Paper III

Convergence of environment polarization effects in multiscale modeling of excitation energies

M. T. P. Beerepoot, A. H. Steindal, K. Ruud,
J. M. H. Olsen and J. Kongsted

Convergence of environment polarization effects in multiscale modeling of excitation energies

CrossMark

Maarten T.P. Beerepoot [a,*], Arnfinn Hykkerud Steindal [a], Kenneth Ruud [a], Jógvan Magnus Haugaard Olsen [b], Jacob Kongsted [b]

[a] Centre for Theoretical and Computational Chemistry, Department of Chemistry, University of Tromsø, N-9037 Tromsø, Norway
[b] Department of Physics, Chemistry and Pharmacy, University of Southern Denmark, DK-5230 Odense M, Denmark

ARTICLE INFO

ABSTRACT

We present a systematic investigation of the influence of polarization effects from a surrounding medium on the excitation energies of a chromophore. We use a combined molecular dynamics and polarizable embedding time-dependent density functional theory (PE-TD-DFT) approach for chromophores in proteins and in homogeneous solvents. The mutual polarization between the chromophore and its surroundings is included in the PE-TD-DFT approach through the use of induced dipoles, placed on all atoms in the classical region, and self-consistent optimization of the quantum and classical polarizable regions. By varying the subset of sites in the environment for which atomic polarizabilities are included, we investigate to what distance from the quantum region explicit polarization effects need to be taken into account in order to provide converged excitation energies. Our study gives new insight into the range of polarization interactions for chromophores in different chemical environments. We find that the rate of convergence of excitation energies with respect to polarization cut-off is much slower for chromophores in an ordered environment such as a protein than for chromophores in a homogeneous medium such as a solvent. We show that this in part is related to the (partial) charges in the protein. Our results provide insight into how to define a representation of complex environments of different kinds in an accurate and affordable way.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

The importance of multiscale modeling is firmly established by the Nobel Prize in Chemistry to Karplus, Levitt and Warshel [1]. The basic assumption behind multiscale models is that a molecular system can be divided into smaller subsystems, each of which can be treated using different methods. The most prominent example of a multiscale model is the combined quantum mechanics and molecular mechanics (QM/MM) method [2–4]. Embedding models are an example of the class of focused multiscale models where we pay particular attention to a central part of the system. The simplest embedding models use an (infinite) continuum description of the environment and can be very efficient for solute–solvent systems [5]. However, the extension to more heterogeneous environments, e.g. proteins, is not well defined. Moreover, continuum models suffer from a poor description of more specific solute–solvent interactions such as hydrogen bonds. The QM/MM methods are a generally applicable alternative to the continuum models since the atomistic structure of the environment is retained.

One of the more important aspects of embedding models is the coupling between the central part and the environment. This coupling can be divided into three subclasses [4]: mechanical embedding (ME), electrostatic embedding (EE) and polarizable embedding (PE). The coupling in the ME scheme is performed on a purely classical level and it is therefore only suitable for ground-state energy calculations. The EE scheme, in contrast, includes one-electron operators in the electronic Hamiltonian that describe the interactions between the permanent charge distribution of the environment and the particles, i.e. electrons and nuclei, in the central subsystem. This will directly affect (polarize) the electron density of the central part and thus also the calculated molecular properties. The environment is normally represented by atomic partial charges or multipole moments. The PE scheme is currently the most advanced QM/MM type embedding scheme. Here, the polarization effects in the environment are also taken into account. It is worth noting that the first study by Warshel et al. also included polarization effects [2]. Despite this fact, PE schemes have not seen

* Corresponding author. Tel.: +47 77623103.
E-mail address: maarten.beerepoot@uit.no (M.T.P. Beerepoot).

widespread use, most likely because of the complexity of such implementations and also partly due to the added computational costs. Polarization is a many-body effect and thus requires a self-consistent solution. It is therefore necessary to update the electronic Hamiltonian according to the electron density of the central subsystem. The polarization effects, as modeled in QM/MM embedding models, are most frequently based on an induced dipoles model [2,6–15]. There are, however, also other ways to include classical polarization in embedding models. Some recent examples are the fluctuating charges model [16,17] and the classical Drude oscillator model [18].

Most QM/MM implementations use standard MM force fields where the electrostatic components have been parametrized to implicitly include polarization effects. The EE schemes can therefore, in principle, also describe the same averaged effects as an explicitly polarizable model on the electronic ground state. The EE schemes cannot, however, describe differential polarization effects between ground and excited states. This effect becomes important when there is a significant rearrangement of the electronic structure upon electronic excitation [13,19].

In this study we use the PE model by Olsen et al. [13,20] as implemented in the PE library [21] which has been interfaced with the Gen1Int integral library [22] and the Dalton program [23,24]. This is a QM/MM-type embedding model that focuses on the calculation of molecular properties, including the use of accurate embedding potentials derived from *ab initio* calculations for each structure explicitly. Currently, the polarization part of this model includes anisotropic polarizabilities that lead to induced dipole moments that can be determined self-consistently with respect to either the ground or the excited states of the central subsystem. The embedding potentials for the proteins are generated using the molecular fractionation with conjugate caps (MFCC) method by Zhang and Zhang [25] as applied to localized properties by Söderhjelm and Ryde [26]. All embedding potential parameters used in this study are derived using DFT for each structure explicitly without using averaged (force field) parameters.

Polarization is generally considered to be a short-range effect, in particular compared to electrostatic interactions. The idea of including polarization only for a subset of atoms in the classical region (those that are closest to the quantum region) has been proposed on several occasions [27–31]. Indeed, Osted et al. argue that this can reduce the computational cost while retaining the quality of the results compared to including polarization for the full classical region [27,28]. Osted et al. calculated several properties of liquid water using a combined coupled-cluster/molecular mechanics (CC/MM) approach on a large number of molecular dynamics snapshots [28]. They found that the optimum balance of quality versus computational cost was to include water molecules within a threshold of 10 Å of the quantum region and polarization of the water molecules within a radius of 7 Å from the quantum region.

Söderhjelm et al. investigated the distance dependence of several embedding potential parameters on the lowest excitation energy of rhodopsin using CASPT2/CASSCF for the QM region. By removing the polarizabilities or their anisotropy outside a certain threshold, their interaction range was analyzed. The effect of including both the polarizabilities in the first place but in particular also their anisotropy was found to be strong within the first 10 Å from the QM region and levelling out at longer distances [29]. In another study, Söderhjelm et al. found that removal of polarizabilities or their anisotropy does not lead to converged protein–ligand interaction energies below a threshold of 20 Å with convergence defined as an absolute error in the interaction energy below 4 kJ/mol [30].

Curutchet et al. studied the electronic properties of a light-harvesting protein using the QM/MMpol method. They found the

errors to be small with a polarization cut-off of 18 Å compared to using a cut-off of 30 Å. However, this was found to lead to a four-fold decrease in the computational time [31].

These studies [27–31] suggest that the influence of polarization extends over a much longer range in proteins than in simple homogeneous media. The aim of this study is to investigate this in a systematic manner by exploring to what distance from the quantum region the influence of polarization needs to be taken into account and how far purely electrostatic effects need to be included. We will in particular focus on how this threshold depends on the type of environment by considering both chromophores in proteins and in solution. We will study in more detail the solvent-induced shift of the $n \rightarrow \pi^*$ transition of acetone in different solvents. The goal is to give clear guidelines for how PE calculations should be designed to ensure that the calculations are both accurate and computationally efficient.

In Section 2 we give the computational details for the results presented in Section 3. In Section 4 we discuss the consequences of our computational results on how to design efficient and accurate polarizable embedding calculations for different environments, before we give some concluding remarks in Section 5.

2. Computational details

2.1. Preparation of the structures

We will study two different classes of systems: proteins as models of structured systems with charged side groups, and solute–solvent systems as models for chromophores in a homogeneous environment. The molecular structures for the green fluorescent protein (GFP) and rhodopsin were prepared from the crystal structures 1GFL [32] and 1U19 [33], respectively. 50 snapshots for GFP solvated in water were taken from a 15 ns molecular dynamics (MD) simulation using the CHARMM27 force field [34]. Details of this MD simulation are provided in Ref. [35]. For the PE-TD-DFT calculations, the protein was extracted together with a solvation shell consisting of all water molecules with one of the atoms within 8 Å from one of the protein atoms. Starting structures for the solute–solvent systems (uracil and the GFP chromophore in water, acetone in various solvents) were carefully minimized and equilibrated. Structures of the solute in a solvent sphere with radius 20 Å (unless otherwise specified) were subsequently obtained from the MD simulations at intervals of 200 ps. In the solute–solvent calculations, solvent molecules were included in the classical region if one or more of its atoms were within the threshold. For the proteins, only atoms within the threshold were included. In all cases, the chromophore was geometry optimized with QM/MM in the frozen environment of the protein or solvent. A detailed description of the structure preparation is given in Section 1 of the Supporting Information.

2.2. Generation of the embedding potential

The embedding potentials for the polarizable embedding (PE) calculations include coordinates, QM-derived multipole moments up to quadrupoles and QM-derived anisotropic polarizabilities for all atoms outside the QM region. Multipole moments and polarizabilities were calculated for each snapshot separately with DFT using the LoProp [36] approach implemented in Molcas [37,38]. The generation of the potential was facilitated by the Polarizable Embedding Assistant Script (PEAS), a python script developed by one of the authors [21]. The details vary slightly between the proteins (GFP and rhodopsin) and solute–solvent structures (uracil and GFP chromophore in water and acetone in different solvents). For the proteins, multipole moments and polarizabilities were

obtained from QM calculations on protein fragments from the molecular fragmentation with the conjugated-caps (MFCC) procedure [25] as applied to localized properties in Ref. [26] and described in more detail in Ref. [39]. The parameters were calculated with the B3LYP functional [40–43] and the 6-31+G* basis set [44–46]. For the solute–solvent structures, potential parameters were calculated from DFT calculations on each solvent molecule separately using the B3LYP functional [40–43] with the aug-cc-pVDZ basis set [47]. In all cases, the basis set was recontracted to an atomic natural orbital type basis as required for the LoProp approach [36].

2.3. Excitation energy calculations

Vertical excitation energies were calculated with polarizable embedding time-dependent density functional theory (PE-TD-DFT) [13], which has recently become available in the 2013 release of the Dalton program system [23,24]. The PE-TD-DFT calculations were performed with the CAM-B3LYP exchange–correlation functional [48] in combination with the 6-31+G* basis set [44–46] for the chromophores in the proteins and the aug-cc-pVDZ basis set [47] for the chromophores in the solvents. These basis sets were chosen to be consistent with the basis sets that were used to derive the embedding potentials. Double-zeta basis sets have been shown to perform well for excitation energies of biological chromophores as long as diffuse functions are included in the basis set [49].

In order to investigate the range of polarization interactions, separate PE-TD-DFT calculations were run with different thresholds $(1, 2, \ldots \text{Å})$ for the inclusion of polarization effects. Polarizabilities (and induced dipoles) were only included for atoms within a certain distance between an atom in the environment and any atom in the chromophore, determined by the polarization threshold R_{pol}. Electrostatic interactions between the environment and the charge density of the quantum region were calculated for all permanent multipoles, i.e. no electrostatic cut-off threshold (R_{el}) was applied, unless otherwise specified.

3. Results

The convergence of the lowest excitation energy with different polarization thresholds is shown in Fig. 1 for a solute–solvent system (the $\pi \rightarrow \pi^*$ transition of uracil solvated in water) and for two proteins (GFP and rhodopsin). It is clear from Fig. 1 that the influence of polarization in the classical region extends over a longer range in the proteins than in water only. Indeed, the excitation energy is stable above a polarization cut-off of 20 Å for both GFP and rhodopsin, but already above 10 Å for uracil in water. This is clearly demonstrated by the results in Fig. 2, which shows the faster convergence for the GFP chromophore in water relative to that of the same chromophore in the native protein environment. For GFP, the same rate of convergence is observed for calculations on the crystal structure and for averages over MD snapshots (Fig. 3). Moreover, the type of embedding potential (order of the multipole moment expansion and isotropic vs. anisotropic polarizabilities) does not affect the convergence either (see Figs. SI-2 and SI-3).

Charged residues are a possible reason for the longer range of polarization interactions in proteins relative to solute–solvent systems [50]. Although the direct effect of polarization is rather short-range, a charged residue at a distance of 20 Å might affect the quantum region via other induced dipoles, since these are determined self-consistently and thus include many-body effects. In order to verify this, we made an artificial system where all charged amino acid residues in the classical region of GFP were changed into neutral residues by adding or removing a proton. We then made a new embedding potential for this 'neutralized GFP' and
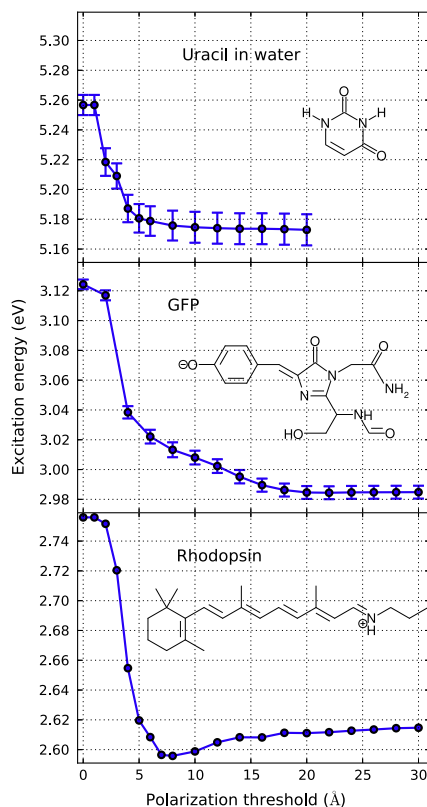

Fig. 1. Lowest excitation energies from PE-TD-DFT calculations on different systems using different polarization cut-off thresholds (in Å). Electrostatic interactions were fully included for all atoms in the systems. The results for anionic GFP and uracil in water are averages over 50 snapshots extracted from an MD simulation with standard errors shown as error bars. The results for rhodopsin are based on the crystal structure. Chromophore structures of the quantum region of the PE-TD-DFT calculations are shown in the insets.

repeated the excitation energy calculations with different polarization cut-off thresholds. The result is a less pronounced effect from the polarization in the classical region (Fig. 4). In particular, the effect of adding induced dipoles of the classical sites between 10 and 20 Å is more than twice as large when the charged residues are kept (0.023 vs. 0.009 eV). The relatively high number of charged side groups in GFP (59 out of 230 residues: 26%) compared to rhodopsin (38 out of 238: 11%)—a membrane protein with therefore many uncharged hydrophobic residues—could explain why the effect of the polarizabilities in the 10 to 20 Å region is almost twice as large for GFP (0.023 eV) as for rhodopsin (0.012 eV). The results for rhodopsin (Fig. 1) agree well with a similar study by Söderhjelm et al. that showed that the lowest excitation energy is mostly affected by the polarizabilities within a threshold of 10 Å [29].

The solvent shift of the $\pi \rightarrow \pi^*$ transition in uracil solvated in water (-0.21 eV using $\Delta E_{vac}^{\pi \rightarrow \pi^*} = 5.384$ eV from Ref. [51]) is in much better agreement with experiment (-0.31 eV, see Ref. [13]) than the results reported in Ref. [13] (-0.12 eV). Since the method (CAM-B3LYP), basis set (aug-cc-pVDZ) and embedding potential (QM-derived multipole moments up to quadrupoles and dipole–dipole polarizabilities) are the same, it is possible that the improvement comes from a better description of the indirect
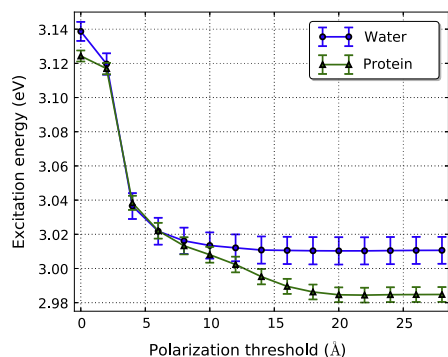
Fig. 2. Excitation energies from PE-TD-DFT calculations on the chromophore of GFP in both the native protein environment and solvated in water using different polarization cut-off thresholds (in Å). The results are averages over 50 snapshots extracted from an MD simulation with standard errors shown as error bars. The GFP protein was explicitly solvated in water in both the MD simulation and the PE-TD-DFT calculations. Electrostatic interactions were included for all atoms in the protein and for all water molecules within a sphere with radius 30 Å around the chromophore. Note that the side chains of the chromophore are different in the two models (see Fig. SI-1).
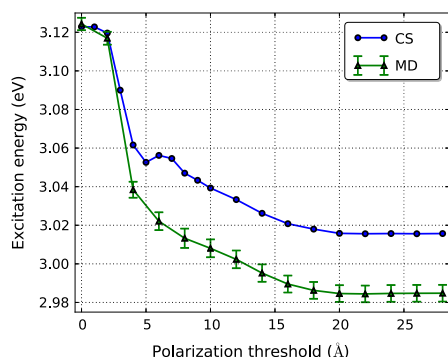

Fig. 3. Excitation energies from PE-TD-DFT calculations on anionic GFP using different polarization cut-off thresholds (in Å). The convergence of the excitation energy is compared for calculations based on the crystal structure (CS) and as an average over 50 snapshots extracted from an MD simulation (MD) with standard errors shown as error bars. Electrostatic interactions were fully included for all atoms in the protein.

solvent effects, i.e. the effect of the solvent on the geometry of the solute. Indeed, the structures in Ref. [13] were obtained from an MD simulation with a frozen solute whereas the solute was allowed to relax in this work, followed by a QM/MM geometry optimization *in the solvent.*

We also consider the convergence of the lowest excitation energy in uracil solvated in water ($\pi \rightarrow \pi^*$) with the threshold for which electrostatic interactions are calculated (i.e. the size of the system since $R_{pol} \leqslant R_{el}$). Electrostatic interactions are less computationally demanding to include in the PE-TD-DFT calculations compared to polarization effects, but an increased size of the system also leads to increased computational costs through the explicit QM calculation of the embedding potential parameters for each solvent molecule. Errors in the truncation of electrostatic and polarization interactions can cancel each other, or other errors, to *accidentally* give the right answer, thus a systematic approach is needed. The excitation energies and relative errors (compared to
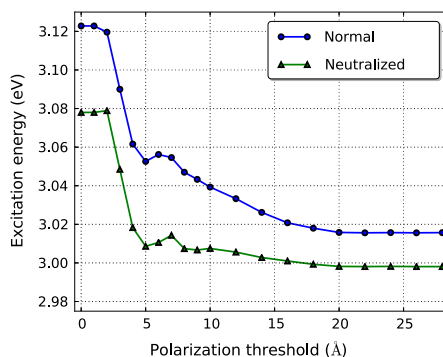

Fig. 4. Excitation energies from PE-TD-DFT calculations on the anionic green fluorescent protein (GFP) using different polarization cut-off thresholds (in Å). Calculations were performed both on the crystal structure and on a modified structure in which all the charged amino acid residues were neutralized. Electrostatic interactions were included for all atoms in the protein.

including electrostatic and polarization interactions up to 20 Å) are shown in Table 1. We observe that the errors beyond $R_{el} = 5$ Å and $R_{pol} = 5$ Å are below 0.01 eV. Moreover, an increase of the thresholds beyond $R_{el} = 10$ Å and $R_{pol} = 10$ Å does not lead to significant changes. The differences (on the order of 0.01 eV) are smaller than other sources of error in the calculations. Nevertheless, these small differences are often of interest when considering for instance the effect of different solvents on molecular properties (see below). In these cases, errors in the method (e.g. overestimation of the excitation energy by the CAM-B3LYP exchange–correlation functional) partly cancel and reliable relative solvent shifts can be obtained.

The accuracy of the method should balance the added computational costs, and we therefore also consider the computational times needed for calculations with different polarization cut-off thresholds (Fig. 5). The main factor determining the computational time of a QM/MM calculation is of course the size of the quantum region. In fact, without polarization effects (electrostatic embedding) the calculations on the GFP chromophore (39 atoms) in the protein used approximately seven times as much time as the calculations on uracil (10 atoms) in water. Including polarization in the classical region does not lead to an extraordinary increase in the computational time. On the contrary, the calculations take only

Table 1
The lowest excitation energy (ΔE in eV) in uracil solvated in water calculated with different electrostatic (R_{el} in Å) and polarization (R_{pol} in Å) cut-off thresholds. The difference ($\Delta\Delta E$ in eV) to the most accurate calculation ($R_{el} = 20$ Å, $R_{pol} = 20$ Å) is shown as a measure of the convergence. The results are averages over 50 snapshots extracted from an MD simulation.

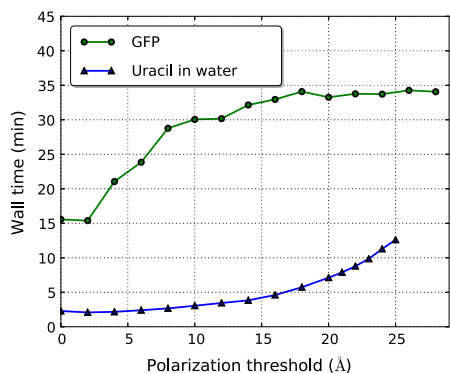| R_{el}	R_{pol}	ΔE	$\Delta\Delta E$
5	0	5.257	+0.084
5	5	5.177	+0.004
10	0	5.258	+0.085
10	5	5.178	+0.005
10	10	5.171	−0.002
15	0	5.258	+0.085
15	5	5.179	+0.006
15	10	5.173	+0.000
15	15	5.174	+0.001
20	0	5.257	+0.084
20	5	5.181	+0.008
20	10	5.175	+0.002
20	15	5.174	+0.001
20	20	5.173	-

Fig. 5. Computational time needed to calculate the lowest four excitation energies in a snapshot of uracil solvated in water and in the anionic GFP chromophore embedded in the protein (crystal structure) using different polarization cut-off thresholds (in Å). All calculations were performed on 64 cores (4 nodes) on the same machine. Electrostatic interactions were fully included for water molecules up to a distance of 25 Å from uracil and for all atoms in the GFP. The computational time needed to generate the potential parameters is not included here.

Fig. 6. Excitation energies from PE-TD-DFT calculations on acetone ($n \rightarrow \pi^*$) in different solvents using different polarization cut-off thresholds (in Å). The results are averages over 50 snapshots extracted from MD simulations with standard errors shown as error bars. Electrostatic interactions were fully included for solvent atoms up to a distance of 20 Å from acetone in all calculations. The differences between the different solvents in the excitation energies at $R_{pol} = 0$ Å are thus due to electrostatic effects and indirect solvent effects on the geometry of acetone.

twice as long when the polarization of water is included up to 15 Å from uracil and only slightly more when polarization is included for the whole GFP (compared to electrostatic embedding only). The computational time increases more rapidly with increasing polarization threshold for uracil in water than for GFP. Indeed, the number of atoms in the classical region increases sharply for uracil in water for thresholds above 15 Å, whereas it levels out for GFP (see Fig. SI-4). This is related to the shape of the system: uracil has a *sphere* of water around whereas solvated GFP is shaped like a *barrel*. Thus, the computational time that can be gained by choosing the right polarization threshold depends on the shape of the systems, the largest gain being for spherical systems as they are truly three-dimensional and thus have the largest volume increase with increasing cut-off thresholds. This also includes proteins embedded in a sphere of water molecules. We note that a smaller region including explicit polarization also reduces the computational time needed for explicit calculation of the potential parameters, which is however not the focus in this study.

A remaining question is how the convergence of an excitation energy with varying polarization cut-off threshold depends on the type of solvent. To investigate this, the $n \rightarrow \pi^*$ transition of acetone in five different solvents was calculated as an average over 50 snapshots extracted from MD trajectories. Explicit inclusion of polarization and statistical averaging have been shown to be mandatory to obtain reliable solvent shifts for this system [52]. Our results are shown in Fig. 6.

The excitation energy converges slowest for acetone in water. In fact, a polarization threshold of 5 Å is enough for dimethyl sulfoxide (DMSO), diethyl ether (DEE) and hexane, but not for water. We note that comparison with bulky solvents, such as diethyl ether and hexane, should be done with caution. If one of the solvent atoms is within the threshold then the whole molecule is included, extending much farther into the classical region than the threshold. Indeed, only a few extra hexane molecules are included when the threshold is enlarged from 3 to 5 Å. Still, including polarization using $R_{pol} = 5$ captures almost all polarization effects (0.0221 eV/178 cm^{-1} of the total polarization solvent shift of 0.0237 eV/191 cm^{-1}) of acetone in a hexane solvent.

The standard errors (and the standard deviations) are largest for the hydrogen-bonding solvents, i.e. water and methanol. In these cases, the number of hydrogen bonds to acetone can change
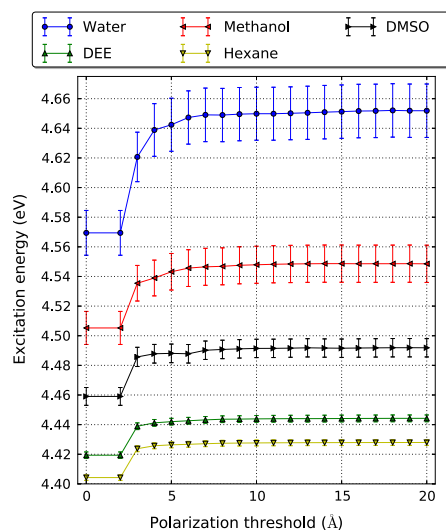
between snapshots, leading to a significant difference in electron density and hence excitation energy. The size of the standard errors (standard deviation of the excitation energy divided by the square root of the number of snapshots) is below the differences between the solvents, indicating that reliable relative solvent shifts can be obtained based on 50 snapshots.

The solvent shifts are separated into indirect solvent effects (comparing a gas-phase calculation with QM calculations on acetone at the 50 solvent geometries), direct electrostatic effects (comparing the QM calculations on the solvent geometries to PE-TD-DFT calculations with $R_{pol} = 0$) and direct polarization effects (comparing PE-TD-DFT calculations with $R_{pol} = 0$ to $R_{pol} = 20$) and compared to experimental solvent shifts [53] in Table 2. The calculated excitation energies (in eV) are tabulated in Section 3 of the Supporting Information.

The order of the solvent shifts is in agreement with experiment. In fact, the calculated solvent shifts are systematically overestimated by values ranging from 91 (diethyl ether) to 280 (DMSO) cm^{-1}. This can most likely be improved upon by including non-classical interactions, such as dispersion interactions and exchange repulsion, which are especially important for non-polar solvents

Table 2

Calculated and experimental solvents shifts (all in cm^{-1}) for the $n \rightarrow \pi^*$ excitation in acetone in five different solvents. The calculated solvent shifts ($\Delta\omega_{tot}$) are separated into indirect ($\Delta\omega_{ind}$), direct electrostatic ($\Delta\omega_{elec}$) and direct polarization ($\Delta\omega_{pol}$) contributions. Electrostatic and polarization interactions were included for solvent atoms up to a distance of 20 Å from acetone. The experimental values ($\Delta\omega_{exp}$) are from Ref. [53].

Solvent	$\Delta\omega_{ind}$	$\Delta\omega_{elec}$	$\Delta\omega_{pol}$	$\Delta\omega_{tot}$	$\Delta\omega_{exp}$ [53]
Hexane	−14	−38	191	139	−35 ± 5
Diethyl ether (DEE)	−119	188	201	271	180 ± 25
Dimethyl sulfoxide (DMSO)	−344	734	265	655	375 ± 10
Methanol	−304	1067	349	1111	941 ± 10
Water	−393	1673	665	1945	1785 ± 7

such as diethyl ether and hexane. This can be done straightforwardly by including the closest solvent molecules in the quantum region. This is, however, computationally expensive. Furthermore, it has been shown that a correct geometry (in particular the C=O bond length) is crucial in order to get an accurate solvent shift for the excitation energy of acetone in water [52]. Improvement of the geometries by more accurate QM/MM geometry optimization is possible in several ways: using a more accurate method (here: B3LYP), using a more flexible basis set (here: 6-31+G∗) and including the environment in a more accurate way (here: electrostatic embedding with standard force field charges). Thus, a more rigorous treatment of solvent effects on the $n \rightarrow \pi^*$ excitation in acetone will include at least a treatment of non-classical interactions and use more accurate geometries.

The $n \rightarrow \pi^*$ excitation energy of acetone in different solvents has also been calculated using the same cut-off for electrostatic and polarization interactions (Table SI-II). It is clear that a threshold of 10 Å is enough, not only for the polarization cut-off (as shown in Fig. 6), but also for the electrostatic cut-off. Indeed, errors in the calculated solvent-induced shifts are 1–2% (⩽0.005 eV) compared to including both types of interactions up to 20 Å. Errors of calculated solvent-induced shifts with both cut-offs at 8 Å range from 3% to 5% (⩽0.007 eV), still low enough for many purposes.

4. Discussion

A point that emerges from the results presented in this work is that the convergence of excitation energies with polarization cutoff is much slower for chromophores in an ordered medium, such as a protein, than for chromophores in a homogeneous medium, such as a solvent. This is clearly demonstrated in Fig. 2 by the faster convergence of the GFP chromophore in water (converged around 10 Å) compared to that of the same chromophore in the protein (converged around 20 Å). We have also shown that this in part is related to the charges in the protein (see Fig. 4) and therefore depends on the type of protein, i.e. the abundance of (partial) charges. Indeed, convergence of excitation energies is slower for solvated GFP than for rhodopsin, which is a membrane protein with therefore relatively few (partial) charges. These findings agree well with the work of Söderhjelm et al. on rhodopsin, in which explicit polarization in the embedding region was found to be especially relevant within a distance of 10 Å from the chromophore [29]. The results indicate that truncation of an atomistic model of a protein without losing accuracy is advisable only for relatively large proteins with parts that extend further than about 20 Å from the chromophore. Thus, for big proteins—especially with a large solvation shell or embedded in a membrane—truncation can lead to a significant speed-up without noteworthy loss of accuracy. For solute–solvent systems, including solvent molecules up to 10 Å from the chromophore is enough for accurate embedding calculations.

A smaller cut-off radius for the explicit polarization than for the electrostatics is worthwhile to consider, since the polarization interactions are computationally more demanding than the electrostatic interactions and since the polarization interactions decay faster. In order for such a three-layer ($R_{el} \neq R_{pol}$) approach to be worthwhile, it needs to satisfy two criteria: (1) faster convergence of the molecular property with respect to R_{pol} than with respect to R_{el} and (2) significant savings in computational time. As far as the first point is concerned, we have shown for uracil in water (Table 1) and for acetone in various solvents that choosing $R_{el} > R_{pol}$ does not lead to an improved convergence of the excitation energies. In fact, induced dipoles relatively far from the chromophore cannot influence the chromophore *directly*, but can do so *via other induced dipoles*. This holds even more for proteins, where the influence of

polarization extends to longer distances from the chromophore as shown in this work. Convergence of excitation energies with R_{pol} is thus not faster compared to R_{el}. As far as the second criterion is concerned, for uracil in water there is a possible gain in computational time only for thresholds beyond those that are needed to include the environment in an accurate way. Indeed, including interactions up to 10 Å (which is enough for an accurate description of the environment) leads to a speed-up of a factor of two compared to a 20 Å cut-off. For GFP and rhodopsin, however, including polarization interactions up to 20 Å does not lead to a speed-up of more than 10% compared to inclusion of the whole protein. This is related to the efficient implementation of the PE method that is being used [13], which is different from earlier implementations in which a cut-off of polarization interactions below 10 Å could lead to an increased efficiency [28]. Thus, neither of the two criteria is satisfied and we find that the use of different thresholds for the inclusion of electrostatic and polarization interactions is not necessary for an optimal balance of accuracy and computational efficiency.

We note that the results for excitation energies discussed in this work are not necessarily transferable to other molecular properties. As an example, we computed the one-photon absorption (OPA) oscillator strength and two-photon absorption (TPA) cross section associated with the $\pi \rightarrow \pi^*$ transition in the GFP chromophore (Figs. SI-5 and SI-6). The convergence with increasing polarization threshold is faster for the TPA cross-section than for the OPA oscillator strength. Similarly, the convergence of the excitation energy—effectively an energy *difference*—may be faster than the convergence of an *absolute* energy. Indeed, Söderhjelm et al. found no clear convergence below 20 Å for the protein–ligand interaction energy when truncating the polarizability in a similar way [30]. It is therefore advisable to do a similar test to decide the size of the classical region needed when another property is being investigated.

We have shown that QM/MM calculations with inclusion of explicit polarization in the classical region are not necessarily very time consuming. A considerable amount of computational time can be won by carefully examining to what extent inclusion of the interactions is needed. Accurate inclusion of all interactions with the surroundings in a computationally efficient way can be done in other ways as well. First, a three-layer QM/MM/PCM model has been developed in our groups [54]. This combines explicit inclusion of the closest molecules in the environment with the advantages of continuum models and reduces the cost of generating embedding potential parameters and explicit QM–MM interactions in the PE-TD-DFT calculations. Second, explicit calculation of the embedding potential parameters for each unique structure can be avoided by using average parameters for all or for a subset of the classical sites, as previously proposed by Söderhjelm et al. [29]. This is relatively straightforward for solute–solvent systems but more involved for e.g. proteins since it requires a general force field tailored towards the calculation of molecular properties. Third, the induced dipoles of the outer region can be frozen at their ground-state values during the response calculation, resulting in a significant speed-up of the PE-TD-DFT calculation. This constitutes an improvement over the approach discussed in this work since the polarization of the outer region is included instead of being truncated. However, the accuracy may be affected if there is a significant difference between the electron densities in the ground and excited states. Fourth, one could define a three-layer approach with inclusion of polarization through induced dipoles up to a certain threshold and *implicit* polarization in the outer region. This would be similar to the approach taken in this work, but with a better description of the polarization of the outer region, with the possible gain of being able to use a smaller threshold. Implicit inclusion of polarization can be done by electrostatic embedding

with re-parametrized electric multipoles that capture a part of the polarization effects. All of these approaches rely on the fact that a relatively small region around the quantum region is responsible for most of the perturbation by the environment, which is clearly shown by the results in this work. This observation allows for an efficient scheme for including the rest of the classical region without losing much accuracy.

5. Conclusion

We have systematically investigated to what extent the polarization of the environment has an influence on the excitation energies of a chromophore both in highly ordered and in homogeneous environments. We have found that convergence of excitation energies with polarization cut-off is much slower for an ordered environment, such as a protein, than for chromophores in a homogeneous medium, such as a solvent. Especially for proteins containing many (partial) charges, atoms in the classical region can contribute to the properties of the quantum region through polarization interactions up to a distance of 20 Å. For chromophores in homogeneous solvents, truncation of the environment polarization can be done even below 10 Å as demonstrated by the solvent-induced shift of the $n \rightarrow \pi^*$ transition in acetone in different solvents. Moreover, we have demonstrated that truncating polarization and electrostatic interactions with different cut-offs is not necessary for an optimal balance of accuracy and computational efficiency. Finally, we have discussed other approaches to increase the computational efficiency of polarizable embedding calculations without compromising the accuracy. This work is therefore valuable for the design of computational approaches to include a complex environment in an accurate and affordable way.

Acknowledgements

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.comptc.2014.03.022.

References

[1] The Nobel Prize in Chemistry 2013 – Press Release, Nobelprize.org. Nobel Media AB 2013, Web, 5 Jan 2014.
[2] A. Warshel, M. Levitt, Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme, J. Mol. Biol. 103 (1976) 227–249.
[3] H. Lin, D.G. Truhlar, QM/MM: what have we learned, where are we, and where do we go from here?, Theor Chem. Acc. 117 (2007) 185–199.
[4] H.M. Senn, W. Thiel, QM/MM methods for biomolecular systems, Angew. Chem. Int. Ed. 48 (2009) 1198–1229.
[5] J. Tomasi, B. Mennucci, R. Cammi, Quantum mechanical continuum solvation models, Chem. Rev. 105 (2005) 2999–3093.
[6] M.A. Thompson, QM/MMpol: a consistent model for solute/solvent polarization. Application to the aqueous solvation and spectroscopy of formaldehyde, acetaldehyde, and acetone, J. Phys. Chem. 100 (1996) 14492–14507.

[7] P.N. Day, J.H. Jensen, M.S. Gordon, S.P. Webb, W.J. Stevens, M. Krauss, D. Garmer, H. Basch, D. Cohen, An effective fragment method for modeling solvent effects in quantum mechanical calculations, J. Chem. Phys. 105 (1996) 1968–1986.
[8] M.S. Gordon, M.A. Freitag, P. Bandyopadhyay, J.H. Jensen, V. Kairys, W.J. Stevens, The effective fragment potential method: a QM-based MM approach to modeling environmental effects in chemistry, J. Phys. Chem. A 105 (2001) 293–307.
[9] L. Jensen, P. Th. van Duijnen, J.G. Snijders, A discrete solvent reaction field model within density functional theory, J. Chem. Phys. 118 (2003) 514–521.
[10] J. Kongsted, A. Osted, K.V. Mikkelsen, O. Christiansen, Coupled cluster/molecular mechanics method: implementation and application to liquid water, J. Phys. Chem. A 107 (2003) 2578–2588.
[11] C.B. Nielsen, O. Christiansen, K.V. Mikkelsen, J. Kongsted, Density functional self-consistent quantum mechanics/molecular mechanics theory for linear and nonlinear molecular properties: applications to solvated water and formaldehyde, J. Chem. Phys. 126 (2007) 154112.
[12] C. Curutchet, A. Muñoz-Losa, S. Monti, J. Kongsted, G.D. Scholes, B. Mennucci, Electronic energy transfer in condensed phase studied by a polarizable QM/MM model, J. Chem. Theory Comput. 5 (2009) 1838–1848.
[13] J.M. Olsen, K. Aidas, J. Kongsted, Excited states in solution through polarizable embedding, J. Chem. Theory Comput. 6 (2010) 3721–3734.
[14] K. Sneskov, T. Schwabe, J. Kongsted, O. Christiansen, The polarizable embedding coupled cluster method, J. Chem. Phys. 134 (2011) 104108.
[15] E.D. Hedegård, N.H. List, H.J. Aa. Jensen, J. Kongsted, The multi-configuration self-consistent field method within a polarizable embedded framework, J. Chem. Phys. 139 (2013) 044101.
[16] F. Lipparini, V. Barone, Polarizable force fields and polarizable continuum model: a fluctuating charges/PCM approach. 1. Theory and implementation, J. Chem. Theory Comput. 7 (2011) 3711–3724.
[17] F. Lipparini, C. Cappelli, V. Barone, Linear response theory and electronic transition energies for a fully polarizable QM/classical Hamiltonian, J. Chem. Theory Comput. 8 (2012) 4153–4165.
[18] E. Boulanger, W. Thiel, Solvent boundary potentials for hybrid QM/MM computations using classical drude oscillators: a fully polarizable model, J. Chem. Theory Comput. 8 (2012) 4527–4538.
[19] N.H. List, J.M.H. Olsen, H.J. Aa. Jensen, A.H. Steindal, J. Kongsted, Molecular-level insight into the spectral tuning mechanism of the DsRed chromophore, J. Phys. Chem. Lett. 3 (2012) 3513–3521.
[20] J.M.H. Olsen, J. Kongsted, Molecular properties through polarizable embedding, Adv. Quantum Chem. 61 (2011) 107–143.
[21] J.M.H. Olsen, 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.
[22] B. Gao, A.J. Thorvaldsen, K. Ruud, GEN1INT: a unified procedure for the evaluation of one-electron integrals over Gaussian basis functions and their geometric derivatives, Int. J. Quantum Chem. 111 (2011) 858–872.
[23] K. Aidas, C. Angeli, K.L. Bak, V. Bakken, R. Bast, L. Boman, O. Christiansen, R. Cimiraglia, S. Coriani, P. Dahle, E.K. Dalskov, U. Ekström, T. Enevoldsen, J.J. Eriksen, P. Ettenhuber, B. Fernández, L. Ferrighi, H. Fliegl, K. Frediani, K. Hald, A. Halkier, C. Hättig, H. Heiberg, T. Helgaker, A.C. Hennum, H. Hettema, E. Hjertenæs, S. Høst, I.-M. Høyvik, M.F. Iozzi, B. Jansík, H.J.Aa. Jensen, D. Jonsson, P. Jørgensen, J. Kauczor, S. Kirpekar, T. Kjærgaard, W. Klopper, S. Knecht, R. Kobayashi, H. Koch, J. Kongsted, A. Krapp, K. Kristensen, A. Ligabue, O.B. Lutnæs, J.I. Melo, K.V. Mikkelsen, R.H. Myhre, C. Neiss, C.B. Nielsen, P. Norman, J. Olsen, J.M.H. Olsen, A. Osted, M.J. Packer, F. Pawlowski, T.B. Pedersen, P.F. Provasi, S. Reine, Z. Rinkevicius, T.A. Ruden, K. Ruud, V.V. Rybkin, P. Sałek, C.C.M. Samson, A.S. de Merás, T. Saue, S.P.A. Sauer, B. Schimmelpfennig, K. Sneskov, A.H. Steindal, K.O. Sylvester-Hvid, P.R. Taylor, A.M. Teale, E.I. Tellgren, D.P. Tew, A.J. Thorvaldsen, L. Thøgersen, O. Vahtras, M.A. Watson, D.J.D. Wilson, M. Ziolkowski, H. Ågren, The Dalton quantum chemistry program system, WIREs Comput. Mol. Sci. (2013). Doi: 10.1002/wcms.1172.
[24] Dalton, A Molecular Electronic Structure Program, Release DALTON2013.0, 2013, See <http://daltonprogram.org/>.
[25] D.W. Zhang, J.Z.H. Zhang, Molecular fractionation with conjugate caps for full quantum mechanical calculation of protein–molecule interaction energy, J. Chem. Phys. 119 (2003) 3599–3605.
[26] P. Söderhjelm, U. Ryde, How accurate can a force field become? A polarizable multipole model combined with fragment-wise quantum-mechanical calculations, J. Phys. Chem. A 113 (2009) 617–627.
[27] A. Osted, J. Kongsted, K.V. Mikkelsen, O. Christiansen, Linear response properties of liquid water calculated using CC2 and CCSD within different molecular mechanics methods, J. Phys. Chem. A 108 (2004) 8646–8658.
[28] A. Osted, J. Kongsted, K.V. Mikkelsen, P.-O. Åstrand, O. Christiansen, Statistical mechanically averaged molecular properties of liquid water calculated using the combined coupled cluster/molecular dynamics method, J. Chem. Phys. 124 (2006) 124503.
[29] P. Söderhjelm, C. Husberg, A. Strambi, M. Olivucci, U. Ryde, Protein influence on electronic spectra modeled by multipoles and polarizabilities, J. Chem. Theory Comput. 5 (2009) 649–658.
[30] P. Söderhjelm, F. Aquilante, U. Ryde, Calculation of protein–ligand interaction energies by a fragmentation approach combining high-level quantum chemistry with classical many-body effects, J. Phys. Chem. B 113 (2009) 11085–11094.

[31] C. Curutchet, V.I. Novoderezhkin, J. Kongsted, A. Muñoz-Losa, R. van Grondelle, G.D. Scholes, B. Mennucci, Energy flow in the cryptophyte PE545 antenna is directed by bilin pigment conformation, J. Phys. Chem. B 117 (2013) 4263–4273.

[32] F. Yang, L.G. Moss, G.N. Philips Jr., The molecular structure of green fluorescent protein, Nat. Biotechnol. 14 (1996) 1246–1251.

[33] T. Okada, M. Sugihara, A.-N. Bondar, M. Elstner, P. Entel, V. Buss, The retinal conformation and its environment in rhodopsin in light of a new 2.2 Å crystal structure, J. Mol. Biol. 342 (2004) 571–583.

[34] B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan, M. Karplus, CHARMM: a program for macromolecular energy, minimization, and dynamics calculations, J. Comput. Chem. 4 (1983) 187–217.

[35] M.T.P. Beerepoot, A.H. Steindal, J. Kongsted, B.O. Brandsdal, L. Frediani, K. Ruud, J.M.H. Olsen, A polarizable embedding DFT study of one-photon absorption in fluorescent proteins, Phys. Chem. Chem. Phys. 15 (2013) 4735–4743.

[36] L. Gagliardi, R. Lindh, G. Karlström, Local properties of quantum chemical systems: the LoProp approach, J. Chem. Phys. 121 (2004) 4494–4500.

[37] G. Karlström, R. Lindh, P.-Å. Malmqvist, B.O. Roos, U. Ryde, V. Veryazov, P.-O. Widmark, M. Cossi, B. Schimmelpfennig, P. Neogrady, L. Seijo, MOLCAS: a program package for computational chemistry, Comput. Mater. Sci. 28 (2003) 222–239.

[38] F. Aquilante, L. De Vico, N. Ferré, G. Ghigo, P.-Å. Malmqvist, P. Neogrády, T.B. Pedersen, M. Pitoňák, M. Reiher, B.O. Roos, L. Serrano-Andrés, M. Urban, V. Veryazov, R. Lindh, MOLCAS 7: the next generation, J. Comput. Chem. 31 (2010) 224–247.

[39] A.H. Steindal, J.M.H. Olsen, K. Ruud, L. Frediani, J. Kongsted, A combined quantum mechanics/molecular mechanics study of the one-and two-photon absorption in the green fluorescent protein, Phys. Chem. Chem. Phys. 14 (2012) 5440–5451.

[40] A.D. Becke, Density-functional thermochemistry. III. The role of exact exchange, J. Chem. Phys. 98 (1993) 5648–5652.

[41] C. Lee, W. Yang, R.G. Parr, Development of the Colle–Salvetti correlation-energy formula into a functional of the electron density, Phys. Rev. B 37 (1988) 785–789.

[42] S.H. Vosko, L. Wilk, M. Nusair, Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis, Can. J. Phys. 58 (1980) 1200–1211.

[43] P.J. Stephens, F.J. Devlin, C.F. Chabalowski, M.J. Frisch, Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields, J. Phys. Chem. 98 (1994) 11623–11627.

[44] W.J. Hehre, R. Ditchfield, J.A. Pople, Self-consistent molecular orbital methods. XII. Further extensions of Gaussian-type basis sets for use in molecular orbital studies of organic molecules, J. Chem. Phys. 56 (1972) 2257–2261.

[45] T. Clark, J. Chandrasekhar, G.W. Spitznagel, P. Von Ragué Schleyer, Efficient diffuse function-augmented basis sets for anion calculations. III. The 3 − 21 + G basis set for first-row elements, Li–F, J. Comput. Chem. 4 (1983) 294–301.

[46] P.C. Hariharan, J.A. Pople, The influence of polarization functions on molecular orbital hydrogenation energies, Theor. Chim. Acc. 28 (1973) 213–222.

[47] T.H. Dunning Jr., Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen, J. Chem. Phys. 90 (1989) 1007–1023.

[48] T. Yanai, D.P. Tew, N.C. Handy, A new hybrid exchange–correlation functional using the Coulomb-attenuating method (CAM-B3LYP), Chem. Phys. Lett. 393 (2004) 51–57.

[49] N.H. List, J.M. Olsen, T. Rocha-Rinza, O. Christiansen, J. Kongsted, Performance of popular XC-functionals for the description of excitation energies in GFP-like chromophore models, Int. J. Quantum Chem. 112 (2012) 789–800.

[50] P. Söderhjelm, Polarization effects in protein–ligand calculations extend farther than the actual induction energy, Theor. Chem. Acc. 131 (2012) 1159.

[51] J.M. Olsen, K. Aidas, K.V. Mikkelsen, J. Kongsted, Solvatochromic shifts in uracil: a combined MD-QM/MM study, J. Chem. Theory Comput. 6 (2010) 249–256.

[52] K. Aidas, J. Kongsted, A. Osted, K.V. Mikkelsen, O. Christiansen, Coupled cluster calculation of the $n \rightarrow \pi^*$ electronic transition of acetone in aqueous solution, J. Phys. Chem. A 109 (2005) 8001–8010.

[53] I. Renge, Solvent dependence of n-π^* absorption in acetone, J. Phys. Chem. A 113 (2009) 10678–10686.

[54] A.H. Steindal, K. Ruud, L. Frediani, K. Aidas, J. Kongsted, Excitation energies in solution: the fully polarizable QM/MM/PCM method, J. Phys. Chem. B 115 (2011) 3027–3037.

Electronic Supplementary Information for
"Convergence of environment polarization effects in
multiscale modeling of excitation energies"

Maarten T.P. Beerepoot*, Arnfinn Hykkerud Steindal, Kenneth Ruud

Centre for Theoretical and Computational Chemistry, Department of Chemistry, University of Tromsø, N-9037 Tromsø, Norway.

Jógvan Magnus Haugaard Olsen, Jacob Kongsted

Department of Physics, Chemistry and Pharmacy, University of Southern Denmark, DK-5230 Odense M, Denmark.

Contents

*E-mail: maarten.beerepoot@uit.no; Tel: +47 77623103

1. Preparation of the structures

1.1. Green fluorescent protein

The 1GFL crystal structure [1] (containing a bond length pattern that is characteristic for a neutral chromophore [2]) for the green fluorescent protein (GFP) was obtained from the protein data bank [3] and prepared with an anionic chromophore as also described in Ref. 2. Only chain A was used from the 1GFL crystal structure, which is a dimer. Crystal waters were included based on the crystal structure. The protonation states of the ionizable residues were predicted with the Schrödinger protein preparation wizard [4]. As a result of this the following residues were charged: Ala1, all lysine, arginine and aspartic acid residues and all glutamic acid residues with the exception of Glu222, in line with experimental evidence for the anionic chromophore [5]. Six histidines were protonated at N_δ (His139, His148, His169, His181, His199 and His217) and three histidines were protonated at N_ϵ (His25, His77, His81).

To model the environment of the *anionic* chromophore, the N–C_α–C_β–O dihedral angle from Thr203 was set to 66°, which is the average value from the 2Y0G and 1EMA crystal structures (where the chromophore is anionic). The hydrogen-bonding network around the chromophore (OH in Ser65, CH_2OH in Ser205, COOH in Glu222 and H_2O between Ser205 and the phenolate from the chromophore) was relaxed (conjugate gradient optimization with a convergence criterium of gradient changes less than 0.05 kcal/mol/Å between consecutive steps) using the OPLS force field [6, 7] with the rest of the protein present with the constraint of being frozen. Thr203 and its closest water molecule were also relaxed in the otherwise frozen protein in order to relieve steric strain between Thr203 and the water molecule.

The chromophore and parts of its surroundings were subsequently geometry optimized using DFT (B3LYP [8–11]/6-31+G* [12–14]) within the electrostatic field of the rest of the protein using the QSite QM/MM program [15] and the OPLS force field [6, 7] (electrostatic embedding). The QM part consisted of i) the conjugated system of the chromophore (using hydrogen caps to cut the bonds to the rest of the protein), ii) six residues involved in hydrogen-bonding with the chromophore: Gln94, Arg96, Gln148, Thr203, Ser205 and Glu222 (all using frozen orbital cuts between C_α and C_β) and iii) three water molecules that form hydrogen bonds directly to the chromophore. The positions of nearby protein atoms (from residues 65 and 67) were allowed to move during the optimization in order to keep the link between the QM and MM regions stable.

The chromophore cut out as the QM region for the calculations in the protein consisted of residues 65-67 including a carbonyl group at the N-terminus and a nitrogen at the C-terminus that come from residues 64 and 68, respectively, in agreement with earlier work [2, 16]. Both sides were capped with an extra hydrogen atom. In the same manner, residue Tyr203 was cut out and capped with hydrogen atoms at the linking position to the protein backbone.

50 snapshots for GFP solvated in water were taken from a 15 ns molecular dynamics (MD) simulation using the CHARMM27 force field [17]. Details of this MD simulation are provided in Ref. 2. For the PE-TD-DFT calculations,

the protein was extracted together with a solvation shell consisting of all water molecules with one of the atoms within 8 Å from one of the protein atoms.

An extra 1 ns MD simulation using the same procedure [2] was run with Desmond [18] on the GFP chromophore in water described by the TIP3P force field [19] to investigate the influence of a different environment on the same chromophore. For this purpose, the conjugated system of the chromophore was capped with methyl groups. The chemical structures are shown in Figure SI–1 for the chromophore in the protein (left) and in water (right). Both orbitals of the $\pi \rightarrow \pi^*$ excitation that is considered here are localized on the conjugated system consisting of the two rings and the bridge between them. Parts of the side chains are included in the PE-TD-DFT calculations on the whole protein to minimize the influence of the QM/MM boundary on the calculated property.
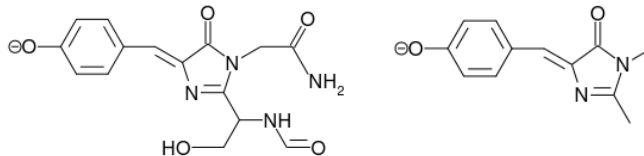

Figure SI–1: Comparison of the structure of the GFP chromophore used in the PE-TD-DFT calculations in the protein (left) and in water (right).

1.2. Rhodopsin

The 1U19 crystal structure [20] (resolution 2.2 Å) was prepared using the protein preparation wizard of the Schrödinger suite [4]. The system studied includes chain A of the dimer (including the retinal chromophore which is co-valently bonded to Lys296) and two crystal water molecules coordinating to Glu113 and Glu181, respectively. A disulfide bond between Cys110 and Cys187 was made [21]. Glu113 [20, 21] and Glu181 [21, 22] were both manually de-protonated. Hydrogen atoms were added and hydrogen-bonding networks were optimized by changing the orientation and protonation state of all histidine, asparagine and glutamine residues as well as the the protonation states for the other ionizable residues using the protein preparation wizard [4]. As a result of this, all arginine and lysine residues were positively charged and all histidine residues neutral. All glutamic acid and aspartic acid residued were negatively charged except for Asp83 and Glu122, which were neutral. Three histidines were protonated at N_δ (His65, His100, His278) and three histidines were proto-nated at N_ϵ (His152, His195, His211). Particular attention was taken to have the correct hydrogen-bonding network around Glu113 as evidenced by QM/MM molecular dynamics [20, 23]. The hydroxyl group of Thr94 was oriented such that it donates a hydrogen bond to the carboxylate of Glu113. The hydrogen-bonding network (proton of Schiff base, $CH_2CH_2COO^-$ of Glu113, $CHCH_3$

3

of Thr94 and the two hydrogens of the water molecule) were subsequently relaxed (conjugate gradient optimization with a convergence criterium of gradient changes less than 0.05 kcal/mol/Å between consecutive steps) with the OPLS force field [6, 7] and with the rest of the protein present with the constraint of being frozen. No changes were made to the hydrogen-bonding network around Glu181, where Tyr286 and a water molecule coordinate to the one and Tyr192 to the other oxygen of the carboxylate group. The water also donates a hydrogen bond to Ser186.

All hydrogens in the protein were optimized with the OPLS force field [6, 7]. The retinal chromophore and the side chains of Lys296 and Glu113 were subsequently geometry optimized using DFT (B3LYP [8–11]/6-31+G* [12–14]) within the electrostatic field of the rest of the protein using the QSite QM/MM program [15] and the OPLS force field [6, 7] (electrostatic embedding). The two QM/MM boundaries were chosen between C_α and C_β using hydrogen caps and frozen orbitals for Lys296 and Glu113, respectively. The positions of nearby protein atoms were allowed to move during the optimization in order to keep the link between the QM and MM regions stable.

The negative charge on Glu113 stabilizes the positive charge on the Schiff base [20]. The distance between the Schiff base N and the closest carboxylate O from Gly113 is 3.5 Å (chain A) / 3.3 Å (chain B) in the 1U19 crystal structure and at least 3.1 Å in the other available crystal structures [20]. Theoretical calculations, however, give a significantly shorter distance of 2.6–2.7 Å [20, 23]. Also in this study, the distance of 3.5 Å decreases to 3.0 Å during the classical relaxation of the hydrogen-bonding network and further to 2.8 Å during QM/MM geometry optimization.

We note that our work lacks a couple of factors that are taken into account in e.g. the thorough approach of Valsson et al. [23], where the dimer with all the cofactors is embedded in a membrane, Gly113 is included in the QM region of the excitation energy calculation and where QM/MM molecular dynamics is performed. However, we believe that our approach of taking one structure of the monomer is adequate in order to investigate the extent of polarization effects in rhodopsin. This is partially supported by our observation that the rate of convergence is the same for calculations based on the crystal structure and those based on molecular dynamics snapshots for the GFP (Figure 3 in the main article).

1.3. Solute–solvent structures

Solute–solvent configurations were made by putting one solute molecule in a cubic box with edges of 60 Å and filling the rest of the box with solvent molecules. The following solute–solvent systems were made in this way: uracil in water and acetone in water, methanol, dimethyl sulfoxide (DMSO), diethyl ether (DEE) and hexane.

The systems were minimized and equilibrated using molecular mechanics with Gromacs [24] using three-dimensional periodic boundary conditions and the OPLS force field [6, 7]. The SPC model [25] was used for water. Non-bonded interactions were treated with a cut-off radius of 15 Å. Electrostatic

interactions beyond this threshold were treated with the smooth particle-mesh Ewald method [26] with a tolerance of 10^{-5}. The minimization consisted of 20 steps of steepest descent followed by 1000 steps of conjugate gradient. To optimize the density of the system, an NPT equilibration of 500 ps was run with the Berendsen [27] temperature (298 K) and pressure (1 bar) coupling. Initial velocities were obtained from a Maxwell distribution at 298 K. The lengths of the cubic simulation boxes were at least 57 Å for all simulation boxes after NPT equilibration. Subsequently, the systems were equilibrated for 2 ns in the NVT ensemble using the Berendsen thermostat at 298 K and a time step of 1 fs.

Snapshots for the solute–solvent systems were obtained from molecular dynamics simulations. The systems were simulated for 10 ns in the NVT ensemble with the same time step, non-bonded interactions and other parameters as in the equilibration. A snapshot was made every 200 ps, generating 50 non-correlated snapshots in total. For uracil in water it has previously been shown that fluctuations in the excitation energy of the $\pi \rightarrow \pi^*$ transition are small, so that 50 snapshots suffice [28].

The QSITE QM/MM program [15] was used to optimize the solute molecules using DFT (B3LYP [8–11]/6-31+G*[12–14]) within the frozen environment of the solvent molecules (electrostatic embedding). In this case, an accurate geometry for the solute is obtained without altering the sampling of solvent configurations by the force field. The solvent environment was described by the OPLS force field [6, 7].

The structures used for the PE-TD-DFT calculations consisted of the solute (in the QM region) and all solvent molecules with at least one atom within 20 Å from the solute (in the embedding region). Only for one snapshot of uracil in water the threshold was chosen to be 25 Å for comparison of different types of embedding parameters (Figure SI-2).

2. Additional figures

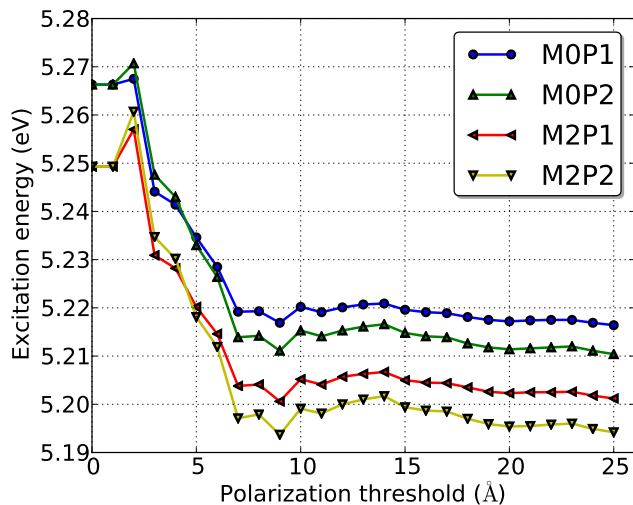*2.1. Influence of different force fields*

Figure SI–2: Excitation energies from PE-TD-DFT calculations on a snapshot of **uracil in water** using different thresholds (in Å) to include polarization. The different graphs represent different types of embedding with distributed multipole moments up to either charges (M0) or quadrupoles (M2) and either isotropic (P1) or anisotropic (P2) polarizabilities. Electrostatic interactions were fully included for water atoms up to a distance of 25 Å from uracil.
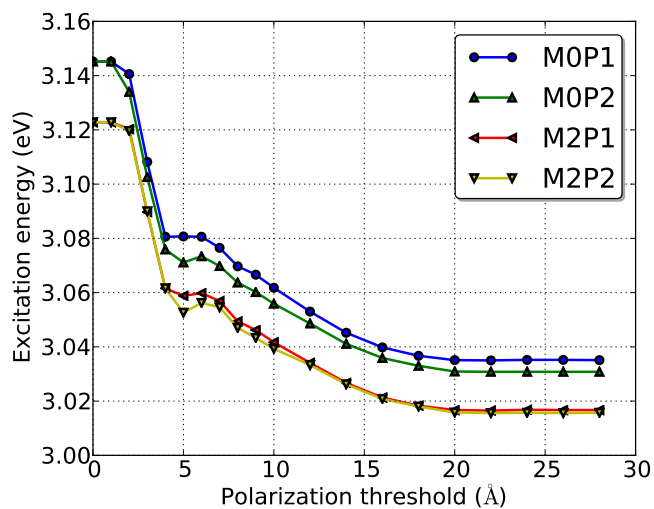
Figure SI–3: Excitation energies from PE-TD-DFT calculations on the **anionic GFP crystal structure** using different thresholds (in Å) to include polarization. The different graphs represent different types of embedding with distributed multipole moments up to either charges (M0) or quadrupoles (M2) and either isotropic (P1) or anisotropic (P2) polarizabilities. Electrostatic interactions were fully included for all atoms in the protein.
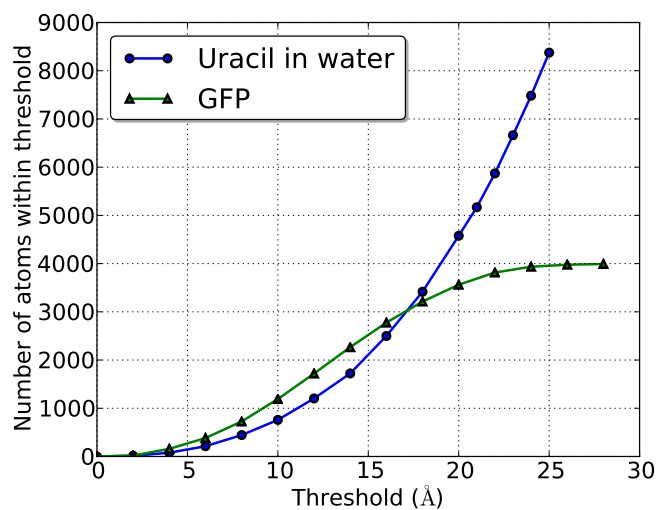
Figure SI–4: Number of atoms within a certain distance from the chromophore for a snapshot of uracil in water and for the anionic GFP crystal structure. The uracil snapshot contains water molecules with one atom within 25 Å from the solute and is more or less spherical. The GFP snapshot contains water molecules within 8 Å from the protein and is barrel-shaped.
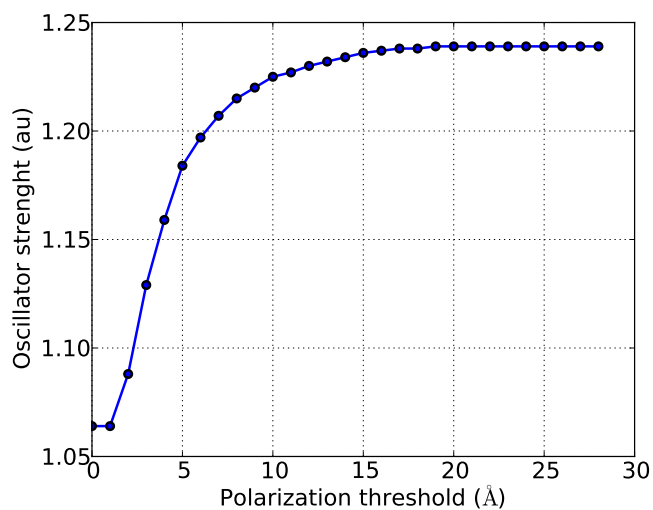
Figure SI–5: One-photon oscillator strengths associated with the lowest $(\pi \rightarrow \pi^*)$ excitation of the anionic GFP chromophore. The results are obtained from PE-TD-DFT calculations on the crystal structure using different thresholds (in Å) to include polarization. Electrostatic interactions were fully included for all atoms in the protein.

Figure SI–6: Two-photon absorption cross sections associated with the lowest $(\pi \to \pi^*)$ excitation of the anionic GFP chromophore. The results are obtained from PE-TD-DFT calculations on the crystal structure using different thresholds (in Å) to include polarization. Electrostatic interactions were fully included for all atoms in the protein.
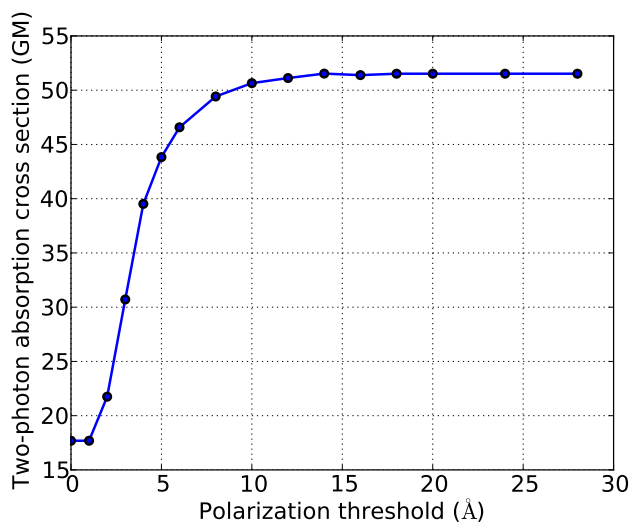
3. Excitation energies for acetone in different solvents

3.1. Seperation into different contributions

Table SI–I: Calculated $n \to \pi^*$ excitation energies for acetone in five different solvents. The vacuum value (ΔE_{vac} in eV) is for the gas phase geometry optimized (B3LYP/6-31+G*) structure. The TD-DFT excitation energy in the solvent geometry (ΔE_{sol} in eV), the PE-TD-DFT excitation energy with only electrostatics (ΔE_{el} in eV) and the PE-TD-DFT excitation energy including also polarization (ΔE_{pol} in eV) are averages over 50 snapshots extracted from an MD simulation in which acetone is QM/MM geometry optimized (B3LYP/6-31+G*) in the frozen solvent. Excitation energies are calculated with CAM-B3LYP/aug-cc-pVDZ in all cases. Electrostatic and polarization interactions are cut off at 20 Å in the PE-TD-DFT calculations.

solvent	ΔE_{vac}	ΔE_{sol}	ΔE_{el}	ΔE_{pol}
hexane	4.411	4.409	4.404	4.428
diethyl ether (DEE)	4.411	4.396	4.419	4.444
dimethylsulfoxide (DMSO)	4.411	4.368	4.459	4.492
methanol	4.411	4.373	4.505	4.549
water	4.411	4.362	4.569	4.652

Table SI–II: Calculated $n \to \pi^*$ excitation energies (ΔE in eV) and solvent shifts (ω in cm^{-1}) for acetone in five different solvents using different system sizes. The devation of ω w.r.t. the calculations on the largest system (R_{el} = R_{pol} = 20 Å) is also shown as a measure of the convergence (deviation in %). The excitation energies are averages over 50 snapshots extracted from an MD simulation in which acetone is QM/MM geometry optimized (B3LYP/6-31+G*) in the frozen solvent. Excitation energies are calculated with the CAM-B3LYP/aug-cc-pVDZ in all cases.

solvent	ΔE	ω	deviation
R_{el} = R_{pol} = 5 Å			
hexane	4.426	126	-7.7 %
diethyl ether (DEE)	4.440	232	-12.7 %
dimethylsulfoxide (DMSO)	4.478	540	-16.3 %
methanol	4.532	975	-11.2 %
water	4.629	1760	-7.5 %
R_{el} = R_{pol} = 8 Å			
hexane	4.427	135	-3.2 %
diethyl ether (DEE)	4.443	263	-2.9 %
dimethylsulfoxide (DMSO)	4.488	622	-5.0 %
methanol	4.544	1079	-3.0 %
water	4.645	1893	-2.7 %
R_{el} = R_{pol} = 10 Å			
hexane	4.428	137	-1.9 %
diethyl ether (DEE)	4.444	267	-1.5 %
dimethylsulfoxide (DMSO)	4.491	646	-1.3 %
methanol	4.547	1100	-1.1 %
water	4.647	1905	-2.1 %
R_{el} = R_{pol} = 20 Å			
hexane	4.428	139	
diethyl ether (DEE)	4.444	271	
dimethylsulfoxide (DMSO)	4.492	655	
methanol	4.549	1112	
water	4.652	1945	

References

[1] F. Yang, L. G. Moss, G. N. Philips Jr., The molecular structure of green fluorescent protein, Nat. Biotechnol. 14 (1996) 1246–1251.

[2] M. T. P. Beerepoot, A. H. Steindal, J. Kongsted, B. O. Brandsdal, L. Frediani, K. Ruud, J. M. H. Olsen, A polarizable embedding DFT study of one-photon absorption in fluorescent proteins, Phys. Chem. Chem. Phys. 15 (2013) 4735–4743.

[3] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, The Protein Data Bank, Nucl. Acids Res. 28 (2000) 235–242.

[4] Schrödinger Suite 2013 Protein Preparation Wizard; Epik version 2.5, Schrödinger, LLC, New York, NY, 2013; Impact version 6.0, Schrödinger, LLC, New York, NY, 2013; Prime version 3.3, Schrödinger, LLC, New York, NY, 2013.

[5] R. Y. Tsien, The green fluorescent protein, Annu. Rev. Biochem. 67 (1998) 509–544.

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

[7] G. A. Kaminski, R. A. Friesner, J. Tirado-Rives, W. L. Jorgensen, Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides, J. Phys. Chem. B 105 (2001) 6474–6487.

[8] S. H. Vosko, L. Wilk, M. Nusair, Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis, Can. J. Phys. 58 (1980) 1200–1211.

[9] C. Lee, W. Yang, R. G. Parr, Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density, Phys. Rev. B 37 (1988) 785–789.

[10] A. D. Becke, Density-functional thermochemistry. III. The role of exact exchange, J. Chem. Phys. 98 (1993) 5648–5652.

[11] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, M. J. Frisch, Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields, J. Phys. Chem. 98 (1994) 11623–11627.

[12] W. J. Hehre, R. Ditchfield, J. A. Pople, Self-Consistent Molecular Orbital Methods. XII. Further Extensions of Gaussian-type Basis Sets for use in Molecular Orbital Studies of Organic Molecules, J. Chem. Phys. 56 (1972) 2257–2261.

[13] T. Clark, J. Chandrasekhar, G. W. Spitznagel, P. Von Ragué Schleyer, Efficient diffuse function-augmented basis sets for anion calculations. III. the 3-21+G basis set for first-row elements, Li–F, J. Comput. Chem. 4 (1983) 294–301.

[14] P. C. Hariharan, J. A. Pople, The influence of polarization functions on molecular orbital hydrogenation energies, Theor. Chim. Acc. 28 (1973) 213–222.

[15] QSite, version 5.8, Schrödinger, LLC, New York, NY, 2012.

[16] A. H. Steindal, J. M. H. Olsen, K. Ruud, L. Frediani, J. Kongsted, A combined quantum mechanics/molecular mechanics study of the one-and two-photon absorption in the green fluorescent protein, Phys. Chem. Chem. Phys. 14 (2012) 5440–5451.

[17] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, M. Karplus, CHARMM: A program for macromolecular energy, minimization, and dynamics calculations, J. Comput. Chem. 4 (1983) 187–217.

[18] Desmond Molecular Dynamics System, version 3.0, D. E. Shaw Research, New York, NY, 2011. Maestro-Desmond Interoperability Tools, version 3.0, Schrödinger, New York, NY, 2011.

[19] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, Comparison of simple potential functions for simulating liquid water, J. Chem. Phys. 79 (1983) 926–935.

[20] T. Okada, M. Sugihara, A.-N. Bondar, M. Elstner, P. Entel, V. Buss, The Retinal Conformation and its Environment in Rhodopsin in Light of a New 2.2 Å Crystal Structure, J. Mol. Biol. 342 (2004) 571–583.

[21] K. Palczewski, T. Kumasaka, T. Hori, C. A. Behnke, H. Motoshima, B. A. Fox, I. Le Trong, D. C. Teller, T. Okada, R. E. Stenkamp, M. Yamamoto, M. Miyano, Crystal Structure of Rhodopsin: A G Protein-Coupled Receptor, Science Signaling 289 (2000) 739–745.

[22] M. N. Sandberg, T. L. Amora, L. S. Ramos, M.-H. Chen, B. E. Knox, R. R. Birge, Glutamic Acid 181 is Negatively Charged in the Bathorhodopsin Photointermediate of Visual Rhodopsin., J. Am. Chem. Soc. 133 (2011) 2808–2811.

[23] O. Valsson, P. Campomanes, I. Tavernelli, U. Rothlisberger, C. Filippi, Rhodopsin Absorption from First Principles: Bypassing Common Pitfalls, J. Chem. Theory Comput. 9 (2013) 2441–2454.

[24] B. Hess, C. Kutzner, D. Van Der Spoel, E. Lindahl, GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation, J. Chem. Theory Comput. 4 (2008) 435–447.

[25] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, J. Hermans, in: B. Pullman (Ed.), Intermolecular Forces, Reidel, Dordrecht, 1981.

[26] T. Darden, D. York, L. Pedersen, Particle mesh Ewald: An N·log (N) method for Ewald sums in large systems, J. Chem. Phys. 98 (1993) 10089–10092.

[27] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, J. R. Haak, Molecular dynamics with coupling to an external bath, J. Chem. Phys. 81 (1984) 3684–3690.

[28] J. M. Olsen, K. Aidas, K. V. Mikkelsen, J. Kongsted, Solvatochromic Shifts in Uracil: a Combined MD-QM/MM Study, J. Chem. Theory Comput. 6 (2010) 249–256.