Electronic Absorption Spectra from MM and *ab initio* QM/MM Molecular Dynamics: Environmental Effects on the Absorption Spectrum of Photoactive Yellow Protein

Christine M. Isborn[a,b,c], **Andreas W. Götz**[d], **Matthew A. Clark**[d], **Ross C. Walker**[d,e], and **Todd J. Martínez**[a,b,*]

[a]PULSE Institute and Department of Chemistry, Stanford University, Stanford, CA 94305

[b]SLAC National Accelerator Laboratory, Menlo Park, CA 94025

[d]San Diego Supercomputer Center, University of California San Diego, La Jolla, CA 92093

[e]Department of Chemistry and Biochemistry, University of California San Diego, La Jolla, CA 92093

Abstract

We describe a new interface of the GPU parallelized TᴇʀᴀCʜᴇᴍ electronic structure package and the Aᴍʙᴇʀ molecular dynamics package for quantum mechanical (QM) and mixed QM and molecular mechanical (MM) molecular dynamics simulations. This QM/MM interface is used for computation of the absorption spectra of the photoactive yellow protein (PYP) chromophore in vacuum, aqueous solution, and protein environments. The computed excitation energies of PYP require a very large QM region (hundreds of atoms) covalently bonded to the chromophore in order to achieve agreement with calculations that treat the entire protein quantum mechanically. We also show that 40 or more surrounding water molecules must be included in the QM region in order to obtain converged excitation energies of the solvated PYP chromophore. These results indicate that large QM regions (with hundreds of atoms) are a necessity in QM/MM calculations.

Introduction

Combining quantum mechanical (QM) and classical force field molecular mechanical (MM) methods creates the hybrid QM/MM approach.[1–9] This framework has been used to study many complex reactions including enzymatic catalysis,[10–12] ligand-protein binding,[13] and the excited states of chromophores embedded in a protein matrix.[14–17] The QM/MM approach will doubtless be applied in the future to modeling many complex systems, including charge transfer at interfaces, electronic excitation in polymeric materials, and fluorescence of biomolecules. The QM/MM technique is traditionally used to treat a chemically important part of the system, often where bond breaking or electronic excitation occurs, with QM, while the remainder is modeled with an MM force field. It is then hoped that the MM environment of atomic point charges (usually fixed) provides a sufficiently accurate electronic description for the chemically important QM region, and that any potential artifacts at the QM/MM interface do not disturb the chemical reactivity. Although

it is clearly desirable to have a large QM region to reduce effects due to the QM/MM interface, there is a necessary balance between increasing the accuracy of the QM method and increasing the size of the QM region. This is especially true when going beyond single point energy calculations to *ab initio* QM/MM molecular dynamics (MD).

TERACHEM is a recently introduced QM electronic structure package[18] that uses GPU parallelism[19,20] to compute the ground[21] and excited states[22] of molecules of unprecedented size at previously unattainable speeds. The code is thus a natural choice for *ab initio* QM/MM MD simulations. The AMBER MD software package[23] provides many options for MD simulations, features an efficient semi-empirical QM/MM implementation,[24] and has recently introduced a general, extensible interface for *ab initio* QM/MM MD with external QM software packages.[25] There are various ways to couple QM and MM methods in the QM/MM framework. QM/MM MD simulations in AMBER support both a simple mechanical embedding scheme as well as electronic embedding based on an additive scheme. In the electronic embedding approach, used throughout this work, the MM point charges are included in the QM Hamiltonian and thus polarize the QM region throughout the self-consistent field process. Bonds crossing the QM/MM boundary are capped with hydrogen link atoms.

In this work, we describe efforts towards large scale *ab initio* QM/MM MD through the use of a robust interface between the TERACHEM and AMBER codes that supports data communication via the message passing interface[26] (MPI) standard. This QM/MM interface is applied to the computation of the UV/Vis absorption spectra of the photoactive yellow protein (PYP) chromophore in vacuum, aqueous solution, and protein environments.

Molecular dynamics provides many configurations within the temperature accessible potential energy landscape, and thus provides a much richer picture of molecular structure than a single geometrical configuration. Sampling of the vibrational profile and local environment is key to computing accurate properties that can be compared directly to experimental data. In this work we compare the spectra computed from dynamics performed on both the MM and QM potential energy surfaces (PESs). The computed absorption spectrum acts as a reporter for the ground and excited state PES. Broadening of spectral lines occurs naturally through vibrational motion of the chromophore and fluctuations in the surrounding environment.

While the electronic excitation is localized to the chromophore, the position and shape of the spectra are affected by the environment. The description of the environment given by a fully QM Hamiltonian differs from a QM/MM Hamiltonian, which is reflected in differing absorption spectra. This work examines the effect on the computed absorption spectrum when varying the size of the QM region around the chromophore, both in solution and in the protein. With increasing capabilities of QM computation, checking the convergence of a molecular property with QM region size is more feasible than previously possible. Convergence with respect to QM region size in the context of solvation and intermolecular interaction energies has recently been explored in single snapshot QM/MM calculations by Fox et al,[27] Meier et al have investigated how the size of the QM region affects the strength of hydrogen bonding within a photoreceptor protein active site,[28] and the Ochsenfeld and Johnson groups have recently examined the convergence of NMR shieldings and shifts with QM region size.[29,30] These first studies disappointingly suggest that rather large QM regions are needed to achieve convergence for ground state properties such as solvation energies. As will be shown below, we find that this conclusion also holds for excitation energies and absorption spectra.

For PYP, the electronic excitation and solvatochromic shifts in vacuum, water, and protein environments are of great interest. In the general class of photoreceptor proteins,[31] PYP is relatively small: 125 residues. The protein acts as photosensor for blue light in the bacterium *Halorhodospira halophila*, having a strong peak with an absorption maximum at 446 nm (2.78 eV) from the $\pi \rightarrow \pi$ * transition in the *para*-coumaric acid chromophore.[32–41] The excited state dynamics of the chromophore reveal a *trans-cis* isomerization upon photoabsorption, which leads to signal transduction and a negative phototactic response of the bacterium to possibly harmful blue light. The PYP chromophore is covalently linked to the protein via a thioester bond to cysteine-69, and is known to be deprotonated when inside the hydrophobic pocket of the protein.[33,34] The negative charge of the phenolate is stabilized by hydrogen bonding with three nearby residues: GLH-46 (glutamic acid in the protonated form), TYR-42, and THR-50. Additional stabilization is also provided by the adjacent positive charge on ARG-52, which affects the ionization potential of the chromophore, and π-stacking interactions with PHE-96.

In an attempt to measure the absorption of the PYP anionic chromophore without any complicating environmental effects, Nielsen et al[42] performed vacuum electrospray ionization action spectroscopy on deprotonated *trans*-thiophenyl-*p*-coumarate (pCT⁻, see Figure 1), an analog of the PYP chromophore. The absorption maximum of 395 nm (3.14 eV) in basic aqueous buffer solution (pH 10.2 to keep the phenolate deprotonated as in the protein) is significantly blue shifted from the measured vacuum absorption maximum at 460 nm (2.70 eV). However, because the bright state is embedded in the ionization continuum for the anionic PYP chromophores in the gas phase,[43] there are some complications in the experimental interpretation, and this vacuum absorption maximum must be viewed with caution.

The excitation energies of the PYP chromophore have proven to be quite challenging to model theoretically.{Gromov, 2007 #23} This is partially due to the excited state of the anionic chromophore lying above or near the first ionization potential.[44–47] Krylov and co-workers recently performed an extensive computational study of the excitation energies of the phenoxide and carboxylate forms of the *para*-coumaric acid (pCA⁻) model of the PYP chromophore using a variety of high-level methods.[48] They found that the phenoxide excited state, which is the form similar to that in the native protein state, is dominated by a single one-electron excitation and that both single- and multi-reference approaches give similar excitation energies. However, the inclusion of triples also proved important, decreasing the excitation energy by 0.2 eV, and the addition of diffuse functions decreased the excitation energy by an additional 0.16 eV. Their best estimate of the first excitation energy of pCA⁻ was 3.0 eV as predicted by the CC3/6-31+G(d,p) and SS-CASPT2/ANO-RCC-VTZP methods.

While the excited states of the PYP chromophore in vacuum can be modeled with high levels of *ab initio* theory, the excited states of the chromophore in solution and in the protein are also of great interest, and must be treated with more affordable excited state techniques.[49] A conductor-like polarizable continuum model with time-dependent density functional theory (TDDFT) using B3LYP/aug-cc-pVDZ was employed by Wang et al[50] to compute the solvatochromic shift of another analog of the PYP chromophore. In going from vacuum to aqueous solvation, they computed that the excitation energy would be red shifted by 0.14 eV (3.17 eV $\rightarrow$ 3.03 eV). This is in contrast to the experimental finding and it is unknown if the cause of this incorrect red shifting is due to a problem with the simplicity of the implicit solvation model or with the TDDFT level of theory.

Gromov, et al. have systematically studied the effects of including surrounding protein residue side chains (amino acid fragments) on the low-lying excited states of the

chromophore using the second-order approximate coupled cluster singles and doubles (CC2) method.[51] They confirm that the $\pi \rightarrow \pi$ * excitation is dominated by a single configuration, even when some of the surrounding residues are modeled quantum mechanically. When using the same CC2 method as Gromov et al, Rocha-Rinza et al computed the excitation energy of the phenoxide form of pCA$^-$ to be 3.10 eV, while their TD-CAM-B3LYP calculations gave a value of 3.40 eV. Both methods overestimate the excitation energy compared to the experimental absorption maximum of 2.88 eV.[52]

Although TDDFT has well known problems in modeling some classes of excited states, including those involving double excitation, charge-transfer, and Rydberg character, it is currently the method of choice for computing the valence excited states of large molecular systems simply because it is computationally tractable.[53–57] The TDDFT charge-transfer issue is particularly problematic in aqueous environments,[58,59] but recently developed range-corrected functionals such as ωPBE and CAM-B3LYP show promise in resolving the charge-transfer deficiencies of TDDFT.[60] In general, TDDFT calculations consistently overestimate the excitation energy of the anionic PYP chromophore.[61] Thus, even though pCA$^-$ is dominated by a single electron excitation, auto-ionizing near-resonance states likely affect the accuracy of TDDFT excitations.[43]

The protein environment around the PYP chromophore has also been modeled via MM methods.[61–65] However, in order to explore the effects of including a substantial portion of the protein surrounding the chromophore in the QM region, it is necessary to select suitable QM approaches. TDDFT is one of the few QM excited state methods that is computationally tractable enough for this size of computation. The method has previously been used to compute the excitation of the PYP chromophore with the inclusion of the side chains of some of the nearby residues in the QM region.[65–68]

With few exceptions,[62–64,69] previous work on excited states of PYP (in vacuum, aqueous, and protein environments) has been restricted to static structures, with no consideration of dynamics. The present work takes the vast literature of PYP excitation energy calculations beyond static structures to include dynamic effects. While exploring the effects of the PYP environment, we report on how computed excitation energies vary within MM and QM/MM ensembles of configurations and how the size of the QM region affects the excitation energy. Since the photoresponse of the PYP chromophore is a model for other photoactive and signal transduction proteins, such as those in the rhodopsin family, it is important to understand how the computational treatment of the dynamics of the chromophore in the protein and solution environment tune the spectral response.[41] This work is also the first application of the TeraChem / Amber QM/MM interface for robust computation of QM/MM dynamics of large QM regions. The interface is used to compute the spectra of the PYP chromophore in successively more complicated dynamic environments: vacuum, aqueous solution, and in the protein. The spectra from MM and QM/MM dynamics are compared for each environment.

Computational Details: MM and QM/MM system setup, dynamics

MM and QM/MM dynamics were run with the TeraChem (QM) and Amber (MM) programs utilizing the interface described in the following section. Because of the availability of both vacuum and solution phase experimental data, pCT$^-$ (Figure 1) was chosen as the PYP chromophore for the gas and aqueous solution phase investigations. The PYP protein (PDB ID 1NWZ[70]) and pCT$^-$ were solvated in a sphere of SPC/Fw[71] flexible three site point charge waters of 32 Å radius (3395 water molecules). The water solvation sphere was created with the `solvatecap' command in the leap utility of AMBERTools, which uses a soft half-harmonic restraining potential beyond the 32 Å radius. The ff99SB force field[72] was

used for the MM region of the protein with general A$_{MBER}$ force field[73] (GAFF) parameters for the *p*-coumaric acid chromophore. Protonation of the GLU-46 residue was enforced by converting it to GLH-46. Standard protonation states of the PYP residues were verified with the pKa webserver H++[74], and the resulting charge of −6 on the protein was neutralized with six Na+ counter ions.

For each environment, the whole system was minimized, heated, and equilibrated at the MM level of theory. Minimization included 50 steps of steepest decent, followed by 950 steps of conjugate gradient. The time step used throughout the MM and QM/MM dynamics was 0.5 fs, and the center of mass motion was removed every 2000 steps. After minimization, the systems were slowly heated via Langevin dynamics[75] using a friction coefficient of 5 ps^{-1} from 10 K to 300 K over a period of 40 ps, followed by a short equilibration at 300 K for 10 ps using Langevin dynamics and 50 ps of dynamics with the weak coupling Berendsen[76] thermostat using a time constant of 1 ps for the heat bath coupling. From this point, NVT (T=300 K) dynamics were either continued with MM or switched to QM/MM. When we switched to QM/MM dynamics, another 25 ps of each trajectory was discarded from further analysis in order to let the system equilibrate and adjust for the change in Hamiltonian.

The entire protein and surrounding water droplet is included in both the MM and QM/MM PYP ground state dynamics and excited state computations without truncation of any nonbonded interactions. Ground state dynamics and excited state calculations were run with increasingly larger portions of the protein surrounding the chromophore included in the QM region. Two of the larger QM regions used for excited state computation are shown in Figures 2 and 3. The QM regions used are (with number of basis functions listed for 6-31G*):

QM region 1) chromophore + CYS-69 (including as QM the S - C from the cysteine); 22 atoms, 217 basis functions; charge = −1;

QM region 2) Additionally including the side chains of the two residues that may participate in H-bonding with the phenolate: GLH-46 and TYR-42; 49 atoms, 440 basis function; charge = −1;

QM region 3) Additionally including the sides chains of nearby residues ARG-52, VAL-66, THR-50, TYR-98; 104 atoms, 823 basis functions; charge = 0;

QM region 4) Additionally including the side chains of residues PHE-62, PHE-96, PRO-68, ALA-67, THR-70; 159 atoms, 1206 basis functions; charge = 0.

QM region 5) Include entire residues 42–70 and 92–106, including the protein backbone, along with either C-O or N-H atoms on adjacent residues to keep from cutting through the peptide bond; 723 atoms, 6277 basis functions; charge = 0.

Entire protein + counter ions) 1935 atoms, 16827 basis functions; charge = 0.

For ground state QM and QM/MM MD we use the B3LYP[77,78] density functional with the 6-31G basis set for the QM region, with some additional benchmarking using the 6-31G* basis set as noted. Charge-transfer transitions can be particularly problematic in an aqueous and protein environment since nearby waters and residues can act as charge donors or acceptors; the long-range corrected density functional ωPBE[79–81] was therefore used for all excited state computations with the 6-31G* basis set, with some benchmarking with larger basis sets as described in the text. A recent TDDFT study showed that the results of pure DFT functionals compared better to experiment with increasing system size, but the trend for hybrid functionals was in the opposite direction.[82] It is thus not at all clear how DFT/TDDFT will perform with increasing size of QM region, and whether one can assume that the behavior is consistent across system size. It is also known that the amount of double

excitation character increases for increasing length of conjugation in a hydrocarbon,[83] which suggests that TDDFT may become less accurate with increasing system size if the amount of double excitation character in the ground state increases. As the PYP chromophore is not part of an extended delocalized hydrocarbon network, it is not anticipated that the amount of double excitation character will substantially vary as more of the surrounding environment (water or protein) is taken into consideration.

The QM *ab initio* dynamics technique used here invokes the Born-Oppenheimer approximation, thus, only the electrons are treated quantum mechanically, and classical nuclear trajectories are evolved on this quantum electronic PES. As this model ignores the effects of vibrational progressions, no vibrational fine structure is computed for the absorption spectra. Upon computation of the TDDFT vertical excited states for each snapshot, each `stick' in the resulting spectrum is convoluted with a Gaussian of width 0.04 eV, and the Gaussians from all states for all snapshots are summed together to create the average spectrum. Errors in λ_{max} due to sampling are reported as $+/- 2\times$ the standard deviation (2σ) of 100 bootstrapped values of the average computed spectrum, which would contain ~95% of the computed averages, assuming a normal distribution.

Interface of TₑᵣₐCₕₑₘ and Aₘᵦₑᵣ

The QM/MM interface employed here is currently available as part of the version 12 release of the AMBER software package and is described in detailed elsewhere.[24,25] The energy of the QM/MM system can be decomposed into the energy of the QM region, the energy of the MM region, and the interaction energy between the QM and MM regions:

$$E_{QM/MM} = E_{QM} + E_{MM} + E_{int}. \quad (1)$$

A QM/MM calculation thus requires not only a choice for the QM and MM potentials, but also for the interaction energy E_{int}. In the simplest approach, electronic coupling between the QM and MM regions is neglected, and all non-bonded interactions, both electrostatic and van der Waals (vdW), are treated using the underlying MM potential. This "mechanical embedding" approach is useful to impose steric constraints on the embedded QM system. This approximation is inappropriate for considering solvent effects on electronic excitations and instead an electronic embedding approach is required which allows for polarization of the embedded QM region. In this case the interaction energy reads

$$E_{int} = \sum_{J=1}^{N_{QM}} \sum_{k=1}^{N_{MM}} \frac{Z_J q_k}{|\mathbf{R}_J - \mathbf{R}_k|} - \sum_{k=1}^{N_{MM}} \int q_k \frac{\rho(\mathbf{r})}{|\mathbf{r} - \mathbf{R}_k|} d\mathbf{r} + \sum_{J=1}^{N_{QM}} \sum_{k=1}^{N_{MM}} \varepsilon_{Jk} \left[\left(\frac{\sigma_{Jk}}{|\mathbf{R}_J - \mathbf{R}_k|} \right)^{12} - \left(\frac{\sigma_{Jk}}{|\mathbf{R}_J - \mathbf{R}_k|} \right)^6 \right], \quad (2)$$

where the first term is the Coulomb interaction of the QM nuclei of charge Z_J with the MM fixed point charges q_k, the second term is the Coulomb interaction of the QM electron density $\rho(\mathbf{r})$ with the MM point charges, and the final term is the Lennard-Jones interaction of the QM atoms with the MM atoms to account for vdW dispersion forces, where ε and σ are the parameters describing the interaction energy and interatomic distance for each pair of atoms (derived from atomic parameters using the usual Lennard-Jones combination rules).

TₑᵣₐCₕₑₘ computes the first two terms in which the external potential of the Hartree-Fock / Kohn-Sham Hamiltonian is augmented by the external potential generated by the embedding MM point charges as provided by the ₛₐₙᵈₑᵣ MD engine of Aₘᵦₑᵣ. The energy of the QM system interacting with these point charges is thus solved in self-consistently and the MM point charges provide an electrostatic environment that polarizes the QM region. The Lennard-Jones contribution to the interaction energy is calculated by ₛₐₙᵈₑᵣ together with the non-bonded terms of the MM region. TₑᵣₐCₕₑₘ calculates the gradients with respect to

atomic positions of both the QM and MM regions that result from E_{QM} and the first two terms of Eq. (2), while SANDER computes the remaining gradients. SANDER then uses the combined gradients to propagate the dynamics of the total system.

Two different communication methods are available for data exchange between AMBER and TERACHEM in the interface for QM/MM MD simulations[25]: File based and MPI-2[26] client/ server based. In both cases, AMBER drives the computation with all settings for the QM/MM simulation being defined in the AMBER input file. The data that needs to be exchanged during each MD step consists of the coordinates and atom types of the QM region and the coordinates and point charges of the MM region (sent from SANDER to TERACHEM), and the QM contribution to the energy and the forces as discussed above (received by SANDER from TERACHEM). The file based data exchange is implemented as a Fortran90 module of the extensible AMBER interface[25] for QM/MM simulations with external QM programs that was recently released with AMBER 12. In each MD step, SANDER writes an input file for TERACHEM, executes TERACHEM via a system call and then collects the energy and forces from the TERACHEM output file.

The MPI approach to data exchange has several advantages over communication through files and system calls. It subverts the initialization/shutdown of TERACHEM that is otherwise required during each MD step which, due to the GPU initialization, can take a substantial percentage of the computation time in QM/MM dynamics with small QM regions (fewer than ~50 atoms). In the MPI-based approach, TERACHEM is initialized once and then continues to receive atom and point charge coordinates and pass forces until given an exit signal by AMBER. Not only does this avoid the repeated initialization of the GPUs, it also ensures full accuracy in the exchanged forces and energies (as opposed to the limited accuracy normally available when formatted input and output files are used). Furthermore, it opens new possibilities for enhancements to the way QM/MM interactions are handled with *ab initio* QM methods. Since data exchange is possible while both the QM and MM programs are running, advanced interaction models become feasible that require communication during each step of the wave function or electron density optimization in the SCF procedure. For example, this could be useful in generalized Born (GB) solvation models or the treatment of long-range QM/MM electrostatics under periodic boundary conditions via QM compatible particle mesh Ewald (PME) approaches.[24] The MPI-2 interface is easy to maintain, robust to changes in the formats of input and output files, and can easily be extended to support other software packages. Further details of the application programming interface are included in the Supporting Information.

Performance and Accuracy

Timings (number of ps/day for the dynamics) comparing the file based and MPI based communication for PYP QM/MM dynamics with QM regions 1 and 3 are given in Table 1 for 1, 2, and 4 GPUs. We also compare results for generating the initial guess Fock matrix from scratch using atomic densities at each MD step and for starting the SCF process from the converged wave function from the previous step. Reuse of the wave function yields substantial computational savings, but can introduce a systematic bias that results in an energy drift within NVE dynamics.[84] For many applications this small drift can be tolerated, in particular if a thermostat is employed.

Table 1 shows that for the smaller QM region 1 in which GPU initialization takes up a larger percentage of the calculation time, MPI communication speeds up the dynamics by 30–50%. Without MPI communication for QM region 1, running the QM calculation across multiple GPUs no longer accelerates the calculation due to the latency in sequentially initializing each GPU. The time savings with MPI communication are less dramatic for the larger QM

region 3. For larger QM regions, the GPU acceleration outweighs the slow down due to initialization time. The highest computational throughput is achieved with MPI communication and re-using the wavefunction from the previous timestep. This enables 5–6 ps/day for QM region 1 with 139 basis functions and 1–2 ps/day for QM region 3 with 541 basis functions. The last row of Table 1 also includes the timings for dynamics without the cost of computing the one-electron interaction between the QM electron density and the MM point charges. The small increase in time shows that the inclusion of MM point charges in the QM Hamiltonian within electronic embedding does not have a large affect on the performance of the MD simulations.

The accuracy of the interface between TeraChem and Amber was tested by considering the energy conservation during microcanonical (NVE) QM/MM dynamics. The total energy along a 10 ps QM/MM trajectory of PYP using B3LYP/6-31G in the QM region showed excellent energy conservation, comparable to that seen in classical MD simulations, with a very small energy drift of -1.15×10^{-3} kT/ns/degree of freedom (Figures in Supporting Information). Standard settings of Amber and TeraChem were used for this test and the wave function from previous steps was not reused.

Maximizing the number of statistically independent snapshots

Even with GPU acceleration, QM/MM dynamics remains much slower than MM dynamics. Thus, it is not possible to run MD simulations on the ns time scale as is typically done with MM force fields. Because of the smaller sampling time in QM/MM dynamics, it may be difficult to obtain enough statistically independent configurations to compute a converged spectrum.

We will compute the absorption spectrum by sampling configurations from ground state classical dynamics and histogramming the excitation energies. This does not take into account any of the quantum mechanical character of the nuclei. An alternative semiclassical model would sample configurations from the appropriate Wigner distribution.[85–87] This would include quantum effects of the nuclei, but can be challenging to carry out except within a harmonic approximation (which is questionable in solution). Mixed Wigner/classical sampling methods have been devised,[17] as well as methods to construct a Wigner distribution from information limited to the equilibrium density, which can be obtained from path integral or semiclassical techniques.[88–90] However, in the present work, we limit our treatment to sampling of the ground state equilibrium density through classical dynamics and thus quantum mechanical effects of the nuclei associated with the equilibrium density will be absent. In this work we obtain MM and QM(/MM) geometric configurations from running a single NVT MD trajectory for each environment.

Herein, the excitation energy of the bright state (the state with the largest oscillator strength) is used as a measure of the correlation of the N configurations sampled. The autocorrelation time τ of the transition energy b_n based on the normalized fluctuation autocorrelation function C_t is

$$C_t = \frac{\langle b_n b_{n+t} \rangle - \langle b_n \rangle^2}{\langle b_n^2 \rangle - \langle b_n \rangle^2}, \quad (3)$$

$$\tau = \sum_{t=1}^{N-1} \left(1 - \frac{t}{N} \right) C_t. \quad (4)$$

The statistical inefficiency g of this sample is given by[91]

$$g = 1 + 2\tau. \quad (5)$$

In this work, the statistical inefficiency measure g is used in two ways: 1) to determine how frequently statistically independent excitation energies can be computed from N snapshots, and 2) to determine the effective number of uncorrelated configurations N_{eff} in our sample of N snapshots, where $N_{eff} = N / g$.

The statistical inefficiency can be affected by how the dynamics are thermostatted: slow or weak coupling with the thermostat will lead to less perturbed dynamics, but it also yields snapshots that are more correlated than with strong or random coupling. The Berendsen weak coupling thermostat[76] causes little perturbation of collective motions, in particular if the time constant τ_T for coupling to the heat bath is chosen to be large such that the dynamics approach the microcanonical ensemble. Langevin stochastic dynamics[75] may lead to greater independence of configurations and more efficient sampling.

To determine how these two thermostats affect the computed absorption spectra, 500 ps of gas phase NVT MM dynamics of pCT$^-$ were performed with three different thermostat settings: 1) Langevin dynamics with a relatively strong collision frequency $\gamma = 5.0$ ps^{-1}, 2) Berendsen with moderate coupling using a time constant $\tau_T = 1.0$ ps, and 3) Berendsen with weak coupling using a time constant $\tau_T = 10.0$ ps. The excited states from configurations sampled at every 0.25 ps (2000 snapshots for each absorption spectrum) were then computed. The resulting ωPBE/6-31G* spectra for the three sets of dynamics are nearly identical. The computed spectra are also nearly identical for settings 1) and 3) for pCT$^-$ in solution and for the PYP protein (spectra available in Supporting Information). This indicates that the stronger perturbation of the Langevin thermostat does not disturb the computed absorption spectrum.

While the absorption spectra computed from MM dynamics using the Langevin and Berendsen thermostats are very similar, the statistical inefficiencies computed for the bright transition time series in vacuum are quite different: $g = 1.0$ for Langevin, $g = 1.4$ for Berendsen dynamics with tautp= 1.0 ps, $g = 1.8$ for tautp = 10.0 ps. This increase in inefficiency is due to the weak global coupling of the Berendsen thermostat, which results in more correlated snapshots. With a statistical inefficiency larger than $g = 1.0$, the N configurations within the timeseries are not statistically independent. Cutting the sampling time in half for Langevin dynamics (sampling every 0.125 ps) leads to a statistical inefficiency of $g = 1.7$. This shows that, for Langevin dynamics, sampling every 0.25 ps leads to nearly the maximum number of independent snapshots for vacuum simulations.

Unlike in a general bath model of a solution or a protein environment, the explicit interactions between the chromophore and water molecules / protein residues are included in the MM and QM/MM dynamics. In solution, the chromophore phenolate forms hydrogen bonds with nearby waters. In the protein, the chromophore phenolate forms hydrogen bonds with GLH-46 and TYR-42. Hydrogen bonding with the carboxyl group of the chromphore is also possible, and will affect the electronic structure of the chromophore. These specific interactions with the local environment lead to an increase in the correlation time of the excitation energy compared to vacuum.

The computed statistical inefficiency for the bright state time-series when sampling every 0.25 ps for 500 ps in aqueous solution with Langevin dynamics is $g = 2.1$ and with the Berendsen thermostat (tautp=10.0 ps) it is $g = 5.1$. Thus, there are twice as many independent configurations when using the Langevin thermostat. The solution phase MD therefore uses Langevin dynamics, with the sampling frequency is decreased to every 0.5 ps.

For the chromophore in the protein, the 0.25 ps sampling from MM dynamics is highly correlated: with Langevin dynamics over 500 ps giving $g = 20.6$. The Berendsen thermostat gives a similar, but somewhat smaller, statistical inefficiency of $g = 17.5$. For the protein MM MD in this work, the Berendsen thermostat is used, sampling every 10 ps.

The high correlation of configurations in the protein MD leaves very few independent snapshots when sampling from more expensive QM/MM dynamics. With so few statistically independent snapshots, a reliable computed absorption spectrum is impossible. For the protein QM/MM MD in this work, Langevin dynamics are performed, sampling configurations every 1 ps, allowing them to retain some dependence. With these correlated configurations, the total number of snapshots N, and the number of effective snapshots $N_{eff} = N / g$ are reported.

Absorption spectra of the PYP chromophore

In the next sections, the computed absorption spectra of the PYP chromophore in three different environments are presented: in vacuum, in aqueous solution, and in the protein. In each case the spectra computed based on snapshots from MM and QM/MM MD are compared. For QM vacuum and QM/MM solution dynamics, the pCT$^-$ chromophore is treated as QM. For QM/MM protein dynamics, there are three increasingly larger QM regions 1–3, as detailed previously.

Absorption spectrum of the PYP chromophore in vacuum

The computed average spectrum of pCT$^-$ from 2000 snapshots of 500 ps of vacuum MM Langevin dynamics, sampling every 0.25 ps, is shown in red in Figure 4. Computed spectra for 25 and 100 ps of sampling, along with the computed spectra from three individual snapshots are available in the Supporting Information. Comparison of the spectra generated from 25–100ps of MM dynamics (Figure S5 in the Supporting Information) shows that λ_{max} is already converged to within 0.05eV with 25ps of dynamics, suggesting that the statistical errors are much smaller than the errors inherent in the MM force field and the TDDFT method.

The bright state excitation energies for snapshots chosen from 500 ps of MM dynamics ranged from 2.97 eV to 3.63 eV, with an average value of 3.30 eV, and a standard deviation of 0.10 eV. Dihedrals between the phenolate and double bond, the sulfur and the phenyl, and C-O and C-C bond lengths in the phenolate were examined for correlation with bright state excitation energy, but no obvious pattern emerged. The variation in λ_{max} is therefore likely from small geometrical changes due to nonspecific vibrational motion of the structure.

After discarding the first 25 ps of QM dynamics for MM $\rightarrow$ QM equilibration, the average vacuum spectrum from the remaining 125 ps (500 snapshots) of QM B3LYP/6-31G Langevin dynamics of pCT$^-$ was computed, and is shown in Figure 4 in blue. Switching from MM to QM dynamics gives a λ_{max} of 3.51 +/− 0.01 eV, which leads to an MM $\rightarrow$ QM blue shift of 0.23 eV (3.28 eV $\rightarrow$ 3.51 eV). The absorption peak is also significantly narrower with QM dynamics compared to MM dynamics. The bright state excitation energies from the snapshots used to generate this spectrum ranged from 3.27 eV to 3.65 eV, with an average of 3.50 eV and a standard deviation of 0.06 eV.

Both this MM $\rightarrow$ QM blue shift and the wider MM peak can be rationalized by examining the energy scatter of MM and QM configurations along the MM $\rightarrow$ QM minimum energy path (MEP). The ωPBE/6-31G* minimum energy path (MEP) geometries connecting the QM B3LYP/6-31G S_0 minimized structure and the MM GAFF S_0 minimized structure is shown in Figure 5 in which the x-axis quantifies the sum of the absolute value of the

deviation in bond lengths from the B3LYP/6-31G minimum. The S_0 potential (black) rises more steeply than the TD-ωPBE/6-31G* bright S_1 potential (gray) along this path. This leads to a 0.28 eV blue shift in λ_{max} when going from the MM → QM minimum, in close accord with the 0.23 eV blue shift from MM and QM dynamics. Even this small difference can be accounted for by the difference in the average ωPBE/6-31G* S_0 potential energies of the two ensembles, which is 0.44 eV, while a 0.50 eV difference separates the two minima. Thus the higher energy λ_{max} from sampling from QM configurations can be traced back to stabilization of the ground state surface with QM relative to MM. With the MM dynamics propagating on a higher energy region of the ωPBE/6-31G* S_0 PES, the potential is also steeper, leading to a larger energy scatter. The standard deviation of the ωPBE/6-31G* S_0 potential energies from MM dynamics is 0.24 eV, and from QM dynamics is 0.18 eV. This broader range of S_0 energies leads to a broader absorption peak from the MM dynamics.

As changes in the energetics and geometries around the sulfur atom could be significant in going from 6-31G to 6-31G* ground state dynamics, additional QM dynamics with the larger 6-31G* basis set which includes *d*-functions on the non-hydrogen atoms was performed. The computed spectra are quite similar for QM B3LYP/ 6-31G and 6-31G* ground state dynamics (spectrum in Figure S6 of Supporting Information), suggesting that the addition of polarizing *d*-functions on the heavy atoms has a small effect on the sampled configuration space.

What are the possible sources of error in the computed absorption spectrum? The error due to limited sampling of the MD trajectory can be estimated via the 2σ values reported with λ_{max}. One clear possible source of error is that the 6-31G* basis set is not adequate for treating anionic electronic transitions in vacuum, in which the excited state may be quite diffuse. Calculations of the bright state of pCA$^-$ by the Krylov group[48] show that the addition of a diffuse function gives a significant red shift: the EOM-EE-CCSD/6-31G** excitation energy is 3.35 eV, the EOM-EE-CCSD/6-31+G** value is 3.19 eV. To check the effects of adding a diffuse function to the heavy atoms, the excited states of pCT$^-$ were recomputed at the TD-ωPBE/6-31+G* level for the MM Langevin dynamics. The resulting absorption spectrum (Figure S7 in Supporting Information) red shifts λ_{max} by 0.13 eV (3.28 eV → 3.15 eV). Further checks of the basis set effect for 13 configurations were made and it was found that including diffuse functions on hydrogen with 6-31++G* decreased the bright state excitation energy by at most an additional 0.006 eV, and on average 0.004 eV, while going from double zeta to triple zeta quality with 6-311+G* also decreased the excitation energy, with an average decrease of 0.011 eV. Thus, a further red shift of λ_{max} of 0.015 eV, or a total red shift of ~0.15 eV can be expected from improving the 6-31G* basis set.

Another source of error in the computed spectra is the ωPBE density functional. While range-correction provides a better treatment of charge-transfer transitions, it does not help with the lack of double and higher excitations, or in the inclusion of multi-reference character. Rocha-Rinza, et al. also obtained an overly blue shifted excitation energy for the PYP chromophore in vacuum:[52] their computed excitation energy of 3.40 eV for pCA$^-$ with the range-corrected functional CAM-B3LYP/aug-cc-pVDZ is blue shifted 0.52 eV from their experimentally measured λ_{max} value of 2.88 eV. Changing to the RI-CC2 method decreased their computed excitation energy to 3.10 eV, and adding multi-reference character with MRMP2 decreased the value even further to 2.79 eV, bringing the theoretically predicted value much more in accord with experiment. One might also therefore expect a very large red shift of ~0.6 eV in λ_{max} for pCT$^-$ in going from ωPBE to a multi-reference method with dynamical correlation for double and triple excitations. When combined with the ~0.1 eV red shift predicted when increasing the quality of the basis set, the total red shift of ~0.7 eV puts the computed TD-ωPBE/6-31G* 3.5 eV λ_{max} from QM dynamics in good agreement with the experimental λ_{max} for pCT$^-$ of 2.7 eV.[42]

Absorption spectrum of the PYP chromophore in water

An aqueous environment will stabilize both the ground and excited states of the anionic chromophore. The bright state transition moves charge from the oxygen phenolate into the center of the molecule, thus the excited state dipole moment is smaller than the ground state. The ground state is therefore more strongly stabilized by the polar solvent, resulting in a blue shift of λ_{max} when going from vacuum to aqueous solvent. Using Nielsen's experimental absorption energy for pCT$^-$ in solution of 3.14 eV,[42] the solvatochromic blue shift in going from vacuum to solution is 0.44 eV. The hydrogen bonding of water to the phenolate, along with the general friction provided by solvent, leads to greater correlation between snapshots and a larger statistical inefficiency than in vacuum as previously discussed. For the computed spectra in this section, configurations from MM and QM/MM Langevin dynamics were used, decreasing the sampling frequency to every 0.5 ps.

The spectra computed from 500 ps (1000 snapshots) of MM (red) and 125 ps (250 snapshots) of QM/MM (blue) dynamics of pCT$^-$ solvated with SPC/Fw waters are shown in Figure 6. Similar to the vacuum MM $\rightarrow$ QM results, the solution QM/MM dynamics blue shifts λ_{max} with respect to the MM value, in this case by 0.14 eV (3.43 +/− 0.02 eV $\rightarrow$ 3.57 +/− 0.02 eV). Neither MM nor QM/MM dynamics with solvation of the chromophore by MM waters yields the very broad absorption peak observed experimentally. Compared to the experimental blue shift of 0.44 eV, the computed vacuum to water blue shift is too small in both cases: 0.15 eV for MM dynamics, and 0.06 eV for QM(/MM) dynamics.

To test the error due to the 6-31G* basis set, the excitation energies from MM Langevin dynamics were recomputed with 6-31+G* (Figure S8 in Supporting Information). The red shift in going from 6-31G* to 6-31+G* is 0.13 eV (3.41 eV $\rightarrow$ 3.28 eV), the same basis set shift obtained in vacuum (Figure S7 in Supporting Information). Checking 10 configurations with the larger 6-31++G* and 6-311+G* basis sets gave an average additional shift to the bright state of only 0.0008 eV and 0.008 eV, respectively. Thus, there is not a large difference in basis set effects for vacuum and solution phase pCT$^-$.

Since basis set effects cannot explain the unexpectedly small solvatochromic blue shift in the calculations (compared to the experimental blue shift of 0.44eV), remaining possibilities to explain the discrepancy include 1) inaccuracies of the QM/MM solvation (mutual polarization/charge transfer needs to be taken into account, possibly through a larger QM region), 2) the experimental λ_{max} in vacuum is incorrect due to complications in the electrospray ionization absorption measurement, or 3) errors in the ωPBE density functional are quite different for the vacuum and solution environments of pCT$^-$. This last hypothesis could be tested with a high-level post-SCF QM/MM method. It is reasonable to expect that solvation of the anionic chromophore could significantly alter the electronic structure from vacuum, but it is difficult to test for a change in contributions from double and triple excitations, or an increase/decrease in multi-reference character without a multi-reference CI method with QM/MM capabilities. These calculations are beyond the scope of this work, but a preliminary test comparing EOM-CCSD[92,93] and TD-ωPBE calculations with a small STO-3G basis set for two configurations was performed using the Q-Chem 3.2 software program.[94] These calculations showed that the inclusion of the water point charge environment decreased the difference in excitation energies for these two methods by ~0.2 eV, suggesting that the error in the TDDFT excitation energies computed with the ωPBE functional is indeed less for aqueous environments, and that one might not expect such a large shift as in vacuum when going from TD-ωPBE to a higher level of theory. In the next section, the errors in the solvation QM/MM treatment are explored by converting the nearest waters to the chromophore from MM to QM.

Solvation of pCT⁻ by QM waters

In addition to the previously discussed sources of error, the solution phase absorption spectrum includes error in the treatment of solvation of the QM chromophore by the MM SPC/Fw waters. Previous work has shown that surrounding a QM region with MM point charges can over-polarize the QM electron density, leading, for example, to over-binding of the MM and QM regions and overly negative QM/MM free energies of hydration.[13,95–97] In addition, polarization of the environment and charge transfer between the solute and solvent is completely absent but has been shown to be important even for purely ground electronic state properties.[98–100] While testing the effects of changing the waters from MM to a QM treatment for the ground state dynamics is beyond the scope of the current work, it is simple to convert the nearest waters around the QM chromophore to a QM representation to test the effect on the computed absorption spectrum of going from solvation with MM waters to solvation with a QM water shell surrounded by MM waters.

For QM/MM dynamics, the nearest waters to the pCT⁻ chromophore were converted from an MM to QM treatment in the excited state computation. The remaining sphere of waters was still treated as MM fixed point charges. As the number of QM waters is increased, the computed absorption spectrum becomes red shifted and broader (see Figure 7, and additional spectra for MM dynamics in Supporting Information). Additionally, a high-energy tail appears, in qualitative agreement with the experimental solution spectra (cf. Figure 6). Inclusion of 40 QM waters seems to be approaching convergence of the spectrum, although there are still noticeable differences when the number of QM waters is further increased to 200.

The bright state was the first excited state in more than 90% of the excited state computations that included 200 waters, indicating that spuriously low-energy water → chromophore charge-transfer states do not generally lie below the bright state transition. To double check that the TDDFT shift in λ_{max} was not due to problems with charge-transfer states, configuration interaction singles (CIS) calculations were performed. The CIS red shift in going from solvation with all MM waters to solvation with a 40 QM water shell is only 0.08 eV, much smaller than the ωPBE red shift of 0.21 eV. Being unable to benchmark the effects of many QM waters with high-level excited state methods, it is difficult to know which of range-corrected DFT or non-correlated CIS is more trustworthy.

The red shift upon solvation with a QM water shell signifies that solvation of the QM charge-separated anionic phenolate ground state by fixed point charge MM waters does indeed over-polarize the ground state, providing excess stabilization of the anion. As the switch from MM to QM waters has a large effect on the electronic absorption spectrum, the electronic structure of the ground state will differ when solvated by MM and QM waters. It is thus likely that including the nearest waters as QM in the ground state dynamics could sample significantly different configurations than with MM or QM/MM dynamics with only MM waters. It is possible that sampling from dynamics with QM waters could further broaden the absorption spectrum, bringing it closer to the experimental width. With polarization playing such a substantial role in the computed solution absorption spectrum, it is also possible that the TDDFT method is not able to fully capture this large effect, and a method that includes higher order excitations is necessary to capture to full broadening of the spectrum in aqueous solution.

A full study of how basis set size affects convergence of the computed absorption spectrum with QM region size is beyond the scope of this work. However, examination of the behavior for a few snapshots suggests that the trends in bright state excitation energy are quite similar across a range of basis sets. Results of bright state excitation energy vs number

of QM waters for one representative snapshot with the 6-31G*, 6-311G*, 6-31+G*, and 6-311+G* basis sets are shown in Figure 8. The addition of a diffuse function to the non-hydrogen atoms decreases the value of the excitation energy more than increasing the basis set size from double-zeta to triple-zeta. Minimal changes are computed for going from 6-31+G* to 6-311+G*.

Solvation of pCT⁻ with MM waters gives a small solvatochromic blue shift in going from vacuum to water environments for pCT⁻, but the red shift upon solvation with a QM water shell decreases the MM dynamics solvatochromic blue shift to only 0.01 eV (3.28 +/− 0.01 eV in vacuum → 3.29 +/− 0.02 eV with 40 QM waters), and transforms the QM/MM dynamics solvatochromic shift into a red shift of 0.15 eV (3.51 +/− 0.01 eV in vacuum → 3.36 +/− 0.05 eV with 40 QM waters). These values are summarized in Table 2. While the solvatochromic shift in λ_{max} is worse with the QM water shell compared to the experimental shift, the agreement of TD-ωPBE/6-31G* with the solution phase λ_{max} is better than the vacuum λ_{max} from the experimental value (blue shifted by 0.2 eV instead of 0.8 eV), suggesting that TD-ωPBE/6-31G* is better able to capture the electronic structure of the solvated anionic chromophore compared to the more diffuse electron density of the anionic chromophore in vacuum.

To summarize, many surrounding waters (40–200) must be included in the excited state TDDFT computation of pCT⁻ to achieve converged excitation energies. Treating the waters as QM additionally appears to broaden the computed absorption spectrum, bringing it closer in line with the experimental spectrum.

Absorption spectrum of PYP: MM dynamics, QM/MM excited states

In this section results are presented for computation of the spectrum of photoactive yellow protein in solution, based on 3 ns of Berendsen (tautp=10.0 ps) MM dynamics, sampling every 10 ps (300 snapshots). The computed QM region 1 spectra from 500 ps of Langevin MM dynamics (2000 snapshots) and 3 ns of Berendsen MM dynamics (300 snapshots) are identical within the computed error (see Figure S4 in Supporting Information). The 300 configurations are identical for the spectra computed in this section; only the QM and MM regions are altered.

To examine the effect of the dynamic electronic protein environment around the chromophore, the computed spectra for QM region 1 is first compared with and without the protein point charges (Figure 9). The electrostatic influence of the protein leads to a small blue shift in λ_{max}: from λ_{max} = 3.32 +/− 0.05 eV without point charges to λ_{max} = 3.38 +/− 0.02 eV including the electrostatic influence of the protein environment. This small shift suggests that the *electrostatics* of the hydrophobic pocket of the protein do not strongly alter the electronic environment of the chromophore.

The surrounding environment is next gradually switched from an MM to QM treatment, with increasingly larger QM regions. Changing nearby residues from MM to QM goes *beyond* electrostatics to include exchange, charge-transfer, and polarization effects, which are known from the previous section to be substantial in solution. The computed absorption spectra for QM regions 1–5 is shown in Figure 10. The λ_{max} values are given in Tables 2 and 3. Conversion of the H-bonding residues GLH-46 and TYR-42 from MM to QM in going from QM region 1 to 2 gives a very small blue shift in λ_{max}. This does not necessarily indicate that the H-bonding residues play a small role in excitation energy. Their effect may have been primarily accounted for as MM point charges, or their inclusion as QM electron densities may have cancelling effects with movement of the QM/MM boundary region. The doubling of the QM region size, including conversion of cationic ARG-52 from MM to QM, in going from QM region 2 to 3 shows very little change in the absorption spectrum, and one

might mistakenly think that convergence has been achieved with respect to QM region size with QM region 3. However, the larger QM region 4, which includes some aromatic residues that might participate in π-stacking interactions with the chromophore, shifts λ_{max} to higher energy by almost 0.1eV.

Surprisingly, the largest spectral change does not occur with conversion of the nearest residues from MM to QM, but in going from QM region 4 to 5: λ_{max} is strongly red shifted by 0.3 eV. QM region 5 no longer includes capped side chains of the nearby residues, but instead includes entire stretches of the protein in the area surrounding the chromophore. The size of the QM region is greatly increased in going from the capped side-chains of twelve residues to forty-two entire residues (104 atoms to 723 atoms). Perhaps just as important as increasing the QM region size is decreasing the number of QM and MM boundary regions that may have electrostatic artifacts at the interface of the two methods. With the chromophore plus eleven side chains, QM region 4 contains twelve regions of QM and MM interface, while QM region 5 has only four such interface regions.

The change in the bright state excitation energy in eV as a function of QM region for the first 1 ns (100 snapshots) of MM dynamics is shown in Figure 11, with the average value displayed as red circles. The excitation energies are centered at the value computed for QM region 1, which ranges from 3.06 to 3.56 eV for the 100 configurations sampled. By increasing the size of the QM system from region 1 to 2, the excitation energies undergo small increases and decreases (+/– 0.1 eV) of equal amounts, yielding no change in the average value. Consistent with the absorption spectrum in Figure 10, the trend when increasing in size from region 3 to 4 is an increase in excitation energy for all snapshots, while the trend is decreasing for all snapshots in increasing the QM region from 4 to 5. These results were checked with CIS for three snapshots, and the general trend was the same, indicating that this behavior is not an artifact of DFT or TDDFT.

To investigate this large change in excitation energy, one representative snapshot was chosen for further study. This snapshot is highlighted with dark blue triangles in Figure 11. Visual inspection of the bright excited-ground state density difference for QM regions 1, 4, and 5 (Figure S11 in Supporting Information) showed fairly localized $\pi \rightarrow \pi^*$ transitions, with no substantial change in going from QM region 4 to 5, suggesting that large changes in delocalization are not responsible for the red shifted excitation energy.

Next, the effects of intermediate QM regions on the calculated bright state excitation energy were examined. Inclusion of complete residues 66–70 (VAL-66, ALA-67, PRO-68, CYC-69, and THR-70) along with the remaining side chains from QM region 4 (giving 188 atoms total) decreases the number of QM and MM interface regions from twelve to nine. This new QM region yielded a bright state excitation energy intermediate between QM regions 4 and 5. This is shown at QM region 4.5 in Figure 11. The QM region was then chosen as only residues 66–70 (84 atoms total), without the side chains of the hydrogen bonding residues GLH-46 and TYR-42 or cationic ARG-52 (resulting in a charge of –1 for this new QM region), creating a QM system with only two QM and MM interfaces. This QM region also shows a red shift in excitation energy, but not nearly as large as that from the previous test, indicating that the QM treatment of the side chain residues does contribute to red shift. Although this QM region does not include the residues from QM region 2, it is represented as QM region 2.75 in Figure 11 since the number of atoms in the QM system is just below that of QM region 3. These tests show that extending the region covalently bonded to the chromophore to allow polarization/charge transfer through covalent bonds and removing artifacts at the QM and MM interface is necessary to achieve a converged excitation energy.

To ensure that the excitation energy at QM region 5 was not an artifact of that particular choice of QM region, the bright state energy treating the entire protein and counter ions with QM was also computed for three configurations (open triangles in Figure 11). There was virtually no change in excitation energy. This result is shown as QM region 6 in the plot for the selected snapshots. Taken together, these results suggest that it is very difficult to determine when a QM region is `converged,' and very large QM regions with few QM to MM interfaces may be required to achieve stable results.

To test the effects of the inclusion of diffuse functions on the computed absorption spectra, the spectra of PYP for the first 1 ns (100 snapshots) of MM dynamics with 6-31+G* was computed for QM regions 1–3. The addition of diffuse functions to the basis set red shifts λ_{max} ~0.1 eV (Figure S9 in Supporting Information), similar to the basis set effects in vacuum and solution. Comparing the bright state results computed with 6-311+G* for six configurations gave a small additional red shift of at most 0.009 eV.

Based on the λ_{max} of 3.4 eV from QM regions 1–3 in PYP, going from a protein to an aqueous environment incorrectly red shifts λ_{max} (assuming the QM water shell value); the experimental blue shift is 0.36 eV. However, using the QM region 5 λ_{max} of 3.17 +/− 0.03 eV gives a blue shift of 0.13 eV in closer accord with experiment. The computed shift from vacuum to protein is in the wrong direction, leaving the previous sources of error: differing electronic structure of the chromophore in the protein and vacuum environments leading to inconsistent errors with the ωPBE functional, errors in the experimental gas phase measurement, and errors in the configurations of PYP sampled by the MM dynamics. In addition to problems in the computed value of λ_{max} for PYP, it is also clear that when comparing to the experimental absorption spectrum, which has a shoulder on the blue side of the peak, that the configurations from MM dynamics yield an incorrect red shoulder for the absorption peak, suggesting that the sampled configurations may indeed be a contributing factor to the spectral error. Because of the possibility of both a rotated GLH and non-protonated GLH in solution affecting the experimental spectrum, preliminary calculations were performed to gauge how these changes altered the spectral shape, and no substantial difference was found. It is thus likely either the ground state dynamics or the excited state methodology that leads to the difference between the computed spectrum and experiment.

In summary, it is found that for MM dynamics of the full protein, very large QM regions are necessary to obtain a converged excitation energy. It appears that an extended network of covalently bonded residues should be included in the QM region, not merely nearest neighbor side chains.

Spectra of PYP: *ab initio* QM/MM dynamics, QM/MM excited states

In this section we examine how ground state dynamics performed with the chromophore and a portion of its surroundings treated with QM compares to the MM dynamics described in the previous section. Because of the limited dynamics for QM region 3 QM/MM dynamics, the full 25 ps of dynamics were not discarded to allow for MM → QM equilibration; here only 12.5 ps of dynamics are discarded. After removing snapshots from this equilibration period, QM/MM Langevin dynamics sampling every 1 ps were performed with QM regions 1, 2, and 3 for 193 ps, 77 ps, and 26 ps. Based on the computed statistical inefficiency of the QM region 1 bright state excitation energy from QM region 1 dynamics of $g = 2.7$, the number of effective statistically independent snapshots for QM regions 1, 2, and 3 is $N_{eff} =$ 71, 28, and 9, respectively.

In agreement with MM → QM/MM dynamics for vacuum and solution, the switch to QM treatment of the chromophore during the dynamics blue shifts λ_{max}. For PYP, QM region 1, the MM → QM/MM blue shift is 0.18 eV (3.38 +/ 0.02 eV → 3.55 +/ 0.01 eV), similar to

the vacuum and solution MM → QM/MM blue shift. However, the blue shift in going from MM → QM/MM ground state dynamics does not hold for the larger QM regions in the excited state computation, due to the competing trend to red shift with increasing QM region size.

While there was no clear trend in λ_{max} for the various QM regions in the computed spectra based on MM dynamics of PYP, QM/MM dynamics with all three QM regions show the same pattern: increasing the QM region size in the excited state calculation red shifts λ_{max} (see computed spectra in Figure 12 and λ_{max} values in Table 3). For QM/MM dynamics, converting the H-bonding residues from MM to QM in going from QM region 1 to 2 red-shifts λ_{max}, in the opposite direction of the shift with MM dynamics. This pattern shows that the configurations adopted when the chromophore is treated with a QM Hamiltonian during the dynamics have substantially different chromophore-protein interactions than when the system is treated with an MM Hamiltonian.

The differences in λ_{max} between QM regions 1–3 for QM/MM dynamics are within the sampling error bars, which are quite large with the small sample size from QM/MM dynamics. No certain conclusions in λ_{max} can be made about increasing the QM region for ground state dynamics. However, because a large QM region is clearly important for an accurate excited state computation, the configurations sampled from larger dynamics with larger QM regions may be important for computation of an accurate spectral shape.

Conclusions

A GPU accelerated QM/MM MD program has been developed by interfacing the TeraChem and Amber software packages. This new interface goes beyond existing interfaces in Amber in that it supports data exchange between the MD engine Sander and TeraChem based on the MPI-2 standard as opposed to files and system calls. While this required changes to both the QM and the MM programs, it provides added stability, performance, portability and extensibility. The interface was used in this work to examine some of the key issues related to computing absorption spectra from MM and QM/MM MD simulations. Some of these issues are related to performing MD with a limited sample size, while others are due to the behavior of a QM wave function within an MM environment.

Due to the high computational cost of the QM dynamics, the sample size for spectral computation was much smaller with QM and QM/MM dynamics than with sampling from MM dynamics. Thus, we explored how to maximize the number of effective configurational snapshots. Comparing the stochastic Langevin and Berendsen weak coupling thermostats shows that the thermostats provided extremely similar absorption spectra in the limit of many snapshots, but that Langevin dynamics generally has a smaller correlation time, thus providing more statistical independence of configuration snapshots.

In comparing spectra computed from MM and QM dynamics of the PYP chromophore in vacuum, environmental effects external to the chromophore were removed. The spectral differences are thus solely from the changes in the PES for the two methods. The range of computed ground state and bright excited states was much larger for MM dynamics than QM dynamics, leading to a wider absorption peak. The blue shift of λ_{max} in going from MM to QM dynamics indicates a more accurate PES for the chromophore ground state with QM dynamics. This MM → QM blue shifting was observed for all chromophore environments, and is likely a general trend for many systems, absent any systematic errors in the MM or QM Hamiltonian.

For the PYP chromophore in solution, a 0.21 eV red shift was computed with TD-wPBE/ 6-31G* when converting the nearest 40 waters to the chromophore from MM to QM. This

indicates that QM/MM models with small QM regions (e.g. with QM treatment restricted to a solute or chromophore) may not be able to capture electrostatic, exchange, polarization, and charge-transfer phenomena that are essential in the excitation energy.

When increasing the size of the QM region in computation of the excited states of PYP, the largest change in excitation energy was surprisingly not from converting the closest residues to the chromophore from MM to QM, but from inclusion of the entire long-range protein environment surrounding the chromophore. This very large QM region (QM region 5: 723 atoms) red shifted λ_{max} by 0.3 eV from the λ_{max} computed with the QM treatment of the surrounding residue side chains (QM region 4: 159 atoms). As examination of the ground to excited state density difference for these two QM regions showed very little change, the large shift is likely due to the change in the long-range electrostatic environment of the chromophore. It was verified that the results did not change when the entire protein is included in the QM region. Thus, very large QM regions are required to achieve spectral convergence. It is likely that the inclusion of only residue side chains in the QM region, rather than covalently bonded full residues, will lead to unconverged results. A polarizable MM method may provide better convergence behavior and will be explored in future work.

There is poor agreement with all of the computed spectral solvatochromic shifts of the PYP chromophore compared to experiment. With QM and QM/MM dynamics there is an incorrect red shift in going from vacuum to solution, while experimentally there is a large blue shift. Using $\lambda_{max} = 3.25$ eV for PYP computed for QM region 5 based on QM/MM dynamics, the computed solvatochromic shift in going from protein to solution is not blue shifted enough: ~0.1 eV instead of the experimental value of 0.36 eV. The protein to vacuum shift is in the wrong direction: 0.3 eV blue shifted instead of 0.08 eV red shifted. A number of possible reasons for the poor agreement are:

-- Unequal treatment of errors by the ωPBE functional for the various electronic environments. This could be tested with multi-reference configuration interaction QM/MM calculations. We believe this is the largest source of error.

-- Poor treatment of aqueous solvation by the QM/MM technique. Many surrounding QM waters may be required to accurately treat solvation.

-- Problems with charge-transfer transitions with the ωPBE functional when using a large amount of QM waters. The shift in going from MM waters to QM waters with CIS is 0.13 eV less than with TD-ωPBE.

-- Imprecise measurement of the experimental vacuum λ_{max} due to ionization complications.

-- Inaccurate molecular dynamics for the chromophore-protein interactions. As the excited state calculations show the need for large QM regions for spectral convergence, large QM regions may also be required to sample the correct ground state configurations.

Although future work should further isolate the sources of the errors, the results we present here certainly argue that large QM regions are needed in QM/MM calculations, at variance with standard practice.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

(1). Warshel A, Levitt M. J. Mol. Bio. 1976; 103:227. [PubMed: 985660]

(2). Monard, G. r.; Merz, KM. Acc. Chem. Res. 1999; 32:904.

(3). Bakowies D, Thiel W. J. Phys. Chem. 1996; 100:10580.

(4). Singh UC, Kollman PA. J. Comput. Chem. 1986; 7:718.

(5). Sherwood, P. Hybrid Quantum Mechanics/Molecular Mechanics Approaches. In: Grotendorst, J., editor. Modern Methods and Algorithms of Quantum Chemistry. 2 ed.. Vol. Vol. 3. John von Neumann Institute for Computing; Jülich: 2000. p. 285

(6). Gao J, Truhlar DG. Ann. Rev. Phys. Chem. 2002; 53:467. [PubMed: 11972016]

(7). Lin H, Truhlar D. Theo. Chem. Acc. 2007; 117:185.

(8). Senn HM, Thiel W. Ang. Chem., Int. Ed. 2009; 48:1198.

(9). Zhang, R.; Lev, B.; Cuervo, JE.; Noskov, SY.; Salahub, DR. A Guide to QM/MM Methodology and Applications. In: John, RS.; Erkki, B., editors. Advances in Quantum Chemistry. Vol. Volume 59. Academic Press; 2010. p. 353

(10). Aaqvist J, Warshel A. Chem. Rev. 1993; 93:2523.

(11). Bathelt CM, Zurek J, Mulholland AJ, Harvey JN. J. Am. Chem. Soc. 2005; 127:12900. [PubMed: 16159284]

(12). Rod TH, Rydberg P, Ryde U. J. Chem. Phys. 2006; 124:174503. [PubMed: 16689579]

(13). Beierlein FR, Michel J, Essex JW. J. Phys. Chem. B. 2011; 115:4911. [PubMed: 21476567]

(14). Shim S, Rebentrost P, Valleau S. p. Aspuru-Guzik A. n. Biophys. J. 2012; 102:649. [PubMed: 22325289]

(15). Walker RC, de Souza MM, Mercer IP, Gould IR, Klug DR. J. Phys. Chem. B. 2002; 106:11658.

(16). Cho BM, Walker RC, Amer H, Mercer I, Klug DR, Gould IR. J. Phys. Chem. B. 2005; 109:5954. [PubMed: 16851650]

(17). Virshup AM, Punwong C, Pogorelov TV, Lindquist BA, Ko C, Martínez TJ. J. Phys. Chem. B. 2008; 113:3280. [PubMed: 19090684]

(18). PetaChem, LLC; http://www.petachem.com

(19). Ufimtsev IS, Martinez TJ. J. Chem. Theory Comput. 2008; 4:222.

(20). Ufimtsev IS, Martinez TJ. J. Chem. Theory Comput. 2009; 5:1004.

(21). Ufimtsev IS, Martinez TJ. J. Chem. Theory Comput. 2009; 5:2619.

(22). Isborn CM, Luehr N, Ufimtsev IS, Martinez TJ. J. Chem. Theory Comput. 2011; 7:1814. [PubMed: 21687784]

(23). Case, DA.; Darden, TA.; T.E.; Cheatham, I.; Simmerling, CL.; Wang, J.; Duke, RE.; Luo, R.; Walker, RC.; Zhang, W.; Merz, KM.; Roberts, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Swails, J.; Götz, AW.; Kolossvari, I.; Wong, KF.; Paesani, F.; Vanicek, J.; Wolf, RM.; Liu, J.; Wu, X.; Brozell, SR.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M-J.; Cui, G.; Roe, DR.; Mathews, DH.; Seetin, MG.; Salomon-Ferrer, R.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, PA. AMBER 12. University of California; San Francisco: 2012.

(24). Walker RC, Crowley MF, Case DA. J. Comput. Chem. 2008; 29:1019. [PubMed: 18072177]

(25). Götz AW, Clark MC, Walker RC. 2012 in preparation.

(26). MPI: A Message-passing Interface Standard. Version 2.2. High-Performance Computing Center, University of Stuttgart; Stuttgart: 2009.

(27). Fox SJ, Pittock C, Fox T, Tautermann CS, Malcolm N, Skylaris C-K. J. Chem. Phys. 2011; 135:224107. [PubMed: 22168680]

(28). Meier K, Thiel W, van Gunsteren WF. J. Comput. Chem. 2012; 33:363. [PubMed: 22180225]

(29). Johnson ER, DiLabio GA. J. Mol. Struct. (THEOCHEM). 2009; 898:56.

(30). Flaig D, Beer M, Ochsenfeld C. J. Chem. Theory Comput. 2012; 8:2260.

(31). van der Horst MA, Hellingwerf KJ. Acc. Chem. Res. 2003; 37:13. [PubMed: 14730990]

(32). Sprenger WW, Hoff WD, Armitage JP, Hellingwerf KJ. J. Bacter. 1993; 175:3096.

(33). Baca M, Borgstahl GEO, Boissinot M, Burke PM, Williams DR, Slater KA, Getzoff ED. Biochem. 1994; 33:14369. [PubMed: 7981196]

(34). Hoff WD, Dux P, Hard K, Devreese B, Nugteren-Roodzant IM, Crielaard W, Boelens R, Kaptein R, Beeumen JV, Hellingwerf KJ. Biochem. 1994; 33:13959. [PubMed: 7947803]

(35). Genick UK, Devanathan S, Meyer TE, Canestrelli IL, Williams E, Cusanovich MA, Tollin G, Getzoff ED. Biochem. 1997; 36:8. [PubMed: 8993312]

(36). Groenhof G, Lensink MF, Berendsen HJC, Snijders JG, Mark AE. Proteins: Struct., Funct., Bioinf. 2002; 48:202.

(37). Premvardhan LL, van der Horst MA, Hellingwerf KJ, van Grondelle R. Biophys. J. 2003; 84:3226. [PubMed: 12719252]

(38). Borucki B, Otto H, Meyer TE, Cusanovich MA, Heyn MP. J. Phys. Chem. B. 2004; 109:629. [PubMed: 16851055]

(39). Lee I-R, Lee W, Zewail AH. Proc. Natl. Acad. Sci. 2006; 103:258. [PubMed: 16407155]

(40). Carroll EC, Song S-H, Kumauchi M, van Stokkum IHM, Jailaubekov A, Hoff WD, Larsen DS. J. Phys. Chem. Lett. 2010; 1:2793. [PubMed: 20953237]

(41). Philip AF, Nome RA, Papadantonakis GA, Scherer NF, Hoff WD. Proc. Natl. Acad. Sci. 2010; 107:5821. [PubMed: 20220103]

(42). Nielsen IB, Boyé-Péronne S, El Ghazaly MOA, Kristensen MB, Bröndsted Nielsen S, Andersen LH. Biophys. J. 2005; 89:2597. [PubMed: 16040745]

(43). Bravaya KB, Grigorenko BL, Nemukhin AV, Krylov AI. Acc. Chem. Res. 2011; 45:265. [PubMed: 21882809]

(44). He Z, Martin CH, Birge R, Freed KF. J. Phys. Chem. A. 2000; 104:2939.

(45). Gromov EV, Burghardt I, Köppel H, Cederbaum LS. J. Phys. Chem. A. 2005; 109:4623. [PubMed: 16833801]

(46). Epifanovsky E, Polyakov I, Grigorenko B, Nemukhin A, Krylov AI. J. Chem. Theory Comput. 2009; 5:1895.

(47). Epifanovsky E, Polyakov I, Grigorenko B, Nemukhin A, Krylov AI. J. Chem. Phys. 2010; 132:115104. [PubMed: 20331319]

(48). Zuev D, Bravaya KB, Crawford TD, Lindh R, Krylov AI. J. Chem. Phys. 2011; 134:034310. [PubMed: 21261356]

(49). Gromov EV, Burghardt I, Köppel H, Cederbaum LS. J. Am. Chem. Soc. 2007; 129:6798. [PubMed: 17474743]

(50). Wang Y, Li H. J. Chem. Phys. 2010; 133:034108. [PubMed: 20649309]

(51). Gromov EV, Burghardt I, Hynes JT, Köppel H, Cederbaum LS. J. Photochem. Photobio. A. 2007; 190:241.

(52). Rocha-Rinza T, Christiansen O, Rajput J, Gopalan A, Rahbek DB, Andersen LH, Bochenkova AV, Granovsky AA, Bravaya KB, Nemukhin AV, Christiansen KL, Nielsen M. B. n. J. Phys. Chem. A. 2009; 113:9442. [PubMed: 19642665]

(53). Casida, ME. Recent Advances in Density Functional Methods. World Scientific; Singapore: 1995.

(54). Hirata S, Head-Gordon M. Chem. Phys. Lett. 1999; 314:291.

(55). Hirata S, Head-Gordon M, Bartlett RJ. J. Chem. Phys. 1999; 111:10774.

(56). Grabo T, Petersilka M, Gross EKU. J. Mol. Struct. (THEOCHEM). 2000; 501–502:353.

(57). Dreuw A, Head-Gordon M. Chem. Rev. 2005; 105:4009. [PubMed: 16277369]

(58). Lange A, Herbert JM. J. Chem. Theory Comput. 2007; 3:1680.

(59). Walker RC, Mercer IP, Gould IR, Klug DR. J. Comput. Chem. 2007; 28:478. [PubMed: 17186476]

(60). Rohrdanz MA, Martins KM, Herbert JM. J. Chem. Phys. 2009; 130:054112. [PubMed: 19206963]

(61). Gonzalez EM, Guidoni L, Molteni C. Phys. Chem. Chem. Phys. 2009; 11:4556. [PubMed: 19475175]

(62). Groenhof G, Bouxin-Cademartory M, Hess B, de Visser SP, Berendsen HJC, Olivucci M, Mark AE, Robb MA. J. Am. Chem. Soc. 2004; 126:4228. [PubMed: 15053611]

(63). Groenhof G, Schäfer LV, Boggio-Pasqua M, Grubmüller H, Robb MA. J. Am. Chem. Soc. 2008; 130:3250. [PubMed: 18293978]

(64). Ko C, Virshup AM, Martinez TJ. Chem. Phys. Lett. 2008; 460:272.

(65). Rocha-Rinza T, Sneskov K, Christiansen O, Ryde U, Kongsted J. Phys. Chem. Chem. Phys. 2011; 13:1585. [PubMed: 21132197]

(66). Sergi A, Grüning M, Ferrario M, Buda F. J. Phys. Chem. B. 2001; 105:4386.

(67). Thompson MJ, Bashford D, Noodleman L, Getzoff ED. J. Am. Chem. Soc. 2003; 125:8186. [PubMed: 12837088]

(68). Chiba M, Fedorov DG, Kitaura K. J. Comput. Chem. 2008; 29:2667. [PubMed: 18484637]

(69). Ko C, Levine B, Toniolo A, Manohar L, Olsen S, Werner H-J, Mart nez TJ. J. Am. Chem. Soc. 2003; 125:12710. [PubMed: 14558810]

(70). Getzoff ED, Gutwin KN, Genick UK. Nat. Struct. Biol. 2003; 10:663. [PubMed: 12872160]

(71). Wu YJ, Tepper HL, Voth GA. J. Chem. Phys. 2006; 124

(72). Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. Proteins: Struct., Funct., Bioinf. 2006; 65:712.

(73). Wang JM, Wolf RM, Caldwell JW, Kollman PA, Case DA. J. Comput. Chem. 2004; 25:1157. [PubMed: 15116359]

(74). Gordon JC, Myers JB, Folta T, Shoja V, Heath LS, Onufriev A. Nucleic Acids Res. 2005; 33:W368. [PubMed: 15980491]

(75). Loncharich RJ, Brooks BR, Pastor RW. Biopolymers. 1992; 32:523. [PubMed: 1515543]

(76). Berendsen HJC, Postma JPM, Vangunsteren WF, Dinola A, Haak JR. J. Chem. Phys. 1984; 81:3684.

(77). Becke AD. J. Chem. Phys. 1993; 98:5648.

(78). Stephens PJ, Devlin FJ, Chabalowski CF, Frisch MJ. J. Phys. Chem. 1994; 98:11623.

(79). Tawada Y, Tsuneda T, Yanagisawa S, Yanai T, Hirao K. J. Chem. Phys. 2004; 120:8425. [PubMed: 15267767]

(80). Vydrov OA, Heyd J, Krukau AV, Scuseria GE. J. Chem. Phys. 2006; 125:074106. [PubMed: 16942321]

(81). Vydrov OA, Scuseria GE. J. Chem. Phys. 2006; 125:234109. [PubMed: 17190549]

(82). Jacquemin D, Wathelet V, PerpeÃÄte EA, Adamo C. J. Chem. Theory Comput. 2009; 5:2420.

(83). Starcke JH, Wormit M, Schirmer J, Dreuw A. Chem. Phys. 2006; 329:39.

(84). Niklasson AMN. Phys. Rev. Lett. 2008; 100:123004. [PubMed: 18517861]

(85). Reimers JR, Wilson KR, Heller EJ. J. Chem. Phys. 1983; 79:4749.

(86). Ben-Nun M, Martínez TJ. J. Phys. Chem. A. 1999; 103:10517.

(87). Hillery M, O'Connell RF, Scully MO, Wigner EP. Physics Reports. 1984; 106:121.

(88). Liu J. J. Chem. Phys. 2011; 134:194110. [PubMed: 21599047]

(89). Liu J, Miller WH. J. Chem. Phys. 2011; 134:104102. [PubMed: 21405151]

(90). Liu J, Miller WH. J. Chem. Phys. 2011; 134:104101. [PubMed: 21405150]

(91). Chodera JD, Swope WC, Pitera JW, Seok C, Dill KA. J. Chem. Theory Comput. 2006; 3:26.

(92). Stanton JF, Bartlett RJ. J. Chem. Phys. 1993; 98:7029.

(93). Krylov AI. Ann. Rev. Phys. Chem. 2008; 59:433. [PubMed: 18173379]

(94). Shao Y, Molnar LF, Jung Y, Kussmann J, Ochsenfeld C, Brown ST, Gilbert ATB, Slipchenko LV, Levchenko SV, O'Neill DP, DiStasio RA Jr, Lochan RC, Wang T, Beran GJO, Besley NA, Herbert JM, Yeh Lin C, Van Voorhis T, Hung Chien S, Sodt A, Steele RP, Rassolov VA, Maslen PE, Korambath PP, Adamson RD, Austin B, Baker J, Byrd EFC, Dachsel H, Doerksen RJ, Dreuw A, Dunietz BD, Dutoi AD, Furlani TR, Gwaltney SR, Heyden A, Hirata S, Hsu C-P, Kedziora G, Khalliulin RZ, Klunzinger P, Lee AM, Lee MS, Liang W, Lotan I, Nair N, Peters B, Proynov EI, Pieniazek PA, Min Rhee Y, Ritchie J, Rosta E, David Sherrill C, Simmonett AC, Subotnik JE, Lee Woodcock Iii H, Zhang W, Bell AT, Chakraborty AK, Chipman DM, Keil FJ, Warshel A, Hehre WJ, Schaefer Iii HF, Kong J, Krylov AI, Gill PMW, Head-Gordon M. Phys. Chem. Chem. Phys. 2006; 8:3172. [PubMed: 16902710]

(95). Laio A, VandeVondele J, Rothlisberger U. J. Chem. Phys. 2002; 116:6941.

(96). Das D, Eurenius KP, Billings EM, Sherwood P, Chatfield DC, Hodoscek M, Brooks BR. J. Chem. Phys. 2002; 117:10534.

(97). Senthilkumar K, Mujika JI, Ranaghan KE, Manby FR, Mulholland AJ, Harvey JN. J. Roy. Soc. Interface. 2008; 5:207.

(98). Park K, Götz AW, Walker RC, Paesani F. J. Chem. Theory Comput. 2012 accepted.

(99). van der Vaart A, Merz KM. J. Am. Chem. Soc. 1999; 121:9182.

(100). Ufimtsev IS, Luehr N, Martinez TJ. J. Phys. Chem. Lett. 2011; 2:1789.
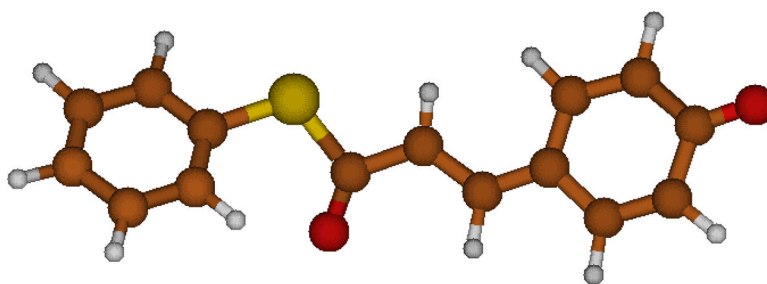
Figure 1.
t*rans*-thiophenyl-*p*-coumarate (pCT⁻). In this model system the cysteine linkage of the p-coumaric acid to the protein is replaced by a thiophenyl group (sulfur is yellow, oxygen is red).
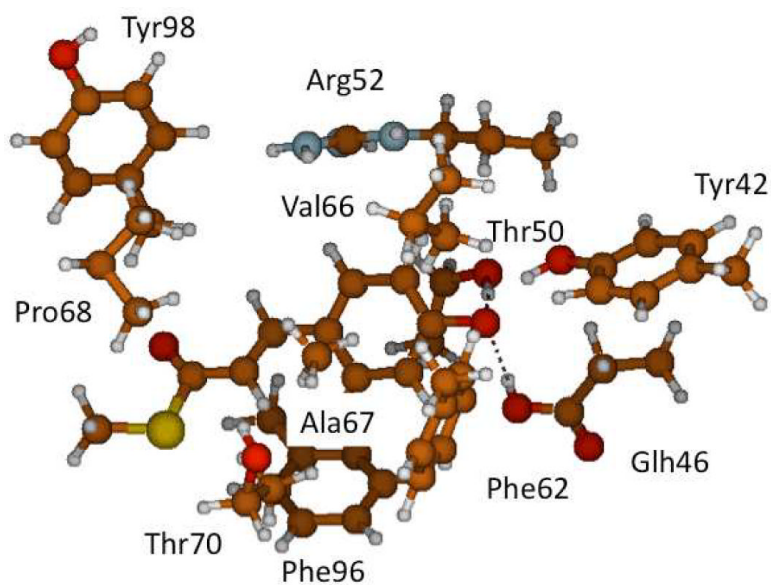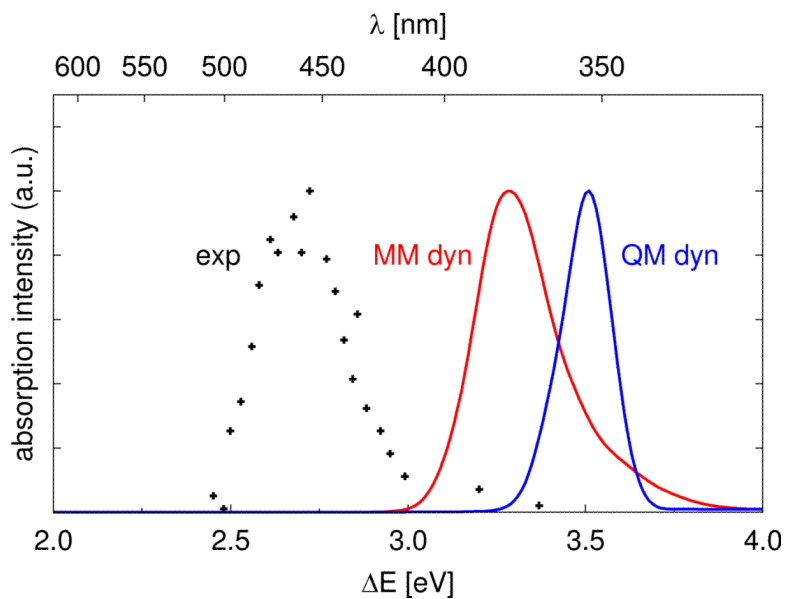
Figure 2.
QM region 4: the PYP chromophore and the side chains of 11 surrounding residues. The MM region (not shown) includes the protein backbone of the 11 residues shown, the remaining protein, and an SPC/Fw water shell. Carbon atoms are orange, oxygen atoms are red, nitrogen atoms are blue, and sulfur atoms are yellow.
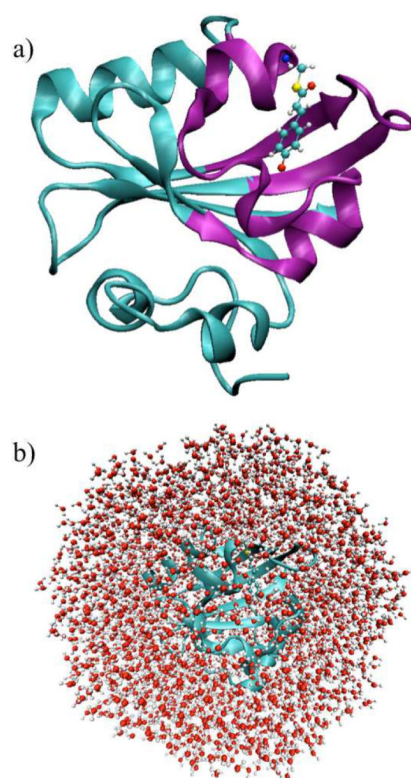
Figure 3.
a) Photoactive yellow protein (PYP) with the coumarate chromophore shown in CPK representation and the protein in ribbon representation. QM region 5 is highlighted in purple. (b) Additionally showing the water sphere (32 Å radius) included in the MM region.

Figure 4.
TD-ωPBE/6-31G* computed spectra for pCT⁻ in vacuum from snapshots of MM (red) and QM (blue) Langevin dynamics, sampling every 0.25 ps. MM dynamics sample over 500 ps (2000 snapshots), QM dynamics over 125 ps (500 snapshots). Experimental results from Nielsen et al.[42] are shown in black for comparison. Absorption intensity is given in arbitrary units (a.u.).
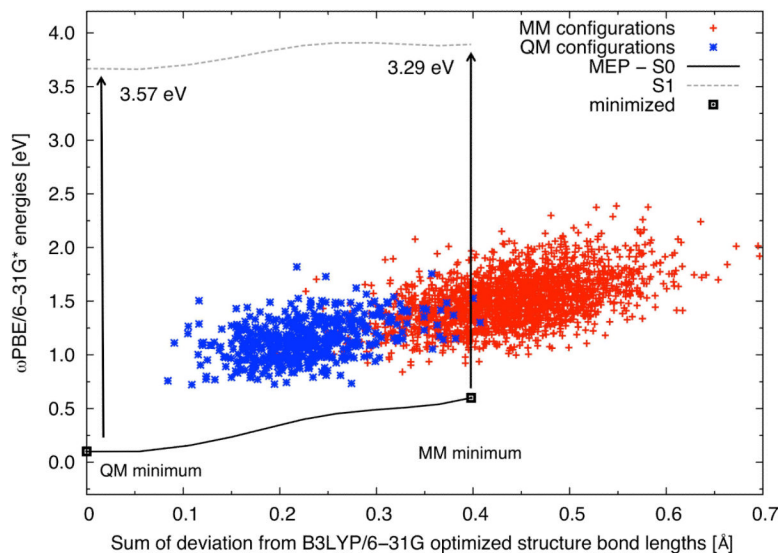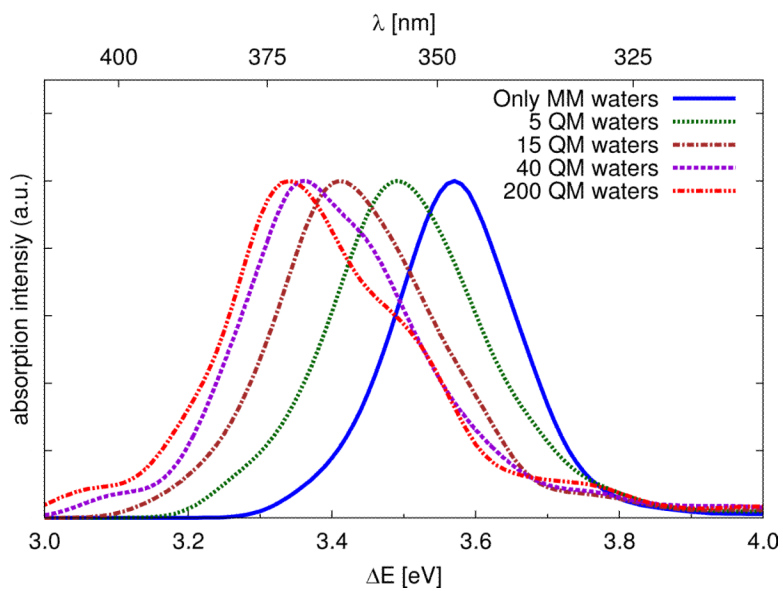
Figure 5.
The ωPBE/6-31G* minimum energy path (MEP) between the QM B3LYP/6-31G minimized S_0 structure and the MM GAFF S_0 minimized structure is shown in black. The TD-wPBE/6-31G* bright state for S_1 along this MEP is shown in gray, with the excitation energies from minimized structures noted with the arrows. The x-axis gives the sum of the absolute value of the deviation of the bond lengths from the B3LYP/6-31G minimum. The MM GAFF and QM B3LYP/6-31G configurations are shown along this axis to give an idea of the scatter in energy and structures for the two ensembles.
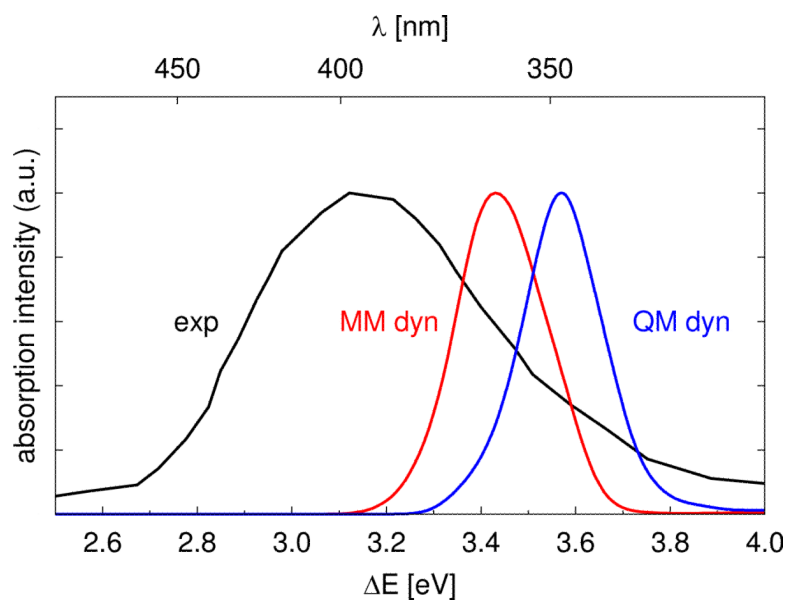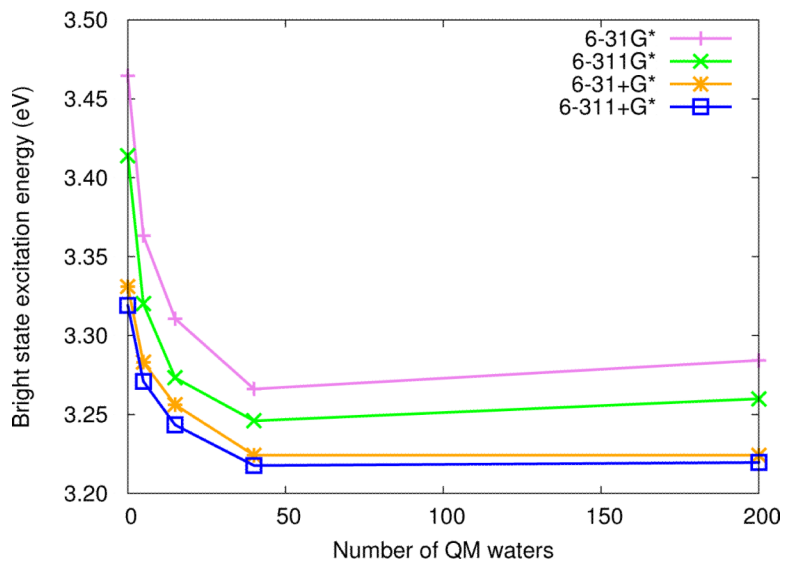
Figure 6.
TD-ωPBE/6-31G* computed spectra for pCT⁻ solvated with SPC/Fw waters from snapshots of MM (red) and QM/MM (blue) Langevin dynamics. MM dynamics sample over 500 ps (1000 snapshots), QM/MM dynamics over 125 ps (250 snapshots). Experimental results from Nielsen et al.[42] are shown in black for comparison.
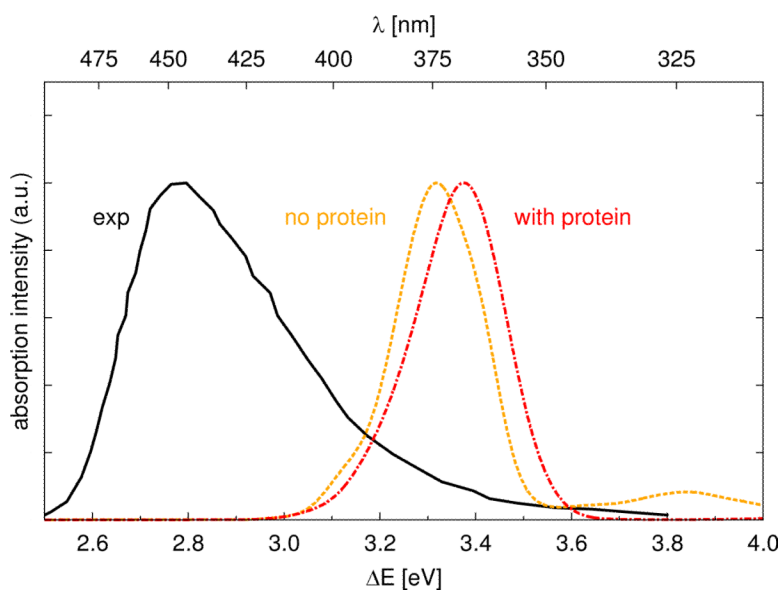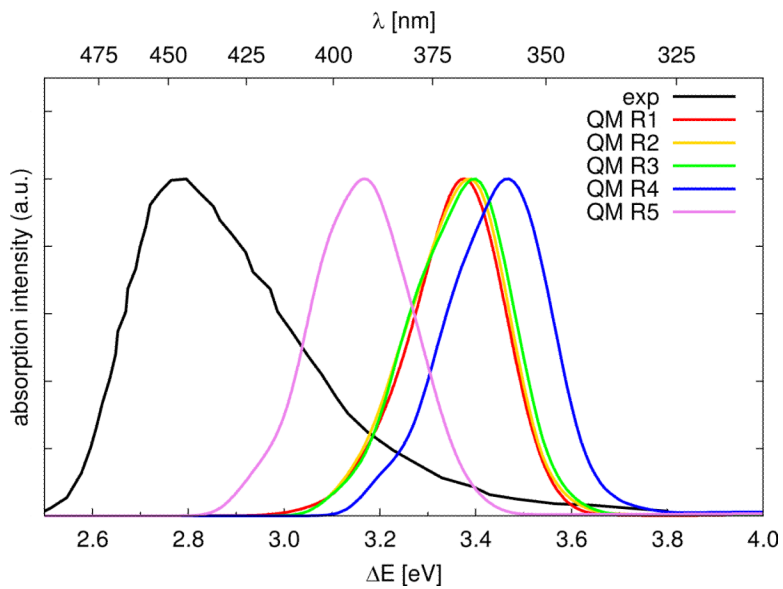
Figure 7.
TD-ωPBE/6-31G* computed spectra for pCT⁻ solvated with a shell of 0, 5, 15, 40, and 200 QM waters surrounded by SPC/Fw waters. Configurations are from 125 ps of QM/MM dynamics (250 snapshots).
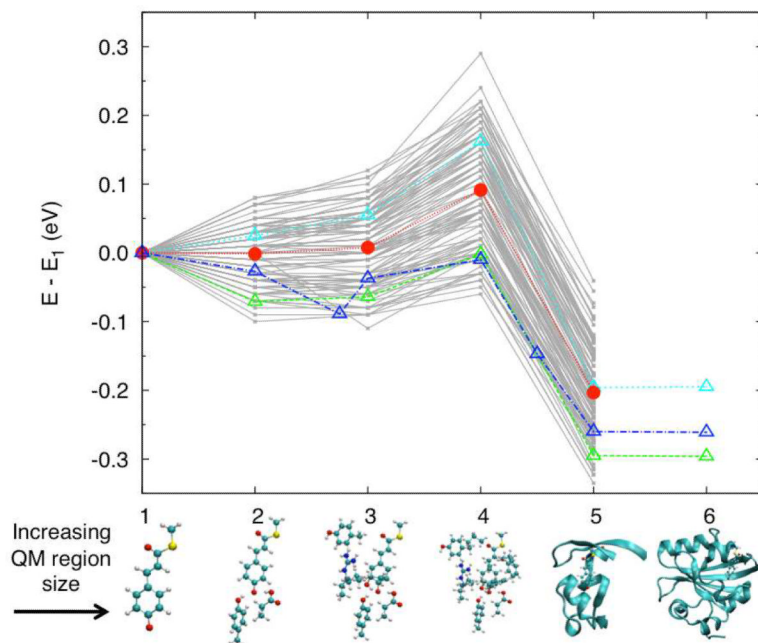
Figure 8.
TD-ωPBE/6-31G* excitation energies for one snapshot of pCT⁻ solvated with a shell of 0, 5, 15, 40, and 200 QM waters surrounded by SPC/Fw waters. The trend towards decreasing excitation energy are similar for the 6-31G*, 6-311G*, 6-31+G*, and 6-311+G* basis sets.
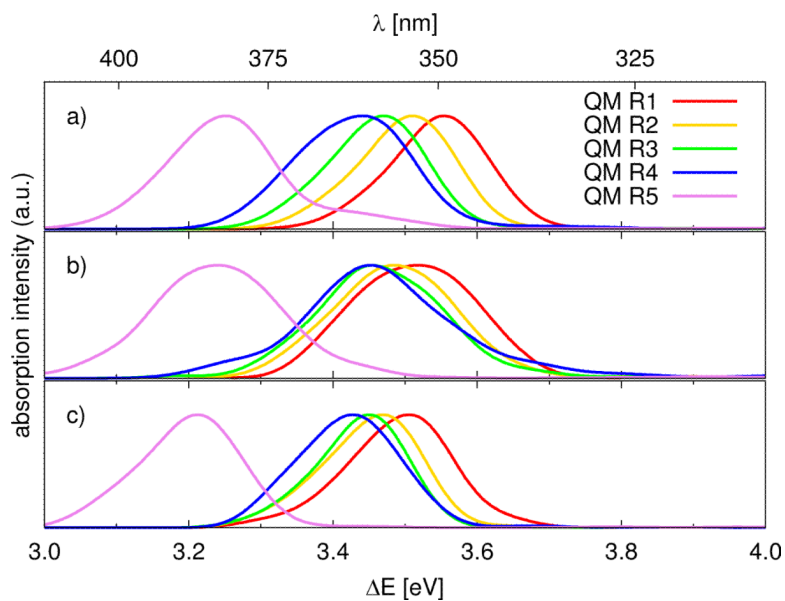
Figure 9.
TD-ωPBE/6-31G* computed spectra (QM region 1) for PYP from 300 snapshots of MM Berendsen dynamics, with and without the protein point charge environment. MM dynamics are sampled over 3 ns. The experimental absorption spectrum[38] is shown for comparison.

Figure 10.
TD-ωPBE/6-31G* computed spectra for PYP from 300 snapshots of MM Berendsen dynamics for the various QM regions. MM dynamics are sampled over 3 ns. The experimental absorption spectrum from ref[38] is shown for comparison.

Figure 11.
Excitation energy difference of the computed TD-ωPBE /6-31G* bright state transition with increasing QM region size for 100 snapshots of MM dynamics of PYP. The average value is shown in red circles, three snapshots are highlighted with colored triangles.

Figure 12.
TD-ωPBE/6-31G* spectra computed for PYP for the various QM regions based on snapshots from QM/MM dynamics: a) from 193 ps of QM region 1 dynamics (193 snapshots), b) from 77 ps of QM region 2 dynamics (77 snapshots), c) from 26 ps of QM region 3 dynamics (26 snapshots).

Table 1

Computational throughput (ps/day) of QM/MM TeraChem/AMBER molecular dynamics of PYP using a 0.5fs time step. File and MPI based communication are compared for QM regions 1 (R1) and 3 (R3) over 1, 2, and 4 GPUs (Tesla C2050 GPU cards). The QM region is modeled with the B3LYP/6-31G method. Calculations were performed on a dual Intel Xeon X5570 workstation with 72 GB RAM. Results are presented for both generating an initial wave function guess from atomic densities and for using the wave function from the previous time step as an initial guess in the SCF process.

# of GPUs	QM R1 (139 basis fxns)			QM R3 (541 basis fxns)		
	1	2	4	1	2	4
File, guess generated	2.75	2.80	2.40	0.48	0.76	0.96
MPI, guess generated	3.63	4.49	4.63	0.50	0.85	1.25
File, reuse guess	3.44	3.20	2.11	0.75	1.11	1.24
MPI, reuse guess	4.82	5.61	5.82	0.80	1.33	1.90
MPI, guess read, no MM	5.00	5.99	6.34	0.84	1.39	2.01

Table 2

TD-ωPBE/6-31G* absorption spectrum λ_{max} (eV) of the PYP chromophore in vacuum, aqueous, and protein environments. Experimental vacuum and aqueous λ_{max} is from Nielson et al[42], PYP λ_{max} is from Borucki et al[38]. QM/MM dynamics for PYP is based on QM region 1.

	vacuum	aqueous (all MM waters)	aqueous (shell of 40 QM waters)	PYP QM R1	PYP QM R5
MM dynamics	3.28	3.43	3.29	3.38	3.17
QM/MM dynamics	3.51	3.57	3.36	3.55	3.25
Experiment	2.70	3.14		2.78	

Table 3

TD-ωPBE/6-31G* absorption spectrum λ_{max} (eV) of PYP from snapshots of MM and QM/MM dynamics, with +/− 2σ bootstrap values reported.

dynamics	R1	R2	R3	R4	R5
MM	3.38 +/− 0.02	3.39 +/− 0.02	3.40 +/− 0.03	3.47 +/− 0.03	3.17 +/− 0.03
QM(R1)/MM	3.55 +/− 0.01	3.51 +/− 0.01	3.47 +/− 0.01	3.44 +/− 0.02	3.25 +/− 0.01
QM(R2)/MM	3.52 +/− 0.05	3.49 +/− 0.05	3.46 +/− 0.04	3.45 +/− 0.03	3.24 +/− 0.05
QM(R3)/MM	3.51 +/− 0.04	3.47 +/− 0.03	3.45 +/− 0.02	3.43 +/− 0.04	3.21 +/− 0.03