Efficient Determination of Accurate Atomic Polarizabilities for Polarizeable Embedding Calculations

Heiner Schröder and Tobias Schwabe*

We evaluate embedding potentials, obtained via various methods, used for polarizable embedding computations of excitation energies of *para*-nitroaniline in water and organic solvents as well as of the green fluorescent protein. We found that isotropic polarizabilities derived from DFTD3 dispersion coefficients correlate well with those obtained via the LoProp method. We show that these polarizabilities in conjunction with appropriately derived point charges are in good agreement with calculations employing static multipole moments up to quadrupoles and anisotropic polarizabilities for both computed systems. The (partial) use of these easily-accessible parameters drastically reduces the computational effort to obtain accurate embedding potentials especially for proteins. © 2016 The Authors. Journal of Computational Chemistry Published by Wiley Periodicals, Inc.

DOI: 10.1002/jcc.24425

Introduction

Spectroscopic methods are widely used and of great importance in every part of experimental chemistry. The understanding of the underlying photophysical processes and the interpretation of spectra require insights from highly accurate theoretical methods. However, most of the interesting systems are not isolated molecules in vacuum, but such that have a strong coupling to their respective environment, like molecules in solution or chromophores in proteins. These systems contain at the very least a thousand electrons,[1,2] which makes highly accurate QM computations on the whole system impractical. However, only a small part of the system is usually directly involved in photophysical processes and thus an obvious opportunity is to use a (high level) QM method only for this part and a more approximate one for the environment. A highly efficient but yet accurate possibility is to treat the environment with classical molecular mechanics (MM). Of course, this can be done in varying complexity.[3] Alternatively, a QM description of small fragments making up the environment is also possible, but computationally more demanding.[4–7]

It has been shown that a polarizable QM/MM model, which allows for the mutual interaction between the QM and MM region, is crucial for accurate excitation energies.[8–13] The method used and discussed in this work is the polarizable embedding method by Kongsted and coworkers.[2,14] It is an explicit solvent model, for which the electrostatic potential (ESP) of a fragment is characterized by static multipole moments (up to hexadecapoles) centered at the atom positions and the responsiveness of the system to the QM potential and other MM sites by induceable dipoles through isotropic or anisotropic polarizability. Unfortunately, a lot of effort has to be put in obtaining accurate parameters.

A multitude of schemes for the derivation of point charges[15–19] alone or point charges and higher multipole moments[20–22] has been developed. Point-charge models, which are fitted to the quantum-mechanically derived ESP, can mimic the effects of higher multipole moments to some extend when additional constraints are considered.[12,23] For the more elaborate task of obtaining dipole–dipole polarizabilities only a few schemes exist, which are discussed, for example, in Ref. [22].

In a standard experimental setup, a multitude of molecules in different configurations contribute to the experimentally measured excitation energies. According to the ergodic hypothesis, energies have to be averaged over time for a single particle to obtain a representative multitude of configurations to cover these effects.[24] Beerepoot et al. found that sampling is not necessary for computing the one-electron excited state properties via TDDFT for the green fluorescent protein (GFP),[25] while significant effects for solvated molecules were found.[26] Ideally, geometry-specific PE parameters would be obtained for every single configuration through quantum chemical calculations. The whole environment is still too large for a single calculation with a simpler QM method and has therefore to be fragmented. While for a solvent environment the obvious choice for a fragment are the solvent molecules, a reasonable choice has to be made for more structured environments. For proteins this can be done by the molecular fractionation with conjugate caps method by Zhang.[27]

Söderhjelm et al.[9] found that highly accurate embedding potentials including static multipoles up to quadrupoles and

H. Schröder, T. Schwabe
Center for Bioinformatics and Institute of Physical Chemistry, University of Hamburg, Bundesstraße 43, Hamburg, 20146, Germany
E-mail: schwabe@zbh.uni-hamburg.de

anisotropic polarizabilities are only needed for the MM region close to the QM part to obtain accurate excited state properties and suggested to use less accurate parameters for the rest. In these cases one would prefer to use a set of precalculated parameters from a standard force field like AMBER[15] or OPLS,[16] but polarizabilities, which are crucial for accurate embedding potentials, are not as easily accessible so far. A way of minimizing the computational effort is to drastically reduce the systems degrees of freedom, for example, using rigid solvent molecules in the molecular dynamic (MD) simulation. This way, only a single computation of the solvent potential parameters has to be performed and those can then be placed accordingly around the QM region. Obviously, this approximation lacks any geometry-specific effects due to flexibility, but gives yet accurate results in some cases.[28]

The aim of this work is to investigate, whether a high level determination of atomic parameters for polarizable embedding is necessary for accurate excitation energies and moreover, how much of the system needs to be described that way at most. Finding a set of easily accessible parameters for the major part of the environment and thus minimizing the amount of QM calculations to obtain the embedding potential, would allow for routine computation of several protein- or solvent configurations. Our ansatz is to use high-level parameters for the inner MM region only and isotropic polarizabilities obtained from D3 dispersion coefficients with standard force field point charges for the outer MM region.

Para-nitroaniline (PNA) in four different solvents and the chromophor in GFP in its neutral (protonated) and anionic (deprototnated) form Ref. [29] have been chosen for this study, since both are well investigated model systems for solvatochromism and enzymichromism, respectively.

Methods

Based on the suggestions of Söderhjelm et al.,[9] we divide the MM environment into an inner shell, which is described by static multipoles up to quadrupoles and anisotropic polarizabilities obtained by the LoProp method,[22] and an outer shell described by point charges alone or in combination with isotropic dipole–dipole polarizabilities α^A of an atom A derived from DFT-D3[30] dispersion coefficients C_6^{AA} according to the equation

$$\alpha^A = \sqrt{C_6^{AA}} \tag{1}$$

as introduced for Grimme's quantum-mechanically derived force field.[31] The cutoff distance is gradually increased in size to figure out, when convergence to the full LoProp treatment with multipole moments up to quadrupoles and anisotropic polarizabilities is reached.

Since the DFT-D3 dispersion coefficients are derived from molecular polarizabilities,[30] it seems likely to use them to obtain approximate polarizabilities. In DFT-D3 the dispersion-coefficient C_6^{AA} of an atom A in the homoatomic pair AA is calculated from the respective dynamic polarizabilities $\alpha^A(iv)$ at an imaginary frequency v using the Casimir–Polder-formula[32]

$$C_6^{AA} = \int_0^\infty \alpha^A(iv)\alpha^A(iv)dv\,. \tag{2}$$

According to Tang,[33] the dynamic polarizabilities may be approximated as

$$\alpha^A(iv) \approx \frac{\alpha_0^A}{1+(\frac{v}{\eta^A})}\,, \tag{3}$$

where α_0^A is the static polarizability of atom A and η_A is an empirical constant. Plugging this approximation into eq. (2) yields

$$C_6^{AA} = (\alpha_0^A)^2 \int_0^\infty \left(\frac{1}{1+(\frac{v}{\eta_A})}\right)^2 dv \tag{4}$$

The integral of eq. (4) has the finite value η^A and thus the static polarizability is proportional to the squareroot of the dispersion coefficient.

$$\alpha_0^A = \sqrt{\frac{C_6^{AA}}{\eta^A}} \tag{5}$$

Thus the static polarizability of an atom A may be easily approximated as the square root of the self-dispersion-coefficient times a constant factor. The quality of this approximation as well as a value for the proportionality factor will be evaluated later.

In a recent study,[10] the influence of the size of the protein surrounding on the excitation energies of GFP was investigated by gradually increasing its size for a single conformation and comparing it to the respective full quantum mechanical treatment. However, in this work we will work with the same structure but use the whole protein throughout.

Recently and independently from us, Beerepoot and coworkers took the same distance-dependent approach[19] for the computation of molecular properties of PNA in different solvents via TDDFT. In contrast to our work they use point charges and isotropic polarizabilities, that are averaged over 1000 solvent configurations from a molecular dynamics simulation. The individual point charges are fitted to the respective ESPs, while the polarizabilities are obtained via the LoProp method. The detailed procedure to obtain these parameters, which are termed QP1, is described in their work.[19] The QP1 solvent parameters are also evaluated in this work.

In this work we denote the PE force fields with MXPY, where *X* is the order of the highest multipole moment and *Y* indicates that no (0), isotropic (1), or anisotropic (2) polarizabilities were used.

Computational Details

PNA in solution

Optimized solvent structures and static multipoles up to quadrupoles as well as anisotropic dipole–dipole polarizabilities obtained via the LoProp[22] method were taken from previous

studies.[28,34] Charge Model 5[17] (CM5) point charges were obtained on the B3LYP/aug-cc-pVTZ-level with the UltraFine-Grid in Gaussian09.[35] Gasteiger[18] (Gas) point charges were obtained with Open Babel V.2.3.1.[36,37] Isotropic dipole–dipole polarizabilities were calculated from the D3 dispersion coefficients according to eq. (5) with $\eta^A = 1$. D3 dispersion coefficients were obtained with a local Python implementation for the computation of dispersion corrections, which has been developed in our group for a recent revision of the DFTD3[30] approach.[38] QP1 point charges and isotropic dipole–dipole polarizabilities were taken directly from Ref. [19]. All pure M0P1 embedding potentials were setup using the WHIRPOOL program[39] as described in Ref. [28].

The combination of embedding potentials was based on the minimal distance between the QM and MM region. For every atom in the QM region, the distance to every MM atom was computed. A fragment was described by the M2P2 LoProp potential, if any of its atoms was within the cutoff distance, else with the given M0P1 potential. The merging of embedding potentials was performed by a local Python implementation. For the electronic excitation computations, the PERI-CC2 model has been applied,[40] which is a slight modification of the standard CC2 approach combined with the polarizable embedding approach[14] to exploit the efficient implementation of RI-CC2 model.[41] All PERI-CC2 calculations have been carried out using a local development version of TURBOMOLE[42] using the (aug-)cc-pVDZ[43,44] with the corresponding auxiliary basis sets.[45] Diffuse functions were added only to non-hydrogen atoms to reduce the computational cost. Polarizabilities and multipole moments at MM atoms closer than 2.5 a.u. to the QM region were transferred to the nearest neighbor atom to avoid overpolarization. Polarizabilities and multipole moments were therefore removed from the respective atoms and added component-wise to its nearest neighbor. The results were averaged over 121 snapshots from an MD simulation, also described in Ref. [28].

GFP

Protein structures were taken from a previous study.[10] Static multipoles up to quadrupoles and anisotropic dipole–dipole polarizabilities were obtained by the LoProp-method[22] as described in Ref. [10]. The M0P1 potential used was a combination of Amber94[15] point charges and isotropic dipole–dipole polarizabilities derived from D3 dispersion coefficients as described above. Element-averaged isotropic polarizabilities (AvgLoP) were calculated from the anisotropic LoProp polarizabilities.

The def2-SVPD basis set[46] was used for all calculation on GFP. All PE-DFT[2] calculations were performed with Dalton2015[47,48] using the long-range corrected CAM-B3LYP functional.[49–52] PERI-CC2[40] calculations have been carried out with the RICC2 module of a local TURBOMOLE[42] version subsequent to a PE-Hartree–Fock calculation with the DSCF module.[53] Polarizabilities and multipole moments at MM atoms closer than 1.4 Å (2.6456 a.u.) to the QM region were transferred to the nearest neighbor atom for all PE-calculations as described above. A slightly different distance than for the PNA calculations was used so that both are consistent with previous studies. The difference is very small though and not expected to have any significant effect.

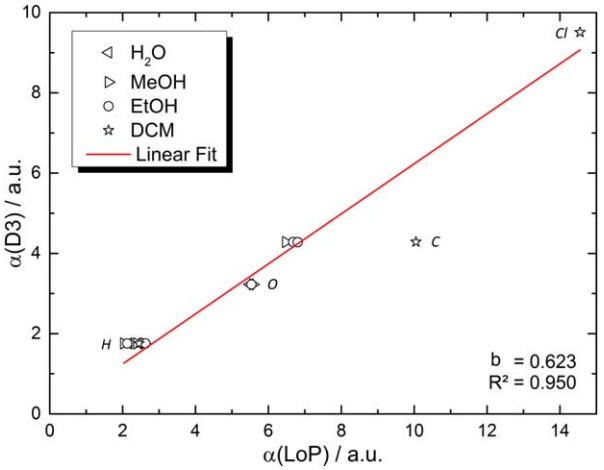

Figure 1. Correlation-plot of isotropic polarizabilities obtained via LoProp (y) or D3 dispersion coefficients (x) for various solvents. Polarizabilities in a.u. Additionaly, R^2 and slope b of the linear fit line are given. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Results and Discussion

Isotropic polarizabilities

In the first part of this section, the relation between the D3 polarizabilities and the ones obtained via LoProp, which act as a reference throughout this work, are checked. Therefore, we plotted the respective isotropic polarizabilities of every solvent atom against each other (Fig. 1) and of the protein surroundings taken from the neutral chromophore system of our GFP model (Fig. 2). Additionally, a linear fit-line through the origin

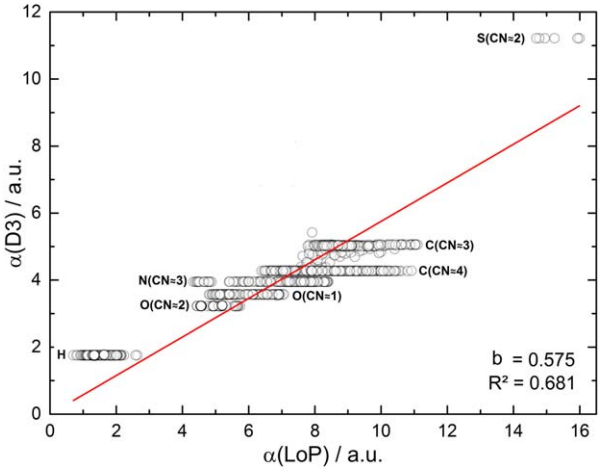

Figure 2. Correlation-plot of isotropic polarizabilities obtained via LoProp (y) and D3 dispersion coefficients (x) for the protein surrounding of the neutral form of the cromophore in GFP containing 3990 atoms. Polarizabilities in a.u. Additionaly, R^2, the slope b of the linear fit line, and the atom type with the coordination number (CN) are given. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

is added, with the goal of obtaining a simple linear scaling factor.

The polarizabilities from D3 and LoProp correlate well for the solvents (Pearson's $r = 0.977$). Atoms of the same element with the same coordination number (CN) can be found on a parallel line to the *x*-axis, since they are assigned very similar dispersion coefficients and thus the same D3 polarizability. Because of that, they cannot cover effects from different chemical environments with the same CN. This can be seen on the basis of the sp^3–carbon in dichloromethane (DCM; single outlier) which is assigned the same value as the sp^3–carbons in the alcohols for D3, whereas LoProp finds a 50% higher value (see Fig. 1).

The range of LoProp polarizabilities is even higher in the protein, due to the multitude of different chemical environments. Data points between the linear carbon data cluster are mostly carbon atoms with noninteger CNs, for which the dispersion coefficients are interpolated between the parametrized integer values.

As an example, the 296 sp^3–nitrogen atoms in the GFP MM region have polarizabilities in the range of $\sim$ 4.3 – 8.4 a.u. (Fig. 2). However, the majority of data points nevertheless yield a still good correlation coefficient (Pearson's $r = 0.971$). The single outlier in the mid of Figure 2 is a carbon atom with a CN of $\sim$2, which is an artifact from the crystal structure at the outside of the protein.

Both correlations resulted in very similar slopes of the best-fit line and sufficient R^2 values for the fitted linear functions can be found (solvents: $R^2 = 0.950$, GFP: $R^2 = 0.681$) Even though according to eq. (5) each atom should have an individual scaling factor $k^A = \sqrt{\frac{1}{\eta^A}}$, we choose an average atom-independent global scaling factor of $k = \frac{5}{3}$ for simplicity. The scaled D3 polarizabilities are denoted scD3 hereinafter. From our experience, excitation energies in PE computations are not very sensitive to the quality of polarizabilities, so we are only looking for a rough estimate.

LoProp polarizabilities used in this study have been obtained with different basis sets—an ANO-type recontraction of the aug-cc-pVTZ basis set for the solvents,[28,34] and an ANO-type recontraction of the 6-31 + G*[54–56] basis set for the protein.[10] The basis set dependence of the polarizabilities is thus negligible for our scaling parameter within the aspired precision.

Calculation of excitation energies

The purpose of embedding parameters is to obtain accurate molecular properties in PE-calculations, which will be evaluated in the following section. In the first part of this section, M0P1 embedding potentials from different sources are compared among each other by means of their convergence behavior to the full M2P2 embedding potential for the excitation energies of PNA in four different solvents. In the second part, the same is done in a finer fashion for the neutral and anionic state of the chromophore in GFP.

We use M0P1 embedding potentials with parameters from different sources, namely LoProp (LoP), charge model 5 (CM5), Gasteiger point charges (Gas), geometry-averaged parameters from Beerepoot et al. (QP1) and polarizabilities from D3 dispersion coefficients: scaled (scD3) as well as unscaled ones (D3). We use the notation A/B, where A indicates the source of the point charges and B the source of the polarizabilities. For example, CM5/LoP means, that CM5 point charges and LoProp polarizabilities are employed. M2P2 parameters are always LoP/LoP.

Table 1. Excitation energies of the brightest PNA excitation in different solvents.

Cutoff	LoP/ LoP	QP1/ QP1	QP1/ scD3	CM5/ QP1	CM5/ scD3	CM5/ D3	Gas/ scD3
				Water			
0	3.45	3.41	3.39	3.45	3.42	3.58	3.66
2.5	3.36	3.35	3.34	3.36	3.35	3.43	3.45
3.5	3.34	3.33	3.33	3.34	3.33	3.38	3.41
5.0	3.32	3.31	3.31	3.32	3.31	3.34	3.36
Full M2P2	3.30						
→ exp	3.29						
				MeOH			
0	3.52	3.49	3.42	3.58	3.52	3.67	3.65
2.5	3.48	3.47	3.44	3.50	3.47	3.56	3.52
3.5	3.48	3.47	3.45	3.49	3.48	3.52	3.51
5.0	3.47	3.47	3.46	3.48	3.47	3.49	3.49
Full M2P2	3.47						
→ exp	3.37						
				EtOH			
0	3.56	3.52	3.47	3.59	3.54	3.70	3.67
2.5	3.52	3.50	3.48	3.53	3.51	3.59	3.54
3.5	3.52	3.49	3.50	3.52	3.51	3.54	3.53
5.0	3.51	3.51	3.51	3.52	3.51	3.53	3.52
Full M2P2	3.51						
→ exp	3.36						
				DCM			
0	3.82	3.81	3.81	3.82	3.82	3.82	3.86
2.5	3.82	3.81	3.81	3.82	3.82	3.82	3.85
3.5	3.81	3.80	3.8	3.81	3.81	3.81	3.81
5.0	3.81	3.81	3.8	3.81	3.81	3.81	3.81
Full M2P2	3.81						
→ exp	3.58						

Solvent molecules within the cutoff radius are described by a M2P2 potential obtained via LoProp, the rest of the solvent shell up to 12 Å by various M0P1 potentials. Energies were averaged over 121 snapshots. Cutoff in Å, all other values in eV.

Solutions

Excitation energies of the bright $\pi-\pi^*$-state of PNA in water, methanol, ethanol, and DCM are shown in Table 1 for different embedding potentials as well as mixed M2P2/M0P1 potentials. The brightest of the three lowest excitations was chosen for every configuration.

As discussed in detail previously,[28] full M2P2 results are in good agreement with the experimental values for water and are sufficient for methanol with a deviation of 0.1 eV. A deviation of 0.15 eV can be found for ethanol. The higher deviation of the alcohols may be caused by the rigid solvent model, but this has to be investigated in a further study. At least, it has been shown that partial atomic charges in alcohols varied strongly with the configuration.[19] The highest deviation of more than 0.2 eV can be found for dichlormethane.

However, the focus of this study is not on the excitation energies themselves, but on the performance of the different embedding potentials in comparison to the full M2P2 potential and their convergence behavior when the higher potential is used in the inner MM region. Considering the pure potentials, QP1/QP1, CM5/scD3, and LoP/LoP all perform well. Excitation energies are already in good agreement with full M2P2 (deviations < 0.05 eV) with the exception of water, for which deviations of 0.10–0.15 eV are found. A ranking of the three potentials would be in the order as listed but the differences are not significant and the amount of data is too small for any definite conclusion.

CM5/D3 and Gas/scD3 show larger deviations of up to 0.36 eV (Gas/scD3 for water). Although Gasteiger charges have the advantage that they can be generated from structural data directly, the deviations are not acceptable. DCM is an exception and almost similar results are obtained with all potentials.

As mentioned, the DCM results are not in good agreement with experiment but their insensitivity to the applied potential might give a hint that there is already some problem with the underlying solute/solvent configurations. Computing reliable structures might now be the most challenging part in theoretical studies on solvatochromism.[28]

One interesting finding can be derived from the QP1/scD3 and CM5/QP1 results. It seems that the agreement of QP1/QP1 and CM5/scD3 is caused by a compensation of opposing effects. When exchanging either QP1 point charges or polarizabilities with their respective counterpart from the CM5/scD3 potential, excitation energies are always lower for QP1/scD3 than for both original potentials and always higher for CM5/QP1. The shift is more pronounced for the point charges but overall, the combined potentials still yield acceptable results.

The actual reason for this behavior of the combined potentials has not been found. It might be that the averaged QP1 point charges and the averaged QP1 polarizabilities have some dependency. But as they are only derived from the same geometries but not fitted at the same time and also with different approaches, it is not obvious where the dependency is coming from.

Turning now to the convergence behavior, it can be found that the full M2P2 result is almost reached for the well performing M0P1 potentials when the parameters of solvent molecules within only 2.5 Å of the QM region are replaced. Only for water, the convergence is a little bit slower and a cutoff radius of 5.0 Å is required for the same accuracy. As expected, the less reliable CM5/D3 and Gas/scD3 variants converge much slower. But even for these cases, full M2P2 can almost be recovered by a cutoff radius of 5.0 Å. Because the solute/solvent configurations are obtained with a 12-Å-cutoff radius for the inclusion of solvent molecules, the inner MM region accounts only for about 25–30% with the 5-Å-cutoff radius.

GFP

Encouraged from the results of the previous section, we now turn to a related problem: the environmental effects of the protein surroundings on the spectral properties of the GFP chromophore. In the PE setup, it is even more beneficial to have an efficient way to obtain the MM parameters for a protein embedding potential because their determination is normally computationally much more demanding than for an isolated solvent molecule. Here, GFP serves as a test case for the pragmatic way of deriving isotropic polarizabilities from D3 coefficients. As additional test, PE-CAM-B3LYP results are also computed.

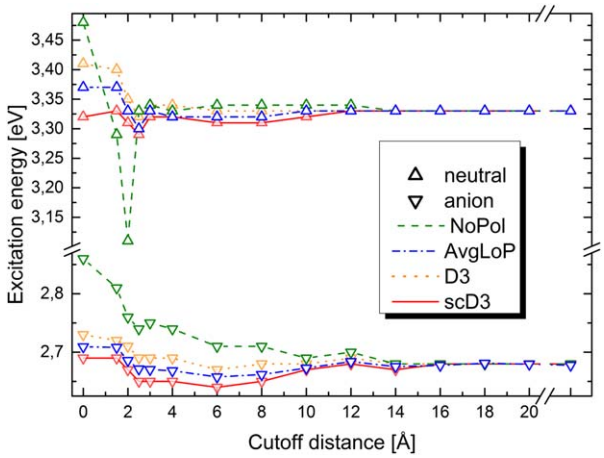

Figure 3. PERI-CC2 excitation energies of the $\pi-\pi^*$-state in neutral and anionic chromophore in GFP in dependence of the cutoff distance for different embedding potentials. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Amber94 point charges are chosen to model the electrostatic part of the embedding potential. This choice avoids additional QM computations to derive appropriate parameters, the force field parameters are well tested, and in a previous study, they have yielded very similar results to a M2P0 potential.[10] The point charges are combined with scD3 polarizabilities, unscaled D3 polarizabilities, and element-wise-averaged isotropic polarizabilities from LoProp (AvgLoP). The latter approach has been applied previously in the context of protein–ligand binding.[57] Based on increasing values for the cutoff radius, the more approximate potential parameters are

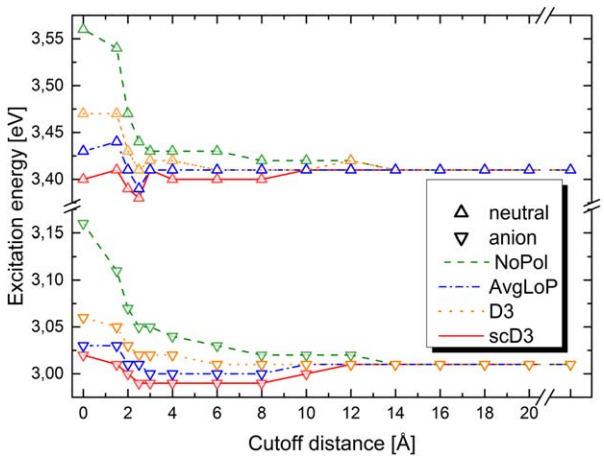

Figure 4. PE-TDDFT (CAM-B3LYP) excitation energies of the $\pi-\pi^*$-state in the neutral and anionic chromophore in GFP in dependence of the cutoff distance for different embedding potentials. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Table 2. PERI-CC2 (E_{exc}^{CC}) and PE-CAM-B3LYP (E_{exc}^{DFT}) excitation energies of the $\pi-\pi^*$-excitation in GFP obtained with a mixed M2P2(LoP)/M0P1(Amber94/scD3)-embedding potential in dependance of the cutoff distance.

	Neutral			Anionic			Shift	
Cutoff	%	E_{exc}^{DFT}	E_{exc}^{CC}	%	E_{exc}^{DFT}	E_{exc}^{CC}	ΔE_{exc}^{DFT}	ΔE_{exc}^{CC}
0.0	0.0	3.40	3.32	0.0	3.02	2.69	−0.39	−0.63
1.5	1.8	3.41	3.33	1.8	3.01	2.69	−0.39	−0.64
2.0	4.8	3.39	3.31	5.9	3.00	2.67	−0.40	−0.64
2.5	8.3	3.38	3.29	8.3	2.99	2.65	−0.39	−0.63
3.0	11.6	3.41	3.32	11.6	2.99	2.65	−0.42	−0.67
4.0	16.1	3.40	3.32	15.9	2.99	2.65	−0.41	−0.67
6.0	27.1	3.40	3.31	27.1	2.99	2.64	−0.41	−0.66
8.0	40.3	3.40	3.31	40.3	2.99	2.65	−0.41	−0.66
10.0	54.5	3.41	3.32	54.4	3.00	2.67	−0.41	−0.66
12.0	65.8	3.41	3.33	65.7	3.01	2.68	−0.40	−0.65
14.0	74.5	3.41	3.33	74.8	3.01	2.67	−0.40	−0.65
16.0	87.4	3.41	3.33	87.4	3.01	2.68	−0.40	−0.65
18.0	94.2	3.41	3.33	94.3	3.01	2.68	−0.40	−0.65
20.0	97.4	3.41	3.33	97.3	3.01	2.68	−0.40	−0.65
∞	100.0	3.41	3.33	100.0	3.01	2.68	−0.40	−0.65
→ exp.		3.05/3.12			2.63/2.59		−0.42/−0.53	

The respective percentage of MM atoms described by the M2P2 potential and relative shift (ΔE_{exc}) between excitation energies in both GFP systems is also shown. Experimental values (at 1.6 K and 295 K) are taken from Ref. [58]. Energies are given in eV, distances in Å.

replaced by their M2P2 counterparts. To check whether induced dipoles are necessary in the outer MM region, point charge only results (NoPol) are also computed. All data points are plotted in Figure 3 for PERI-CC2 and Figure 4 for PE-CAM-B3LYP.

It is instantly noticeable that the neutral and anionic form as well as PERI-CC2 and PE-CAM-B3LYP show qualitatively very similar results among each other. The excitation energies obtained with pure M0P1 potentials already show good results with respect to the full M2P2 potential whereby the deviations for scD3 and AvgLoP are less than 0.05 eV for all four cases, and less than 0.10 eV for D3 polarizabilities. When increasing the size of the M2P2 region, the excitation energies are hardly affected although the convergence is not smooth so that deviations for combined potentials might be larger than for the pure M0P1 potential. This might be an artifact of how the protein fragments with different potentials are mixed. For example, the NoPol plot for the neutral GFP system treated with PERI-CC2 shows such an outlier for the cutoff distance $R_{CO} = 2$ Å. It cannot be deduced if such artifacts do not appear for larger cutoff distances or if they just do not influence the excitation energies. In any case, the maximum deviation for the two more reliable M0P1 potentials never exceeds a deviation of 0.04 eV to the full M2P2 results.

The question, if the distinction of different atom types based on the CNs for the computation of the scD3 polarizabilities is benefical and causes the difference to the AvgLoP results, cannot be answered from the data. The deviation with respect to M2P2 is lower for small cutoff radii $R_{CO} \leq 3$ Å, especially for $R_{CO} = 0$ Å, but this might as well be coincidental.

When only point charges are used for the whole protein, deviations go up to about 0.15 eV. Roughly a quarter of the whole protein ($R_{CO} \sim 6$ Å) needs to be described by the M2P2 potential to give similar results to combined M2P2/M0P1 potentials, while about a half ($R_{CO} \sim 10$ Å) is needed for convergence.

It should be noted that all potentials show an almost constant relative shift between the neutral and anionic form. For scD3 results, this has been exemplarily compiled in Table 2. The shift for PE-CAM-B3LYP (∼−0.40 eV) is close to the experimental range of −0.42 (at 1.6 K) to −0.53 eV (at 295 K),[58] while PERI-CC2 (∼−0.65 eV) overestimates the shift. For absolute excitation energies, the converged PERI-CC2 values are within the aspired 0.1 eV error margin to the experiment for the excitation energy in the anionic form but overestimates the excitation energy of the neutral form by 0.2 eV, while PE-CAM-B3LYP strongly overestimates both by 0.3–0.4 eV.

Using the pure Amber94/scD3 practically nullifies the computational effort to obtain the embedding potential for GFP. It is hard to say though, if these findings are easily transferable to other proteins (or even completely different environments) and not just a coincidence. More experience for this is required. However, the excitation energies get relatively stable for a cutoff distance of $R_{CO} \geq 3$ Å for all embedding models. Thus static multipoles up to quadrupoles and isotropic polarizabilities should be employed for this region to obtain reliable results. This cutoff distance corresponds to 12% of the whole GFP environment, which still immensely reduces the computational effort to obtain embedding potentials to the same fraction.

Conclusions

We found that isotropic polarizabilities obtained from D3 dispersion coefficients correlate well with those obtained via the LoProp method. Using these polarizabilities in combination with appropriately derived point charges for the calculation of PNA excitation energies in different solvents gave comparable energies to those obtained via the LoProp method. While the best pure M0P1 potentials already give reasonable results,

M2P2 parameters for the innermost solvent shell ($\leq$5 Å) are needed for very high accuracy though.

Using the scaled D3 polarizabilities in combination with standard force field point charges from Amber94 for the calculation of GFP's excitation energies, yields excellent results. The use of a pure M0P1 embedding potential already reproduces the energies obtained with a full M2P2 potential. One should, however, take care when transferring these results to other proteins. For the innermost shell ($\leq$3 Å) a geometry-specific M2P2 embedding potential can be recommended as a save estimate for already obtaining results close to a full M2P2 treatment.

Similar results can also be found for average element-wise LoProp parameters. In previous studies concerning polarizable force fields for protein–ligand interactions, a higher sensitivity to the quality of the polarizabilities has been found[57] but this problem also requires a higher accuracy than the envisaged error of 0.1 eV of this study. Of course, more refined parameters can be (and have been) assigned based on a better discrimination of different atom types. Such a force field-like approach looses the generality of the present alternative which can be applied with a minimum of curation by the user.

The procedure introduced here is principally transferable to all environments—proteins, solutions, or even other cases like surfaces or polymers— provided that appropriate point charges are accessible. D3 polarizabilities can be easily obtained from the respective structures. Using the scaled D3 polarizabilities with appropriate point charges for the major part of the environment significantly reduces the computational effort to obtain embedding potentials, while almost keeping the accuracy of the full M2P2 LoProp embedding potential for PE-calculations.

Employing these parameters makes the computation of multiple protein- and solvent configurations feasible.

Acknowledgments

We are indebted to Beerepoot et al. for letting us see in manuscript a forthcoming paper on the same general problem and providing their data for our use in advance.

Keywords: *ab initio* calculations · density functional calculations · polarizable embedding · solvatochromism · fluorescent protein

How to cite this article: H. Schröder, T. Schwabe. *J. Comput. Chem.* **2016**, *37*, 2052–2059. DOI: 10.1002/jcc.24425

[1] J. Kongsted, A. Osted, K. V. Mikkelsen, P. O. Åstrand, O. Christiansen, *J. Chem. Phys.* **2004**, *121*, 8435.
[2] J. M. Olsen, K. Aidas, J. Kongsted, *J. Chem. Theory Comput.* **2010**, *6*, 3721.
[3] D. Bakowies, W. Thiel, *J. Phys. Chem.* **1996**, *100*, 10580.
[4] K. Ohta, Y. Yoshioka, K. Morokuma, K. Kitaura, *Chem. Phys. Lett.* **1983**, *101*, 12.
[5] K. Kitaura, E. Ikeo, T. Asada, T. Nakano, M. Uebayasi, *Chem. Phys. Lett.* **1999**, *313*, 701.
[6] W. Xie, J. Gao, *J. Chem. Theory Comput.* **2007**, *3*, 1890.
[7] C. R. Jacob, J. Neugebauer, L. Visscher, *J. Comput. Chem.* **2007**, *29*, 1011.
[8] K. Sneskov, T. Schwabe, J. Kongsted, *Phys. Chem. Chem. Phys.* **2011**, *13*, 18551.
[9] P. Söderhjelm, C. Husberg, A. Strambi, M. Olivucci, U. Ryde, *J. Chem. Theory Comput.* **2009**, *5*, 649.
[10] T. Schwabe, M. T. P. Beerepoot, J. M. H. Olsen, J. Kongsted, *Phys. Chem. Chem. Phys.* **2015**, *17*, 2582.
[11] J. R. J. R. Lewandowski, J. N. N. Dumez, U. Akbey, S. Lange, L. Emsley, H. Oschkinat, *J. Phys. Chem. Lett.* **2011**, *2*, 2205.
[12] T. Schwabe, J. M. H. Olsen, K. Sneskov, J. Kongsted, O. Christiansen, *J. Chem. Theory Comput.* **2011**, *7*, 2209.
[13] C. Daday, C. Curutchet, A. Sinicropi, B. Mennucci, C. Filippi, *J. Chem. Theory Comput.* **2015**, *11*, 4825.
[14] K. Sneskov, T. Schwabe, J. Kongsted, O. Christiansen, *J. Chem. Phys.* **2011**, *134*, 104108.
[15] W. D. Cornell, P. Cieplak, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, P. A. Kollman, *J. Am. Chem. Soc.* **1995**, *117*, 5179.
[16] W. L. Jorgensen, D. S. Maxwell, J. Tirado-Rives, *J. Am. Chem. Soc.* **1996**, *118*, 11225.
[17] A. V. Marenich, S. V. Jerome, C. J. Cramer, D. G. Truhlar, *J. Chem. Theory Comput.* **2012**, *8*, 527.
[18] J. Gasteiger, M. Marsili, *Tetrahedron* **1980**, *36*, 3219.
[19] M. T. P. Beerepoot, A. H. Steindal, N. H. List, J. Kongsted, J. M. H. Olsen, *J. Chem. Theory Comput.* **2016**, 1684.
[20] A. J. Stone, *J. Chem. Phys. Lett.* **1981**, *83*, 233239.
[21] A. J. Stone, M. Alderton, *Mol. Phys.* **2002**, *100*, 221.
[22] L. Gagliardi, R. Lindh, G. Karlström, *J. Chem. Phys.* **2004**, *121*, 4494.
[23] U. C. Singh, P. A. Kollman, *J. Comput. Chem.* **1984**, *5*, 129.
[24] Boltzmann, L. *Vorlesungen über Gastheorie, Bd. 2.*; Verlag Barth, Leipzig, **1898**, *34*, 96.
[25] M. T. P. Beerepoot, A. H. Steindal, J. Kongsted, B. O. Brandsdal, L. Frediani, K. Ruud, J. M. H. Olsen, *Phys. Chem. Chem. Phys.* **2013**, *15*, 4735.
[26] J. P. Zobel, J. J. Nogueira, L. Gonza, *J. Phys. Chem. Lett.* **2015**, *6*, 3006.
[27] D. W. Zhang, J. Z. H. Zhang, *J. Chem. Phys.* **2003**, *119*, 3599.
[28] T. Schwabe, *J. Phys. Chem. B* **2015**, *119*, 10693.
[29] M. Chattoraj, B. A. King, G. U. Bublitz, S. G. Boxer, *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 8362.
[30] S. Grimme, J. Antony, S. Ehrlich, H. Krieg, *J. Chem. Phys.* **2010**, *132*, 154104.
[31] S. Grimme, *J. Chem. Theory Comput.* **2014**, *10*, 4497.
[32] H. B. G. Casimir, D. Polder, *Phys. Rev.* **1948**, *73*, 360.
[33] G. Karlström, *Theor. Chim. Acta* **1982**, *60*, 535.
[34] M. T. P. Beerepoot, A. H. Steindal, K. Ruud, J. M. H. Olsen, J. Kongsted, *Comput. Theor. Chem.* **2014**, 1040, *304*.
[35] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, D. J. Fox, Gaussian 09, Rev. D.01. Gaussian Inc. Wallingford CT, **2009**.
[36] N. M. O'Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch, G. R. Hutchison, *J. Cheminform.* **2011**, *3*, 33.
[37] The Open Babel Package, version 2.3.1. Available at: http://openbabel.org, accessed Jan 24, 2016.
[38] H. Schröder, A. Creon, T. Schwabe, *J. Chem. Theory Comput.* **2015**, *11*, 3163.
[39] Aidas, K. Whirlpool, a QM/MM analysis program, 1.0. **2010**.
[40] T. Schwabe, K. Sneskov, J. M. Haugaard Olsen, J. Kongsted, O. Christiansen, C. Hättig, *J. Chem. Theory Comput.* **2012**, *8*, 3274.
[41] C. Hättig, F. Weigend, *J. Chem. Phys.* **2000**, *113*, 5154.

[42] TURBOMOLE, V6.6, **2014**, developed by University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, TURBOMOLE GmbH, since 2007; Available at: http://www.turbomole.com, accessed Jan 24, 2016.
[43] T. H. Dunning, *J. Chem. Phys.* **1989**, *90*, 1007.
[44] D. E. Woon, T. H. Dunning, Jr., *J. Chem. Phys.* **1995**, *103*, 2975.
[45] F. Weigend, A. Köhn, C. Hättig, *J. Chem. Phys.* **2002**, *116*, 3175.
[46] D. Rappoport, F. Furche, *J. Chem. Phys.* **2010**, *133*, 134105.
[47] K. Aidas, C. Angeli, K. L. Bak, V. Bakken, R. Bast, L. Boman, O. Christiansen, R. Cimiraglia, E. K. Dalskov, U. Ekstr, S. Coriani, T. Enevoldsen, J. J. Eriksen, P. Ettenhuber, B. Fern, L. Ferrighi, H. Fliegl, L. Frediani, E. Hjertenæs, S. Høst, I. Høyvik, D. Jonsson, P. Jørgensen, J. Kauczor, S. Kirpekar, T. Kjærgaard, W. Klopper, S. Knecht, R. Kobayashi, H. Koch, J. Kongsted, A. Krapp, K. Kristensen, A. Ligabue, O. B. Lutnæs, J. I. Melo, K. V. Mikkelsen, R. H. Myhre, C. Neiss, C. B. Nielsen, P. Norman, M. H. Olsen, A. Osted, P. F. Provasi, S. Reine, Z. Rinkevicius, K. Sneskov, A. H. Steindal, K. O. Sylvester-hvid, P. R. Taylor, A. M. Teale, E. I. Tellgren, D. P. Tew, A. J. Thorvaldsen, L. Thøgersen, O. Vahtras, M. A. Watson, D. J. D. Wilson, *WIREs Comput. Mol. Sci.* **2014**, *4*, 269.
[48] Dalton, a molecular electronic structure program, Release Dalton2015.X (**2015**). Available at: http://daltonprogram.org, accessed Jan 24, 2016.
[49] A. D. Becke, *J. Chem. Phys.* **1993**, *98*, 5648.
[50] C. Lee, W. Yang, R. G. Parr, *Phys. Rev. B* **1988**, *37*, 785.
[51] Y. Tawada, T. Tsuneda, S. Yanagisawa, T. Yanai, K. Hirao, *J. Chem. Phys.* **2004**, *120*, 8425.
[52] T. Yanai, D. P. Tew, N. C. Handy, **2004**, *393*, 51.
[53] M. Häser, R. Ahlrichs, *J. Comput. Chem.* **1989**, *10*, 104.
[54] W. J. Hehre, R. Ditchfield, J. A. Pople, *J. Chem. Phys.* **1972**, *56*, 2257.
[55] P. C. Hariharan, J. A. Pople, *Theor. Chim. Acta* **1973**, *28*, 213.
[56] T. Clark, J. Chandrasekhar, G. W. Spitznagel, P. V. R. Schleyer, *J. Comput. Chem.* **1983**, *4*, 294.
[57] P. Söderhjelm, J. Kongsted, U. Ryde, *J. Chem. Theory Comput.* **2011**, *7*, 1404.
[58] T. M. H. Creemers, A. J. Lock, V. Subramaniam, *Nat. Struct. Biol.* **1999**, *6*, 557.

Received: 28 January 2016
Revised: 3 May 2016
Accepted: 23 May 2016
Published online on 18 June 2016