

# Polarizable Density Embedding for Large Biomolecular Systems

Peter Reinholdt,<sup>\*,†</sup> Frederik Kamper Jørgensen,<sup>†</sup> Jacob Kongsted,<sup>†</sup> and Jógvan Magnus Haugaard Olsen<sup>\*,‡,¶</sup>

<sup>†</sup>*Department of Physics, Chemistry and Pharmacy, University of Southern Denmark, DK-5230 Odense M, Denmark*

<sup>‡</sup>*Hylleraas Centre for Quantum Molecular Sciences, Department of Chemistry, UiT The Arctic University of Norway, N-9037 Tromsø, Norway*

<sup>¶</sup>*Department of Chemistry, Aarhus University, DK-8000 Aarhus C, Denmark*

E-mail: reinholdt@sdu.dk; jmho@chem.au.dk

## Abstract

We present an efficient and robust fragment-based quantum–classical embedding model capable of accurately capturing effects from complex environments such as proteins and nucleic acids. This is realized by combining the molecular fractionation with conjugate caps (MFCC) procedure with the polarizable density embedding (PDE) model at the level of Fock matrix construction. The Fock matrix of the core region is constructed using the local molecular basis of the individual fragments rather than the supermolecular basis of the entire system. Thereby, we avoid complications associated with the application of the MFCC procedure on environment quantities such as electronic densities and molecular-orbital energies. Moreover, the computational cost associated with solving self-consistent field (SCF) equations of the core region remains unchanged from that of purely classical polarized embedding models. We analyze the

performance of the resulting model in terms of the reproduction of the electrostatic potential of an insulin monomer protein and further in the context of solving problems related to electron spill-out. Finally, we showcase the model for the calculation of one- and two-photon properties of the Nile Red molecule in protein environments. Based on our analyses, we find that the combination of the MFCC approach with the PDE model is an efficient, yet accurate approach for calculating molecular properties of molecules embedded in structured biomolecular environments.

Quantum chemical methods enable calculations of a wide range of molecular properties with high accuracy.<sup>1</sup> However, the high computational cost and steep scaling of conventional quantum chemical methods limit practical applications to single molecules or smaller molecular clusters. Fortunately, in many molecular systems, including biomolecular ones, the electrons are fairly localized. For such systems, it becomes feasible to calculate local molecular properties by treating only a smaller part of a composite system with a quantum-mechanics-based method. For example, a large part of the electronic and vibrational spectrum of a solvated chromophore can be obtained from calculations where it is only the chromophore itself that is described by quantum mechanics. Nonetheless, the environment can substantially alter the properties of the central part and, therefore, cannot be neglected in general.

Environmental effects can be included effectively through the use of embedding methods. Here the total system is partitioned into two subsystems: the core part, which is the system of main interest, and its environment. The central component of an embedding method is the embedding potential produced by the environment, through which the wavefunction of the embedded core part interacts with its environment. Many different embedding methods exist, and these can be roughly divided into three groups according to their description of the environment: structure-less continuum embedding,<sup>2,3</sup> classical atomistic embedding,<sup>4-6</sup> and quantum embedding.<sup>7-9</sup> Continuum models have proven to be highly successful for capturing bulk solvent effects due to their efficiency and ease of use. However, they are less suited for heterogeneous environments and systems that involve highly directional interactions be-

tween the core and its environment, such as hydrogen bonds or  $\pi$ -interactions, which are typical characteristics of biomolecular systems. For such systems, retaining an atomistic environment will perform better. This is the strategy used by quantum mechanics/molecular mechanics (QM/MM)-type embedding models<sup>10</sup> as well as more advanced variants, such as the hybrid quantum mechanics/effective fragment potential (QM/EFP) model<sup>11-15</sup> and the polarizable embedding (PE) model.<sup>16-19</sup> One of the main limitations of classical atomistic models is the lack of exchange repulsion, which results in an inadequate description of short-range interactions. It manifests itself as electron spill-out and spurious molecular orbitals.<sup>20-26</sup> To avoid these issues, quantum embedding models, which include exchange repulsion in some form, can be used. However, their applicability to biomolecular systems is limited by a comparably high computational cost; i.e., while such models are very efficient compared to conventional quantum chemistry methods, they are still too expensive compared to the requirements of, e.g., spectroscopic applications on large biomolecular systems, which easily consist of several tens or even hundreds of thousands of atoms and, moreover, require statistical sampling and thus evaluation of a large number of structures. There are thus limits on the size of systems that can be treated and the level of theory that can be used to describe the environment. The main bottleneck stems from the fact that either a quantum-mechanics-based calculation on the full system is required or, when relying on fragmentation procedures to split the environment into smaller fragments, an iterative procedure is needed in order to include environment polarization. The latter iterative procedure involves successive quantum-mechanics-based calculations on the individual fragments (often referred to as freeze-and-thaw cycles), and, while it is not strictly required, it is often important in order to obtain good accuracy.

The polarizable density embedding (PDE) model<sup>22</sup> is a fragment-based mixed quantum-classical embedding model where the permanent charge distributions of environment fragments are described quantum mechanically by their electronic densities and nuclear charges, while the induced charge distributions are described classically through an induced-dipole

model based on localized dipole–dipole polarizabilities. It was introduced as an extension of the PE model in order to improve electrostatic interactions and to avoid electron spill-out. The classical description of polarization mitigates the high cost of pure quantum embedding and provides a relatively accurate description of polarization. Exchange repulsion is modeled by a Huzinaga–Cantu-like projection operator<sup>27</sup> which effectively solves the electron spill-out problem.<sup>24</sup> The total embedding operator thus consists of three terms that describe the electrostatic, induction, and exchange-repulsion interactions between the core region and its environment.

The electrostatic operator ( $\hat{v}^{\text{es}}$ ) describes Coulomb interactions between the core region and the permanent charge densities of the environment. It is constructed from the fragment-based density matrices ( $\mathbf{D}$ ) as well as nuclear charges ( $Z$ ) and positions ( $\mathbf{R}$ ), together with the corresponding two-electron repulsion integrals ( $v_{\mu\nu,\gamma\delta}$ ) and one-electron nuclear attraction integrals ( $v_{\mu\nu}(\mathbf{R})$ ). The contribution of this operator to the Fock matrix in an atomic orbital (AO) basis is defined as

$$v_{\mu\nu}^{\text{es}} = \sum_{f=1}^F \sum_{\gamma\delta \in f} D_{\gamma\delta} v_{\mu\nu,\gamma\delta} - \sum_{f=1}^F \sum_{m \in f}^{M_f} Z_m v_{\mu\nu}(\mathbf{R}_m) \quad (1)$$

where  $\mu$  and  $\nu$ , here and in the following equations, are indices of AOs belonging to the core region and  $\gamma$  and  $\delta$  belong to AOs of fragments in the environment. The first term of eq 1 accounts for the repulsion between electrons in the core region and electrons in the environment, while the second term accounts for the attraction between electrons in the core region and nuclei in the environment. The density matrices are optimized for each fragment in isolation and then kept frozen during the optimization of the core-region electronic density, consistent with the fact that the polarization is described classically (as described below).

The induction operator ( $\hat{v}^{\text{ind}}$ ) takes into account the polarization of the environment through localized dipole–dipole polarizabilities ( $\boldsymbol{\alpha}$ ) which are calculated for each fragment in isolation. Apart from the polarizabilities, the electronic electric-field integrals ( $\mathbf{t}_{\mu\nu}(\mathbf{R})$ )



are also needed to construct the operator. Its contribution to the Fock matrix is given by

$$v_{\mu\nu}^{\text{ind}} = - \sum_{f=1}^F \sum_{s \in f}^{S_f} \boldsymbol{\mu}_s \mathbf{t}_{\mu\nu}(\mathbf{R}_s) \quad (2)$$

where the induced dipoles ( $\boldsymbol{\mu}$ ) are determined by the total electric field ( $\mathbf{F}$ ) exerted on the polarizable sites, i.e.  $\boldsymbol{\mu}_s = \boldsymbol{\alpha}_s \mathbf{F}(\mathbf{R}_s)$ . Mutual polarization between the core region and environment is included by updating the induced dipoles during the optimization if the core-region electronic density.

The exchange-repulsion operator ( $\hat{v}^{\text{rep}}$ ) is based on a Huzinaga–Cantu-like projection<sup>27</sup> using energy-weighted density matrices ( $\mathbf{W}$ ) and orbital overlaps ( $\mathbf{S}$ ). The corresponding Fock matrix element is written as

$$v_{\mu\nu}^{\text{rep}} = - \sum_{f=1}^F \sum_{\gamma\delta \in f} W_{\gamma\delta} S_{\mu\gamma} S_{\nu\delta} \quad (3)$$

where  $W_{\gamma\delta}$  is an element of the density matrix of a fragment in the environment weighted by the occupied molecular orbital (MO) energies ( $\varepsilon$ ), i.e.,  $W_{\gamma\delta} = \sum_{i=1}^{N_{\text{occ}}^f} \varepsilon_i C_{\gamma i} C_{\delta i}$  with  $C$  being MO coefficients. This operator models the effect of exchange repulsion and thus effectively prevents the electrons in the core region from penetrating into the environment by modeling the effect of exchange repulsion. It is approximate because we do not solve the Huzinaga–Cantu equations,<sup>27</sup> which would require either a calculation on the full system or an iterative procedure. Instead, the fragment densities are determined in isolation, and the repulsion is dampened by a transferable prefactor.<sup>22,28</sup> Exchange repulsion is a crucial component of the PDE model (and generally in quantum embedding) needed to avoid variational collapse, whereas it can sometimes be neglected when using conventional classical embedding, though with the risk of erroneous results.

In practice, the quantities in eqs (1) and (3) are computed in the local molecular basis of the individual fragments (and not the supermolecular basis). The evaluation of these

operators therefore scales linearly with the number of fragments in the environment. In the present implementation, the evaluation of the induction operator in eq (2) scales quadratically with the number of polarizable sites. However, this can be reduced to linear scaling by introducing, e.g., a fast multipole method<sup>29,30</sup> for the evaluation of the electric fields from the induced dipoles.

The PDE model has so far only been considered for solute–solvent systems.<sup>22–25,28,31,32</sup> When tackling large molecular systems, fragmentation of the environment can be used to circumvent the poor scaling of the underlying methods used to derive the embedding potential. This is the approach used in both the PE and PDE models, as detailed above for the latter. For solute–solvent systems, the selection of individual molecules as fragments is straightforward, but more care is needed when treating typical biological macromolecules such as proteins, which often contain several thousands of atoms. For such systems, the fragmentation method needs to handle the cutting of covalent bonds. In the PE model, inspired by the work of Ryde and coworkers,<sup>33,34</sup> the method of molecular fractionation with conjugate caps (MFCC)<sup>35</sup> was adopted to partition larger molecules into small, computationally manageable fragments. Specifically, Söderhjelm and Ryde<sup>33</sup> used the method to calculate localized properties (multipoles and polarizabilities) of proteins in order to derive a highly accurate force field.

Zhang and Zhang<sup>35</sup> first introduced the MFCC method in the context of calculating protein–ligand interaction energies<sup>35</sup> and later it was also used for nucleic-acid–ligand interaction energies.<sup>36</sup> In short, the MFCC method proposes the following: In order to calculate protein–ligand interaction energies, the protein is partitioned into a set of amino-acid-based fragments by cutting the peptide bonds (see figure 1). To saturate the open valencies caused by cutting covalent bonds, the fragments are capped by groups that consist of a number of atoms from the neighboring fragments. The procedure thus leads to a set of (partly) overlapping fragments. The capping groups of adjacent capped fragments are combined to create new fragments called conjugate caps (concap). The total protein–ligand interaction

energy is then approximated as the sum of the capped-fragment–ligand interaction energies, minus the sum of the concap–ligand interaction energies, thus removing double-counting due to the overlapping groups. The accuracy of the method can be systematically improved by increasing the size of the capping groups. Shortly after the original method was presented, it was extended to (real-space) electronic densities (and thereby total energies), electrostatic potentials, and total dipole moments,<sup>37</sup> and later the method was further extended in many different directions.<sup>38–47</sup>

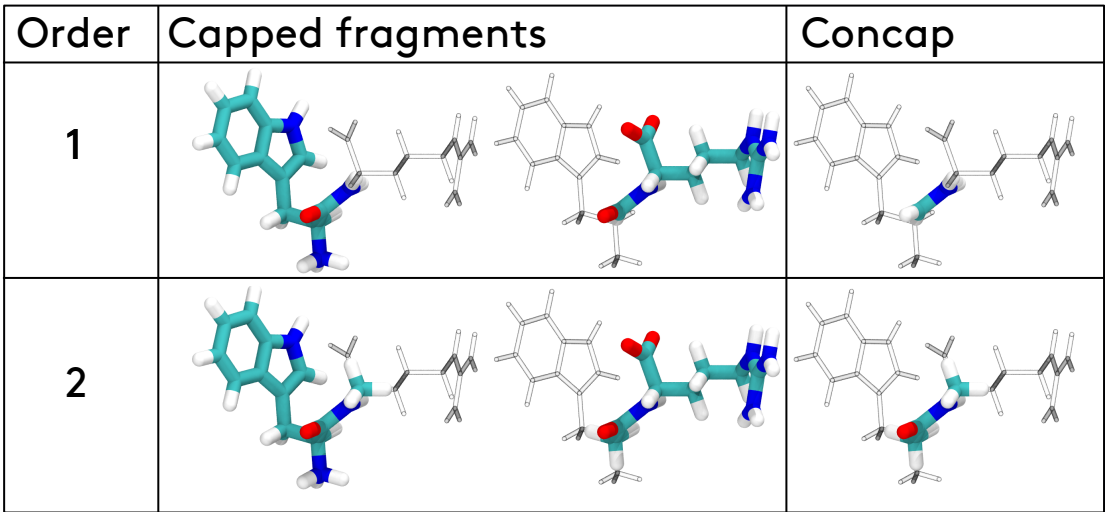


Figure 1: Illustration of molecular fractionation with conjugate caps (MFCC). A dipeptide (TRP-ARG) is split into two capped fragments and a concap molecule (colored sticks). The fragmentation procedure can be systematically improved in terms of an MFCC order, which increases the number of bonded non-hydrogen atoms used for the capping groups (here shown for orders 1 and 2).

Applying the MFCC procedure for the individual quantities of the embedding potential in the PDE model is not straightforward for the construction of the electrostatic and exchange-repulsion operators (eqs 1 and 3 respectively). Whereas MFCC can be directly applied to derive the localized polarizabilities needed for the induction operator (eq 2), it is more involved for fragment density and overlap matrices, and it is entirely unclear how to handle MO energies and coefficients. However, the operators only contain contributions that depend on one fragment,  $f$ , of the environment at a time. Therefore, the fragmentation procedure can be simply realized at the level of Fock matrix contributions, which is the strategy we

present here. The electrostatic and exchange-repulsion contributions to the Fock matrix are first calculated fragment-wise, after which all of the contributions are combined according to the MFCC approach. The Fock matrix contributions are thus obtained according to

$$v_{\mu\nu}^x = \sum_{f=1}^F v_{\mu\nu}^{x,f} - \sum_{c=1}^C v_{\mu\nu}^{x,c} \quad (4)$$

where  $x$  indicates either the electrostatic or exchange-repulsion contribution,  $f$  and  $c$  are indices associated with the capped fragments or concaps, respectively. The matrix elements  $v_{\mu\nu}^{x,f}$  and  $v_{\mu\nu}^{x,c}$  are calculated according to eqs 1 and 3.

The fragmentation and application of the MFCC procedure are implemented in the PyFraME<sup>48</sup> package, while the PDE model is implemented in the Polarizable Embedding library (PElib),<sup>49</sup> which is interfaced to the Dalton program.<sup>50</sup> Our implementation relies on the HDF5 file format<sup>51</sup> for storage and transfer of embedding quantities, which is expected to permit relatively easy incorporation of the model into other quantum-chemistry codes. The implementations are made available together with the Dalton2020 release.

Before moving on, we note that this is not the first use of MFCC in the context of embedding. Apart from the PE model and the work by Söderhjelm et al., as mentioned above, the MFCC procedure has also been applied to frozen density embedding (FDE)<sup>52</sup> by Jacob and Visscher<sup>53</sup> who presented a three-partition FDE (3-FDE) model.<sup>53,54</sup> In this model, the density is mapped on a real-space grid, and the MFCC procedure can, therefore, be straightforwardly applied. However, the density can become negative in the capping-group regions, which is problematic for a model that employs density-functional exchange-correlation kernels. To avoid this issue, special constraints are required to ensure positive densities.<sup>53</sup> Benchmarks of the 3-FDE scheme showed good performance in terms of the quality of electronic densities and electrostatic potentials of peptides and proteins.<sup>55</sup> It has also been used to investigate environmental effects in protein systems.<sup>56,57</sup>

One of the most important aspects of a classical embedding model is the ability to

reproduce the electrostatic potential (ESP) created by the environment. For purely classical embedding, the ESP is the main factor determining the quality of the environment description.<sup>58</sup> Figure 2 shows the errors in the ESP of an insulin monomer relative to a DEC-MP2<sup>59,60</sup>/cc-pVDZ reference<sup>61</sup> for two density-functional approximations and the PE and PDE potentials. The ESP errors are mapped onto a van der Waals (vdW) surface of the protein, with larger errors leading to saturated colors (blue or red). For this example, and in the following, an MFCC order of 2 was adopted (see figure 1), since initial tests revealed that using larger orders did not provide substantial improvement (see the Supporting Information). The PDE- and PE-based ESPs are derived from fragment calculations using CAM-B3LYP<sup>62</sup>/cc-pVDZ, i.e., using the same basis set as was used in the reference MP2 calculation. As reported by Jakobsen et al.,<sup>61</sup> it is immediately apparent that large errors in the ESP are found when using the B3LYP functional. The poor reproduction of the ESP occurs due to severe self-interaction errors leading to incorrect delocalization of electrons, which may be exacerbated by the fact that there is not a stabilizing environment surrounding the protein.<sup>55</sup> This highlights the need for using range-separated exchange–correlation functionals when treating large molecular systems. Accordingly, the range-separated CAM-B3LYP functional yields an ESP that is almost identical to the reference MP2-based ESP.

The ESP errors in Figure 2 are shown on two vdW surfaces, one which is further away from the nuclei (surface scaling factor 2.0) and one which is much closer (surface scaling factor 1.0). For the former, the errors in the reproduction of the ESP are comparable between PE, PDE, and CAM-B3LYP with root-mean-square deviations (RMSDs) of about 6.2, 5.4, and 7.7 kJ/mol, respectively. At this distance, the multipole expansion used in the PE potential is convergent, and the performance of the PE and PDE potentials is therefore similar. Some differences can be expected from the use of different methods, i.e., MP2 for the reference and CAM-B3LYP for the PE and PDE potentials. When moving closer to the nuclei (surface scaling factor of 1.0), the CAM-B3LYP method gives the best reproduction of the ESP (RMSD: 10 kJ/mol), followed by the PDE potential (RMSD: 18 kJ/mol). The

higher error in the latter is mainly due to an incomplete description of polarization (since polarizabilities are truncated at the dipole–dipole level). Meanwhile, the quality of the PE potential has clearly deteriorated (RMSD: 38 kJ/mol) due to the divergence of the multipole expansions at short distances. Thus, we see that avoiding the multipole expansion used in the PE potential and instead retaining the fragment densities as in the PDE potential leads to substantial improvements in terms of the short-range electrostatics. Importantly, introducing a fragmentation procedure (MFCC) does not compromise the inherently good quality of the ESP produced by the PDE potential.

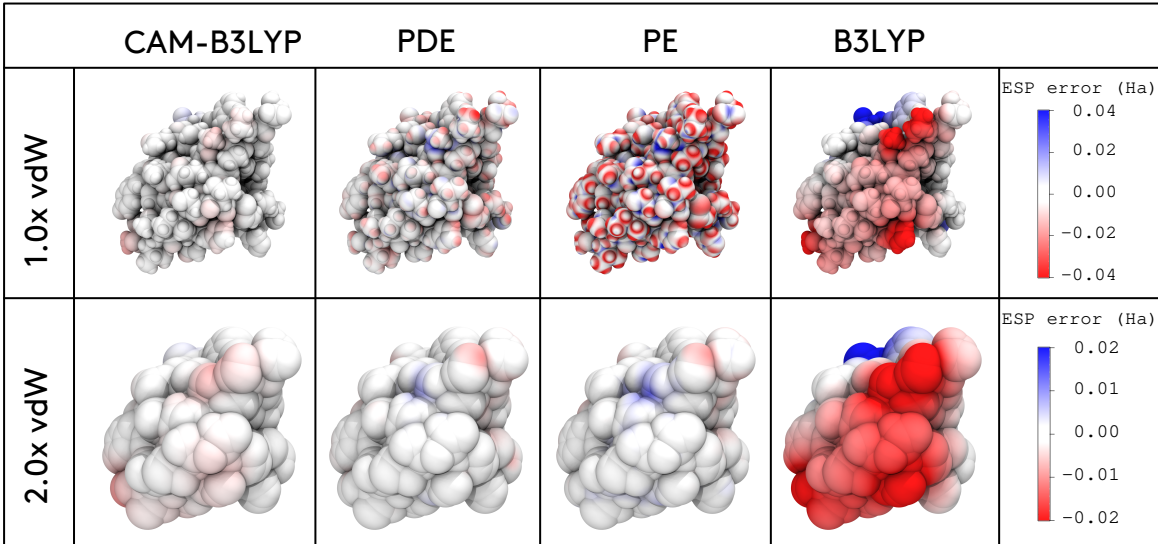


Figure 2: Electrostatic potential (ESP) error maps of insulin, computed using CAM-B3LYP, PE, PDE, or B3LYP. The reference ESP is based on a DEC-MP2 calculation.<sup>61</sup> In all cases, a cc-pVDZ basis set was used. The errors are shown on vdW surfaces using two scaling values: 1.0 (color scale  $\pm 0.04$  Ha) and 2.0 (color scale  $\pm 0.02$  Ha).

Beyond a better reproduction of ESPs, the major advantage of the PDE model over classical embedding models is the inclusion of exchange repulsion. The host–guest complex of 2,3-diazabicyclo[2.2.2]oct-2-en (DBO) chromophore inside cucurbit[7]uril (CB[7]) has previously been shown to suffer from electron spill-out,<sup>26,63</sup> with excitation energies being extremely redshifted when using the PE model to describe CB[7]. For this system, we calculated reference excitation energies based on supermolecular time-dependent Hartree–Fock (TDHF) and compared them to PE-TDHF and PDE-TDHF excitation energies of the DBO

molecule in CB[7] (see table 1). When using a non-augmented basis set (cc-pVDZ), both PE- and PDE-TDHF accurately reproduce the two lowest singlet excitation energies, which appear at around 3.5 and 6.5 eV for the supermolecular TDHF calculation. Both PE- and PDE-TDHF slightly overshoot the excitation energies, more so for the second state than the first (0.10 and 0.15 eV using PE, and 0.05 and 0.10 eV using PDE), but are in good quantitative agreement with the supermolecular reference. Upon augmentation of the basis set with diffuse functions (aug-cc-pVDZ<sup>64</sup>), the supermolecular TDHF excitation energies change only very slightly (less than 0.01 eV). Meanwhile, using PE-TDHF results in two extremely low-lying states at 1.5 and 1.8 eV, which are clearly unphysical. The correct behavior is restored by including exchange repulsion through the PDE model, which gives excitation energies that still are in good agreement with the supermolecular TDHF reference (the errors are the same as for cc-pVDZ). In this case, the problem with the PE model arises only when using diffuse basis functions since compact basis sets simply do not extend far enough into the environment to make the problem apparent.

Table 1: Excitation energies (in eV) of the DBO chromophore in CB[7] calculated using either PE-TDHF, PDE-TDHF, or supermolecular TDHF

Method State/Basis	PE		PDE		Supermolecular	
	cc-pVDZ	+aug	cc-pVDZ	+aug	cc-pVDZ	+aug
1	3.592	1.503	3.544	3.553	3.496	3.504
2	6.617	1.829	6.566	6.562	6.465	6.471

With the introduction of the MFCC method, the PDE model is now applicable to large biomolecular systems. As an illustrative example, we consider the case of the Nile Red chromophore embedded in the binding pocket of the  $\beta$ -lactoglobulin (BLG) protein. This system has been studied previously<sup>65,66</sup> due to the strong solvatochromism of the Nile Red dye. Figure 3 shows simulated one- and two-photon absorption spectra of Nile Red embedded in the protein and in the gas-phase. For the condensed-phase spectra, configurational sampling is included by extracting 125 snapshots from a QM/MM molecular dynamics (MD) trajectory.<sup>66</sup> In all cases, the singlet excitation energies and associated one- and two-photon

transition strengths are computed using the CAM-B3LYP functional along with the 6-31+G\* basis set. Local-field effects are taken into account through the effective external field (EEF) method.<sup>67</sup> The embedding-potential parameters are computed at the same level of theory. The spectra are generated using a Lorentzian broadening function for each transition with a line-width of  $1000\text{ cm}^{-1}$ . The condensed-phase spectra are averaged across 125 snapshots, while the gas-phase spectra are computed on the gas-phase optimized structure of Nile red.

The one-photon spectrum of Nile Red in the gas-phase has a bright transition at 2.95 eV. At higher energies, the gas-phase spectrum shows several distinct peaks with low intensity between 3.5 and 5.3 eV, followed by a major peak (composed of multiple transitions) at 5.7 eV. Introducing the protein and solvent broadens and redshifts the lowest band (by about 0.2 eV), irrespective of the model used to describe the environment. For this transition, the absorption bands are almost identical. For the transitions at higher excitation energies, the results are also similar, with several low-intensity bands between 4–5 eV, followed by an intense band at 5.5 eV (corresponding to a 0.2 eV redshift compared to the gas-phase result). However, unlike for the first transition, there are visible differences between the spectra where the environment is described by PE or PDE. In particular, in the low-intensity regions where the spectrum exhibits broader and less well-defined transition bands when the environment is modeled using PE. This difference can be understood by considering the underlying density of states, as shown in figure 4. The lowest-lying excitations (from 2.5–4.5 eV) are similar in both cases and relatively well-separated. When moving towards higher excitation energies, however, the spectrum based on the PE model becomes plagued by a much higher density of (partially unphysical) states. As a result, the state distribution, when using the PE model, shows a less well-defined structure. Moreover, to reach the same energetic range (i.e., converging the spectrum up to 6 eV) when using the PE model requires the solution of a larger number of states (40) than when using the PDE model (20), which can be seen from the number of color alternations along the  $y$ -axis of the histograms in figure 4. Overall, the PE and PDE models show almost identical performance for the low-energy, bright excitation



in Nile Red, and also yield quite similar one-photon spectra for higher energies, although the underlying physical description is clearly less satisfactory when using PE to model the environment.

The simulated two-photon spectra of Nile Red shown in figure 3 reveals larger differences between the PE and PDE models. The spectra are generated from the five lowest singlet excited states. The gas-phase spectrum is composed of three intense transitions located at 1.5, 1.9, and 2.1 eV. Both embedding models predict that the lowest-energy transition is redshifted by 0.1 eV when including the protein and solvent. Moreover, the intensities of this lowest excited state are almost identical. When moving towards higher energies, the spectrum based on the PDE model retains the basic characteristics of the gas-phase spectrum, although (due to broadening) the two peaks from the gas-phase now appear as one peak with a shoulder. In contrast, the spectrum based on the PE model shows no clearly defined structure and instead shows a broad featureless band starting from 1.6–2.1 eV. A correct description of the two-photon transition probability requires an accurate description of not only the initial and final states involved in the transition, but in principle, also all other excited states.<sup>68,69</sup> Thus, the artificial stabilization of higher-lying states and resulting overly large density of states at higher energies (see figure 4), is the root cause of the broader spectra found when using PE to model the environment. Overall, we find that the two-photon absorption spectrum is clearly more sensitive to the environment description than the one-photon absorption spectrum and that the PDE model yields a more physically reasonable result.

To sum up, we have presented a formulation of the PDE model that is suitable for biomolecular systems, which include large biomolecules such as proteins and nucleic acids. This formulation of PDE was realized based on the MFCC scheme at the level of Fock matrix construction, thus avoiding complications and issues related to the application of the fragmentation procedure for fragment quantities such as density matrices and MO energies. We have shown that the PDE potential produced via the MFCC procedure can accurately repro-

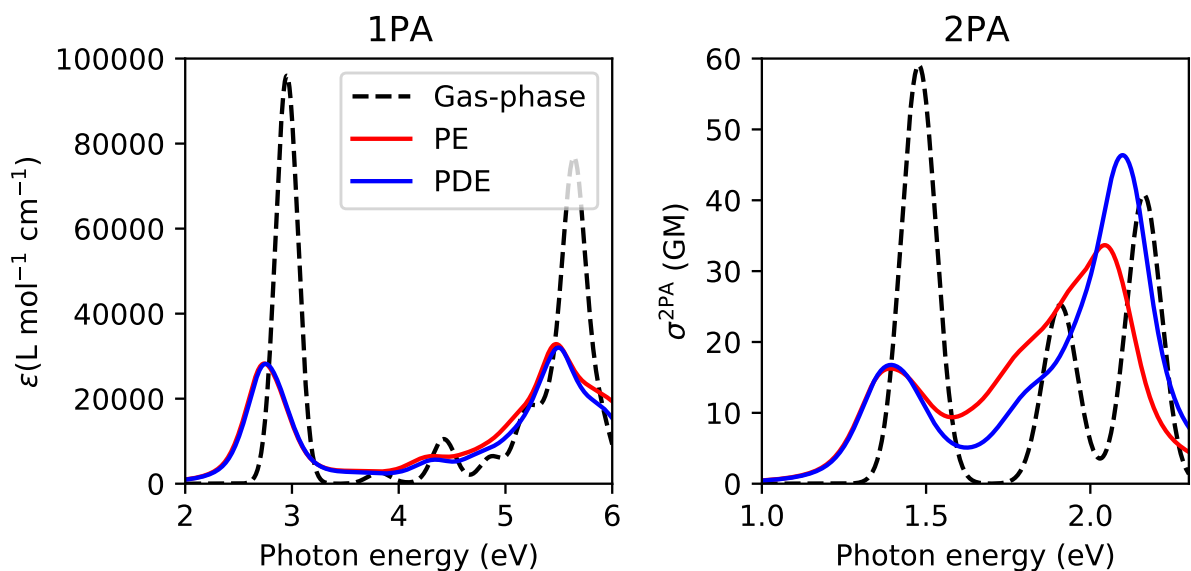


Figure 3: One- and two-photon absorption spectra of Nile Red in BLG. The environment is described using the PE model (dashed lines) or the PDE (solid lines) model.

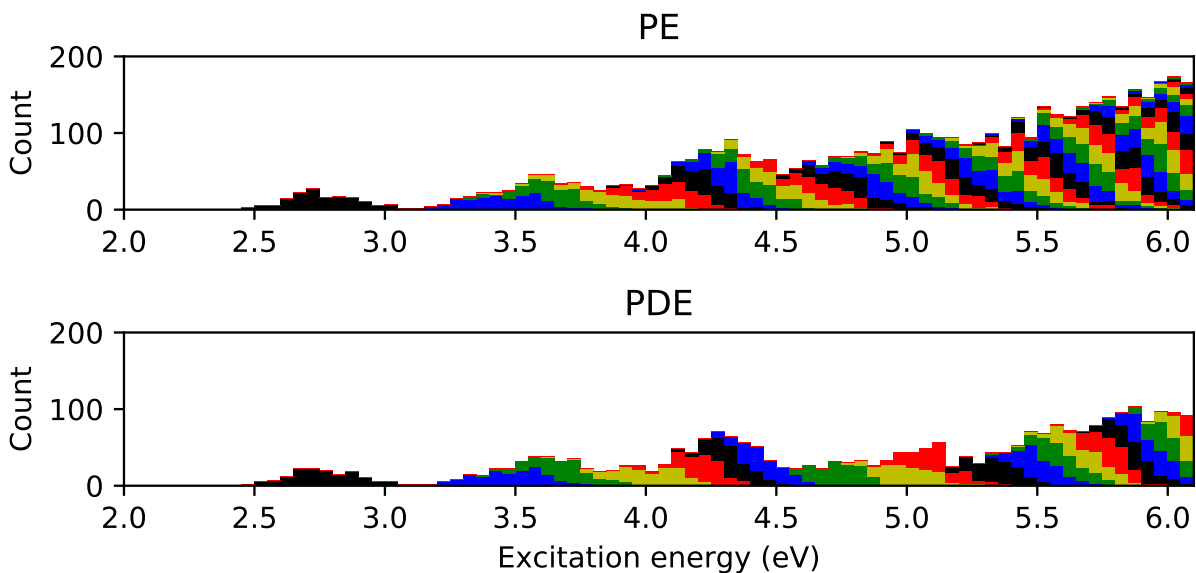


Figure 4: Stacked histograms of the excitation energies of Nile Red in a BLG environment represented by the PE model (top) or the PDE (bottom) model. The different colors correspond to individual excited states, energetically ordered.

duce ESPs of proteins even at short distances. Moreover, through illustrative examples, we demonstrated that this extension of PDE effectively solves electron spill-out issues and that it provides a physically correct and balanced description of excited states and spectroscopic properties such as two-photon absorption.

## Acknowledgement

Computational resources were provided by the DeIC National HPC Center at the University of Southern Denmark. J. M. H. O. acknowledges financial support from the Research Council of Norway through its Centres of Excellence scheme (Project ID: 262695). J. K. acknowledges the Danish Council for Independent Research for financial support (Grant ID: DFF-7014-00050B) and the H2020-MSCA-ITN-2017 COSINE Training network for COmputational Spectroscopy In Natural sciences and Engineering (Project ID: 765739) for financial support.

## Supporting Information Available

Detailed analyses of the influence of the MFCC capping order on the quality of the ESP reproduction.

## References

- (1) Helgaker, T.; Coriani, S.; Jørgensen, P.; Kristensen, K.; Olsen, J.; Ruud, K. Recent Advances in Wave Function-Based Methods of Molecular-Property Calculations. *Chem. Rev.* **2012**, *112*, 543–631.
- (2) Tomasi, J.; Mennucci, B.; Cammi, R. Quantum Mechanical Continuum Solvation Models. *Chem. Rev.* **2005**, *105*, 2999–3094.
- (3) Klamt, A.; Mennucci, B.; Tomasi, J.; Barone, V.; Curutchet, C.; Orozco, M.;

- Luque, F. J. On the Performance of Continuum Solvation Methods. A Comment on “Universal Approaches to Solvation Modeling”. *Acc. Chem. Res.* **2009**, *42*, 489–492.
- (4) Senn, H. M.; Thiel, W. QM/MM Methods for Biomolecular Systems. *Angew. Chem. Int. Ed.* **2009**, *48*, 1198–1229.
- (5) Brunk, E.; Rothlisberger, U. Mixed Quantum Mechanical/Molecular Mechanical Molecular Dynamics Simulations of Biological Systems in Ground and Electronically Excited States. *Chem. Rev.* **2015**, *115*, 6217–6263.
- (6) Morzan, U. N.; de Armiño, D. J. A.; Foglia, N. O.; Ramírez, F.; Lebrero, M. C. G.; Scherlis, D. A.; Estrin, D. A. Spectroscopy in Complex Environments from QM–MM Simulations. *Chem. Rev.* **2018**, *118*, 4071–4113.
- (7) Jacob, C. R.; Neugebauer, J. Subsystem density-functional theory. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2014**, *4*, 325–362.
- (8) Wesolowski, T. A.; Shedge, S.; Zhou, X. Frozen-Density Embedding Strategy for Multilevel Simulations of Electronic Structure. *Chem. Rev.* **2015**, *115*, 5891–5928.
- (9) Sun, Q.; Chan, G. K.-L. Quantum Embedding Theories. *Acc. Chem. Res.* **2016**, *49*, 2705–2712.
- (10) Warshel, A.; Levitt, M. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.* **1976**, *103*, 227–249.
- (11) Day, P. N.; Jensen, J. H.; Gordon, M. S.; Webb, S. P.; Stevens, W. J.; Krauss, M.; Garmer, D.; Basch, H.; Cohen, D. An effective fragment method for modeling solvent effects in quantum mechanical calculations. *J. Chem. Phys.* **1996**, *105*, 1968–1986.
- (12) Gordon, M. S.; Freitag, M. A.; Bandyopadhyay, P.; Jensen, J. H.; Kairys, V.;

- Stevens, W. J. The Effective Fragment Potential Method: A QM-based MM Approach to Modeling Environmental Effects in Chemistry. *J. Phys. Chem. A* **2001**, *105*, 293–307.
- (13) Yoo, S.; Zahariev, F.; Sok, S.; Gordon, M. S. Solvent effects on optical properties of molecules: A combined time-dependent density functional theory/effective fragment potential approach. *J. Chem. Phys* **2008**, *129*, 144112.
- (14) Slipchenko, L. V. Solvation of the excited states of chromophores in polarizable environment: Orbital relaxation versus polarization. *J. Phys. Chem. A* **2010**, *114*, 8824–8830.
- (15) Gordon, M. S.; Smith, Q. A.; Xu, P.; Slipchenko, L. V. Accurate First Principles Model Potentials for Intermolecular Interactions. *Annu. Rev. Phys. Chem.* **2013**, *64*, 553–578.
- (16) Olsen, J. M.; Aidas, K.; Kongsted, J. Excited States in Solution through Polarizable Embedding. *J. Chem. Theory Comput.* **2010**, *6*, 3721–3734.
- (17) Olsen, J. M. H.; Kongsted, J. *Adv. Quantum Chem.*; Elsevier, 2011; pp 107–143.
- (18) List, N. H.; Olsen, J. M. H.; Kongsted, J. Excited states in large molecular systems through polarizable embedding. *Phys. Chem. Chem. Phys.* **2016**, *18*, 20234–20250.
- (19) Steinmann, C.; Reinholdt, P.; Nørby, M. S.; Kongsted, J.; Olsen, J. M. H. Response properties of embedded molecules through the polarizable embedding model. *Int. J. Quantum Chem.* **2019**, *119*, e25717.
- (20) Fradelos, G.; Wesolowski, T. A. The Importance of Going beyond Coulombic Potential in Embedding Calculations for Molecular Properties: The Case of Iso-G for Biliverdin in Protein-Like Environment. *J. Chem. Theory Comput.* **2010**, *7*, 213–222.
- (21) Fradelos, G.; Wesolowski, T. A. Importance of the Intermolecular Pauli Repulsion in Embedding Calculations for Molecular Properties: The Case of Excitation Energies for a Chromophore in Hydrogen-Bonded Environments. *J. Phys. Chem. A* **2011**, *115*, 10018–10026.

- (22) Olsen, J. M. H.; Steinmann, C.; Ruud, K.; Kongsted, J. Polarizable Density Embedding: A New QM/QM/MM-Based Computational Strategy. *J. Phys. Chem. A* **2015**, *119*, 5344–5355.
- (23) Nãbo, L. J.; Olsen, J. M. H.; Holmgaard List, N.; Solanko, L. M.; Wüstner, D.; Kongsted, J. Embedding beyond electrostatics—The role of wave function confinement. *J. Chem. Phys.* **2016**, *145*, 104102.
- (24) Reinholdt, P.; Kongsted, J.; Olsen, J. M. H. Polarizable Density Embedding: A Solution to the Electron Spill-Out Problem in Multiscale Modeling. *J. Phys. Chem. Lett.* **2017**, *8*, 5949–5958.
- (25) Hršak, D.; Olsen, J. M. H.; Kongsted, J. Polarizable Density Embedding Coupled Cluster Method. *J. Chem. Theory Comput.* **2018**, *14*, 1351–1360.
- (26) Khah, A. M.; Khani, S. K.; Hättig, C. Analytic Excited State Gradients for the QM/MM Polarizable Embedded Second-Order Algebraic Diagrammatic Construction for the Polarization Propagator PE-ADC(2). *J. Chem. Theory Comput.* **2018**, *14*, 4640–4650.
- (27) Huzinaga, S.; Cantu, A. Theory of separability of many-electron systems. *J. Chem. Phys.* **1971**, *55*, 5543–5549.
- (28) Hršak, D.; Olsen, J. M. H.; Kongsted, J. Optimization and transferability of non-electrostatic repulsion in the polarizable density embedding model. *J. Comput. Chem.* **2017**, *38*, 2108–2117.
- (29) Rokhlin, V.; Greengard, L. A fast algorithm for particle simulations. *J. Comp. Phys* **1987**, *73*, 325–348.
- (30) Lipparini, F. General Linear scaling implementation of polarizable embedding schemes. *J. Chem. Theory Comput.* **2019**, *15*, 4312–4317.

- (31) Fahleson, T.; Olsen, J. M. H.; Norman, P.; Rizzo, A. A QM/MM and QM/QM/MM study of Kerr, Cotton–Mouton and Jones linear birefringences in liquid acetonitrile. *Phys. Chem. Chem. Phys.* **2018**, *20*, 3831–3840.
- (32) Hršak, D.; Nørby, M. S.; Coriani, S.; Kongsted, J. One-photon Absorption Properties from a Hybrid Polarizable Density Embedding/complex Polarization Propagator Approach for Polarizable Solutions. *J. Chem. Theory Comput.* **2018**, *14*, 2145–2154.
- (33) Söderhjelm, P.; Ryde, U. How Accurate Can a Force Field Become? a Polarizable Multipole Model Combined with Fragment-wise Quantum-mechanical Calculations. *J. Phys. Chem. A* **2008**, *113*, 617–627.
- (34) Söderhjelm, P.; Husberg, C.; Strambi, A.; Olivucci, M.; Ryde, U. Protein Influence on Electronic Spectra Modeled by Multipoles and Polarizabilities. *J. Chem. Theory Comput.* **2009**, *5*, 649–658.
- (35) Zhang, D. W.; Zhang, J. Z. H. Molecular fractionation with conjugate caps for full quantum mechanical calculation of protein–molecule interaction energy. *J. Chem. Phys.* **2003**, *119*, 3599–3605.
- (36) Chen, X. H.; Zhang, J. Z. H. Theoretical method for full ab initio calculation of DNA/RNA–ligand interaction energy. *J. Chem. Phys.* **2004**, *120*, 11386–11391.
- (37) Gao, A. M.; Zhang, D. W.; Zhang, J. Z.; Zhang, Y. An efficient linear scaling method for ab initio calculation of electron density of proteins. *Chem. Phys. Lett.* **2004**, *394*, 293–297.
- (38) Chen, X.; Zhang, Y.; Zhang, J. Z. H. An efficient approach for ab initio energy calculation of biopolymers. *J. Chem. Phys.* **2005**, *122*, 184105.
- (39) He, X.; Zhang, J. Z. H. A new method for direct calculation of total energy of protein. *J. Chem. Phys.* **2005**, *122*, 031103.

- (40) Mei, Y.; Zhang, D. W.; Zhang, J. Z. H. New Method for Direct Linear-Scaling Calculation of Electron Density of Proteins. *J. Phys. Chem. A* **2005**, *109*, 2–5.
- (41) He, X.; Zhang, J. Z. H. The generalized molecular fractionation with conjugate caps/molecular mechanics method for direct calculation of protein energy. *J. Chem. Phys.* **2006**, *124*, 184703.
- (42) Jiang, N.; Ma, J.; Jiang, Y. Electrostatic field-adapted molecular fractionation with conjugated caps for energy calculations of charged biomolecules. *J. Chem. Phys.* **2006**, *124*, 114112.
- (43) Chen, X. H.; Zhang, J. Z. H. Molecular fractionation with conjugated caps density matrix with pairwise interaction correction for protein energy calculation. *J. Chem. Phys.* **2006**, *125*, 044903.
- (44) Wang, X.; Liu, J.; Zhang, J. Z. H.; He, X. Electrostatically Embedded Generalized Molecular Fractionation with Conjugate Caps Method for Full Quantum Mechanical Calculation of Protein Energy. *J. Phys. Chem. A* **2013**, *117*, 7149–7161.
- (45) Jia, X.; Wang, X.; Liu, J.; Zhang, J. Z. H.; Mei, Y.; He, X. An improved fragment-based quantum mechanical method for calculation of electrostatic solvation energy of proteins. *J. Chem. Phys.* **2013**, *139*, 214104.
- (46) Brinkmann, L.; Heifets, E.; Kantorovich, L. Density functional calculations of extended, periodic systems using Coulomb corrected molecular fractionation with conjugated caps method (CC-MFCC). *Phys. Chem. Chem. Phys.* **2014**, *16*, 21252–21270.
- (47) Wang, Z.; Han, Y.; Li, J.; He, X. Combining the Fragmentation Approach and Neural Network Potential Energy Surfaces of Fragments for Accurate Calculation of Protein Energy. *J. Phys. Chem. B* **2020**,



- (48) Olsen, J. M. H. PyFraME: Python framework for Fragment-based Multiscale Embedding. 2020; <https://gitlab.com/FraME-projects/PyFraME>.
- (49) Olsen, J. M. H.; List, N. H.; Steinmann, C.; Steindal, A. H.; Nørby, M. S.; Reinholdt, P. PELib: The Polarizable Embedding library (v1.4.1). 2020; <https://gitlab.com/pe-software/pelib-public>.
- (50) Aidas, K.; Angeli, C.; Bak, K. L.; Bakken, V.; Bast, R.; Boman, L.; Christiansen, O.; Cimiraglia, R.; Coriani, S.; Dahle, P.; Dalskov, E. K.; Ekström, U.; Enevoldsen, T.; Eriksen, J. J.; Ettenhuber, P.; Fernández, B.; Ferrighi, L.; Fliegl, H.; Frediani, L.; Hald, K.; Halkier, A.; Hättig, C.; Heiberg, H.; Helgaker, T.; Hennum, A. C.; Hettner, H.; Hjertenæs, E.; Høst, S.; Høyvik, I.-M.; Iozzi, M. F.; Jansík, B.; Jensen, H. J. Aa.; Jonsson, D.; Jørgensen, P.; Kauczor, J.; Kirpekar, S.; Kjærgaard, T.; Klopper, W.; Knecht, S.; Kobayashi, R.; Koch, H.; Kongsted, J.; Krapp, A.; Kristensen, K.; Ligabue, A.; Lutnæs, O. B.; Melo, J. I.; Mikkelsen, K. V.; Myhre, R. H.; Neiss, C.; Nielsen, C. B.; Norman, P.; Olsen, J.; Olsen, J. M. H.; Osted, A.; Packer, M. J.; Pawłowski, F.; Pedersen, T. B.; Provasi, P. F.; Reine, S.; Rinkevicius, Z.; Ruden, T. A.; Ruud, K.; Rybkin, V. V.; Sałek, P.; Samson, C. C. M.; de Merás, A. S.; Saue, T.; Sauer, S. P. A.; Schimmelpfennig, B.; Sneskov, K.; Steindal, A. H.; Sylvester-Hvid, K. O.; Taylor, P. R.; Teale, A. M.; Tellgren, E. I.; Tew, D. P.; Thorvaldsen, A. J.; Thøgersen, L.; Vahtras, O.; Watson, M. A.; Wilson, D. J. D.; Ziolkowski, M.; Ågren, H. The Dalton quantum chemistry program system. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2014**, *4*, 269–284.
- (51) Folk, M.; Heber, G.; Koziol, Q.; Pourmal, E.; Robinson, D. An overview of the HDF5 technology suite and its applications. Proceedings of the EDBT/ICDT 2011 Workshop on Array Databases. 2011; pp 36–47.
- (52) Wesolowski, T. A.; Warshel, A. Frozen density functional approach for ab initio calculations of solvated molecules. *J. Phys. Chem.* **1993**, *97*, 8050–8053.

- (53) Jacob, C. R.; Visscher, L. A subsystem density-functional theory approach for the quantum chemical treatment of proteins. *J. Chem. Phys.* **2008**, *128*, 04B612.
- (54) Kiewisch, K.; Jacob, C. R.; Visscher, L. Quantum-Chemical Electron Densities of Proteins and of Selected Protein Sites from Subsystem Density Functional Theory. *J. Chem. Theory Comput.* **2013**, *9*, 2425–2440.
- (55) Goetz, A.; Neugebauer, J. Benchmarking Electron Densities and Electrostatic Potentials of Proteins from the Three-Partition Frozen Density Embedding Method. *J. Chem. Theory Comput.* **2016**, *12*, 4843–4855.
- (56) Goetz, A.; Jacob, C. R.; Neugebauer, J. Modeling environment effects on pigment site energies: Frozen density embedding with fully quantum-chemical protein densities. *Comput. Theor. Chem.* **2014**, *1040*, 347–359.
- (57) Goetz, A.; Neugebauer, J. Including protein density relaxation effects in first-principles embedding calculations of cofactor excitation energies. *Mol. Phys.* **2016**, *115*, 526–537.
- (58) Olsen, J. M. H.; List, N. H.; Kristensen, K.; Kongsted, J. Accuracy of Protein Embedding Potentials: An Analysis in Terms of Electrostatic Potentials. *J. Chem. Theory Comput.* **2015**, *11*, 1832–1842.
- (59) Kristensen, K.; Høyvik, I.-M.; Jansik, B.; Jørgensen, P.; Kjærgaard, T.; Reine, S.; Jakowski, J. MP2 energy and density for large molecular systems with internal error control using the Divide-Expand-Consolidate scheme. *Phys. Chem. Chem. Phys.* **2012**, *14*, 15706–15714.
- (60) Høyvik, I.-M.; Kristensen, K.; Jansik, B.; Jørgensen, P. The divide-expand-consolidate family of coupled cluster methods: Numerical illustrations using second order Møller-Plesset perturbation theory. *J. Chem. Phys.* **2012**, *136*, 014105.

- (61) Jakobsen, S.; Kristensen, K.; Jensen, F. Electrostatic Potential of Insulin: Exploring the Limitations of Density Functional Theory and Force Field Methods. *J. Chem. Theory Comput.* **2013**, *9*, 3978–3985.
- (62) Yanai, T.; Tew, D. P.; Handy, N. C. A new hybrid exchange–correlation functional using the Coulomb-attenuating method (CAM-B3LYP). *Chem. Phys. Lett.* **2004**, *393*, 51–57.
- (63) Marefat Khah, A.; Reinholdt, P.; Olsen, J. M. H.; Kongsted, J.; Hattig, C. Avoiding Electron Spill-Out in QM/MM Calculations on Excited States with Simple Pseudopotentials. *J. Chem. Theory Comput.* **2020**,
- (64) Kendall, R. A.; Dunning Jr, T. H.; Harrison, R. J. Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. *J. Chem. Phys.* **1992**, *96*, 6796–6806.
- (65) Murugan, N. A.; Kongsted, J.; Rinkevicius, Z.; Ågren, H. Color modeling of protein optical probes. *Phys. Chem. Chem. Phys.* **2012**, *14*, 1107–1112.
- (66) Hartmann, P.; Reinholdt, P.; Kongsted, J. *QM/MM studies of light-responsive biological systems*; Springer, 2020.
- (67) List, N. H.; Jensen, H. J. A.; Kongsted, J. Local electric fields and molecular properties in heterogeneous environments through polarizable embedding. *Phys. Chem. Chem. Phys.* **2016**, *18*, 10070–10080.
- (68) Beerepoot, M. T.; Friese, D. H.; List, N. H.; Kongsted, J.; Ruud, K. Benchmarking two-photon absorption cross sections: performance of CC2 and CAM-B3LYP. *Phys. Chem. Chem. Phys.* **2015**, *17*, 19306–19314.
- (69) Norman, P.; Ruud, K.; Saue, T. *Principles and practices of molecular properties: Theory, modeling, and simulations*; John Wiley & Sons, 2018.

## Graphical TOC Entry

