The quality of the embedding potential is decisive for minimal quantum region size in embedding calculations – The case of the green fluorescent protein

Lina J. Nåbo,[*,†] Jógvan M. H. Olsen,[†] Todd J. Martínez,[*,‡,¶] and Jacob Kongsted[*,†]

†*Department of Physics, Chemistry and Pharmacy, University of Southern Denmark, Campusvej 55, 5230 Odense M, Denmark*

‡*Department of Chemistry and PULSE Institute, Stanford University, Stanford, California 94305, United States*

¶*SLAC National Accelerator Laboratory, Menlo Park, California 94305, United States*

E-mail: naabo@sdu.dk; todd.martinez@stanford.edu; kongsted@sdu.dk

Abstract

The calculation of spectral properties for photoactive proteins is challenging because of the large cost of electronic structure calculations on large systems. Mixed quantum mechanical (QM) and molecular mechanical (MM) methods are typically employed in order to make such calculations computationally tractable. This study addresses the connection between the minimal QM region size and the method used to model the MM region in the calculation of absorption properties – here exemplified for calculations on the green fluorescent protein. We find that polarizable embedding is necessary for a qualitatively correct description of the MM region, and that this enables the

use of much smaller QM regions compared to fixed charge electrostatic embedding. Furthermore, absorption intensities converge very slowly with system size and inclusion of effective external field effects in the MM region through polarizabilities is therefore very important. Thus, this embedding scheme enables accurate prediction of intensities for systems that are too large to be treated fully quantum mechanically.

1 Introduction

In recent years there has been much interest in the spectroscopic properties of biological systems, which are typically both large and complex. The size of these systems presents a huge challenge for accurate quantum mechanical (QM) calculations that quickly become very costly due to their steep scaling with respect to system size. A popular approach to overcome this problem is to use embedding techniques, which rely on a division of the total system into subsystems that can be treated at different levels of theory. Typically the active part of the system, e.g. the chromophore in a photoactive protein, is treated using a QM approach, while the rest of the system is described with classical methods such as molecular mechanics (MM) force fields.[1–5] The two parts are referred to as the QM and MM regions, respectively. A central aspect of the quality of an embedding calculation is the size of the QM region. The QM region must be big enough to capture the main quantum effects of the system or process without being too costly. Low computational cost is particularly important when dynamical effects are included, since molecular dynamics requires a large number of electronic structure calculations. Furthermore, it is important to choose the QM region carefully. This is not trivial, and it has been shown that the outcome of QM/MM calculations on proteins can be very sensitive to the choice of amino acid residues included in the QM region.[6–9]

An aspect that has not been much discussed in the literature is the importance of the quality of the embedding potential in specific applications – and especially how this quality of the embedding potential is connected to the QM region size.[10,11] Indeed, as we will show in this paper, the necessary QM region size is highly coupled to the level of description of the MM region and depends on the specific properties investigated. So far, studies pointing to the need for large QM regions typically rely on fixed point charges from standard MM force fields to describe the electrostatic perturbation of the environment on the QM region.[7,9,12–14] These point charges are generally not designed for use in combination with electronic structure calculations, and should therefore be used with caution in QM/MM applications. On the other hand, studies that employ more elaborate embedding schemes allowing for mutual

polarization between QM and MM regions have obtained satisfactory results with quite small QM regions.[15] Another critical aspect that needs to be considered is the accuracy of the molecular structure that is used in the embedding calculation. Many spectroscopic properties are very sensitive to small alterations in the geometry, and the use of accurate structures is therefore a prerequisite for obtaining accurate predictions of spectroscopic properties.

In this study we compare the use of electrostatic and polarizable embedding in the calculation of electronic excitation energies and absorption intensities for the challenging case of the green fluorescent protein (GFP). The QM region size is varied and the effects of the two types of embedding are analyzed. The aim is to illustrate how the quality of the embedding potential influences the size of the QM region needed for accurate results. GFP was discovered by Shimomura et al.[16] as a companion protein to the chemiluminescent protein aequorin isolated from *Aequorea* jellyfish. GFP fluoresces strongly and is one of the most widely used fluorescent proteins in biochemistry and cell biology. It is often used as a marker of gene expression and protein targeting in living cells and organisms.[17]

X-ray crystal structures have provided detailed information about the structure of GFP. The chromophore is a *p*-hydroxybenzylideneimidazolinone formed from three amino acid residues (a serine, a tyrosine and a glycine) in the native protein (see Figure 1). The chromophore is surrounded by an 11-stranded β-barrel threaded by an α-helix containing the chromophore. A number of polar residues as well as structured water molecules are found in close proximity of the chromophore and are very important for the optical properties of GFP. Mutations in one or more of these amino acids can shift the absorption and emission maxima considerably.[17] The excitation spectrum of wild-type GFP has two clear peaks: a major peak at 3.14 eV and a minor peak at 2.61 eV.[17] These peaks are associated with two distinct protonation states of the chromophore, with the peak at 3.14 eV stemming from a neutral chromophore and the peak at 2.61 eV arising from an anionic form, where the phenol is deprotonated. Wild-type GFP contains about a 6:1 ratio of neutral-to-anionic forms, but the ratio of the peak heights in the absorption spectrum is only 3:1,[17] indicating

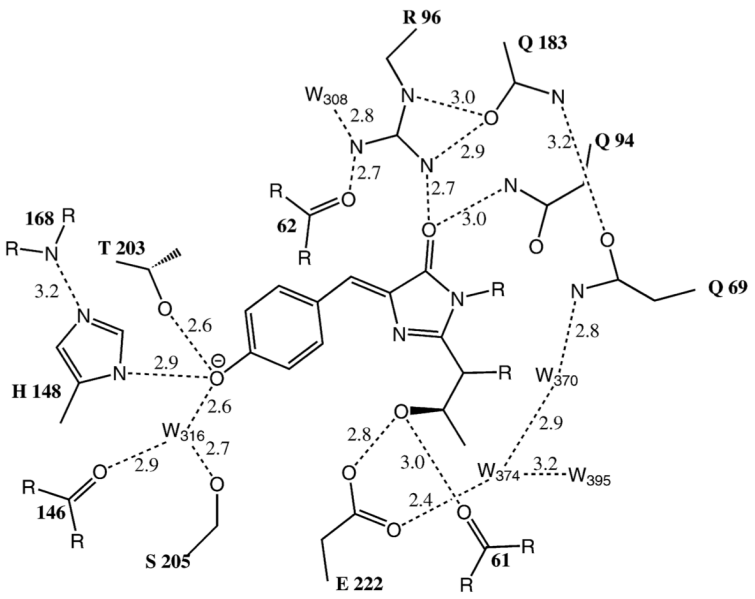


Figure 1: The chemical structure of the wild-type GFP chromophore in its deprotonated state with the closest amino acid residues. The residue numbering corresponds to the Protein Data Bank structure 1EMA. Hydrogen atoms are omitted for clarity. The figure is taken from ref. 17.

that the anionic form has a stronger absorption. Furthermore, excitation of the neutral and anionic chromophores gives almost identical emission maxima because the excited neutral chromophore undergoes excited state proton transfer to form the anionic species.[18] Here, we consider only the anionic form of GFP.

A thorough study of the GFP chromophore–protein coupling was recently published by Daday *et al.*,[15] investigating different electronic structure methods and embedding schemes. Dynamical fluctuations were also included with the aim of reproducing the experimental value for the absorption maximum. In our study, we focus on the importance of the quality of the embedding potential for converging both excitation energies and intensities with respect to QM region size for one equilibrated geometry and within the chosen level of theory. Particularly, the convergence of intensities with system size has not been given much focus in previous literature and is addressed here. The following section explains the computational procedure and briefly describes the two types of embedding used (fixed charge and polarizable representations). The absorption spectra are calculated using the different methods and

presented in the Results and Discussion section.

2 Methods

2.1 Simulations

A requirement for the calculation of accurate absorption energies is the use of geometrically accurate structures. For the generation of the structure used in this study, we use a combined QM and classical approach. The general procedure is presented here, and a detailed description of the molecular dynamics (MD) simulations can be found in the Supporting Information. The anionic form of the GFP is first solvated in water and then simulated using classical MD in order to equilibrate the structure at the desired temperature and pressure. The ff99SBildn Amber force field[19] was used for the ions and protein, the general Amber force field (GAFF)[20] was used for the GFP chromophore, and the water molecules were described with the flexible water model SPC/Fw.[21] In order to obtain an accurate geometry for the chromophore, we then switch to a combined QM/MM potential via the TeraChem[22]/Amber[23] interface,[9] using an electrostatic embedding scheme where the chromophore and nearby protein residues and water molecules (QM region 3 described below) are treated using density functional theory (DFT) at the CAM-B3LYP[24]/6-31G*[25] level of theory. The system is equilibrated for 10 ps, and the last structure, shown in Figure 2, is used in the subsequent analysis.

2.2 Absorption Calculations

Only one geometry, taken from the QM/MM simulation employing QM region 3 described below, is considered for the analysis of the QM region size convergence. A full calculation of the absorption spectrum for comparison with experiment (which is not the goal here) should include a large number of sampled geometries in order to account for disorder in the protein environment. Keeping the geometry fixed, six QM regions of various sizes are constructed

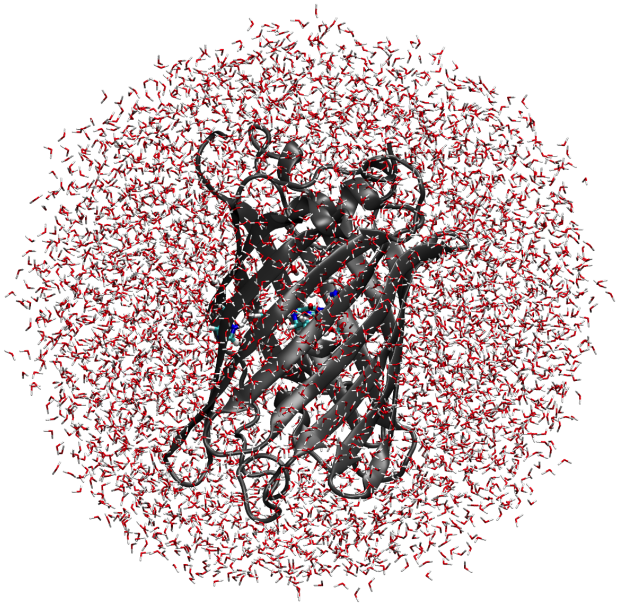
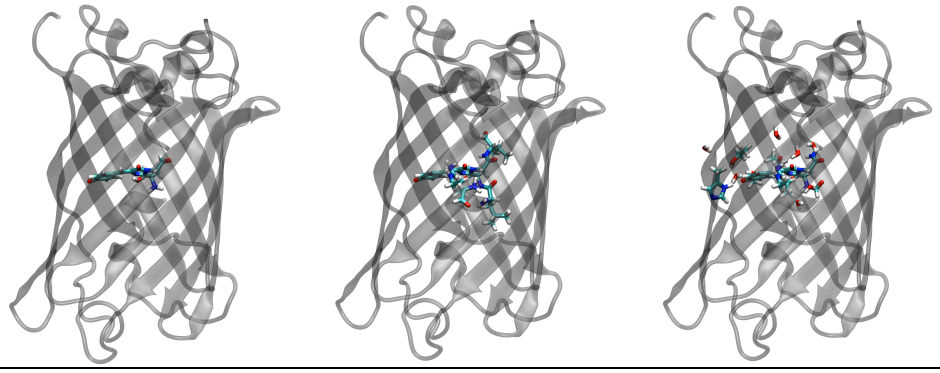
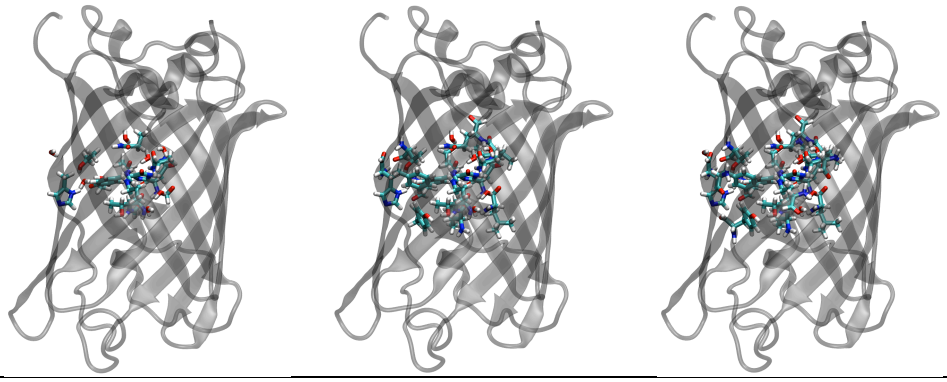


Figure 2: The system simulated using QM/MM. The QM region 3 is shown in stick representation and the rest of the GFP in gray cartoon representation. The sphere of water is shown in stick representation. The figure was produced with the Visual Molecular Dynamics program (VMD).[26]

for the absorption calculations. They are illustrated in Figure 3 together with an overview of the total number of QM atoms in each region as well as how many of these are H link atoms. The smallest region 1 consists of the chromophore only and carries a negative charge. Region 2 includes also the two neighboring amino acids L64 and V68 as well as the positively charged R96 adjacent to the chromophore. Region 2 is thus neutral, and so also are regions 3–6. Region 3, which was used in the QM/MM simulations generating the equilibrated structure, consists of the chromophore, side chains of residues T203, H148, S205, R96 and E222 (protonated), and the seven water molecules closest to the chromophore. Connecting peptide bonds are also included in the QM region. In addition to the residues in region 3, region 4 includes the side chains of residues T62, Q94 and Q183. Region 5 is region 4 plus the backbone of residues T62, Q94 and Q183, as well as residues V61, L64, V68 and the side chains of residues L42, Y145 and F165. The backbone of residues L42, Y145 and F165 are further included in region 6 as well as an additional water molecule.



QM region	1	2	3
Total no. QM atoms	35	96	117
No. link atoms	2	4	7



QM region	4	5	6
Total no. QM atoms	161	281	317
No. link atoms	11	17	22

Figure 3: The six QM regions are shown in stick representation and the rest of the GFP is shown in gray, transparent cartoon representation. Water molecules that are not part of a QM region are omitted for clarity. The number of QM atoms including link atoms as well as the number of link atoms are given for each region.

2.2.1 Electrostatic Embedding

In the electrostatic embedding scheme the MM region is represented by the force field point charges that were used in the MD simulations. These charges electrostatically perturb the charge density of the QM region. The absorption calculations are performed using the TeraChem/Amber interface, similar to the QM/MM simulations. For each QM region size, the 3–10 (depending on the QM region) lowest excitation energies and their respective oscillator strengths were calculated using time-dependent DFT (TD-DFT) at the CAM-B3LYP/6-31G* level of theory using full linear response. Although CAM-B3LYP gives blue-shifted values for excitation energies compared to the experimental absorption maximum, it has been shown to be in excellent agreement with CASPT2 regarding general trends for the anionic chromophore of GFP.[15] In the QM/MM calculation employing electrostatic embedding, QM atoms and any MM atoms involved in link bonds have their force field charges zeroed. To conserve the total charge of the QM/MM system, the charge of the MM region is adjusted by dividing any charge correction among all MM atoms (except for those adjacent to link atoms).

2.2.2 Polarizable Embedding

An alternative way of introducing the effect of the protein and solvent environment into electronic structure calculations is to use a polarizable embedding that allows the QM and MM regions to polarize each other. Here, we employ the polarizable embedding (PE) model,[27,28] which is an approach that is tailored for the calculation of environmental effects on spectroscopic properties (see ref. 29 for a recent perspective). The atoms in the MM region are assigned multipole moments (here up to quadrupole moments) and polarizabilities (here electric dipole–dipole polarizabilities) that give rise to induced dipole moments. The latter are obtained self-consistently, leading to a mutual polarization between the QM and MM regions. Excited state properties are calculated using TD-DFT and the response of the environment upon QM region excitation is therefore modeled using the transition density (this

type of embedding is sometimes denoted polLR[15]).

When a system is subjected to an external electric field ($\mathbf{F}$) as when exposed to irradiation, the whole system – not only the chromophore – is polarized. The local field at the site of the chromophore is modified by the environment (MM region) polarization compared to the external field. Depending on the protein structure, the polarization might lead to an enhancement of the external field in one direction and a reduction in another. This direct polarization of the MM region is accounted for in the effective external field extension of the PE model (PE-EEF).[30] The relation between the effective external field ($\mathbf{F}^{\mathrm{EEF}}$) and the applied field at location $\mathbf{R}$ within the QM region can be expressed in terms of an effective external field tensor[30]

$$\mathbf{L}^{\mathrm{EEF}}(\mathbf{R}) = \frac{\partial \mathbf{F}^{\mathrm{EEF}}(\mathbf{R})}{\partial \mathbf{F}}. \tag{1}$$

In this tensor a diagonal element with a value larger than one thus corresponds to an environment-induced enhancement of the external field in that direction, while a value below one indicates a screening of the external field. The direct MM polarization only affects the strength of the electric field at the chromophore and leaves the frequency unaltered. Hence, excitation energies are unaffected by local field effects and only the oscillator strengths are modified. In this work we compare the PE and PE-EEF models to a full-QM cluster calculation to assess their abilities to reproduce the intensity from the full-QM calculation.

The PE(-EEF)-TD-DFT calculations were performed using a development version of Dalton[31] together with PElib[32] and Gen1Int,[33,34] and using the CAM-B3LYP exchange–correlation functional with the 6-31G* basis set for the QM region. The embedding potentials used in the PE(-EEF)-TDDFT calculations were based on calculations performed on each solvent molecule or protein residue in isolation using the B3LYP[35–37] exchange–correlation functional with the 6-31G* basis set. However, the CAM-B3LYP functional was employed to calculate the potential that was used in the cluster calculations to enable a direct comparison with the reference full-QM calculation. The QM region and embedding potential were constructed in the PE assistant script in combination with FragIt.[38] The po-

larizabilities and multipole moments were distributed to atomic centers using the localized properties (LoProp) method by Gagliardi et al.[39] The embedding potential is thus calculated based on the specific geometry of the protein in the used snapshot and does not include any predefined parameters. To avoid overpolarization, polarizabilities and multipoles of order one and higher on MM atoms closer than 1.3 Å to a QM atom were removed and charges were redistributed to the three closest MM atom neighbors at a distance further than 1.3 Å from the QM region to preserve the total charge in the environment. The redistribution was done using PElib prior to the wave function optimization.

From each calculation, electrostatic or polarizable, the three most intense transitions in the relevant part of the electromagnetic spectrum were used to construct the absorption spectrum. For all PE-calculations, except for the one using the smallest QM region 1, there was only one intense transition and it was found as the lowest. However, with electrostatic embedding, artificial, non-intense charge–transfer (CT) states appear at low energies as the QM region is increased, starting from QM region 5. The intense transition is easily identified amongst the excited states, but the number of CT states at lower energies increase as the size of the QM region increases.

The excitation energies with their respective oscillator strengths were convoluted using a Gaussian function. The standard deviation (σ) determines the width of the Gaussian, i.e. the estimated broadening of the absorption peak related to the dynamics and excited state life time. A σ-value of 0.04 eV is used. The spectra calculated using the different QM regions and embedding methods are compared and discussed in the following section.

3 Results and Discussion

Electrostatic Embedding. We first discuss the convergence of the absorption spectrum with respect to QM region size in the case of electrostatic embedding. A calculation using QM region n and describing the MM region by force field point charges is denoted QMn/Q, e.g.

QM1/Q for an electrostatic embedding calculation with QM region 1. Figure 4 shows the spectra for QM regions 1–6, with (full lines) or without (dashed lines) the rest of the system present and described by point charges. The qualitative changes observed when the MM region is added to the isolated QM region 1 indicates that this region is too small and/or suffers from boundary effects (the covalent bonds to the chromophore's two neighboring amino acid residues are cut and replaced by H link atoms). Region 2 on the other hand gives one well-defined peak in isolation which is merely shifted to higher energy by the inclusion of the rest of the system. Region 2 is therefore defined as the minimal QM region size needed for a qualitatively correct description of this system. It is clear from Figure 4 that increasing the size of the QM region shifts the absorption peak to lower energies. On the other hand, inclusion of the static MM region around QM regions 2 and 3 shifts the absorption to higher energies, i.e. this embedding moves the absorption peak further away from the peak obtained with large QM regions. In the calculations using larger QM regions (regions 4–6) the MM region has little influence on the absorption maximum. Furthermore, it is seen that the absorption intensity also converges slowly with QM region size, and that the electrostatic embedding has no effect on the intensity from QM region 3. As shown in the Supporting Information, increasing the QM region size further to include up to 581 atoms does not change the absorption spectrum significantly, and the absorption maximum for QM6/Q can be considered to be converged for all practical purposes. We therefore use this calculation as our reference in the following analysis.

Polarizable Embedding. Using the PE-EEF approach to account for the presence of the MM region (Figure 5) brings the absorption maximum for a calculation using QM region 2 (yellow full line) only 0.02 eV from the reference result (black dash-dotted line). This calculation with QM region 2 and the MM region described by multipoles up to second order, i.e. quadrupoles (M2), and dipole–dipole electric polarizabilities (P2) is denoted QM2/M2P2. The QM1/M2P2 calculation, like the QM1/Q calculation, shows a splitting of the absorption,

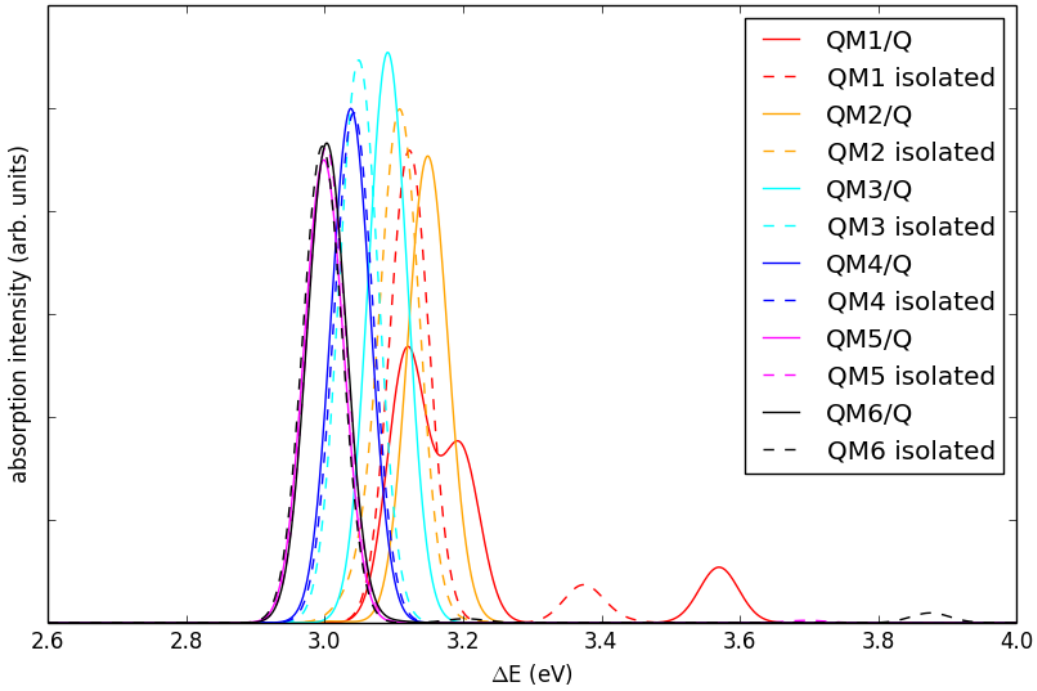


Figure 4: Absorption spectra of anionic GFP calculated at the CAM-B3LYP/6-31G* level of theory from a single geometry using different QM regions with (full lines) or without (dashed lines) the MM region described by point charges. Absorption intensity is given in arbitrary units and $\sigma = 0.04$ eV.

indicating a too small QM region. Increasing the QM region size from QM2 to QM3 gives a red-shift of only 0.03 eV. Further increase of the QM region gives even smaller shifts and the excitation energy is thus converged with the QM3/M2P2 calculation within the used basis set. Notably, the intensity of absorption is converged already using QM2/M2P2 PE-EEF, and this level of theory can therefore be used with good approximation in practical applications.

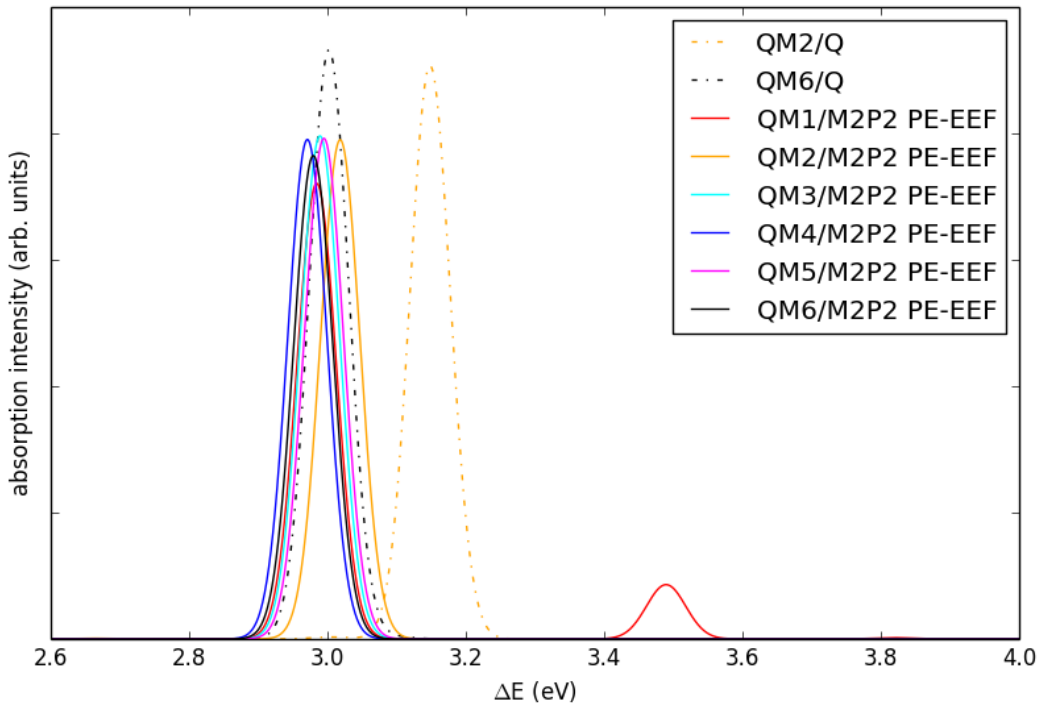


Figure 5: Absorption spectra of anionic GFP calculated at the CAM-B3LYP/6-31G* level of theory using different QM regions with the MM region described using PE-EEF (full lines) or by point charges (dash-dotted lines). Absorption intensity is given in arbitrary units and $\sigma = 0.04$ eV.

The PE-EEF calculations predict a weaker absorption than the QM6/Q calculation. This can be expected since the external field is modified by the whole system in PE-EEF (25506 atoms) but only by the 317 atoms in QM region 6 in the QM6/Q calculation. Inspection of the EEF tensor at the center of mass of QM region 2 in the QM2/M2P2 PE-EEF calculation (see eq. 1 in the Supporting Information) reveals an enhancement of the electric field in

parallel with the barrel (y-direction, see Figure 6) and a reduction perpendicular to the barrel (x- and z-directions). Since the transition dipole moment for the intense transition in

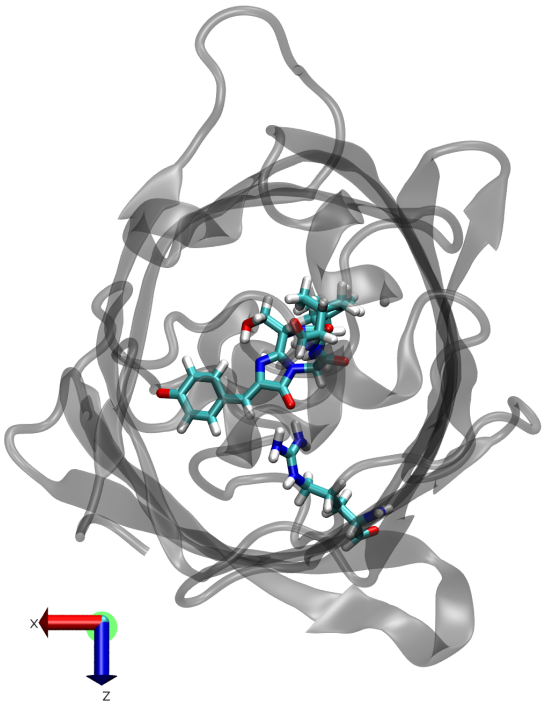


Figure 6: The anionic form of GFP, looking down through the β-barrel. QM region 2 is shown in stick representation and the barrel in transparent cartoon representation. Water is omitted for clarity. The figure was produced with VMD.[26]

QM region 2 is dominated by its x-component (data not shown) the screening of the field in this direction explains the weaker absorption predicted when the whole protein and solvent are included. In summary, these results indicate that the QM6/Q calculation is not yet fully converged with respect to absorption intensity (see also discussion in the Supporting Information).

To assess the ability of the PE-EEF model to reproduce the intensity of a full-QM calculation, further analysis is performed for a cluster consisting of only region 6 – discarding the rest of the system. QM regions 2–4 are used and the rest of the cluster is modeled by PE-EEF (or PE for comparison for QM region 2). As expected, the PE model leads to a slight overestimation of the absorption strength,[30,40] while the PE-EEF model closely reproduces

the full-QM intensity already with QM region 2 and the spectrum is fully converged with QM region 4 (see Figure 7). The EEF tensor in the QM2/M2P2(cluster) PE-EEF calculation (given in eq. 2 of the Supporting Information) is similar to the tensor for the whole system with the same QM region, but has diagonal elements closer to 1, which corresponds to a smaller modification of the external field as can be expected for a smaller system. Furthermore, the excitation energy with QM region 2 is overestimated by only 0.05 eV compared to the full-QM reference.

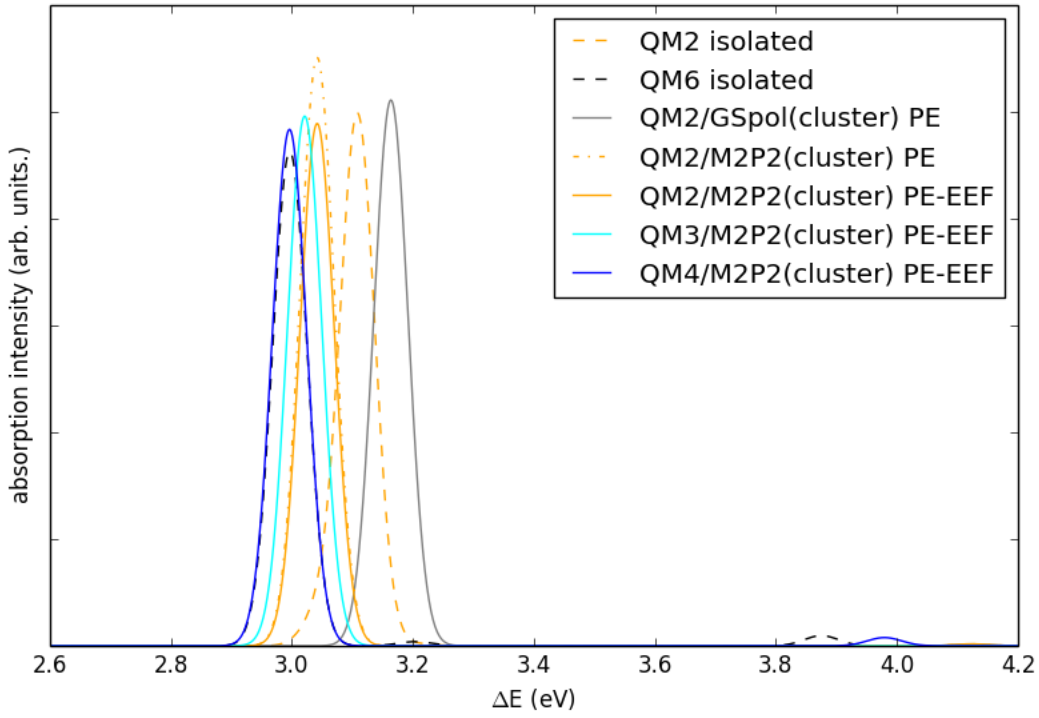


Figure 7: Absorption spectra for the region 6 cluster calculated at the CAM-B3LYP/6-31G* level of theory using a full-QM description (black) or QM region 2 with (magenta and cyan) or without (yellow) the rest of region 6 included using PE or PE-EEF. Absorption intensity is given in arbitrary units and $\sigma = 0.04$ eV.

Next, we deduce whether the red-shift predicted by the QM2/M2P2 calculation compared to the isolated region 2 is caused by the elaborate multipole description of the MM region or is a result of the MM region response to the excitation. To this end, the absorption spectrum for the cluster is calculated using a frozen environment, i.e. the induced dipoles are

converged in the ground state optimization and kept frozen during the TD-DFT calculation (denoted GSpol). The frozen-environment calculation does not produce the expected red-shift compared to the isolated QM region 2, but instead leads to an erroneous blue-shift (see cyan line in Figure 7). It is thus clear that the dynamic polarization of the MM region is needed in order for the model to capture even the qualitative effect of the chromophore surroundings during the excitation process.

4 Conclusions

In this study we have compared two different approaches to account for the environment during electronic excitation of the anionic form of GFP and evaluated the requirements for the size of QM region in the two cases. For smaller QM regions it is found that inclusion of fixed charge electrostatic embedding moves the absorption maximum away from the reference peak, in agreement with previous observations for anionic GFP.[15] Consequently, when a frozen environment is used the QM region must be large enough to include the major polarization effects of the chromophore's surroundings.

In the case of polarizable embedding a much smaller QM region may be employed and still produce accurate results, in agreement with previous studies.[10,15] Until now, many QM/MM studies have focused on excitation energies when evaluating the convergence of QM region sizes. However, the results obtained here show that the absorption intensity also converges slowly with system size. This observation can be rationalized by realizing that the screening term representing the modification of the field enters on the same footing in both the expression for the electronic Hessian and the property gradient (see eqs. 29 and 30 in ref. 41). However, since the intensities are quadratic in the property gradients we might even expect a slower convergence of intensities with respect to system size. In the case of electrostatic embedding or QM-only approaches, huge clusters would therefore be expected to be necessary in order to converge absorption spectra for complex systems. In light of this,

the PE-EEF model provides a very efficient way of obtaining fairly accurate intensities for very large systems even with rather small QM regions. Finally, we stress that the QM region and the embedding method are strongly coupled, as the adequate QM region size largely depends on the quality of the embedding potential used to account for the environment. The choices of QM region and embedding method must therefore be carefully made, and one must consider which types of interactions are important to include in order to answer the question at hand. As shown here, environment polarization is important for the case of electronic excitations.

Acknowledgement

James Snyder is thanked for providing the GFP starting structure and for valuable input through discussions. Computational resources were provided by the DeIC National HPC Center at the University of Southern Denmark. J. K. thanks the Danish Council for Independent Research (the Sapere Aude program) and the Villum Foundation. T. M. thanks the AMOS program within the Chemical Sciences, Geosciences and Biosciences Division of the Office of Basic Energy Sciences, Office of Science, U. S. Department of Energy.

Supporting Information Available

A detailed description of the MD simulations is provided, as well as the discussed EEF tensors and spectrum comparing intensities.

References

(1) Field, M. J.; Bash, P. A.; Karplus, M. A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations. *J. Comput. Chem.* **1990**, *11*, 700–733.

(2) Bakowies, D.; Thiel, W. Hybrid Models for Combined Quantum Mechanical and Molecular Mechanical Approaches. *J. Phys. Chem.* **1996**, *100*, 10580–10594.

(3) Monard, G.; Merz, K. M. Combined Quantum Mechanical/Molecular Mechanical Methodologies Applied to Biomolecular Systems. *Acc. Chem. Res.* **1999**, *32*, 904–911.

(4) Rosta, E.; Klähn, M.; Warshel, A. Towards Accurate Ab Initio QM/MM Calculations of Free-Energy Profiles of Enzymatic Reactions. *J. Phys. Chem. B* **2006**, *110*, 2934–2941.

(5) Lin, H.; Truhlar, D. G. QM/MM: what have we learned, where are we, and where do we go from here? *Theor. Chem. Acc.* **2006**, *117*, 185–199.

(6) Hu, L.; Soderhjelm, P.; Ryde, U. On the Convergence of QM/MM Energies. *J. Chem. Theory Comput.* **2011**, *7*, 761–777.

(7) Kulik, H. J.; Zhang, J.; Klinman, J. P.; Martínez, T. J. How Large Should the QM Region Be in QM/MM Calculations? The Case of Catechol O-Methyltransferase. *J. Phys. Chem. B* **2016**, *120*, 11381–11394.

(8) Karelina, M.; Kulik, H. J. Systematic Quantum Mechanical Region Determination in QM/MM Simulation. *J. Chem. Theory Comput.* **2017**, *13*, 563–576.

(9) Isborn, C. M.; Götz, A. W.; Clark, M. A.; Walker, R. C.; Martínez, T. J. Electronic Absorption Spectra from MM and ab Initio QM/MM Molecular Dynamics: Environmental Effects on the Absorption Spectrum of Photoactive Yellow Protein. *J. Chem. Theory Comput.* **2012**, *8*, 5092–5106.

(10) 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.

(11) Beerepoot, M. T.; 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.

(12) Provorse, M. R.; Peev, T.; Xiong, C.; Isborn, C. M. Convergence of Excitation Energies in Mixed Quantum and Classical Solvent: Comparison of Continuum and Point Charge Models. *J. Phys. Chem. B* **2016**, *120*, 12148–12159.

(13) Flaig, D.; Beer, M.; Ochsenfeld, C. Convergence of Electronic Structure with the Size of the QM Region: Example of QM/MM NMR Shieldings. *J. Chem. Theory Comput.* **2012**, *8*, 2260–2271.

(14) Liao, R. Z.; Thiel, W. Convergence in the QM-Only and QM/MM Modeling of Enzymatic Reactions: A Case Study for Acetylene Hydratase. *J. Comp. Chem.* **2013**, *34*, 2389–2397.

(15) Daday, C.; Curutchet, C.; Sinicropi, A.; Mennucci, B.; Filippi, C. Chromophore–Protein Coupling beyond Nonpolarizable Models: Understanding Absorption in Green Fluorescent Protein. *J. Chem. Theory Comput.* **2015**, *11*, 4825–4839.

(16) Shimomura, O.; Johnson, F.; Saiga, Y. Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, Aequorea. *J. Cell. Comp. Physiol.* **1962**, *59*, 223–239.

(17) Tsien, R. Y. The green fluorescent protein. *Annu. Rev. Biochem.* **1998**, *67*, 509–544.

(18) Stoner-Ma, D.; Jaye, A. A.; Matousek, P.; Towrie, M.; Meech, S. R.; Tonge, P. J. Observation of Excited-State Proton Transfer in Green Fluorescent Protein using Ultrafast Vibrational Spectroscopy. *J. Am. Chem. Soc* **2005**, *127*, 2864–2865.

(19) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation

Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.

(20) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.

(21) Wu, Y.; Tepper, H. L.; Voth, G. A. Flexible simple point-charge water model with improved liquid-state properties. *J. Chem. Phys.* **2006**, *124*, 024503.

(22) `http://www.petachem.com`, (accessed Jan 20, 2015).

(23) Case, D. A.; Darden, T. A.; Cheatham, III, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Swails, J.; Goetz, A. W.; Kolossvry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wolf, R.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M.-J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Salomon-Ferrer, R.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A. AMBER 12. 2012; University of California: San Francisco.

(24) 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.

(25) Hariharan, P. C.; Pople, J. A. The influence of polarization functions on molecular orbital hydrogenation energies. *Theor. Chim. Acta* **1973**, *28*, 213–222.

(26) Humphrey, W.; Dalke, A.; Schulten, K. VMD – Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14*, 33–38.

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

(28) Olsen, J. M. H.; Kongsted, J. Molecular Properties through Polarizable Embedding. *Adv. Quantum Chem.* **2011**, *61*, 107–143.

(29) List, N. H.; Olsen, J. M. H.; Kongsted, J. Excited states in large molecular systems through polarizable embedding. *Phys. Chem. Chem. Phys.* **2016**, *18*, 20234–20250.

(30) List, N. H.; Jensen, H. J. A.; Kongsted, J. Local electric fields and molecular properties in heterogeneous environments through polarizable embedding. *Phys. Chem. Chem. Phys.* **2016**, *18*, 10070–10080.

(31) Aidas, K.; Angeli, C.; Bak, K. L.; Bakken, V.; Bast, R.; Boman, L.; Christiansen, O.; Cimiraglia, R.; Coriani, S.; Dahle, P.; Dalskov, E. K.; Ekström, U.; Enevoldsen, T.; Eriksen, J. J.; Ettenhuber, P.; Fernández, B.; Ferrighi, L.; Fliegl, H.; Frediani, L.; Hald, K.; Halkier, A.; Hättig, C.; Heiberg, H.; Helgaker, T.; Hennum, A. C.; Hettema, H.; Hjertenæs, E.; Høst, S.; Høyvik, I.-M.; Iozzi, M. F.; Jansik, B.; Jensen, H. J. Aa.; Jonsson, D.; Jørgensen, P.; Kauczor, J.; Kirpekar, S.; Kjærgaard, T.; Klopper, W.; Knecht, S.; Kobayashi, R.; Koch, H.; Kongsted, J.; Krapp, A.; Kristensen, K.; Ligabue, A.; Lutnæs, O. B.; Melo, J. I.; Mikkelsen, K. V.; Myhre, R. H.; Neiss, C.; Nielsen, C. B.; Norman, P.; Olsen, J.; Olsen, J. M. H.; Osted, A.; Packer, M. J.; Pawlowski, F.; Pedersen, T. B.; Provasi, P. F.; Reine, S.; Rinkevicius, Z.; Ruden, T. A.; Ruud, K.; Rybkin, V.; Salek, P.; Samson, C. C. M.; de Merás, A. S.; Saue, T.; Sauer, S. P. A.; Schimmelpfennig, B.; Sneskov, K.; Steindal, A. H.; Sylvester-Hvid, K. O.; Taylor, P. R.; Teale, A. M.; Tellgren, E. I.; Tew, D. P.; Thorvaldsen, A. J.; Thøgersen, L.; Vahtras, O.; Watson, M. A.; Wilson, D. J. D.; Ziolkowski, M.; Ågren, H. The Dalton Quantum Chemistry Program System. *WIREs Comput. Mol. Sci.* **2014**, *4*, 269–284.

(32) The PE library developers, PElib: The Polarizable Embedding library (development version). 2016; http://gitlab.com/pe-software/pelib-public.

(33) Gao, B. Gen1Int Version 0.2.1. 2012; http://gitlab.com/bingao/gen1int.

(34) 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.

(35) Becke, A. D. Density-Functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98*, 5648–5652.

(36) Vosko, S. H.; Wilk, L.; Nusair, M. Accurate Spin-Dependent Electron Liquid Correlation Energies for Local Spin Density Calculations: a Critical Analysis. *Can. J. Phys.* **1980**, *58*, 1200–1211.

(37) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti Correlation-Energy Formula Into a Functional of the Electron Density. *Phys. Rev. B* **1988**, *37*, 785–789.

(38) Steinmann, C.; Ibsen, M. W.; Hansen, A. S.; Jensen, J. H. FragIt: A Tool to Prepare Input Files for Fragment Based Quantum Chemical Calculations. *PLoS ONE* **2012**, *7*, e44480.

(39) Gagliardi, L.; Lindh, R.; Karlström, G. Local Properties of Quantum Chemical Systems: The LoProp Approach. *J. Chem. Phys.* **2004**, *121*, 4494–4500.

(40) Nåbo, L. J.; List, N. H.; Steinmann, C.; Kongsted, J. Computational Approach to Evaluation of Optical Properties of Membrane Probes. *J. Chem. Theory Comput.* **2017**, *13*, 719–726.

(41) List, N. H.; Norman, P.; Kongsted, J.; Jensen, H. J. A. A quantum-mechanical perspective on linear response theory within polarizable embedding. *J. Chem. Phys.* **2017**, *146*, 234101.