding (PE-QM/MM).

Structure	λ	R	f
neutral	362	-82.79	0.99
anionic	407	-79.20	1.23
intermediate	411	-71.18	1.24

shifts in the excitation wavelengths and increases in all oscillator and rotatory strengths. The excitation energy is more affected in the case of the neutral state (a red shift of 18 nm) than in the case of the anionic and intermediate states (blue shifts of 5 and 2 nm, respectively). In contrast to the PCM calculations (Table 2), the protein environment is here accounted for explicitly. The main effect of this explicit inclusion is a significant increase in the rotatory strengths. Indeed, the rotatory strengths are three to five times larger in PE-QM/MM calculations than in the PCM results presented in Table 2. The negative sign of the rotatory strengths is in agreement with experiment, while, contrary to what we expected on the basis of the work of Visser *et al*,[25] its magnitude is not higher for the intermediate than for the anionic state. The oscillator strengths are very similar for PCM (Table 2) and PE-QM/MM (Table 3).

Although the directions of the band shift from the anionic to the intermediate structure is reproduced, its magnitude of 4 nm is much less than the experimental 20 nm. This indicates that the difference between the anionic and intermediate forms can be induced by

changing the orientation of Thr203, yet that the computational procedure followed here is not adequate for a more detailed comparison with experiment. An important improvement can be made by including more than one conformation of the chromophore and its surrounding in the analysis, for example through molecular dynamics sampling. Moreover, structural differences between the neutral, anionic and intermediate forms of the protein can be induced by allowing a larger part of the protein to relax in the QM/QM/MM geometry optimization.

QM/QM results

In another series of calculations, we have included the minimal environment of the chromophore (nine amino acid residues and five water molecules) at the same level of theory (CAM-B3LYP/6-31+G*) as the chromophore itself. In one set of calculations (QM/QM/PCM) the remaining part of the protein has been modelled as a dielectric continuum ($\varepsilon = 4$), in the other set (QM/QM) no environment is present outside the cluster. Both sets of calculations are performed on the QM/QM/PCM-optimized structures. The results are shown in Table 4.

Table 4: Absorption wavelengths (λ, in nm), rotatory (R, in 10^{-40} esu $\cdot$ cm $\cdot$ erg $\cdot$ G^{-1}) and oscillator strengths (*f*, dimensionless) calculated for the QM/QM/PCM-optimized system (a cluster containing the GFP chromophore, 9 amino acid residues and 5 water molecules) at the CAM-B3LYP/6-31+G* level of theory with and without PCM.

Structure	Method	λ	R	f
neutral	QM/QM	363	-37.77	0.82
	QM/QM/PCM	366	-41.54	0.92
anionic	QM/QM	405	-45.44	0.99
	QM/QM/PCM	409	-53.33	1.10
intermediate	QM/QM	402	-51.34	0.99
	QM/QM/PCM	408	-53.05	1.10

Compared to the ECD signal of the bare chromophore (Table 1), the main change upon including the environment is an increase in the absolute value of the rotational strength.

Indeed, the rotatory strength increases by more than 50% for the anionic and neutral chromophores. Since the extra residues are not involved in the electronic excitation but change the chirality of the total system, the influence of an explicit description of the environment is larger on the rotatory strength than on the oscillator strength.

The calculated absorption wavelengths are rather similar to the ones calculated with polarizable embedding QM/MM (Table 3). However, the QM/QM calculations do not reproduce the red shift between the anionic and intermediate structures. Indeed, both the excitation wavelength and oscillator strength hardly differ between the anionic and intermediate structures, similarly to what has been observed for the bare chromophores. The rotatory strength is higher for the intermediate than for the anionic structure in the QM/QM calculations, while being of similar magnitude in the QM/QM/PCM calculations. Comparison with Table 1 shows that this is partially a structural effect in the chromophore rather than being related to electrostatic interactions with Thr203. Indeed, the addition of the environment induces a larger increase in the rotatory strength for the anionic state than for the intermediate state. These results thus do not seem to agree with the hypothesis[25] that the environment of the chromophore induces a higher rotatory strength in the intermediate state than in the anionic state. However, sampling of the conformational space of the chromophore and its surroundings is needed to shed light on systematic differences that the environment of the chromophore exerts on the anionic and intermediate forms.

The sign of the rotatory strength in the calculations performed on the structures obtained in QM/QM calculations is negative, which corresponds to the experimental data available for the neutral and anionic forms of the similar eGFP protein.[25]

We point out that a severe limitation of the QM/QM calculations—in comparison with the PE-QM/MM calculations—is the small size of the explicitly treated molecules in the surrounding protein. Indeed, only 115 out of the thousands of atoms in the surrounding protein are explicitly treated. Polarizable embedding calculations on the anionic form of GFP using different threshold distances for the inclusion of electrostatic and polarization interactions

between the chromophore and the protein have shown that both the former and, in particular, the latter extend up to a distance of around 20 Å.[37] Moreover, it has recently been shown that a good agreement between experimental and calculated absorption energies of the same cluster around the GFP chromophore[21] relies on coincidence rather than a correct incorporation of the effect of the protein environment.[54] On top of that, the QM/QM calculations are computationally much more expensive than the PE-QM/MM calculations, with the computational time rising exponentially with the number of basis functions, while the calculations with polarizable embedding QM/MM for this particular system take only roughly twice as much time when the whole protein is included in comparison with a QM/vacuum calculation of the chromophore only. Thus, only PE-QM/MM calculations have been carried out for the other fluorescent proteins.

Influence of water molecules

Water may penetrate close to the chromophore and affect its optical response. So far, water molecules have been included either quantum mechanically in the QM/QM calculations or classically through polarizable embedding (PE-QM/MM). Figure 4 presents the changes of the calculated optical properties caused by transferring the closest five water molecules one by one from the classical to the quantum region. Apart from the chromophore and these closest water molecule(s), the rest of the protein is described here through polarizable embedding. The placement of the five water molecules is shown in Figure 3.

First of all, transferring water molecules from the classical to the quantum region does not change the ordering of the neutral, intermediate and anionic states of GFP in the excitation wavelengths, rotatory strengths and oscillator strength. The only notable difference in this respect is the first water molecule—the one that hydrogen-bonds with the phenol(ate) part of the chromophore—which induces a slight increase in the difference in oscillator strength between the anionic and intermediate forms and a slight decrease in their rotatory strength difference.

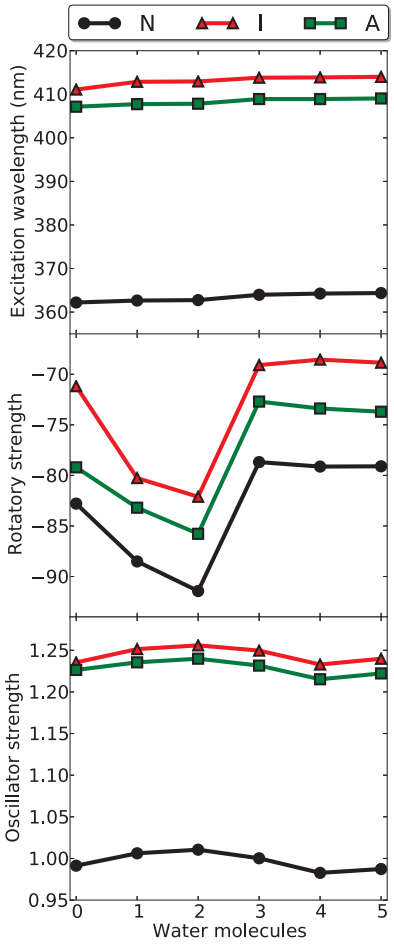

Figure 4: The effect on the excitation wavelength (nm), rotatory (10^{-40} esu $\cdot$ cm $\cdot$ erg $\cdot$ G^{-1}) and oscillator strengths (dimensionless) when transferring water from the classical to the quantum region. All calculations are done using PE-QM/MM with the GFP chromophore and a number of water molecules (0 to 5) in the QM region and the rest of the water molecules and protein environment described through polarizable embedding. Chromophore geometries were optimized with the QM/QM/MM scheme. Results are shown for the neutral (N), anionic (A) and intermediate (I) structures. The numbering of the water molecules is shown in Figure 3.

Furthermore, we clearly see in Figure 4 the success of the polarizable embedding model in describing the influence of the water molecules—including the hydrogen-bonding ones—on the excitation wavelength and oscillator strength of all three species. The rotatory strength, however, *does* change when the first three water molecules are transferred to the QM region. In fact, the first and to a lesser extent also the second water molecule induce an increase in the absolute value of the rotatory strength, while the third water molecule effectively damps this effect. The first and third water molecules are in direct contact with the conjugated system of the chromophore by hydrogen-bonding at the phenol(ate) position and by donating a hydrogen bond to the imidazolidone N, respectively. We note that there is no change in the sign of the rotatory strength and that the relative differences are less than 15% compared to the inclusion of zero water molecules in the QM region. The water molecules are not involved in the electronic excitations giving rise to the peaks in the absorption and ECD properties. However, the hydrogen-bonding water molecules change the electron density of the chromophore and thereby its chirality. The electron density of the surroundings is not explicitly modelled in the polarizable embedding calculations, which thus cannot account for this effect on the rotatory strength. Water molecules that do not directly interact with the chromophore are found to be described adequately by the polarizable embedding approach. We note that the results obtained here with the advanced polarizable embedding QM/MM model do not necessarily hold for any QM/MM approach.

We stress that the results shown in Figure 4 represent only one configuration of the chromophore, its surroundings and the available water molecules. The dynamic effects of water molecules penetrating, moving and leaving the surrounding of the chromophore are not taken into account here. Nevertheless, the results show how water molecules can alter the optical response of the chromophore and to what extent this can be modelled by polarizable embedding QM/MM. Sampling over different structures from e.g. a molecular dynamics simulation can help answering the questions what the net effect of water molecules is on the ECD properties of the chromophore and how well polarizable embedding performs to

describe this averaged effect.

Other fluorescent proteins

Additional calculations have been carried out for another five fluorescent proteins: BFP, eCFP, eGFP, YFP and DsRed. The chromophore structures of the different fluorescent proteins are shown in Figure 2. YFP, like GFP, exists in both neutral and anionic forms, so we have carried out the calculations for both of them. Since residue 203 in YFP is not a threonine but a tyrosine—interacting with the chromophore through π-stacking—the protein does not have the structural distinction between the anionic and intermediate form as GFP has. Calculations on the fluorescent proteins have been carried out on the chromophores with (PE-QM/MM) and without (QM/vacuum) describing the surrounding protein through polarizable embedding QM/MM (Table 5) and with the environment of the chromophore described by PCM (Table 6). The QM/QM calculations have not been carried out on these systems.

Table 5: Absorption wavelengths (λ, in nm), rotatory (R, in 10^{-40} esu $\cdot$ cm $\cdot$ erg $\cdot$ G^{-1}) and oscillator strengths (f, dimensionless) for the fluorescent proteins. The CAM-B3LYP/6-31+G* calculations were performed on the chromophore only (QM/vacuum) and using the polarizable embedding (PE-QM/MM) scheme. Chromophore geometries were optimized with the QM/MM scheme. $\lambda_{\mathbf{max}}$ is the experimental absorption maximum.

	QM/vacuum			PE-QM/MM			Experiment
Protein	λ	R	f	λ	R	f	$\lambda_{\max}$
BFP	328	-12.50	0.28	347	-40.61	0.80	382[16]
	318	-13.23	0.35				
eCFP	372	-17.76	0.64	399	-34.86	0.77	435[15]
eGFP	413	-47.78	1.04	421	-53.08	1.17	488[17]
YFP neutral	343	-38.84	0.87	368	-68.46	1.05	392[19]
YFP anionic	402	-73.65	1.13	421	-71.34	1.17	514[19]
DsRed	491	-137.57	1.16	442	-122.35	1.32	558[14]

The data in Tables 5 and 6 show similar effects as observed for GFP. Including the environment (through PCM or PE-QM/MM) leads to an increase in the oscillator strengths.

Table 6: Absorption wavelengths (λ, in nm), rotatory (R, in 10^{-40} esu $\cdot$ cm $\cdot$ erg $\cdot$ G^{-1}) and oscillator strengths (f, dimensionless) calculated for the PCM-optimized structures of the fluorescent proteins using DFT with CAM-B3LYP/6-31+G* and PCM. λ_{max} is the experimental absorption maximum.

Protein	ε	λ	R	f	λ_{max}
BFP	4	337	-6.89	0.85	382[16]
	15	338	-8.94	0.87	
eCFP	4	380	-21.99	0.84	435[15]
	15	383	-21.56	0.85	
eGFP	4	423	-4.21	1.25	488[17]
	15	419	-3.53	1.24	
YFP neutral	4	358	-34.47	1.00	392[19]
	15	359	-32.83	1.00	
YFP anionic	4	422	-38.44	1.27	514[19]
	15	418	-32.47	1.25	
DsRed	4	507	-59.82	1.25	558[14]
	15	501	-60.51	1.25	

Furthermore, the PCM results are not much affected by the value of the dielectric constant chosen, with a larger change in the excitation wavelength for the anionic chromophores than for the neutral ones. Finally, the rotatory strength is found to be higher when the environment is included explicitly (here: PE-QM/MM) than implicitly (here: QM/PCM).

Experimental ECD spectra of all five fluorescent proteins (Figure 5) under study have negative peaks.[25] The sign of the rotatory strength is reproduced correctly for all proteins with the different computational methods. We note that the relative magnitude of the rotatory strength is reproduced to a certain extent by the PE-QM/MM calculations (see Figure 5 and Table 5), with DsRed showing the highest rotatory strength followed by YFP and then the other fluorescent proteins. QM/PCM seems to perform worse in this respect, underestimating the rotatory strength in BFP and eGFP (compare Figure 5 and Table 6).

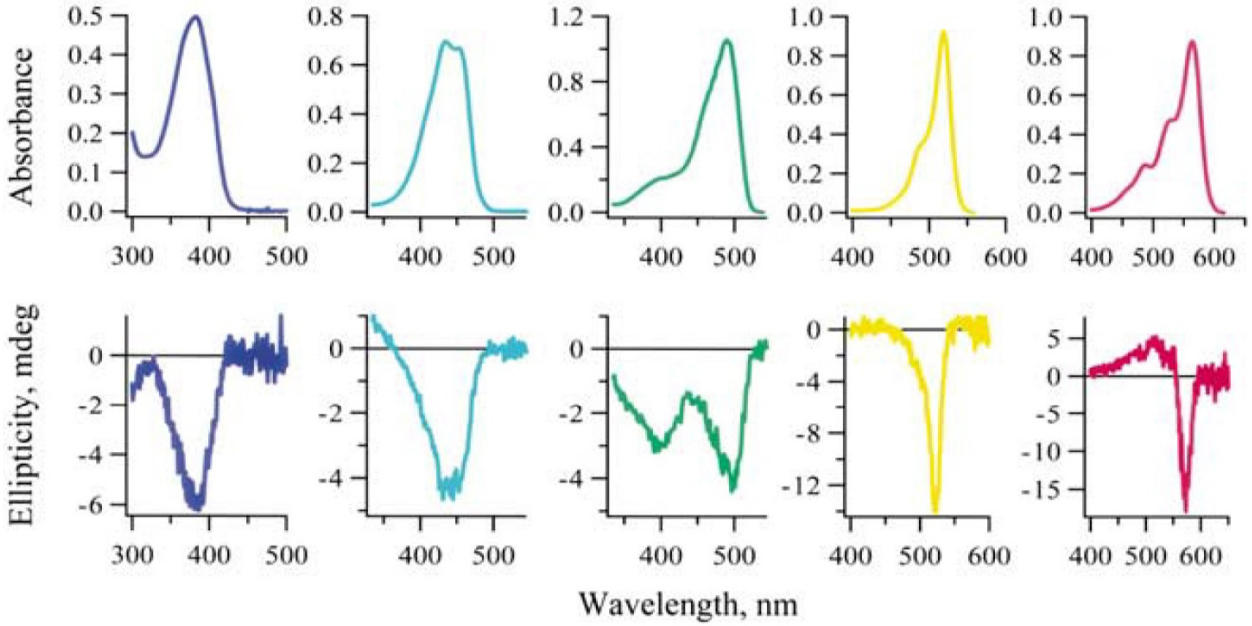

Figure 5: The experimental ECD spectra of BFP, eCFP, eGFP, YFP and DsRed from Visser *et al.*.[25] Reproduced with permission from Elsevier from Ref. 25.

Summary and concluding remarks

The electronic circular dichroism properties of the green fluorescent protein (GFP) and other fluorescent proteins have been modelled with density functional theory calculations for the chromophores in the protein environment simulated by means of different embedding models. The protein environment has been simulated by means of the polarizable continuum model (QM/PCM), the polarizable embedding model (PE-QM/MM) and by a supermolecular model, *i.e.*, treating the minimal environment quantum mechanically at the same footing as the chromophore and the remaining part either neglected (QM/QM) or simulated with PCM (QM/QM/PCM).

Reproducing the correct excitation wavelengths is crucial for a correct treatment of ECD, but has proven to be challenging for the systems under study. This seems, however, related more to the exchange–correlation functional and the quality of the molecular structures than to the type of embedding model employed. In particular, the red shift from the anionic to the intermediate forms of GFP is not quantitatively reproduced using the protein structures

used here. The red shift may be related to other structural changes in the surrounding of the chromophore, such as different orientation of Glu222,[55] which has not been investigated here.

Explicit inclusion of the environment (QM/QM or PE-QM/MM) leads to significant quantitative differences with a PCM treatment, in particular for the rotatory strengths. In general, the observed changes between different embedding models are larger for the rotatory strengths than for the oscillator strengths. Inclusion of the whole protein through polarizable embedding QM/MM is shown not only to give the sign of the rotatory strength in agreement with experiment for all fluorescent proteins under study but also to reproduce to a certain extent quantitative differences in the rotatory strength between different fluorescent proteins. Inclusion of the protein environment through PCM also gives the correct sign of the rotatory strength for all of the protein structures. The rotatory strengths, however, are much smaller than in an explicit treatment and reproduce the experimental quantitative differences less well. When water molecules directly interacting with the chromophore are explicitly included in the QM region, the magnitude of the rotatory strength changes while excitation energies and oscillator strengths are largely unaffected. Unfortunately, the QM/QM cluster calculations suffer from high computational costs even though only a small fraction of the protein is accounted for.

Our findings neither support nor reject the hypothesis of Visser *et al.*[25] that the intermediate form of GFP has a larger rotatory strength than the anionic one. In fact, QM/vacuum and QM/QM calculations show a larger rotatory strength of the intermediate form, which can however be related to structural changes in the chromophore rather than to the different surrounding of the anionic and intermediate forms. However, PE-QM/MM calculations show a slightly larger rotatory strength for the anionic form.

The main limitation in this study is the lack of dynamic effects, *i.e.*, the use of one single structure to model the ECD properties of a fluorescent protein. Indeed, the calculated optical signal of fluorescent proteins have been found to be very sensitive to conformational

changes.[22,24] Sampling over a larger number of structures will allow a better comparison with experiment. Another limitation in our approach is the neglect of vibronic effects, that have been shown to be important for absorption properties of fluorescent proteins.[23] Nevertheless, the findings with the different embedding models contain useful guidance for the modelling of ECD in biomolecular systems.

Acknowledgement

The authors thank Nanna Holmgaard List (University of Southern Denmark) for providing the structure of the DsRed protein, Jógvan Magnus Haugaard Olsen (EPFL Lausanne and University of Southern Denmark) for the preparation of the embedding potentials and Jógvan Magnus Haugaard Olsen and Jacob Kongsted (University of Southern Denmark) for discussion.

This work has received support from the Polish National Science Center (Grant No. 2012/05/B/ST4/01236), the COST-CMTS Action CM1002 COnvergent Distributed Environment for Computational Spectroscopy (CODECS) and from the Wrocław Centre for Networking and Supercomputing and the Norwegian Supercomputing Program (Grant No. NN4654K) through grants of computer time. A.P. was supported by the "Doctorates for Mazovia II", the scholarship programme for the best Ph.D. students at the University of Warsaw in the fields of vital importance for the regional economic development. A. H. S. and M. T. P. B. acknowledge support from the Research Council of Norway through a Centre of Excellence Grant (Grant No. 179568/V30) and from the European Research Council through a Starting Grant (Grant No. 279619).

Supporting Information Available

ECD results calculated with the B3LYP exchange–correlation functional and coordinates for all chromophores and clusters used.

This material is available free of charge via the Internet at `http://pubs.acs.org/`.

References

(1) Sreerama, N.; Woody, R. W. In *Circular Dichroism: Principles and Applications*; Berova, N., Nakanishi, K., Woody, R. W., Eds.; John Wiley & Sons: New York, 2000; Chapter Circular Dichroism of Peptides and Proteins, pp 601–620.

(2) *Comprehensive Chiroptical Spectroscopy: Volume 2 - Applications in Stereochemical Analysis of Synthetic Compounds, Natural Products, and Biomolecules*; John Wiley & Sons: Hoboken, NY, 2012.

(3) Cantor, C. R.; Schimmel, P. R. *Biophysical Chemistry, Part 2: Techniques for the Study of Biological Structure and Function*; W.H. Freeman and Company: New York, 1980.

(4) Rodger, A.; Nordén, B. *Circular Dichroism and Linear Dichroism*; (Oxford Chemistry Masters) Oxford University Press: Oxford New York Tokyo, 1997.

(5) Pecul, M.; Dzwolak, W. In *Optical Spectroscopy and Computational Methods in Biology and Medicine, Challenges and Advances in Computational Chemistry and Physics 14*; Barańska, M., Ed.; Springer Science+Business Media: Dordrecht, 2014; Chapter Electronic Circular Dichroism Spectroscopy in Structural Analysis of Biomolecular Systems, pp 161–177.

(6) Gawroński, J.; Grajewski, J. The Significance of Induced Circular Dichroism. *Org. Lett.* **2003**, *5*, 3301–3303.

(7) Zsila, F. A New Ligand for an Old Lipocalin: Induced Circular Dichroism Spectra Reveal Binding of Bilirubin to Bovine β–Lactoglobulin. *FEBS Lett.* **2003**, *539*, 85–90.

(8) Dzwolak, W.; Pecul, M. Chiral Bias of Amyloid Fibrils Revealed by the Twisted Conformation of Thioflavin T: an Induced Circular Dichroism/DFT Study. *FEBS Lett.* **2005**, *579*, 6601–6603.

(9) Tsien, R. Y. The Green Flourescent Protein. *Annu. Rev. Biochem.* **1998**, *67*, 509–544.

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

(11) Heim, R.; Prasher, D. C.; Tsien, R. Y. Wavelength Mutations and Posttranslational Autoxidation of Green Fluorescent Protein. *Proc. Natl. Acad. Sci.* **1994**, *91*, 12501–12504.

(12) Royant, A.; Noirclerc-Savoye, M. Stabilizing Role of Glutamic Acid 222 in the Structure of Enhanced Green Fluorescent Protein. *J. Struct. Biol.* **2011**, *174*, 385–390.

(13) Wachter, R. M.; Elsliger, M. A.; Kallio, K.; Hanson, G. T.; Remington, S. J. Structural Basis of Spectral Shifts in the Yellow-Emission Variants of Green Fluorescent Protein. *Structure* **1998**, *6*, 1267–1277.

(14) Matz, M. V.; Fradkov, A. F.; Labas, Y. A.; Savitsky, A. P.; Zaraisky, A. G.; Markelov, M. L.; Lukyanov, S. A. Fluorescent Proteins from Nonbioluminescent Anthozoa Species. *Nat. Biotechnol.* **1999**, *17*, 969–973.

(15) Lelimousin, M.; Noirclerc-Savoye, M.; Lazareno-Saez, C.; Paetzold, B.; Le Vot, S.; Chazal, R.; Macheboeuf, P.; Field, M. J.; Bourgeois, D.; Royant, A. Intrinsic Dynamics in ECFP and Cerulean Control Fluorescence Quantum Yield. *Biochemistry* **2009**, *48*, 10038–10046.

(16) Wachter, R. M.; King, B. A.; Heim, R.; Kallio, K.; Tsien, R. Y.; Boxer, S. G.; Remington, S. J. Crystal Structure and Photodynamic Behavior of the Blue Emission Variant Y66H/Y145F of Green Fluorescent Protein. *Biochemistry* **1997**, *36*, 9759–9765.

(17) Cormack, B. P.; Valdivia, R. H.; Falkow, S. FACS-Optimized Mutants of the Green Fluorescent Protein (GFP). *Gene.* **1996**, *173*, 33–38.

(18) Chattoraj, M.; King, B. A.; Bublitz, G. U.; Boxer, S. G. Ultra-Fast Excited

State Dynamics in Green Fluorescent Protein: Multiple States and Proton Transfer. *Proc. Natl. Acad. Sci.* **1996**, *93*, 8362–8367.

(19) Wachter, R. M.; Yarbrough, D.; Kallio, K.; Remington, S. J. Crystallographic and Energetic Analysis of Binding of Selected Anions to the Yellow Variants of Green Fluorescent Protein. *J. Mol. Biol.* **2000**, *301*, 157–171.

(20) Filippi, C.; Buda, F.; Guidoni, L.; Sinicropi, A. Bathochromic Shift in Green Fluorescent Protein: A Puzzle for QM/MM Approaches. *J. Chem. Theory Comput.* **2012**, *8*, 112–124.

(21) Kaila, V. R. I.; Send, R.; Sundholm, D. Electrostatic Spectral Tuning Mechanism of the Green Fluorescent Protein. *Phys. Chem. Chem. Phys.* **2013**, *15*, 4491–4495.

(22) Amat, P.; Nifosì, R. Spectral "Fine" Tuning in Fluorescent Proteins: The Case of the GFP-Like Chromophore in the Anionic Protonation State. *J. Chem. Theory Comput.* **2013**, *9*, 497–508.

(23) Kamarchik, E.; Krylov, A. I. Non-Condon Effects in the One- and Two-Photon Absorption Spectra of the Green Fluorescent Protein. *J. Phys. Chem. Lett.* **2011**, *2*, 488–492.

(24) Beerepoot, M. T. P.; Steindal, A. H.; Kongsted, J.; Brandsdal, B. O.; Frediani, L.; Ruud, K.; Olsen, J. M. H. A Polarizable Embedding DFT Study of One-Photon Absorption in Fluorescent Proteins. *Phys. Chem. Chem. Phys.* **2013**, *15*, 4735–4743.

(25) Visser, N. V.; Hink, M. A.; Borst, J. W.; van der Krogt, G. N.; Visser, A. J. Circular Dichroism Spectroscopy of Fluorescent Proteins. *FEBS Lett.* **2002**, *521*, 31–35.

(26) Goto, H.; Sawada, I.; Nomura, N. Circular Dichroism and Circular Polarized Luminescence from a Green Fluorescent Protein: Initial Research for Chiroptical Emission of Biological Materials. *Int. J. Polym. Mater.* **2010**, *59*, 786–792.

(27) Jayaraman, S.; Haggie, P.; Wachter, R. M.; Remington, S. J.; Verkman, A. S. Mechanism and Cellular Applications of a Green Fluorescent Protein-Based Halide Sensor. *J. Biol. Chem.* **2000**, *275*, 6047–6050.

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

(29) Runge, E.; Gross, E. K. U. Density-Functional Theory for Time-Dependent Systems. *Phys. Rev. Lett.* **1984**, *52*, 997–1000.

(30) Shcherbin, D.; Ruud, K. The Use of Coulomb-Attenuated Methods for the Calculation of Electronic Circular Dichroism Spectra. *Chem. Phys.* **2008**, *349*, 234–243.

(31) Cancès, E.; Mennucci, B.; Tomasi, J. A New Integral Equation Formalism for the Polarizable Continuum Model: Theoretical Background and Applications to Isotropic and Anisotropic Dielectrics. *J. Chem. Phys.* **1997**, *107*, 3032–3041.

(32) 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.

(33) List, N. H.; Olsen, J. M.; Rocha-Rinza, T.; Christiansen, O.; Kongsted, J. Performance of Popular XC-Functionals for the Description of Excitation Energies in GFP-Like Chromophore Models. *Int. J. Quantum Chem.* **2012**, *112*, 789–800.

(34) Brkljača, Z.; Mališ, M.; Smith, D. M.; Smith, A.-S. Calculating CD Spectra of Flexible Peptides: An Assessment of TD-DFT Functionals. *J. Chem. Theory Comput.* **2014**, *10*, 3270–3279.

(35) 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.

(36) 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.

(37) Beerepoot, M. T. P.; Steindal, A. H.; Ruud, K.; Olsen, J. M. H.; Kongsted, J. Convergence of Environment Polarization Effects in Multiscale Modeling of Excitation Energies. *Comput. Theor. Chem.* **2014**, *1040–1041*, 304–311.

(38) Olsen, J. M. H. Development of Quantum Chemical Methods towards Rationalization and Optimal Design of Photoactive Proteins. Ph.D. thesis, University of Southern Denmark, Odense, Denmark, 2012; DOI: 10.6084/m9.figshare.156852.

(39) Gilson, M. K.; Honig, B. Calculation of the Total Electrostatic Energy of a Macromolecular System: Solvation Energies, Binding Energies, and Conformational Analysis. *Proteins: Structure, Function, and Bioinformatics* **1988**, *4*, 7–18.

(40) Schutz, C. N.; Warshel, A. What are the Dielectric "Constants" of Proteins and How to Validate Electrostatic Models? *Proteins: Structure, Function, and Bioinformatics* **2001**, *44*, 400–417.

(41) Li, L.; Li, C.; Zhang, Z.; Alexov, E. On the Dielectric "Constant" of Proteins: Smooth Dielectric Function for Macromolecular Modeling and Its Implementation in DelPhi. *J. Chem. Theory Comput.* **2013**, *9*, 2126–2136.

(42) Tubbs, J. L.; Tainer, J. A.; Getzoff, E. D. Crystallographic Structures of Discosoma Red Fluorescent Protein with Immature and Mature Chromophores: Linking Peptide Bond Trans-Cis Isomerization and Acylimine Formation in Chromophore Maturation. *Biochemistry* **2005**, *44*, 9833–9840.

(43) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.

(44) Ormö, M.; Cubitt, A. B.; Kallio, K.; Gross, L. A.; Tsien, R. Y.; Remington, S. J. Crystal Structure of the Aequorea Victoria Green Fluorescent Protein. *Science* **1996**, *273*, 1392–1395.

(45) Yang, F.; Moss, L. G.; Philips Jr., G. N. The Molecular Structure of Green Fluorescent Protein. *Nat. Biotechnol.* **1996**, *14*, 1246–1251.

(46) Aidas, K.; Angeli, C.; Bak, K. L.; Bakken, V.; Bast, R.; Boman, L.; Christiansen, O.; Cimiraglia, R.; Coriani, S.; Dahle, P. et al. The Dalton Quantum Chemistry Program System. *WIREs Comput. Mol. Sci.* **2014**, *4*, 269–284.

(47) Dalton, a Molecular Electronic Structure Program, Release Dalton2014.0 (2014), see http://daltonprogram.org/.

(48) Olsen, J. M. H. The Polarizable Embedding (PE) Library (Version 1.2). 2014.

(49) Gao, B.; Thorvaldsen, A. J. 2012; Gen1Int Version 0.2.1, http://repo.ctcc.no/projects/gen1int.

(50) Gao, B.; Thorvaldsen, A. J.; Ruud, K. GEN1INT: A Unified Procedure for the Evaluation of One-Electron Integrals over Gaussian Basis Functions and Their Geometric Derivatives. *Int. J. Quantum Chem.* **2011**, *111*, 858–872.

(51) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A. et al. Gaussian 09, Revision A. 02. 2009; Gaussian, Inc. , Wallingford, CT.

(52) QSite, version 5.8, Schrödinger, LLC, New York, NY, 2012.

(53) Murphy, R. B.; Philipp, D. M.; Friesner, R. A. A Mixed Quantum Mechanics/Molecular Mechanics (QM/MM) Method for Large-Scale Modeling of Chemistry in Protein Environments. *J. Comput. Chem.* **2000**, *21*, 1442–1457.

(54) Schwabe, T.; Beerepoot, M. T. P.; Olsen, J. M. H.; Kongsted, J. Analysis of Computational Models for an Accurate Study of Electronic Excitations in GFP. *Phys. Chem. Chem. Phys.* **2015**, *17*, 2582–2588.

(55) Grigorenko, B. L.; Nemukhin, A. V.; Polyakov, I. V.; Morozov, D. I.; Krylov, A. I. First-Principles Characterization of the Energy Landscape and Optical Spectra of Green Fluorescent Protein along the A →I →B Proton Transfer Route. *J. Am. Chem. Soc.* **2013**, *135*, 11541–11549.