Supporting Information for 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, Andreas W. Götz, Matthew A. Clark, Ross C. Walker, and Todd J. Martínez

PULSE Institute and Department of Chemistry, Stanford University, Stanford, CA 94305
SLAC National Accelerator Laboratory, Menlo Park, CA 94025
Department of Chemistry and Chemical Biology, University of California Merced, Merced, CA 95343
San Diego Supercomputer Center, University of California San Diego, La Jolla, CA 92093
Department of Chemistry and Biochemistry, University of California San Diego, La Jolla, CA 92093

This file contains additional information about the implementation of the interface between AMBER and TeraChem, as well as additional figures to support the results discussed in the above paper. The following figures are included:

1. Energy conservation during QM/MM dynamics
2. Computed absorption spectra of pCT^- in vacuum with Berendsen and Langevin thermostatting of the MM dynamics
3. Computed absorption spectra of pCT^- in solution with Berendsen and Langevin thermostatting of the MM dynamics
4. Computed absorption spectra of PYP in solution with Berendsen and Langevin thermostatting of the MM dynamics
5. Computed absorption spectra of pCT^- in vacuum from 25, 100, and 500 ps of MM dynamics, with individual snapshots.
6. Computed absorption spectra of pCT^- in vacuum from snapshots of 6-31G and 6-31G* B3LYP QM dynamics.
7. Computed absorption spectra of pCT^- in vacuum using the 6-31G* and 6-31+G* basis sets from snapshots of MM dynamics.
8. Computed absorption spectra of pCT^- in solution using the 6-31G* and 6-31+G* basis sets from snapshots of MM dynamics.
9. Computed absorption spectra of PYP in solution using the 6-31G* and 6-31+G* basis sets from snapshots of MM dynamics.
10. Computed absorption spectra of pCT^- in solution from snapshots of MM dynamics with a 40 QM water shell, surrounded by MM waters.
11. Excited and ground state density difference plots for a single PYP snapshot for QM regions 1, 4, and 5.

MPI interface of TERACHEM and AMBER:

Two different communication methods are available for data exchange with TERACHEM in the AMBER interface for QM/MM MD simulations:

- File based data exchange
- MPI-2 based client/server model for data exchange.

In both cases AMBER drives the computation with all settings for the QM/MM simulation being defined in the AMBER input file. In what follows we describe the implementation of the client/server model based on the MPI-2 standard.

To run a QM/MM MD simulation using the MPI interface, an MPI version of TeraChem has to be launched with

```
$> terachem –UseMPI [-ID ’ID’]
```

This instructs TERACHEM to start in server mode and to open a port using MPI and publish its name for connection by AMBER‘s MD engine SANDER. The default port name is `terachem_port`, which can be changed by providing the optional `ID` command line argument. This is useful if SANDER needs to communicate with multiple instances of TERACHEM, such as in replica exchange molecular dynamics (REMD) or path integral molecular dynamics (PIMD) simulations.

An MPI version of SANDER can then be started in client mode by setting the `&tc` namelist variable `mpi` in the `mdin` input file to 1:

```
&tc
  …
  mpi = 1
/
```

The QM calculation settings are contained in the `&tc` namelist with variable names and values that reflect TERACHEM input file options, just as for the file-based interface.

When SANDER requires forces for the QM region, it connects as client to the published port. A new MPI communicator is then established and used for all subsequent data exchange, which proceeds via standard MPI send and receive calls. This includes all the QM calculation settings for TERACHEM during the first MD step and all subsequent interchange of atomic coordinates, point charges for electrostatic embedding, and the resulting QM energy and forces. At the end of the simulation, SANDER sends a signal for TERACHEM to exit and disconnects.

For the initialization of TERACHEM in the first MD step, SANDER sends the following data:

- Program settings as keyword/value pairs (character array)

Sending the program settings for TERACHEM as a character array of keyword/value pairs has the advantage of being easily extensible since no assumption about the data type (character, integer number, floating point number) to be transmitted is made and TERACHEM can simply use its input file parser to parse the settings received from SANDER.

During each MD step, SANDER sends the following data in order:

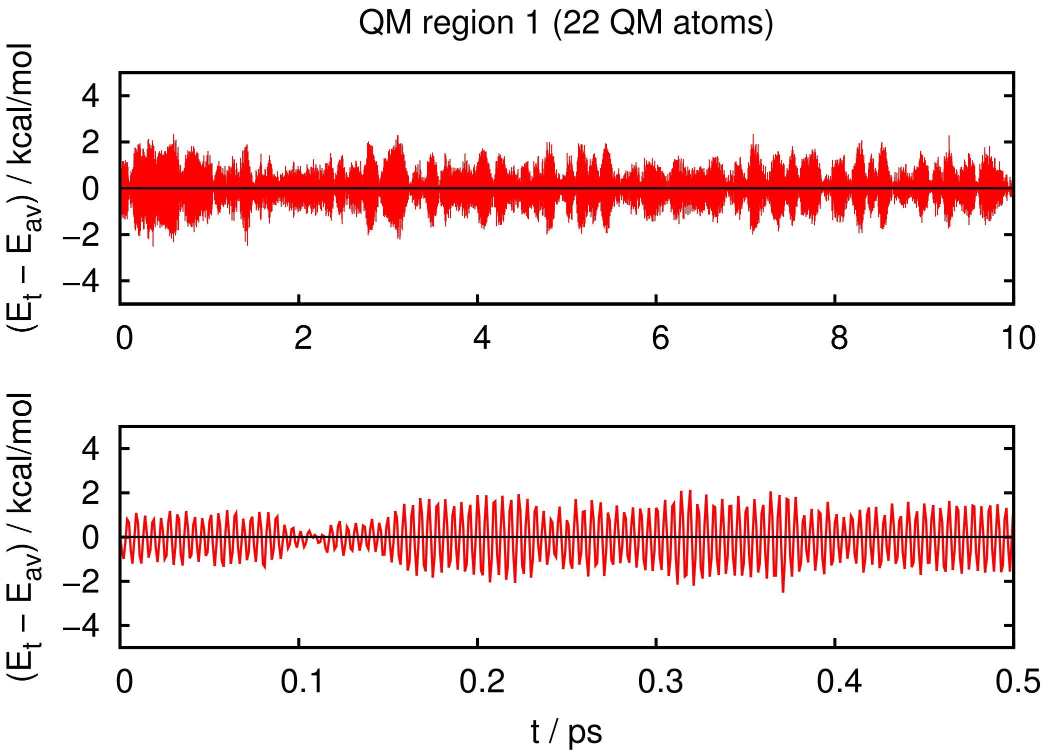
- Number of QM atoms (integer) - *nquant*
- Types of the QM atoms (character array) [*size nquant*]
- QM coordinates (double precision array) [*size 3*nquant*]
- Number of point charges (integer) - *ncharge*
- Point charge values (double precision array) [*size ncharge*]
- Point charge coordinates (double precision array) [*size 3*ncharge*]

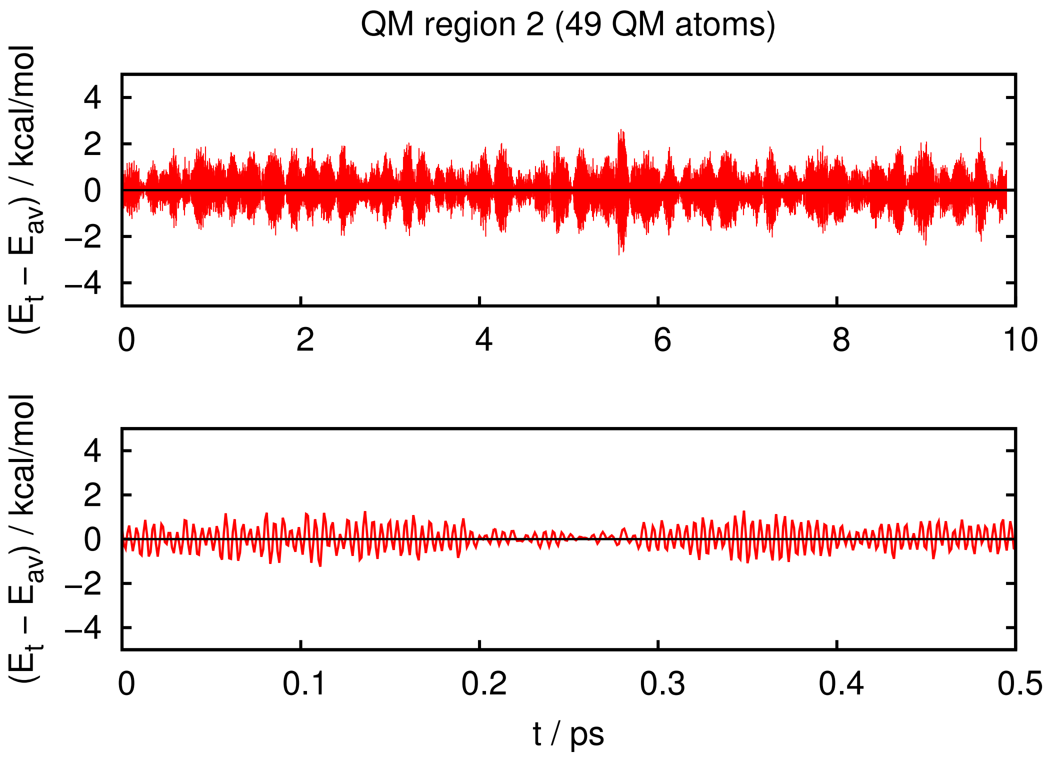
TERACHEM then proceeds to calculate the energy and gradients and sends back the following data to SANDER in order:

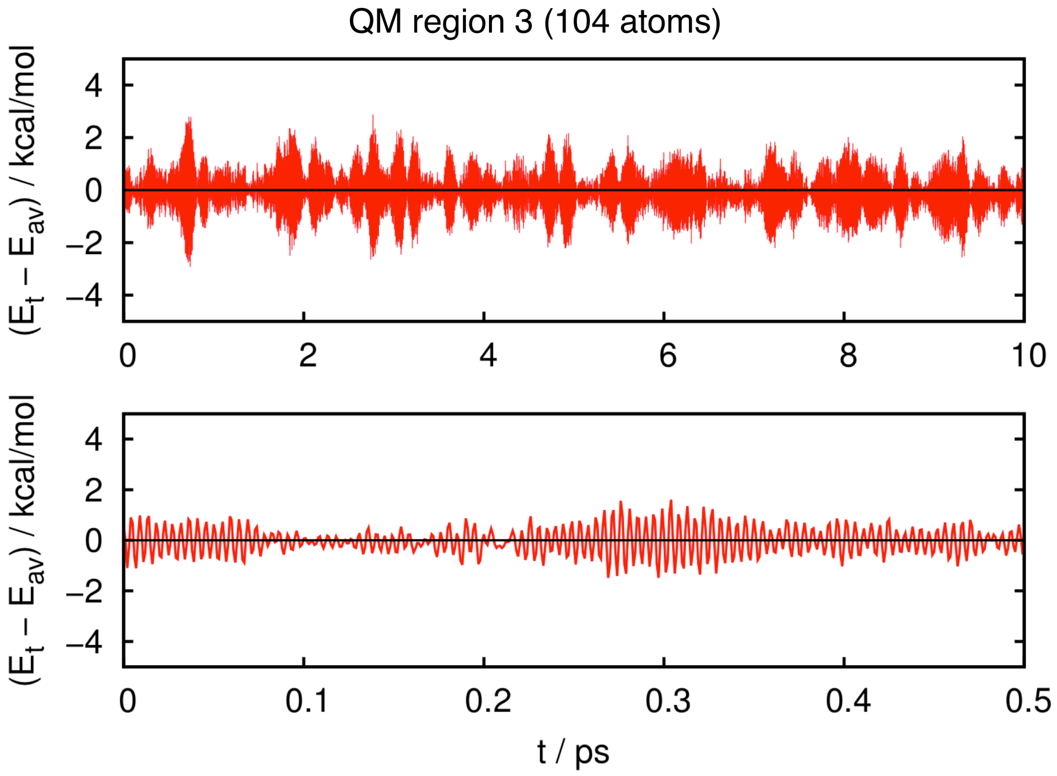
- Energy (double precision)
- Charges from population analysis (double precision array) [*size nquant*]
- QM dipole moment (double precision array) [*size 3*]
- MM dipole moment (double precision array) [*size 3*]
- Total QM/MM dipole moment (double precision array) [*size 3*]
- Gradients on QM atoms and MM atoms (double precision array) [*size 3*nquant + 3*ncharge*]

When SANDER has finished the MD run and does not require any further data from TERACHEM, it invokes an MPI send command with a 0 value tag that instructs TERACHEM to shut down.

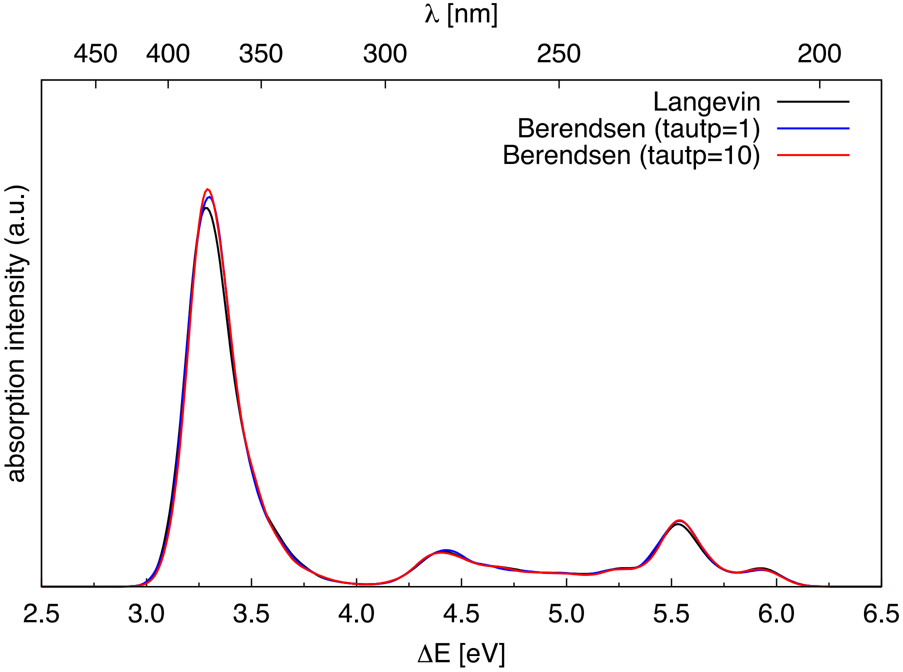
SI - Figure 1. Energy vs time for constant energy (NVE) QM/MM MD trajectories of PYP in an SPC/Fw water droplet of 32 Å radius using a time step of 0.5fs, the B3LYP/6-31G QM Hamiltonian and the ff99SB MM force field, for three different QM region sizes.



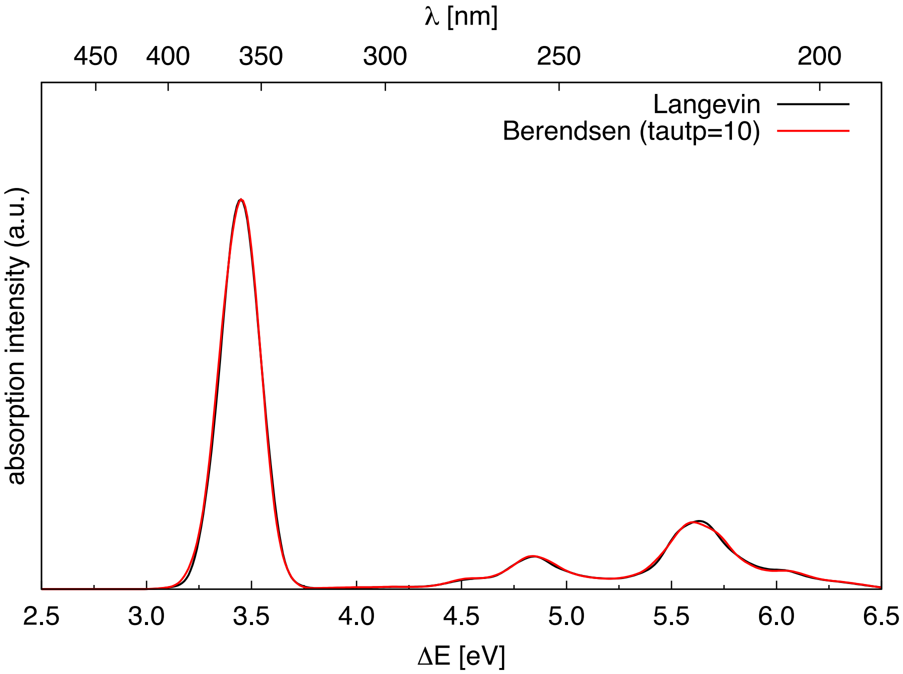

QM region 2 (49 QM atoms)
$(E_t - E_{av})$ / kcal/mol
4
2
0
−2
−4
0
2
4
6
8
10
0
0.1
0.2
0.3
0.4
0.5
t / ps


QM region 3 (104 atoms)
$(E_t - E_{av})$ / kcal/mol
4
2
0
−2
−4
0
2
4
6
8
10
0
0.1
0.2
0.3
0.4
0.5
t / ps

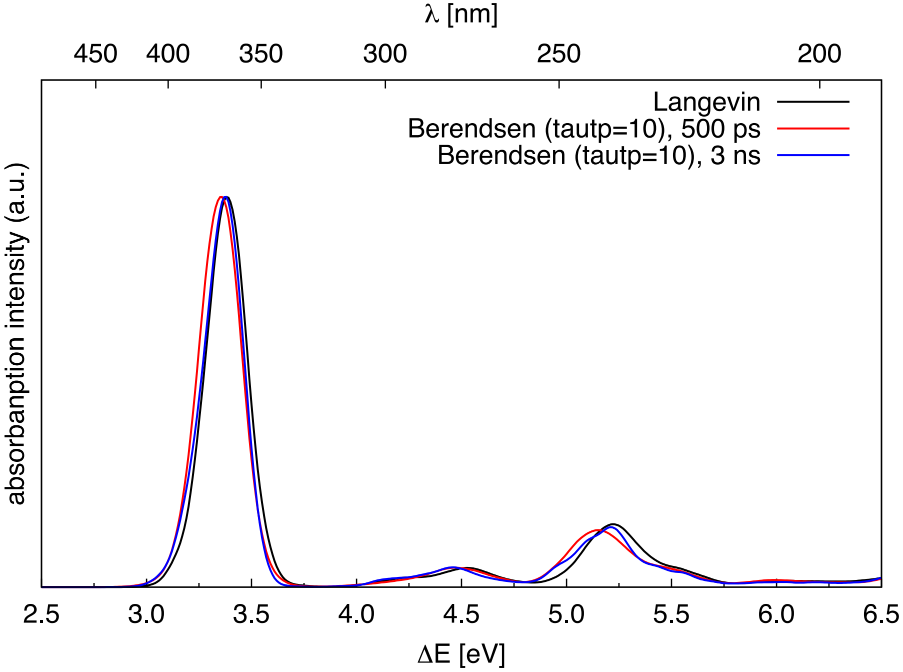
SI - Figure 2. Computed absorption spectra of pCT$^-$ in vacuum from MM MD trajectories using the Berendsen thermostat (tautp=1.0 and tautp=10.0) or Langevin dynamics (gamma_ln=5.0). During 500 ps of dynamics, sampling every 0.25 ps, 2000 snapshots were obtained. Twenty TD-ωPBE/6-31G* excited states were computed for each configuration. Average spectra are plotted together in the first figure, individual spectra are plotted with error bars in the following figures. The computed λ_{max} = 3.28 +/- 0.01 eV. For the Berendsen thermostat with tautp=1.0, the computed λ_{max} = 3.30 +/- 0.01 eV. For Berendsen with tautp=10.0, the computed λ_{max} = 3.29 +/- 0.01 eV.



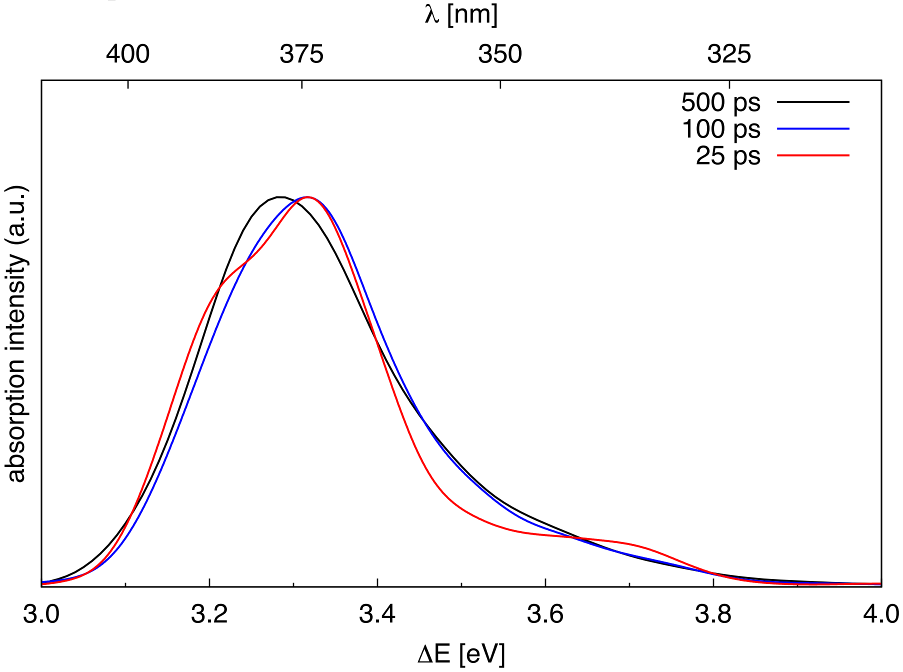
SI - Figure 3. Computed absorption spectra of pCT$^-$ in a SPC/Fw water droplet of 32 Å radius from MM MD trajectories using the Berendsen thermostat (tautp=10.0) or Langevin dynamics (gamma_ln=5.0). During 500 ps of dynamics, sampling every 0.25 ps, 2000 snapshots were obtained. Twenty TD-ωPBE/6-31G* excited states were computed for each configuration. For Langevin dynamics, the computed λ_{max} = 3.45 +/- 0.01 eV. For the Berendsen thermostat with tautp=10.0, the computed λ_{max} = 3.45 +/- 0.01 eV.

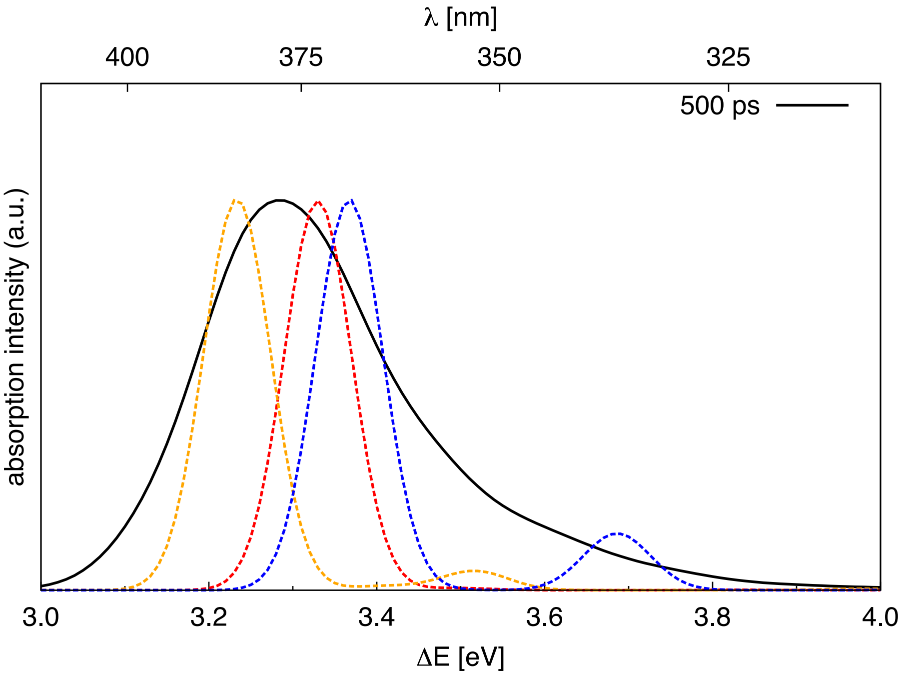


SI - Figure 4. Computed absorption spectra of PYP in an SPC/Fw water droplet of 32 Å radius from MM MD trajectories using the Berendsen thermostat (tautp=10.0) or Langevin dynamics (gamma_ln=5.0). During 500 ps of dynamics, sampling every 0.25 ps, 2000 snapshots were obtained. Additionally, Berendsen MM dynamics were run for 3ns, sampling every 10 ps, giving 300 snapshots. Twenty TD-ωPBE/6-31G* excited states were computed for each configuration using QM region 1. For Langevin dynamics, the computed λ_{max} = 3.38 +/- 0.01 eV. For the Berendsen thermostat sampling every 0.25 ps for 500 ps, the computed λ_{max} = 3.36 +/- 0.01 eV. For the Berendsen thermostat with tautp=10.0 sampling every 10 ps for 3 ns, the computed λ_{max} =3.38 +/- 0.02 eV.



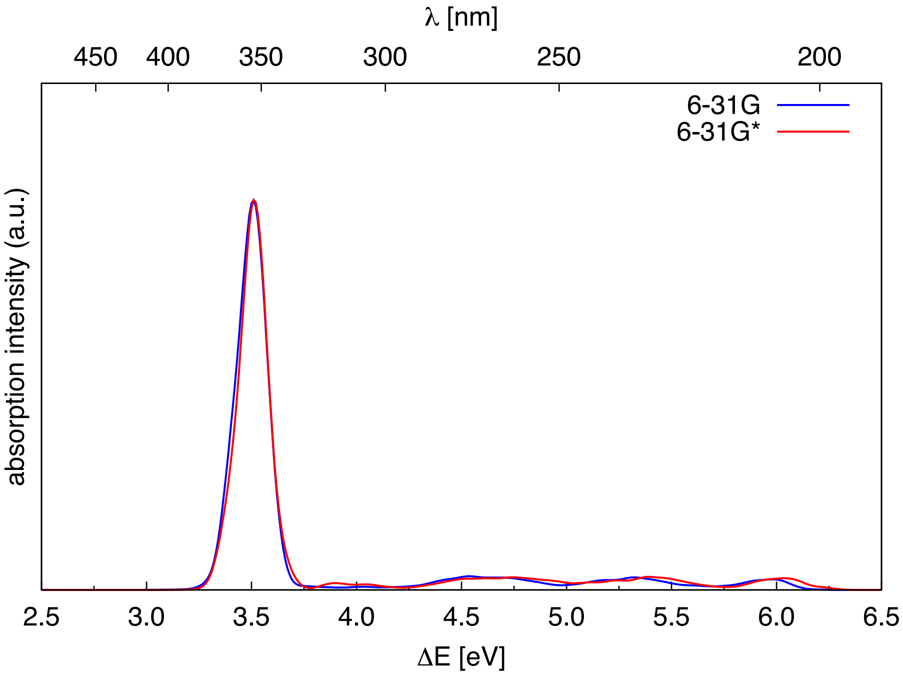
SI - Figure 5. Computed absorption spectra of pCT^- in vacuum from 500, 100, and 25, ps of MM Langevin dynamics, sampling every 0.25 ps. This yielded 2000, 400, and 100 snapshots, respectively. Three individual snapshots are shown in colored lines with the 500 ps average spectrum. λ_{max} varies across this sampling period with values of λ_{max} = 3.32 +/- 0.05 eV for 25 ps, λ_{max} = 3.32 +/- 0.02 eV for 100 ps, and λ_{max} = 3.28 +/- 0.01 eV for 500 ps.



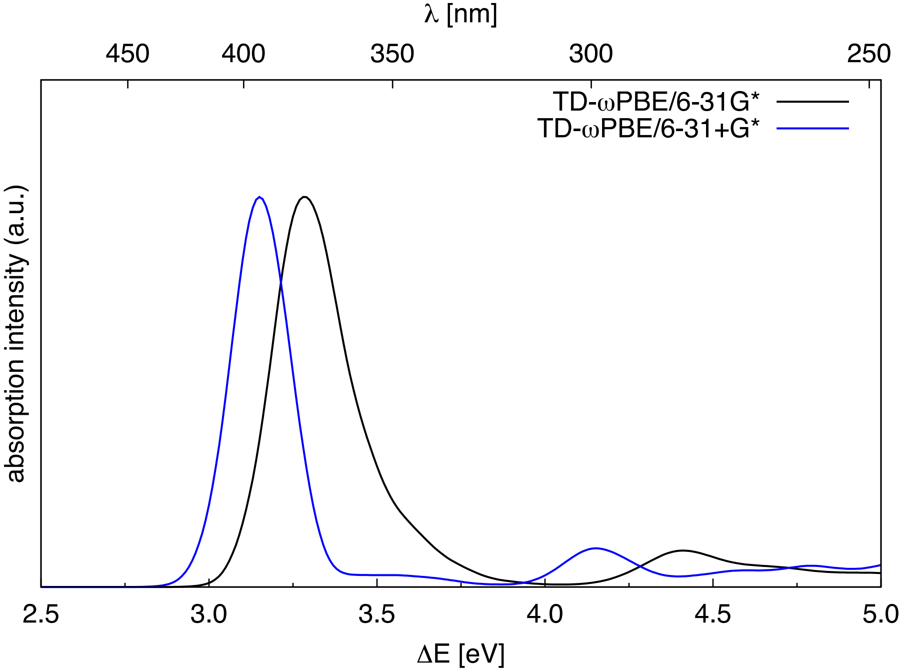


λ [nm]
400
375
350
325
500 ps
absorption intensity (a.u.)
3.0
3.2
3.4
3.6
3.8
4.0
ΔE [eV]

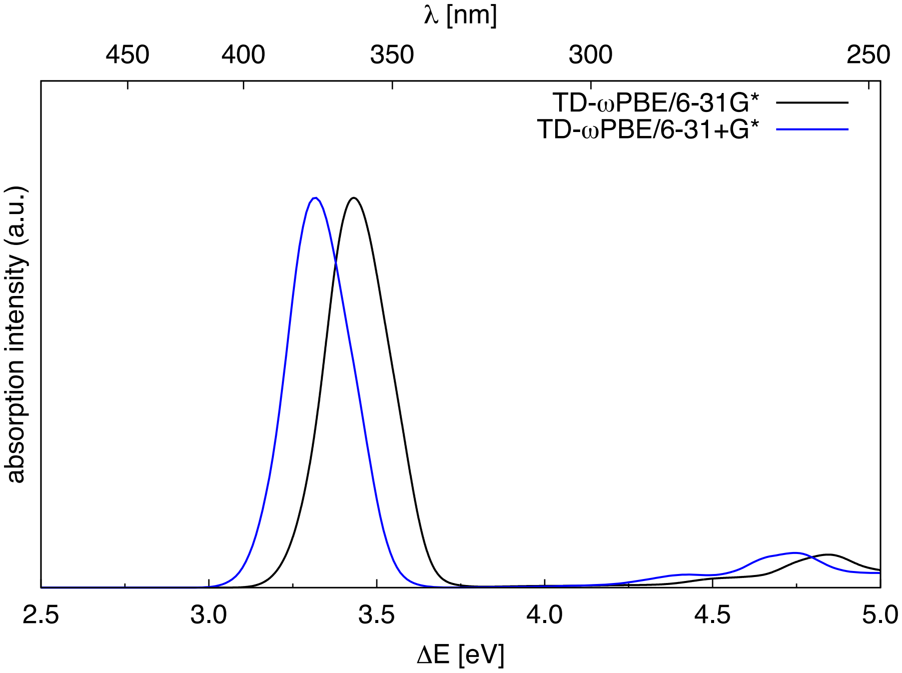
SI - Figure 6. Computed absorption spectra of pCT$^-$ in vacuum from snapshots of 6-31G and 6-31G* B3LYP QM Langevin dynamics, sampling every 0.25 ps. B3LYP/6-31G dynamics were run for 125 ps, resulting in 500 snapshots, and B3LYP/6-31G* dynamics were run for 41 ps, resulting in 165 snapshots. The computed λ_{max} = 3.51 +/- 0.01 eV with B3LYP/6-31G dynamics, and λ_{max} = 3.51 +/- 0.01 eV with B3LYP/6-31G* dynamics.



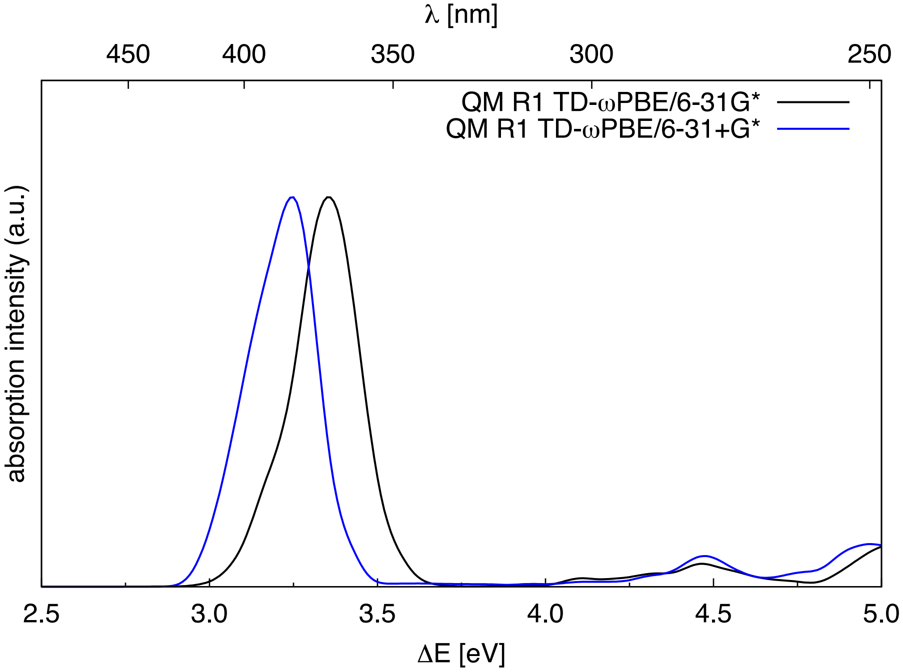
SI - Figure 7. Computed absorption spectra of pCT$^-$ in vacuum using both the 6-31G* and the 6-31+G* basis set from snapshots of MM Langevin dynamics. During 500 ps of dynamics, sampling every 0.25 ps, 2000 snapshots were obtained. Twenty TD-ωPBE/6-31G* and TD-ωPBE/6-31+G* excited states were computed for each configuration. For TD-ωPBE/6-31+G*, the computed λ_{max} = 3.15 +/- 0.01 eV, for TD-ωPBE/6-31G*, the computed λ_{max} = 3.28 +/- 0.01 eV.

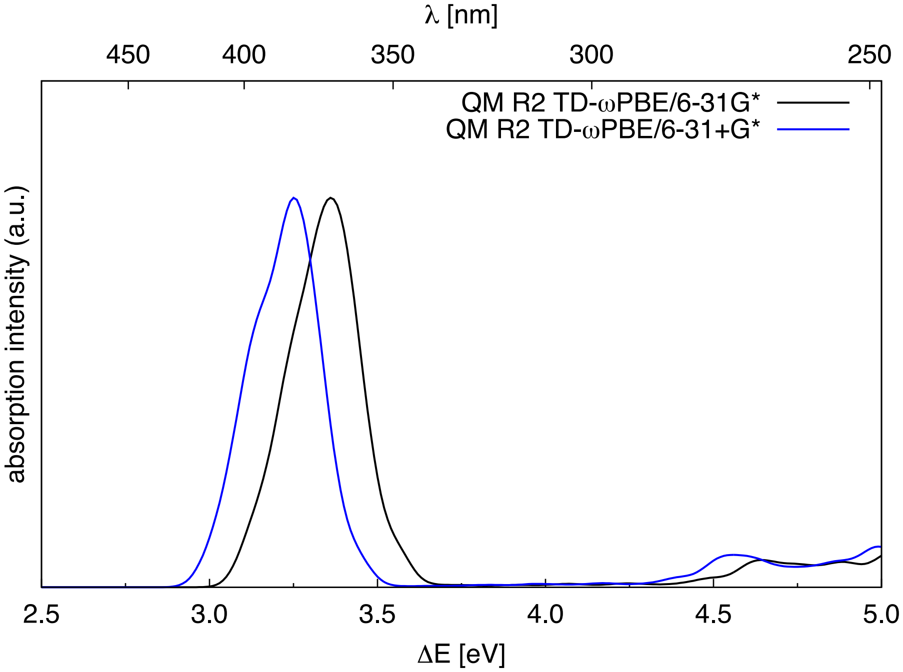


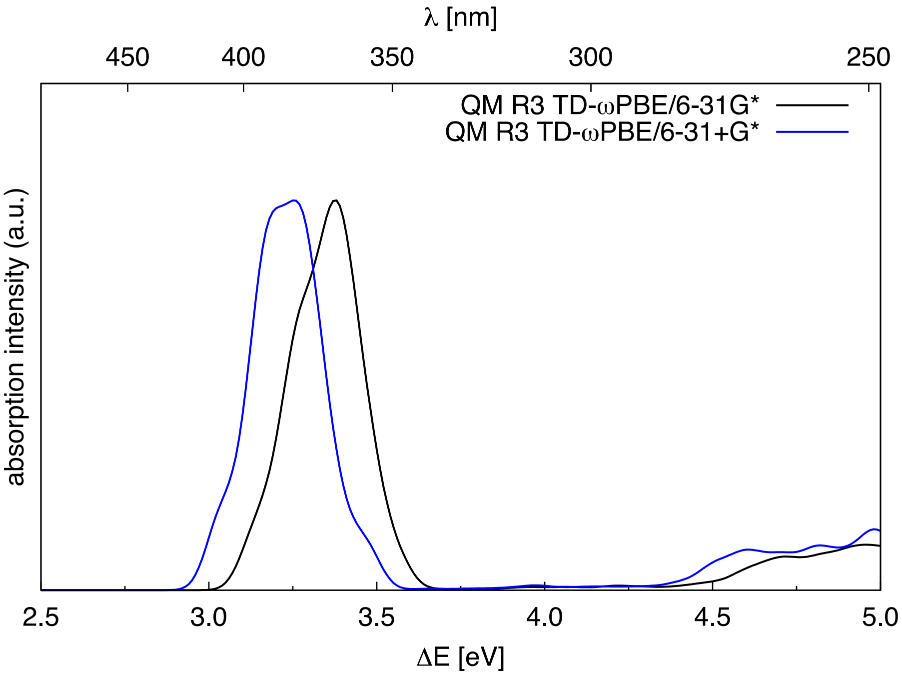
SI - Figure 8. Computed absorption spectra of pCT$^-$ in aqueous solution using both the 6-31G* and the 6-31+G* basis set from snapshots of MM Langevin dynamics. During 500 ps of dynamics, sampling every 0.50 ps, 1000 snapshots were obtained. Twenty TD-ωPBE/6-31G* and TD-ωPBE/6-31+G* excited states were computed for each configuration. The computed TD-ωPBE/6-31+G* λ_{max} = 3.32 +/- 0.01 eV, the computed TD-ωPBE/6-31G* λ_{max} = 3.43 +/- 0.02.



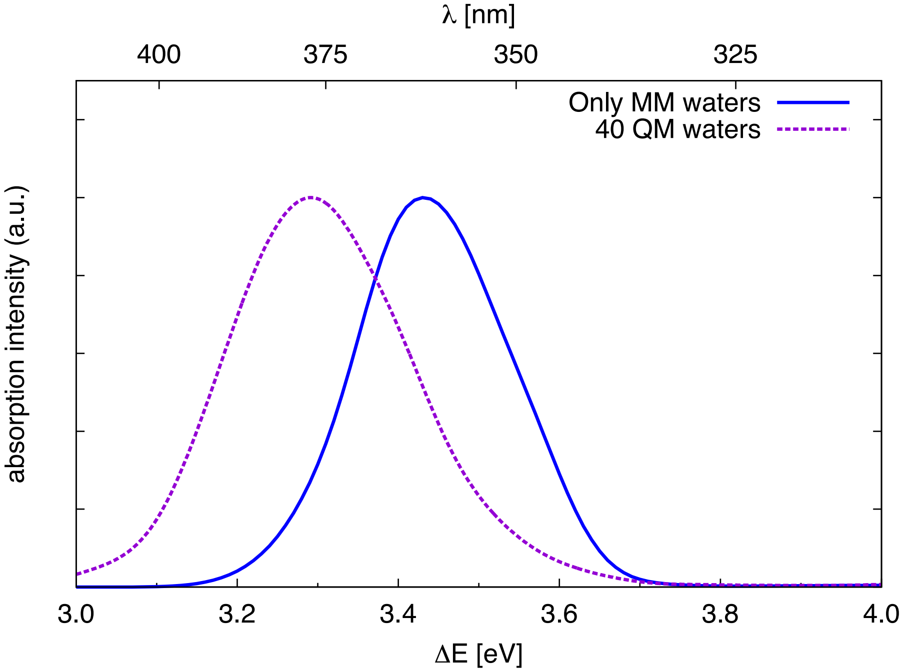
SI - Figure 9. Computed absorption spectra of PYP in aqueous solution using both the 6-31G* and the 6-31+G* basis set from snapshots of MM dynamics. Berendsen MM dynamics were run with tautp = 10. During 1 ns of dynamics, sampling every 10 ps, 100 snapshots were obtained. Twenty TD-ωPBE/6-31G* and TD-ωPBE/6-31+G* excited states were computed for each configuration for QM regions 1-3. The computed TD-ωPBE/6-31G* QM R1 λ_{max} = 3.36 +/- 0.03 eV, QM R2 λ_{max} = 3.36 +/- 0.05 eV, QM R3 λ_{max} = 3.38 +/- 0.05 eV. The computed TD-ωPBE/6-31+G* QM R1 λ_{max} = 3.25+/- 0.05 eV, QM R2 λ_{max} = 3.25 +/- 0.05 eV, QM R3 λ_{max} = 3.25 +/- 0.08 eV.






λ [nm]
450
400
350
300
250
QM R3 TD-ωPBE/6-31G*
QM R3 TD-ωPBE/6-31+G*
absorption intensity (a.u.)
2.5
3.0
3.5
4.0
4.5
5.0
ΔE [eV]

SI - Figure 10. Computed absorption spectra of pCT^- in solution from 1000 snapshots of MM dynamics (500 ps). The 40 nearest waters are treated with QM, the remaining sphere of waters is treated with MM. For only MM waters, the computed TD-ωPBE/6-31G* λ_{max} = 3.43 +/- 0.02 eV, with the 40 QM water shell, the computed λ_{max} = 3.29 +/- 0.02.



SI - Figure 11. Excited and ground state density difference plots for a single PYP snapshot for QM regions 1 (top), 4 (midde), and 5 (bottom). The electron density isovalue is 1.0e-5 electrons / bohr^3.