Cost-Effective Potential for Accurate Polarizable Embedding Calculations in Protein Environments

Reinholdt, Peter; Kjellgren, Erik; Steinmann, Casper; Olsen, Jógvan Magnus Haugaard

Published in:
Journal of Chemical Theory and Computation

DOI (link to publication from Publisher):
10.26434/chemrxiv.8126912
10.1021/acs.jctc.9b00616

Publication date:
2020

Document Version
Accepted author manuscript, peer reviewed version

Cost-effective Potential for Accurate Polarizable Embedding Calculations in Protein Environments

Peter Reinholdt,[†] Erik Rosendahl Kjellgren,[†] Casper Steinmann,[‡] and Jógvan Magnus Haugaard Olsen[*,¶]

[†]*Department of Physics, Chemistry and Pharmacy, University of Southern Denmark, Campusvej 55, DK–5230 Odense M, Denmark*

[‡]*Department of Chemistry and Bioscience, Aalborg University, Fredrik Bajers Vej 7H, DK–9220 Aalborg, Denmark*

[¶]*Hylleraas Centre for Quantum Molecular Sciences, Department of Chemistry, UiT The Arctic University of Norway, Tromsø, N-9037, Norway*

E-mail: jogvan.m.olsen@uit.no

Abstract

The fragment-based polarizable embedding (PE) model combined with an appropriate electronic-structure method constitutes a highly efficient and accurate multi-scale approach for computing spectroscopic properties of a central moiety including effects from its molecular environment through an embedding potential. There is, however, a comparatively high computational overhead associated with the computation of the embedding potential which is derived from first principles calculations on individual fragments of the environment. To reduce the computational cost associated with the calculation of embedding-potential parameters, we developed a set of amino-acid-specific transferable parameters tailored for large-scale PE-based calculations that include proteins. The amino-acid-based parameters are obtained by simultaneously fitting to a set of reference electric potentials based on structures derived

from a backbone-dependent rotamer library. The developed cost-effective polarizable protein potential (CP^3) consists of atom-centered charges and isotropic dipole–dipole polarizabilities of the standard amino acids. In terms of reproduction of electric potentials, the CP^3 is shown to perform consistently and with acceptable accuracy across both small tripeptide test systems and larger proteins. We show, through applications on realistic protein systems, that acceptable accuracy can be obtained by using a pure CP^3 representation of the protein environment, thus altogether omitting the cost associated with the calculation of embedding-potential parameters. High accuracy comparable to the full fragment-based approach can be achieved through a mixed description where the CP^3 is used only to describe amino acids beyond a threshold distance from the central quantum part.

1 Introduction

The use of ab-initio quantum-mechanical methods allows for the calculation of a wide range of molecular properties with high accuracy. A significant drawback of such methods is, however, the steep scaling with respect to the size of the system. This is unfortunate considering that most chemical processes occur in solution or other matrices, such as a protein or cell membrane. In response to this, numerous environment models have been developed. The main difference between the models is their representation of the environment, which ranges from a structure-less continuum,[1] over a classical atomistic potential,[2,3] to a quantum description.[4]

The polarizable embedding (PE)[5] model is an advanced fragment-based quantum–classical embedding scheme where mutual polarization effects are taken explicitly into account in a fully self-consistent manner. The environment is described in terms of distributed multipoles and polarizabilities, both of which are derived from quantum-mechanical calculations on the individual fragments that make up the environment. Extensive analysis in previous works[6,7] have shown that such an approach gives high accuracy when using LoProp-based[8]

atom-centered multipoles up to and including quadrupoles to describe the permanent charge distribution of fragments in the environment, and atom-centered anisotropic dipole–dipole polarizabilities to describe polarization effects.

The fragmentation approach is simple for solute-solvent systems. For example, a water cluster can easily be fragmented into individual water molecules, and the parameters (i.e., multipoles and polarizabilities) are then derived for each molecule in isolation through a quantum-mechanical calculation. In systems that contain molecules that are too large to perform a single quantum-mechanical calculation for the derivation of the embedding parameters, the molecular fractionation with conjugate caps (MFCC) method[9] is used to split the complete system into several smaller fragments, which each has a much smaller computational cost. Using the MFCC approach for deriving atom-centered parameters of proteins was introduced by Söderhjelm and Ryde.[10] This approach has been shown to produce high-quality embedding potentials for proteins[6] as well as nucleic acids[11] and lipids.[12]

Using such an approach, the computational cost is significantly reduced compared to carrying out a full quantum-mechanical calculation, but even so, the calculation of the embedding-potential parameters still carries a significant computational cost when there is a large number of fragments in the environment. The large cost of computing embedding parameters is particularly problematic when the fragments contain many atoms such as in protein environments, where an average capped amino-acid fragment produced in the MFCC approach contains about 30 atoms, with the number of atoms ranging from 19 atoms (in glycine) to 36 atoms (in tryptophan). This becomes even more significant when taking into account configurational sampling, which requires the embedding-potential parameters to be calculated for many configurations of the system. To reduce the computational cost associated with the calculation of embedding-potential parameters, we here present a set of transferable amino-acid-specific parameters, consisting of atom-centered point charges and isotropic dipole–dipole polarizabilities, tailored for use in large-scale PE-based calculations that include proteins. We will denote our parameter set as the cost-effective polarizable

protein potential (CP^3). The CP^3 parameters are validated by comparing electric potentials (EPs) of a diverse set of structures, ranging in size from tripeptides to an insulin monomer protein, to high-quality quantum-mechanical references. Moreover, we also evaluate the performance of the CP^3 in realistic PE-based calculations on chromophores embedded in protein environments.

One of the critical aspects when developing transferable parameters is conformational sampling. One approach is to sample structures from classical molecular dynamics (MD) trajectories.[7,11–13] Obtaining a sufficient sampling of the conformations of all the standard amino acids through MD simulations is non-trivial. Moreover, the choice of force field and simulation conditions may introduce some unwanted bias into the parameters. Instead, we use an alternative approach based on conformations sampled from a rotamer library. Several rotamer libraries[14–18] have been developed, including backbone-dependent ones,[15] facilitated by the abundance of high-quality protein crystal structures. These effectively report the most commonly occurring conformations of amino acids in typical protein environments and thus represent a very focused and efficient way of selecting structures to include in the fitting procedure. Similar approaches of sampling across a set of conformers have been applied in other studies.[19–21] The main difference being that we here use a comparatively large set of conformations sampled from the smoothed backbone-dependent rotamer library by Shapovalov and Dunbrack.[15]

After having selected the set of structures to be used in the parameter derivation, there is still the matter of selecting the method with which to derive the parameters. Numerous approaches have been reported in the literature (see e.g. Ref. 22 for an overview). Our aim is to be able to reproduce the EP as accurately as possible using only atom-centered point charges and isotropic polarizabilities. Based on preliminary testing, we found that EP fitting gave the best performance, and, since it is also rather flexible, it is relatively easy to introduce constraints, e.g. to ensure integer charge within groups of atoms and to equalize parameters on chemically equivalent atoms. Such constraints are useful in the parameterization of

amino-acid residues, which are part of a larger protein structure, and thus require capping groups (typically acetyl and N-methyl groups). Moreover, we can fit simultaneously to the EP across all conformers, as has been explored previously.[21] The reason for doing this is two-fold. First, we expect an improved transferability of the parameters by including a large set of structures that contains conformations representative of those commonly occurring in native proteins. Second, we seek to combat the well-known problem of redundancy in EP fitting.[21,23] Addressing the latter point by including a large number of structures is most likely not sufficient to resolve the problem of ill-conditioned EP fits, so we also reduce the number of free parameters by employing non-redundant fitting and by imposing additional constraints on the parameters.

2 Methodology

2.1 Conformational sampling

It is necessary to sample relevant conformations of each amino-acid residue to generate transferable parameters. In previous works by some of us, conformational sampling was obtained through classical MD simulations.[7,11,12] Here, we employ a more focused strategy instead, using the smoothed backbone-dependent rotamer library of Shapovalov and Dunbrack.[15] This library contains, for each amino acid, conditional probabilities of finding a specific combination of backbone and side-chain dihedral angles based on entries from the Protein Data Bank (PDB).[24] We include rotamers for each amino acid in order of the most commonly occurring. The total number of possible rotamers is rather large. Therefore, for each amino acid, we only included rotamers (based on backbone/side-chain dihedral angle combinations) until the cumulative fraction of covered structures from the database reached 80%. For glycine and alanine, which do not have any side-chain rotamers, the backbone dihedrals from valine were used. For cysteine, lysine, aspartate, glutamate, and histidine, alternative side-chain protonation states were also included, using the same dihedral angles

as the parent amino acid. Specifically, we included the disulfide-bonded cysteine (CYX), deprotonated cysteine (CYD) and lysine (LYD), and the protonated aspartate (ASH) and glutamate (GLH), along with the δ (HID) and ϵ (HIE) protonated variants of neutral histidine. Initial geometries of the amino acids were constructed with the dihedral angles as an input using the FragBuilder tool.[25] The main set of parameters used acetyl and N-methyl capping groups in order to mimic the environment of a typical protein backbone (Fig. 1A). The disulfide-bonded cysteine used in addition a thiomethyl capping group to represent a disulfide-bonded cysteine neighbor. In order to also develop parameters for both charged and neutral N- and C-terminals, a set of structures were also generated with only one capping group being either acetyl and N-methyl, but having the other capping group being either a protonated amine (Fig. 1C), carboxylate (Fig. 1D), amine (Fig. 1E), or carboxyl (Fig. 1F) group. The geometries of the generated structures were optimized with the HF-3c method[26] as implemented in the ORCA program,[27] with the dihedral angles constrained to the values

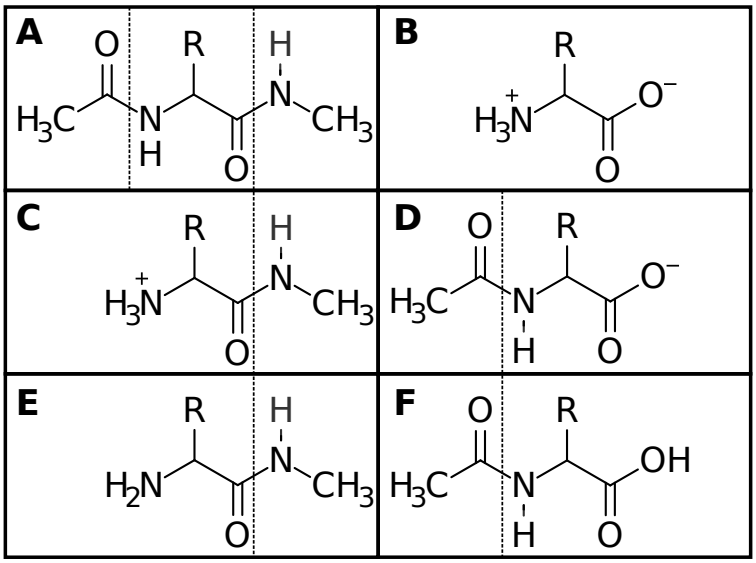

Figure 1: Structures of amino-acid fragments used in the optimization procedure. Dotted lines delimit regions in which integer charges are enforced. Structure **A** models the environment found in the middle of a typical protein chain, whereas **C** and **D** are used to model terminal charged residues, and **E** and **F** are used to model terminal neutral residues. Structure **B** represents a single zwitterionic amino acid which is not covered by the present parameter set.

obtained from the rotamer library. The geometry optimizations include an implicit water environment, modeled with the C-PCM,[28–31] in order to avoid any internal hydrogen transfers. The final set of structures are available at `https://doi.org/10.5281/zenodo.3250016`.[32]

2.2 Fitting procedure

The point charges, q, and isotropic dipole–dipole polarizabilities, α, are derived by fitting to static and induced EPs, respectively. The EP at a point $\mathbf{r}$ due to the electrons and nuclei of a molecule is given as

$$V^{\mathrm{QM}}(\mathbf{r}) = \sum_A \frac{Z_A}{|\mathbf{r} - \mathbf{R}_A|} - \int_\Omega \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \mathrm{d}\mathbf{r}' \tag{1}$$

where Z_A is a nuclear charge, $\mathbf{R}_A$ is a nuclear position, $\rho(\mathbf{r}')$ is the electronic density, and Ω denotes integration over all space.

Our goal is to find the set of atom-centered parameters that provides an optimal reproduction of the EP while at the same time being transferable across a wide range of conformations. We apply two strategies to address both requirements. First, the parameters are fitted to the EP of all conformers of the same molecule simultaneously. Second, we apply a harmonic restraint towards a set of reference parameters to avoid large changes that only marginally affect the quality of the fit (see Fig. S1 in the Supporting Information).

The cost function for the charge fitting can be written as a sum of two terms

$$\chi^q = \chi^q_{\mathrm{EP}} + \chi^q_{\mathrm{restraint}} \tag{2}$$

Here, the first term on the right hand side, χ^q_{EP}, is computed from the weighted mean squared error (wMSE) of the approximate EP computed from the fitted charges relative to the reference EP (eq (1)) evaluated on a set of grid points using all structures of a given

molecule. The wMSE cost function is thus defined as

$$\chi^q_{\mathrm{EP}} = \sum_{s=1}^{N_s} \frac{w_s}{N_{g,s}} \sum_{i=1}^{N_{g,s}} \left(V^{\mathrm{QM}}(\mathbf{r}_i) - \tilde{V}^q(\mathbf{r}_i) \right)^2 \tag{3}$$

where N_s is the number of structures included in the fit, $N_{g,s}$ is the number of grid points in the sth structure, and w_s are the weighting factors obtained from the backbone-dependent rotamer probabilities. The approximate EP from the fitted charges, $\tilde{V}^q$, is computed as

$$\tilde{V}^q(\mathbf{r}) = \sum_{n=1}^{N_{\mathrm{atoms}}} \frac{q_n}{|\mathbf{r} - \mathbf{r}_n|} \tag{4}$$

where N_{atoms} is the number of atoms in the molecule, q_n is an atom-centered charge, and $\mathbf{r}_n$ is its position. The second term in eq (2), $\chi^q_{\mathrm{restraint}}$, is a function that restrains the fitted charges towards a set of reference charges. The restraint cost function is given by

$$\chi^q_{\mathrm{restraint}} = k^q_{\mathrm{restraint}} \sum_{n=1}^{N_{\mathrm{atoms}}} \left(q_n - q_n^{\mathrm{ref}} \right)^2 \tag{5}$$

where $k^q_{\mathrm{restraint}}$ controls the strength of the restraint and q_n^{ref} is a reference charge. In this work, we used the LoProp procedure[8] to produce reference charges based on the most probable conformation of the given molecule. The restraint strength was chosen to be low enough to allow a good fit while at the same time ensure that the fitted charges do not drift far away from the reference charges. More details are provided in Sec. 3.1.

To fit the isotropic dipole–dipole polarizabilities, we apply the same strategy as for the fitting of the charges, but targeting now the reproduction of the induced EP, obtained as the difference between a perturbed and non-perturbed EP

$$V^{\mathrm{QM}}_{\mathrm{ind}}(\mathbf{r}, \mathbf{F}) = V^{\mathrm{QM}}(\mathbf{r}, \mathbf{F}) - V^{\mathrm{QM}}(\mathbf{r}, 0) \tag{6}$$

where the perturbed EP is computed by applying a homogeneous finite electric field, $\mathbf{F}$, on

the molecule. The cost function for fitting polarizabilities is also here a sum of two terms

$$\chi^{\alpha} = \chi^{\alpha}_{\mathrm{EP}} + \chi^{\alpha}_{\mathrm{restraint}} \tag{7}$$

where the first term on the right-hand side, $\chi^{\alpha}_{\mathrm{EP}}$, is the wMSE between the induced EP computed from a quantum-mechanical calculation and the approximate induced EP computed using the fitted polarizabilities. The wMSE can thus be written as

$$\chi^{\alpha}_{\mathrm{EP}} = \sum_{\mathbf{F}} \sum_{s=1}^{N_s} \frac{w_s}{N_{g,s}} \sum_{i=1}^{N_{g,s}} \left(V^{\mathrm{QM}}_{\mathrm{ind}}(\mathbf{r}_i, \mathbf{F}) - \tilde{V}^{\alpha}_{\mathrm{ind}}(\mathbf{r}_i, \mathbf{F}) \right)^2 \tag{8}$$

where the sum over electric fields runs over the applied electric field, which here is $\pm x$, $\pm y$, and $\pm z$. The approximate induced EP, $\tilde{V}^{\alpha}_{\mathrm{ind}}(\mathbf{r}_i, \mathbf{F})$, is computed using the set of induced dipoles that are created in response to the externally applied electric field via the fitted polarizabilities

$$\tilde{V}^{\alpha}_{\mathrm{ind}}(\mathbf{r}, \mathbf{F}) = - \sum_{n=1}^{N_{\mathrm{atoms}}} \boldsymbol{\mu}_n(\mathbf{F}) \cdot \frac{(\mathbf{r} - \mathbf{r}_n)}{|\mathbf{r} - \mathbf{r}_n|^3} \tag{9}$$

where the induced dipoles are obtained as

$$\boldsymbol{\mu}_n(\mathbf{F}) = \alpha_n \mathbf{F} \tag{10}$$

Implicit in this definition is the exclusion of mutual polarization between the fitted polarizabilities. In analogy to eq (5), we apply a harmonic restraint towards a set of reference polarizabilities

$$\chi^{\alpha}_{\mathrm{restraint}} = k^{\alpha}_{\mathrm{restraint}} \sum_{n=1}^{N_{\mathrm{atoms}}} \left(\alpha_n - \alpha^{\mathrm{ref}}_n \right)^2 \tag{11}$$

where $k^{\alpha}_{\mathrm{restraint}}$ controls the strength of the restraint and α^{ref}_n is a reference polarizability. Here we also used the LoProp procedure[8] to produce reference polarizabilities based on the most probable conformation of the given molecule, and the restraint strength was set sufficiently low to allow a good fit while at the same time ensure that the fitted polarizabilities do not

drift far away from the reference polarizabilities. More details are provided in Sec. 3.1.

The atom-centered charges and polarizabilities are derived for amino-acid residues, as shown in Figure 1. The parameters of a given fragment thus intrinsically include, by virtue of the wave-function optimization, polarization among sites that are present in the fragment, i.e., the amino-acid residue itself and part of the backbone of the neighboring residues through the capping groups. Thus, to avoid double-counting of polarization effects, when two neighboring fragments interact, it is appropriate to exclude those interactions accordingly. This interaction scheme is compatible (by design) with the one used for MFCC-based potentials, thus allowing a mixed representation of protein systems.

We reduce the number of free fitting variables in two ways. First, fragments are constrained to carry integer charges, which reduces the number of free parameters by one for the charge fits. For example, the total charge of the central amino acid is constrained to the appropriate charge of the given amino acid, while the capping groups are constrained to have a total charge of zero (see Figure 1). The constraint on the sum of charges can be realized by freely fitting only $N-1$ charges and assigning the final Nth charge to be $q_N = q_{\text{total}} - \sum_{i=1}^{N-1} q_i$. A similar constraint could, in principle, be introduced for the polarizabilities, but the molecular polarizability varies between conformations, unlike the molecular charge, so such a constraint is less appropriate in that case. Second, chemically equivalent atoms are constrained to have identical charge and polarizability. Both types of constraints were enforced by only using non-redundant optimization parameters (rather than Lagrange multipliers). With this, we significantly reduce the number of free parameters in the fit. For example, for the capped valine amino-acid model, which contains 28 atoms, the number of free parameters for the polarizabilities is reduced to 18, while the number of free parameters for the charges is reduced to 15.

3 Computational Details

3.1 Fitting

To compute the static EPs used in the fitting procedure, we first performed single-point energy calculations on the optimized structures using the Gaussian09 program.[33] For this, the PBE0 exchange–correlation functional[34] was used together with the aug-cc-pVDZ basis set[35] for heavy atoms and the cc-pVDZ basis set[36] for hydrogen atoms. In the following, we will denote this mixed basis set as augx-cc-pVDZ. These calculations formed the basis for the static EP evaluations of the capped amino acids. A similar set of calculations were carried out with a finite electric field to obtain induced EPs. Electric fields were applied along the $\pm \{x, y, z\}$ directions with a field strength of 5.0×10^{-3} au, resulting in a total of seven different calculations for each structure (six applied fields and one field-free structure). The EPs were subsequently evaluated on a set of discretized molecular surfaces, using Horton[37] as a backend. The surfaces were generated according to the algorithm of Saff and Kuijlaars,[38] with the radii of the spheres scaled according to the van der Waals (vdW) radii of the atoms. We used two shells in each structure, with scaling factors of 1.4 and 2.0. The construction of the surfaces, as well as the fitting of charges and polarizabilities, has been implemented in the PropertyFit Python package,[39] freely available at `https://github.com/peter-reinholdt/propertyfit`.

The charges and polarizabilities are obtained by minimizing the cost functions (eqs. (2) and (7)) using the sequential least squares programming (SLSQP)[40] optimization algorithm available in SciPy.[41] The reference charges and polarizabilities used for the restraints (see eqs. (5) and (11)) were obtained using the LoProp approach[8] based on the most probable structure of each amino acid, employing the PBE0 functional and the augx-cc-pVDZ basis set which has been recontracted to ANO-type as required by the LoProp procedure. The reference charges and polarizabilities were computed using the Dalton program[42,43] and the LoProp for Dalton Python package.[44] The charges and polarizabilities that are obtained

from the LoProp procedure generally have different values for chemically equivalent atoms, e.g., the charge of hydrogens in a methyl group will not be equal. Therefore we average the reference parameters belonging to such groups of chemically equivalent atoms before they are used in the fitting. The restraint strengths were set to $k^q_{\text{restraint}} = 1.0 \times 10^{-6}$ au and $k^\alpha_{\text{restraint}} = 1.0 \times 10^{-9}$ au to avoid drift as discussed above. The final set of fitted parameters as well as the structures and electric potentials used in the fitting procedure are available at `https://doi.org/10.5281/zenodo.3250016`.[32]

3.2 Validation

In order to evaluate the quality of the CP^3, its ability to reproduce a full quantum-mechanical EP was investigated on a number of test systems. This was done with three sets of increasingly complex test molecules starting with simple tripeptides, over a set of small proteins, and ending with a larger insulin protein. The quality was measured by computing the root-mean-square deviation (RMSD) between the quantum-mechanical EP and the CP^3-based EP. The EPs were evaluated on a set of points located on a surface defined by twice the vdW atomic radii, which is a highly relevant distance for intermolecular interactions. The quality is compared to potentials normally used in PE-based calculations, namely the MFCC-based M2P2 potential, which includes structure-specific charges, dipoles, and quadrupoles as well as anisotropic dipole–dipole polarizabilities, localized to atom centers using the LoProp procedure. The MFCC-based M2P2 potentials were computed using the PyFraME package[45] as detailed in the sections below. We also compare to the quality that would be obtained by a typical force field, using AMBER ff94[46] as a representative example of fixed point charge force fields, and the more advanced AMOEBA polarizable force field,[19,47–49] which includes multipoles up to quadrupoles and isotropic dipole–dipole polarizabilities. We note that the embedding with AMOEBA model has been previously studied,[50–52] but also that this implementation is neither open-source nor publicly available. The EPs produced using CP^3, ff94, and MFCC-based M2P2 potentials were evaluated using in-house Python

script while the EPs based on AMOEBA were computed with the Potential program from the Tinker package.[53]

3.2.1 Tripeptides

The first test set consists of a series of zwitterionic tripeptides using charged glycine terminals. However, this type of tripeptide cannot be used to test the parameters of the disulfide-bonded cysteine (CYX). Instead, a dimer made from two tripeptide units was manually constructed using the Maestro program.[54] Following this, a conformational search was carried out with the ConfGen module,[55,56] keeping the 32 lowest energy conformers. On each of these conformers, single point calculations were carried out at B3LYP[57]/6-31G**[58] level of theory using the Jaguar program.[59,60] The lowest energy conformer was retained for use in the following calculations. Strictly speaking, this molecule consists of six amino acid residues and is therefore not a tripeptide, but we include it in the tests below to be able to address the performance of the CYX parameters. For the remaining tripeptides, geometries were taken from FragBuilder. Geometry optimizations were carried out on all of the tripeptides at PBE0/def2-TZVP[61] level of theory using the ORCA program. The RIJ-COSX[62] approximation was employed with the def2/J[63] auxiliary basis set. Water solvent was included using the C-PCM[28–31] in order to avoid internal hydrogen transfers.

Calculation of the EPs was based on the unrelaxed density from DLPNO-CCSD[64]/aug-cc-pVTZ calculations using ORCA. The aug-cc-pVTZ/C basis set was used as an auxiliary basis set for the correlation calculations and the def2/J basis set for Coulomb fitting. The default truncation thresholds were used for the calculations on all tripeptides.

The MFCC-based M2P2 potentials were derived from PBE0/aug-cc-pVTZ calculations on the amino-acid-based fragments using the LoProp procedure to distribute the parameters to atom centers.

With this series, we get a measure of the quality of the parameters belonging to each of the individual amino acids in the core set, though we should note that quality of the

parameters for GLY will, of course, affect all of the tested structures. The zwitterionic state ensures that the atoms of the central amino acid residue experience rather large electric fields. As such, this serves as a test of both the charges and the polarizabilities.

3.2.2 Small proteins

The second series comprises a series of small proteins. The structures of the proteins were obtained from the study of Kulik et al.[65] We used the structures optimized at ω-PBEh[66–68]/6-31G level of theory. The structures were manually inspected for problems, and proteins with proton transfers or other structural issues were excluded as this is outside of the applicable domain of our CP^3. After also considering computational limitations, a total of 13 small proteins remained the PDB ids of which are: 1LVQ, 1V46, 1Y49, 1YJP, 2JTA, 2OL9, 2ONW, 2OQ9, 2PJV, 3FTK, 3FTR, 3FVA, and 3NVG. The selected proteins contain between 6 and 24 amino acids, which translates to between 70 and 312 atoms.

The EP of the selected proteins was evaluated using unrelaxed DLPNO-CCSD densities calculated using the ORCA program. The cc-pVDZ basis set was employed, along with the cc-pVDZ/C auxiliary basis set for correlation, and the def2/J basis set for Coulomb fitting. The default truncation thresholds were used for the calculations on all of the 13 proteins.

The parameters of the MFCC-based M2P2 potentials were obtained from PBE0/cc-pVDZ calculations on the amino-acid-based fragments using the LoProp procedure to distribute the parameters to atom centers.

3.2.3 Insulin

The final test system is an insulin protein containing two chains with a total of 51 amino acids, which amounts to 787 atoms. For this, we used the DEC-MP2[69]/cc-pVDZ EP from Jakobsen et al.[70] as the reference. The MFCC-based M2P2 potential was derived from PBE0/augx-cc-pVDZ calculations on the amino-acid-based fragments using the LoProp procedure to distribute the parameters to atom centers.

3.3 Performance Benchmarks

To further benchmark the performance of the CP^3 for practical applications, we computed a series of optical properties using three different proteins as examples. For this, we used the green fluorescent protein (GFP), a channelrhodopsin chimera (C1C2), and a flavin-binding dodecin, where the chromophores are an anionic 4-hydroxybenzylidene-1,2-dimethylimidazolinone (HBDI), a protonated retinylidene Schiff base (PRSB), and a lumiflavin (LF) dimer, respectively. The chromophore in GFP and C1C2 is covalently bonded to the protein, whereas the LFs in dodecine are not.

The structure of the GFP system is a single configuration taken from a QM/MM-MD trajectory.[71] It consists of the GFP and a spherical solvent shell made up of 19793 water molecules and five sodium ions.

The C1C2 structure is a QM/MM-optimized geometry from Ref. 72. It was originally based on a crystal structure (PDB ID: 3UG9).[73] The structure includes the protein and 43 crystal water molecules.

The dodecin structure is based on a snapshot from a QM/MM-MD simulation of the LF dimer in dodecin, taken from a previous study.[74] The system contains two π–π stacked LFs embedded in the protein environment. Apart from the LFs and dodecin, the system also contains 8265 water molecules, 44 sodium ions, 12 chloride ions, and two magnesium ions.

The quantum region of the GFP (HBDI including parts of the neighboring residues and hydrogen link atoms), C1C2 (PRSB including most of the lysine side-chain and a hydrogen link atom), and dodecin (two LFs) systems are shown in Figure 2.

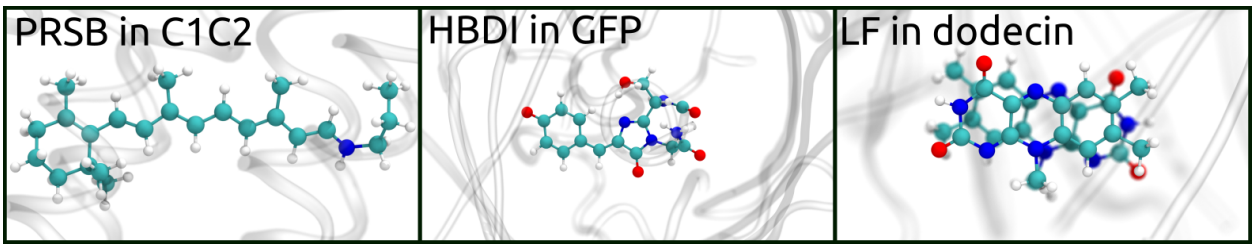

Figure 2: Quantum regions in the GFP, C1C2, and dodecin systems.

The effects of the environment on the chromophore were modeled using the PE model, which we briefly described in the introduction. We investigate the ground-state dipole moment and the excitation energy of the lowest singlet transition together with the associated one- and two-photon absorption strengths. These properties were computed using the CAM-B3LYP[75] exchange–correlation functional and the pcseg-2 basis set.[76] Effective external field (EEF) effects[77] were included in the computation of absorption strengths. The properties were calculated using the Dalton program[42,43] employing PElib[78] and Gen1Int.[79,80]

The reference properties are calculated using the MFCC-based M2P2 potential to describe the protein surroundings which has been shown to be highly accurate.[6] The M2P2 potential consists of atom-centered multipoles up to quadrupoles and atom-centered anisotropic dipole–dipole polarizabilities. In this work, the multipoles and polarizabilities were computed at PBE0/augx-cc-pVDZ level of theory using the LoProp approach[8] to distribute the multipoles and polarizabilities to atomic centers which thus matches the level of theory used to derive the CP^3 parameters. We benchmark properties obtained from calculations where the whole protein is modeled by the CP^3 as well as mixed intermediates where the residues closest to the chromophore (according to a distance criterion) are described by the more accurate M2P2 potential and the remainder by the CP^3. For comparison we also computed properties using the AMOEBA to model the protein surroundings. Here the multipoles and polarizabilities were assigned from the OpenMM[81] implementation of the AMOEBA force field using the parameter set in `amoeba2013.xml`. The amino-acid residues in the GFP and C1C2 systems that are covalently bonded to the quantum region are incomplete due to the fact that parts have been included in the quantum region. The interfacial region between the chromophore and protein was therefore described with an M2P2 potential for these two systems because we were unable to assign parameters to those incomplete residues. Moreover, the protein model of the C1C2 contained a neutral N-terminal ($-NH_2$), which is not present in AMOEBA, and this was also modeled with the M2P2 potential. The exclusion lists were extended to avoid issues with over-polarization. Thus, any atom that is included

in the exclusion list of a protein atom described by an M2P2 potential had their exclusion lists appended as if an M2P2 potential described them. In all cases, all water molecules and all ions were described by the solvent embedding potential (SEP)[7] which consist of charges and isotropic polarizabilities. The PyFraME package[45] was used to generate the potentials employing the Dalton program[42,43] and the LoProp for Dalton script[44] for the individual M2P2 fragment calculations.

In both GFP and C1C2 there are covalent bonds that were cut because they cross the quantum-classical interface. The open valencies in the quantum region were closed using hydrogen link-atoms. The charges on the classical carbon atoms closest to the hydrogen link-atoms were redistributed to the nearest two classical sites while higher-order multipoles and polarizabilities were removed. Thole's exponential damping scheme was used for induced-dipole–induced-dipole interactions in calculations involving CP^3 and M2P2 while Tinker-style Thole damping was used in calculations using AMOEBA.

4 Results and Discussion

4.1 Fitting

The optimized CP^3 charges yield an order-of-magnitude improvement over the initial start guess, while the optimized polarizabilities show more modest improvements (see Fig. S1 in the Supporting Information). Figure 3 compares the numerical values of charges and polarizabilities from CP^3, AMOEBA, and AMBER force fields (ff94 and the more recent ff15ipq[13]).

The CP^3 charges correlate most strongly to the ff15ipq charges ($R^2 = 0.94$), followed by the ff94 charges ($R^2 = 0.81$), and with a much weaker correlation to the AMOEBA charges ($R^2 = 0.68$). The good agreement between CP^3 and ff15ipq charges is somewhat surprising, considering that the latter implicitly include effects of the polarization caused by the solvent. The poor correlation to the AMOEBA charges is to be expected. The AMOEBA charges

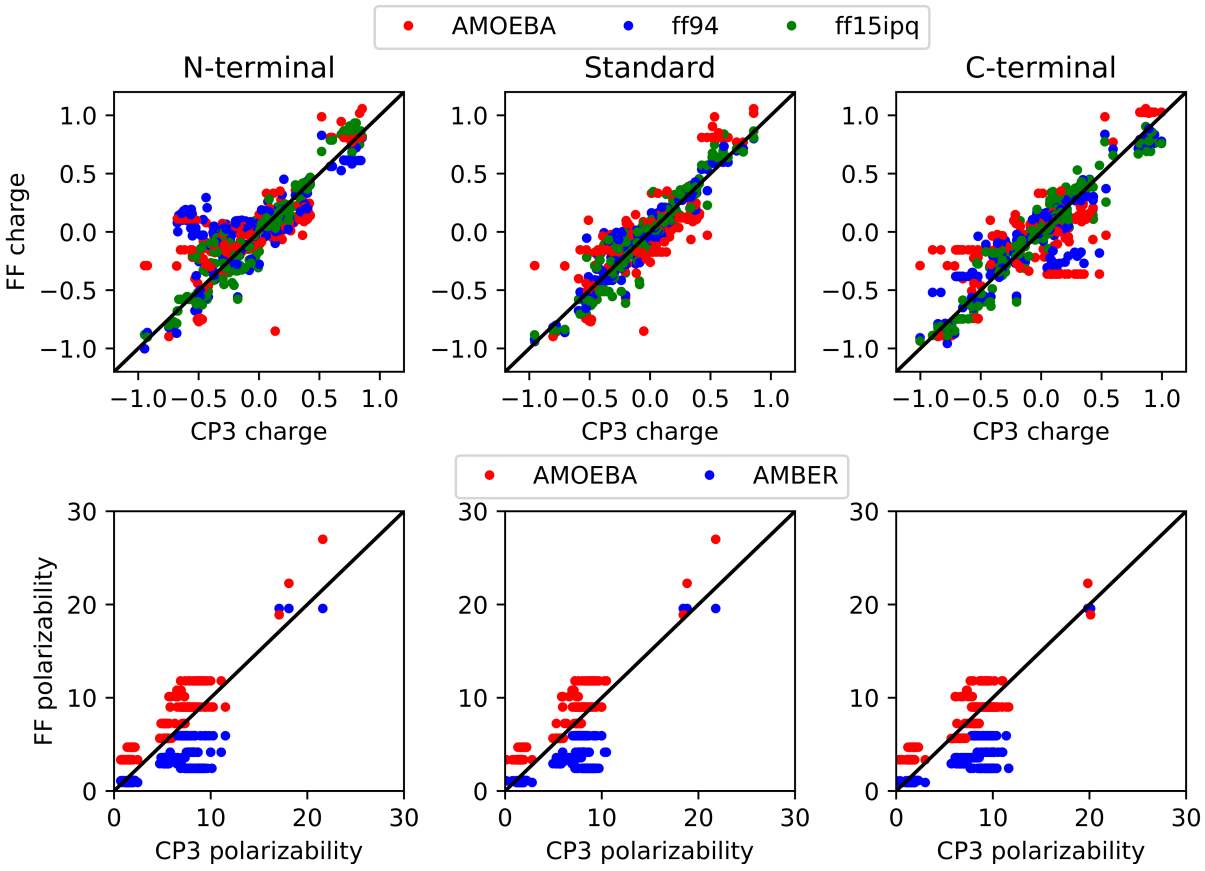

Figure 3: Parameters of CP^3 are compared to those from AMOEBA and AMBER force fields. Top panels compare charges (in au) and lower panels compare polarizabilities (in au). N- and C-terminals refer to the charged terminal amino acid residues. Perfect agreement is illustrated with the black line.

are balanced by the inclusion of higher-order multipoles, whereas the CP^3 charges need to compensate for the lack of higher-order multipoles.

Figure 4 shows the correlation between the magnitudes of the molecular dipoles computed from CP^3 charges and PBE0/augx-cc-pVDZ which is the method used in the fitting procedure. Importantly, there is a very strong correlation ($R^2 = 0.998$) despite the fact that the dipoles were not used in the fitting procedure. This is important because, in a calculation using CP^3, the electric field from the charges is used to induce dipoles in all other fragments via the polarizabilities.

The magnitude of the CP^3 polarizabilities are in general between the AMOEBA and AMBER polarizabilities, being larger than the AMBER polarizabilities, but smaller than

the AMOEBA polarizabilities. A direct comparison of the polarizabilities is, however, complicated by the fact that all three models use different interaction schemes for the polarizabilities. The choice of interaction model can have a large impact on the optimal numerical values of the atomic polarizabilities.[82] The AMOEBA and AMBER force fields have very few polarizabilities, consistent with their interaction model, whereas the CP^3 parameters are residue-specific and thus have different polarizabilities for each atom in each amino-acid residue.

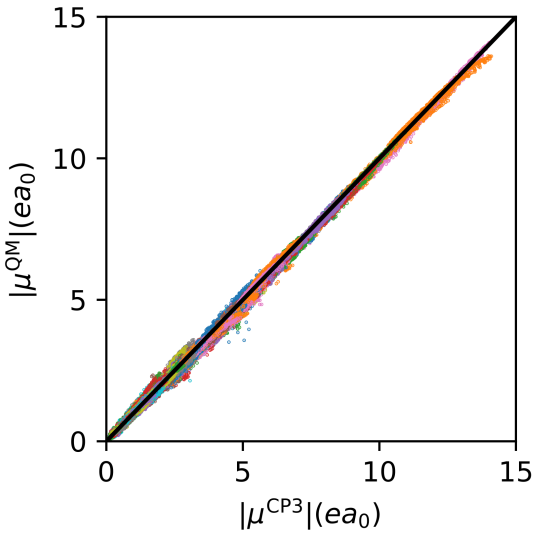

Figure 4: Comparison between the magnitude of the molecular dipoles computed from CP^3 charges and PBE0/augx-cc-pVDZ. Each point corresponds to a single structure used in the fitting procedure. Different colors correspond to different amino acids. Perfect agreement is illustrated with the black line.

4.2 Validation

4.2.1 Tripeptides

The RMSDs of the zwitterionic tripeptide EPs, evaluated on a molecular surface defined by twice the vdW atomic radii, are presented in Figure 5. The average RMSD of the CP^3-based EPs is 13.5 kJ/mol. Most of the individual RMSDs are very close to the average, thus showing that the accuracy is rather consistent. There are a few outliers, namely the HIS

and CYS tripeptides, and the CYX tripeptide dimer, where the RMSD for CYS and HIS tripeptides is somewhat larger, while it is substantially smaller for the CYX tripeptide dimer. The latter deviation is at least partly explained by the fact that the CYX tripeptide dimer is qualitatively different from the remaining tripeptide test systems, since it is a hexapeptide (a dimer consisting of two tripeptides connected by a disulfide bond). Employing a Thole-type damping of the induced-dipole–induced-dipole interactions had very little effect, which means the larger RMSDs for the CYS and HIS tripeptides are not due to over-polarization. Moreover, nothing unusual was observed for these amino acids in the fitting nor in the comparisons presented in the previous section. It is therefore likely that the larger RMSDs are related to the specific geometries of the HIS and CYS tripeptides used here. Here it is important to note that the tripeptide test cases are artificial and extreme, and thus that RMSDs of this magnitude are unlikely to be observed in production calculations.

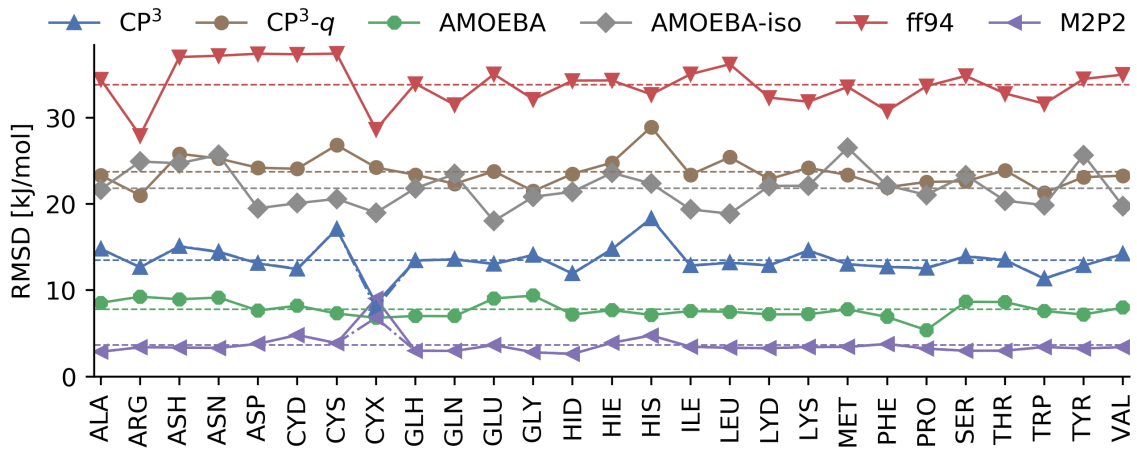

Figure 5: RMSD in kJ/mol of tripeptide EPs modeled by CP^3, AMOEBA, ff94, and M2P2 on a 2.0 × vdW surface relative to a DLPNO-CCSD/aug-cc-pVTZ reference. The tripeptides are zwitterionic GLY-*XXX*-GLY, where *XXX* is the amino acid indicated in the plot, except CYX which consists of two CYX tripeptides linked through a disulfide bond. Dotted lines show the average RMSD of each potential. Dash-dotted lines for CP^3 and M2P2 show the RMSDs obtained when induced dipole interactions are damped.

The average RMSD is relatively large compared to the average RMSD of the M2P2-based EPs (3.6 kJ/mol) which can be considered to be very close to an optimally parametrized

potential. In a single case, the CYX tripeptide dimer, the M2P2 potential produced an EP with a substantially larger RMSD of 9.0 kJ/mol. The larger error is because of very close contact between the two CYX tripeptide monomers. In fact, the hydrogen bonds between the charged terminals are only 1.6 Å which means that there are sites that are close to the divergence sphere of the multipole expansion on other sites. By employing a standard Thole-type exponential damping of the induced dipole–induced dipole interactions, the RMSD was decreased by 2.5 kJ/mol. Note that damping in the remaining M2P2-based and all the CP^3-based calculations only has a negligible effect. Damping is part of the polarization model used in, e.g., AMOEBA, to avoid so-called polarization catastrophe. However, it is not required in MFCC-based potentials because the interactions between neighboring polarizable sites are intrinsically included through the fragment calculations from which the parameters are derived. Thus there are no polarizable sites close enough to trigger a polarization catastrophe. It may be beneficial to include damping, however, when there are close contacts between two fragments that are not covalently bonded (as is the case for the CYX tripeptide dimer), though the overall improvement in most production calculations will be minor.

A comparison to EPs produced by the CP^3 with only charges (CP^3-q) reveals that the isotropic polarizabilities yield a substantial improvement in all cases, reducing the RMSD by roughly 10 kJ/mol on average. The importance of polarization is further emphasized by the large RMSDs associated with the ff94 fixed point-charge potential, which averages at 33.8 kJ/mol. The polarizable AMOEBA potential performs better than the CP^3, with an average RMSD of only 7.8 kJ/mol. The main factor that allows AMOEBA to reach higher accuracy is most likely related to the inclusion of higher-order multipoles. Indeed if the dipoles and quadrupoles are removed from AMOEBA (AMOEBA-iso in Fig. 5) the RMSDs increase to roughly the same range as CP^3-q.

In order to further examine the accuracy that can be gained from higher-order multipole moments, a calculation on the GLY tripeptide was performed using the MFCC-based M0P2 potential (i.e., using structure-specific charges and anisotropic polarizabilities). This resulted

in an RMSD of 9.2 kJ/mol, which should be compared to 2.8 kJ/mol when using the M2P2 potential. It is thus clear that higher-order multipoles provide a substantial improvement. Therefore it appears reasonable to assume that the poorer performance of the CP^3 relative to the AMOEBA potential is, to a large extent, due to the lack of higher-order multipoles.

4.2.2 Small proteins

Figure 6 presents the RMSDs of the small protein EPs evaluated on a discretized surface defined by twice the vdW atomic radii. The average RMSD of the CP^3-based EP remains close to the one observed in the tripeptide test set (i.e., 11.1 kJ/mol compared to 13.5 kJ/mol). The individual RMSDs are also rather close to average, ranging from 7.6 (2PJV) to 15.3 kJ/mol (3FTR). This shows that the CP^3 parameters scale to larger systems with consistent accuracy.

The importance of explicit polarization is also evident here; excluding the polarizabilities (CP^3-q) increases the average RMSD to 24.3 kJ/mol. The average RMSD of the ff94-based EP is in the same range (28.5 kJ/mol). The spread of the individual RMSDs is rather large, including a few relatively low RMSDs indicating that polarization does not have the same

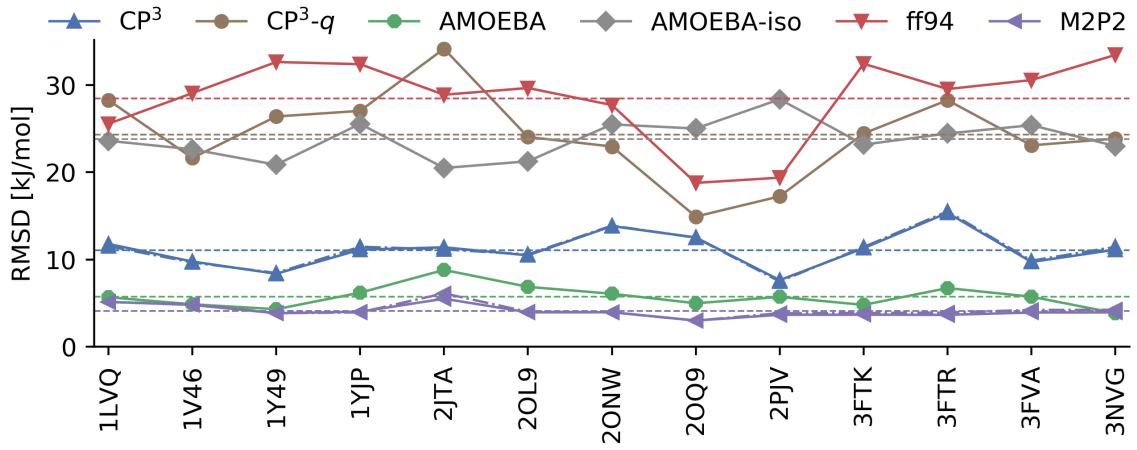

Figure 6: RMSD in kJ/mol of small protein EPs modeled by CP^3, AMOEBA, and M2P2 on a 2.0 × vdW surface relative to a DLPNO-CCSD/cc-pVDZ reference. Dotted lines show the average RMSD of each potential. Dash-dotted lines for CP^3 and M2P2 show the RMSDs obtained when induced dipole interactions are damped.

importance in all cases. Even so, it is clear that polarization is required in order to obtain consistent results.

The AMOEBA potential again performs very well with an average RMSD of only 5.7 kJ/mol, which is quite close to the average RMSD of the M2P2-based EPs (4.1 kJ/mol). This is impressive, considering that the M2P2 parameters are structure-specific, while the AMOEBA potential uses transferable parameters. Removing dipoles and quadrupoles from the AMOEBA potential (AMOEBA-iso) yields substantially larger RMSDs that are in the same range as CP^3 without polarizabilities (i.e., CP^3-q), as was the case for the tripeptide test set, thus further emphasizing the role of higher-order multipoles.

4.2.3 Insulin

The final validation test is performed on the insulin protein, which is more than twice the size of the largest of the smaller proteins. This system is challenging also for some Kohn-Sham density functional approximations.[70] It has been used previously in a study of the

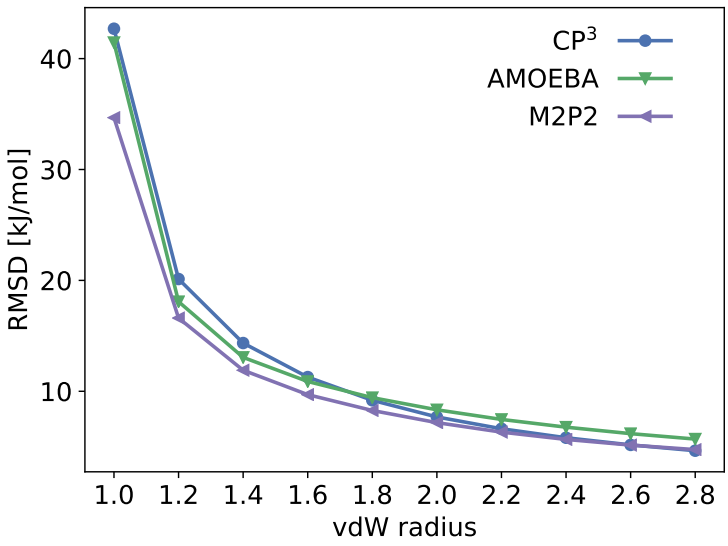

Figure 7: RMSDs in kJ/mol of the insulin EP modeled by CP^3, AMOEBA, ff94, and M2P2 relative to a DEC-MP2/cc-pVDZ reference.[70] The EP is computed on points sampled within volumes defined by two vdW surfaces separated by $\pm$ 0.1 factors of the indicated vdW radius. The markers indicate the midpoint between the surfaces.

MFCC-based LoProp potentials where it was shown that a very good accuracy could be reached as long as polarizable potentials are used.[6] Here we use the same reference as in those two studies (DEC-MP2/cc-pVDZ) to evaluate the performance of the CP^3.

Figure 7 shows the RMSDs of the EPs based on CP^3, AMOEBA, and M2P2, as a function of the distance from the protein. All of the potentials yield comparable accuracy at twice the vdW atomic radii where the RMSDs of the EPs based on CP^3, AMOEBA, and M2P2 are 7.7, 8.3, and 7.2 kJ/mol, respectively. That said, if we go into detail, we find that the M2P2 potential generally has slightly better accuracy than both CP^3 and AMOEBA especially at distances below 2.0 $\times$ vdW atomic radii. The CP^3 produces a marginally lower RMSD than AMOEBA at distances longer than 1.8 $\times$ vdW atomic radii (roughly 1 kJ/mol) whereas it is the opposite at distances shorter than 1.6 $\times$ vdW atomic radii.

4.3 Performance Benchmarks

The main goal of this work is to develop a transferable polarizable protein potential that will reduce the computational cost associated with PE-based calculations of spectroscopic properties of systems containing proteins. This reduction in cost involves a trade-off in terms of a loss of accuracy. To investigate how the accuracy of the CP^3 translates to relevant molecular properties in practice, we computed the ground-state dipole moments and the excitation energy of the lowest singlet transitions together with the associated one- and two-photon absorption (1PA & 2PA) strengths of chromophores in GFP, C1C2, and dodecin. In cases where high accuracy is essential, the CP^3 may also be used to mitigate the high computational cost of the MFCC-based M2P2 potential by using a layered model where the structure-specific M2P2 potential is used only on the fragments closest to the quantum region while the remainder is modeled by the CP^3. The reduction of the computational cost is proportional to the number of fragments described by CP^3 instead of M2P2.

Here we investigate the accuracy of both the pure CP^3 and the mixed M2P2/CP^3 model with varying cutoff distances using properties computed using a full MFCC-based M2P2

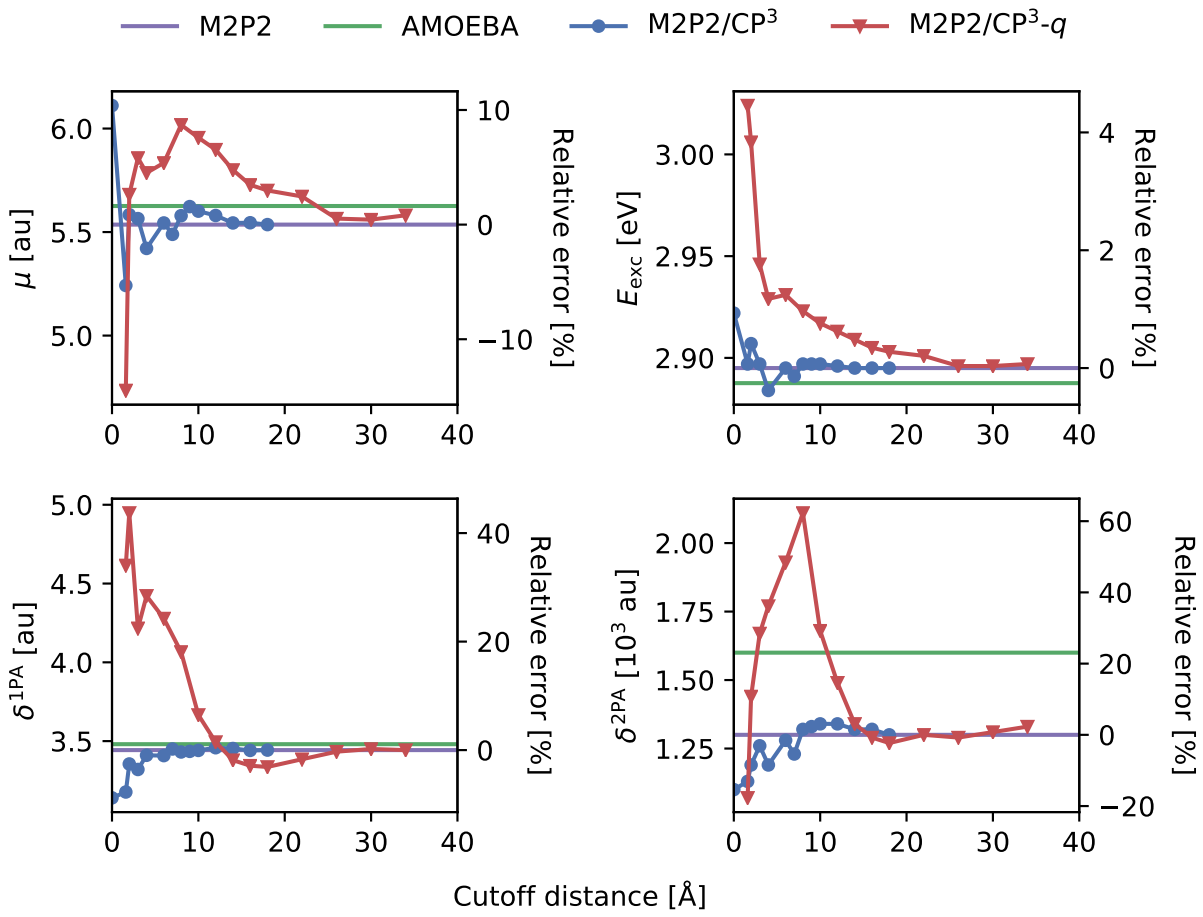

Figure 8: Ground-state dipole moment, lowest singlet excitation energy and associated one- and two-photon absorption strengths of the HBDI chromophore in GFP. The protein environment is modeled using mixed M2P2/CP^3 (blue) or mixed M2P2/CP^3-q (red). The cutoff defines the distance from the quantum region up to which the M2P2 potential is used. All water molecules and ions are in all cases modeled by the SEP. The initial M2P2/CP^3 point at a cutoff distance of 0.0 Å corresponds to pure CP^3. The purple line indicates the reference M2P2 values while the green line indicates values obtained using the AMOEBA potential.

potential as the reference. The results are presented in Figures 8–10. The cutoff in case of mixed M2P2/CP^3 defines the distance from the quantum region up to which the M2P2 potential is used while the CP^3 is used for the remaining amino-acid residues. Thus, a distance of 0.0 Å corresponds to pure CP^3. We also compare to a mixed M2P2/CP^3-q approach where the M2P2 potential is used up to the cutoff distance while the remainder is described by charges only. This is an alternative approach that can be used to reduce computational cost. Here we use the charges from CP^3 but similar results are obtained if

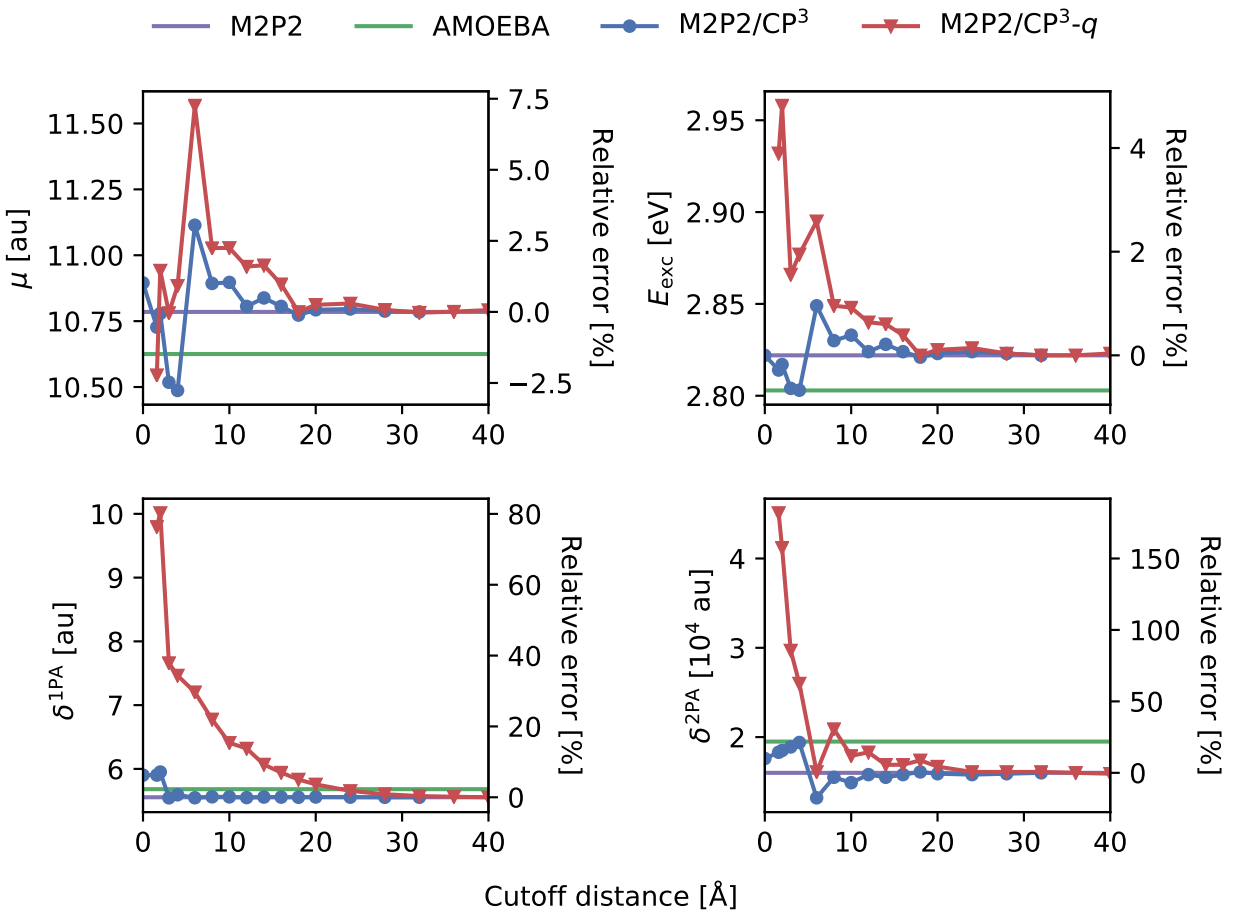

Figure 9: Ground-state dipole moment, lowest singlet excitation energy and associated one- and two-photon absorption strengths of the PRSB chromophore in C1C2. The protein environment is modeled using mixed M2P2/CP^3 (blue) or mixed M2P2/CP^3-q (red). The cutoff defines the distance from the quantum region up to which the M2P2 potential is used. All water molecules are in all cases modeled by the SEP. The initial M2P2/CP^3 point at a cutoff distance of 0.0 Å corresponds to pure CP^3. The purple line indicates the reference M2P2 values while the green line indicates values obtained using the AMOEBA potential.

one uses instead AMBER ff94 charges as can be seen in Figures S2–S4 in the Supporting Information.

In general, we find that a mixed M2P2/CP^3 representation of the protein with an appropriately chosen cutoff can provide properties with essentially the same accuracy as a full M2P2 treatment. However, using a pure CP^3 description of the protein or a mixed M2P2/CP^3 description with a too short cutoff distance can lead to relatively large errors in some cases, especially 2PA strengths. The convergence with respect to the cutoff distance is

not always smooth (see, e.g., Figure 9) indicating that some key amino-acid residues require a high-accuracy potential.

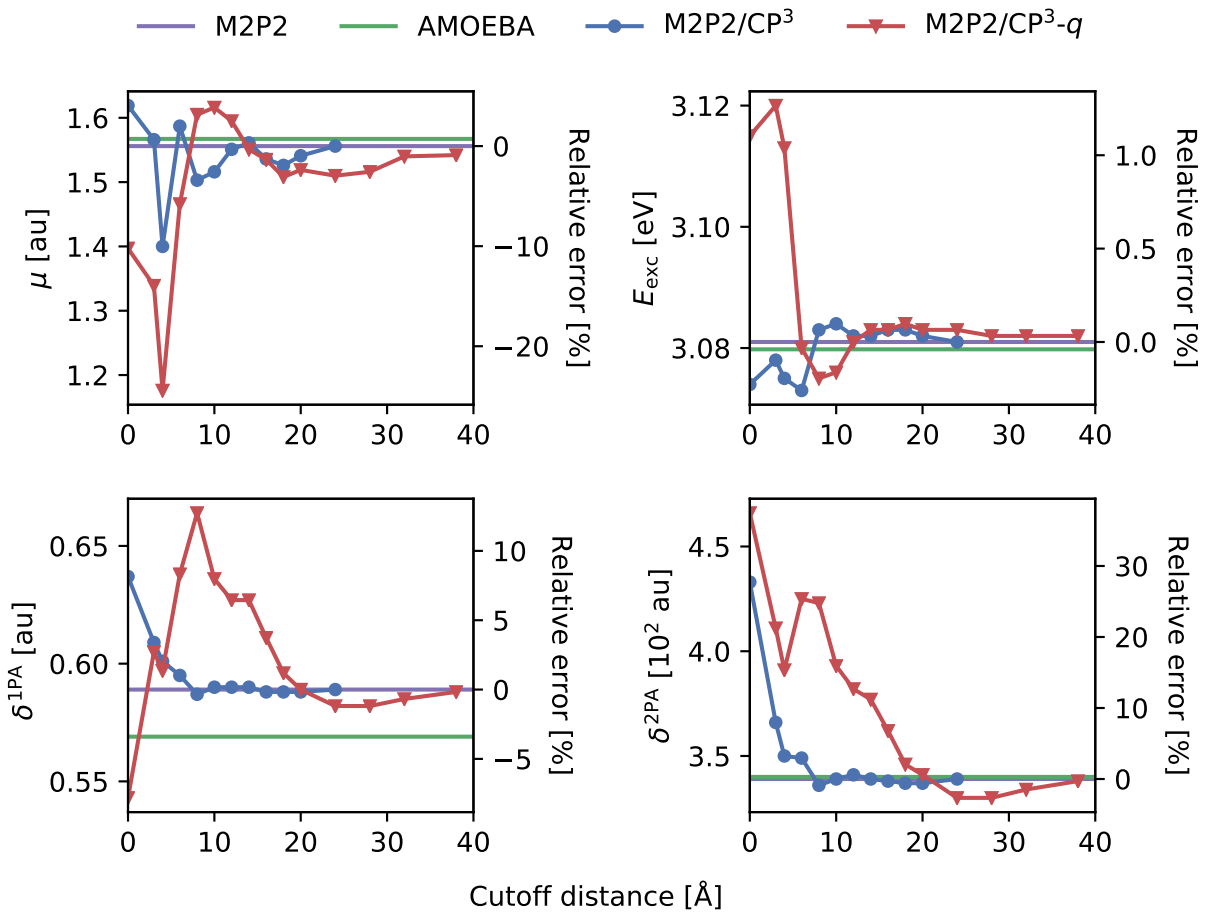

Figure 10: Ground-state dipole moment, lowest singlet excitation energy and associated one- and two-photon absorption strengths of the LF dimers in dodecin. The protein environment is modeled using mixed M2P2/CP^3 (blue) or mixed M2P2/CP^3-q (red). The cutoff defines the distance from the quantum region up to which the M2P2 potential is used. All water molecules and ions are in all cases modeled by the SEP. The initial M2P2/CP^3 point at a cutoff distance of 0.0 Å corresponds to pure CP^3. The purple line indicates the reference M2P2 values while the green line indicates values obtained using the AMOEBA potential.

The maximum absolute error of the dipole moment reaches 0.6, 0.3, and 0.16 au for the GFP, C1C2, and dodecin systems, respectively, which corresponds to a relative error of 9 %, 3 %, and 10 % compared to results obtained using only the MFCC-based M2P2 potential. To get the relative error of the dipole moment below 5% for all systems, a cutoff of 6 Å is needed, thus reducing the computational cost related to the construction of the

embedding potential by more than 80 %. The excitation energies are rather unaffected in all systems with maximum absolute errors that do not exceed 0.03 eV for the GFP and C1C2 systems and 0.01 eV for the dodecin system, which amounts to less than a 1.0 % error. The absorption strengths, on the other hand, are more sensitive with maximum relative errors between about 7–9 % for 1PA strengths and roughly 15–30 % for 2PA strengths. For 1PA strengths, a rather small cutoff of 3.0 Å is enough to reduce the relative error to below 5 % for all systems. This corresponds to a reduction of the computational cost by more than 90 %. Using a cutoff distance of about 12 Å ensures relative errors below 5 % across all properties and systems, including the 2PA strengths. For the systems investigated here this results in a 50 % reduction of the computational cost associated with the calculation of the embedding-potential parameters.. The cutoff distance can be further lowered to 8 Å, if a lower accuracy (i.e. below 10 %) is acceptable, thus cutting the cost by roughly 70 %. It is also entirely viable to use a pure CP^3 description which would altogether remove the computational cost associated with the embedding potential. This can yield relatively large errors for 2PA strengths, as noted above, but for the other properties, the relative error remains at around 10 % or below.

The alternative approach to reduce computational cost, i.e., where the M2P2 potential is used within a cutoff distance and the rest of the system is described by charges only (M2P2/CP^3-q), requires longer cutoffs to reach the same accuracy as the mixed M2P2/CP^3 with short cutoffs. To obtain relative errors below 5 % for all properties and systems, a cutoff of around 20 Å is required, while errors below 10 % can be obtained at roughly 16 Å. If also the charges outside the cutoff are removed, i.e., the systems are described by M2P2 with in the cutoff and everything outside is removed, almost the entire protein needs to be included (see Figures S2–S4 in the Supporting Information).

A third alternative is to use the AMOEBA potential. As can be seen from Figures 8–10, overall it yields small errors compared to results obtained using full M2P2 description of the proteins. The exceptions are 2PA strengths in the GFP and C1C2 systems where the

Table 1: Deviation of properties computed using CP^3 or AMOEBA relative to M2P2

	GFP[a]		C1C2[a]		Dodecin	
Property	CP^3	AMOEBA	CP^3	AMOEBA	CP^3	AMOEBA
μ [au]	-0.29	0.09	-0.06	-0.16	0.06	0.01
E_{exc} [eV]	0.00	-0.01	-0.01	-0.02	-0.01	-0.00
δ^{1PA} [au]	-0.27	0.04	0.35	0.13	0.05	-0.02
δ^{2PA} [au]	-170	300	2300	3500	94.00	1.00

[a] amino-acid residues that are covalently bonded to the chromophore in the quantum region are described by M2P2 in all cases

relative error is just above 20 %. The larger error here may be related to the *ad-hoc* coupling of the M2P2 and AMOEBA potentials for these systems (as explained in the Computational Details). On the other hand, a one-to-one comparison of the errors from using CP^3 and AMOEBA (see Table 1), shows that CP^3 in most cases produces slightly larger errors, as expected based on the validation results presented in the previous subsection. One-to-one here means that M2P2 is used for the amino-acid residues that are covalently bonded to the chromophore in the quantum region while the remainder of the protein is described by CP^3 or AMOEBA. An advantage of CP^3 compared to AMOEBA in this context is the seamless coupling to the M2P2 potential that allows us to straightforwardly extend the cutoff and thus increase the accuracy.

Our results thus show that it is possible to significantly reduce the number of fragment calculations needed to derive the embedding potential for PE-based calculations when targeting spectroscopic properties, while still also retaining very high accuracy. For some applications, namely for the calculation of excitation energies and 1PA strengths, a pure CP^3 description is of acceptable quality. On the other hand, 2PA strengths are much more sensitive to changes in the environment, and consequently, larger cutoff distances of about 8–12 Å are needed to reach better accuracy.

5 Conclusion

We presented a newly developed set of transferable atom-centered charges and isotropic dipole–dipole polarizabilities (CP^3), suitable for use in PE-based spectroscopic property calculations involving protein environments. The set of parameters covers all standard amino acids, including alternate side-chain protonation states. The parameters are amino-acid specific and were derived by simultaneously fitting to the static or induced EPs of a large set of configurations sampled from a backbone-dependent rotamer library. The fitting was performed by minimizing a cost function that contains a probability-weighted sum of contributions from all of the included configurations. We showed that the CP^3 yields consistent accuracy across a wide set of structures, ranging from a series of tripeptides to a series of small proteins, by comparing CP^3-based EPs to references computed using a high-level quantum-mechanical method. In terms of reproducing EPs, the CP^3 outperforms the AMBER ff94 force field mainly due to the inclusion of explicit polarization but performs slightly worse than the polarizable AMOEBA potential.

The CP^3 is fully compatible with MFCC-based potentials, which thus allows straightforward use of a layered model that enables a better cost–accuracy trade-off. The accompanying reduction in computational cost associated with the calculation of the embedding-potential parameters when compared to a full MFCC-based M2P2 potential is proportional to the number of amino-acid residues described by CP^3. The performance of CP^3 and mixed M2P2/CP^3 protein descriptions in practical applications was investigated. We found that excitation energies are reproduced well even with a pure CP^3 description thus completely removing the computational cost associated with the calculation of the embedding-potential parameters. For ground-state dipole moments and 1PA strengths, we observed larger errors but could recover good accuracy with a layered model where the MFCC-based M2P2 potential is used for residues inside a small shell around the chromophore. This corresponds to less than 20 % of the amino-acid residues and thus the computational cost is reduced by 80 % or more. The 2PA strengths were found to be more sensitive to the quality of the potential and

thus require larger shells in order to obtain good accuracy. Nevertheless, the computational cost can be reduced by at least 50 % while still retaining high accuracy (relative error less than 5 %) and the cost can be further reduced to 70 % if a larger error is acceptable (relative error less than 10 %). This will thus make it affordable to perform accurate large-scale spectroscopic studies of molecular systems that include protein environments including a proper configurational sampling.

Acknowledgement

J. M. H. O. acknowledges financial support from the Carlsberg Foundation (Grant ID: CF15-0823) and the Research Council of Norway through its Centres of Excellence scheme (Project ID: 262695). Computational resources were provided by the DeIC National HPC Center at the University of Southern Denmark.

Supporting Information Available

Cost functions related to the fitting of charges and polarizabilities. Tables with RMSD of the EPs of the tripeptide and small protein test cases. Figures showing the ground-state dipole moment, lowest singlet excitation energy and associated one- and two-photon absorption strengths of the HBDI, PRSB, and LF dimer chromophores in GFP, C1C2, and dodecin, respectively, computed where the protein environment is modeled using M2P2, mixed M2P2/CP^3, mixed M2P2/CP^3-q, mixed M2P2/ff94, M2P2/Vacuum, and AMOEBA.

This material is available free of charge via the Internet at `http://pubs.acs.org/`.

References

(1) Tomasi, J.; Mennucci, B.; Cammi, R. Quantum mechanical continuum solvation models. *Chem. Rev.* **2005**, *105*, 2999–3094.

(2) 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.

(3) 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.

(4) Wesolowski, T. A.; Shedge, S.; Zhou, X. Frozen-density embedding strategy for multilevel simulations of electronic structure. *Chem. Rev.* **2015**, *115*, 5891–5928.

(5) Olsen, J. M.; Aidas, K.; Kongsted, J. Excited states in solution through polarizable embedding. *J. Chem. Theory Comput.* **2010**, *6*, 3721–3734.

(6) 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.

(7) Beerepoot, M. T. P.; Steindal, A. H.; List, N. H.; Kongsted, J.; Olsen, J. M. H. Averaged Solvent Embedding Potential Parameters for Multiscale Modeling of Molecular Properties. *J. Chem. Theory Comput.* **2016**, *12*, 1684–1695.

(8) Gagliardi, L.; Lindh, R.; Karlström, G. Local properties of quantum chemical systems: The LoProp approach. *J. Chem. Phys.* **2004**, *121*, 4494–4500.

(9) 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.

(10) 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* **2009**, *113*, 617–627.

(11) Nørby, M. S.; Steinmann, C.; Olsen, J. M. H.; Li, H.; Kongsted, J. Computational Approach for Studying Optical Properties of DNA Systems in Solution. *J. Chem. Theory Comput.* **2016**, *12*, 5050–5057.

(12) Witzke, S.; List, N. H.; Olsen, J. M. H.; Steinmann, C.; Petersen, M.; Beerepoot, M. T. P.; Kongsted, J. An averaged polarizable potential for multiscale modeling in phospholipid membranes. *J. Comput. Chem.* **2017**, *38*, 601–611.

(13) Debiec, K. T.; Cerutti, D. S.; Baker, L. R.; Gronenborn, A. M.; Case, D. A.; Chong, L. T. Further along the Road Less Traveled: AMBER ff15ipq, an Original Protein Force Field Built on a Self-Consistent Physical Model. *J. Chem. Theory Comput.* **2016**, *12*, 3926–3947.

(14) Lovell, S. C.; Word, J. M.; Richardson, J. S.; Richardson, D. C. The penultimate rotamer library. *Proteins: Structure, Function, and Genetics* **2000**, *40*, 389–408.

(15) Shapovalov, M. V.; Dunbrack, R. L. A Smoothed Backbone-Dependent Rotamer Library for Proteins Derived from Adaptive Kernel Density Estimates and Regressions. *Structure* **2011**, *19*, 844–858.

(16) Taghizadeh, M.; Goliaei, B.; Madadkar-Sobhani, A. SDRL: a sequence-dependent protein side-chain rotamer library. *Molecular BioSystems* **2015**, *11*, 2000–2007.

(17) Towse, C.-L.; Rysavy, S. J.; Vulovic, I. M.; Daggett, V. New Dynamic Rotamer Libraries: Data-Driven Analysis of Side-Chain Conformational Propensities. *Structure* **2016**, *24*, 187–199.

(18) Watkins, A. M.; Craven, T. W.; Renfrew, P. D.; Arora, P. S.; Bonneau, R. Rotamer Libraries for the High-Resolution Design of Amino Acid Foldamers. *Structure* **2017**, *25*, 1771–1780.e3.

(19) Shi, Y.; Xia, Z.; Zhang, J.; Best, R.; Wu, C.; Ponder, J. W.; Ren, P. Polarizable atomic multipole-based AMOEBA force field for proteins. *J. Chem. Theory Comput.* **2013**, *9*, 4046–4063.

(20) Yuan, Y.; Mills, M. J.; Popelier, P. L.; Jensen, F. Comprehensive analysis of energy minima of the 20 natural amino acids. *J. Phys. Chem. A* **2014**, *118*, 7876–7891.

(21) Jakobsen, S.; Jensen, F. Systematic improvement of potential-derived atomic multipoles and redundancy of the electrostatic parameter space. *J. Chem. Theory Comput.* **2014**, *10*, 5493–5504.

(22) Jensen, F. *Introduction to Computational Chemistry*; John Wiley & Sons, 2017; Chapter 10.

(23) Jakobsen, S.; Jensen, F. Searching the Force Field Electrostatic Multipole Parameter Space. *J. Chem. Theory Comput.* **2016**, *12*, 1824–1832.

(24) Berman, H. M. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.

(25) Christensen, A. S.; Hamelryck, T.; Jensen, J. H. FragBuilder: an efficient Python library to setup quantum chemistry calculations on peptides models. *PeerJ* **2014**, *2*, e277.

(26) Sure, R.; Grimme, S. Corrected small basis set Hartree-Fock method for large systems. *J. Comput. Chem.* **2013**, *34*, 1672–1685.

(27) Neese, F. The ORCA program system. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2012**, *2*, 73–78.

(28) Klamt, A.; Schrmann, G. COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J. Chem. Soc., Perkin Trans. 2* **1993**, 799–805.

(29) Andzelm, J.; Klmel, C.; Klamt, A. Incorporation of solvent effects into density functional calculations of molecular energies and geometries. *J. Chem. Phys.* **1995**, *103*, 9312–9320.

(30) Barone, V.; Cossi, M. Quantum Calculation of Molecular Energies and Energy Gradients in Solution by a Conductor Solvent Model. *J. Phys. Chem. A* **1998**, *102*, 1995–2001.

(31) Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. Energies, structures, and electronic properties of molecules in solution with the C-PCM solvation model. *J. Comput. Chem.* **2003**, *24*, 669–681.

(32) Reinholdt, P.; Kjellgren, E. R.; Steinmann, C.; Olsen, J. M. H. Dataset for the article "Cost-effective Potential for Accurate Polarizable Embedding Calculations in Protein Environments". 2019; `https://doi.org/10.5281/zenodo.3250016`.

(33) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian 09 Revision E.01. Gaussian Inc. Wallingford CT 2009.

(34) Adamo, C.; Barone, V. Toward reliable density functional methods without adjustable parameters: The PBE0 model. *J. Chem. Phys.* **1999**, *110*, 6158–6170.

(35) Kendall, R. A.; Dunning, 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.

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

(37) Verstraelen, T.; Tecmer, P.; Heidar-Zadeh, F.; González-Espinoza, C. E.; Chan, M.; Kim, T. D.; Boguslawski, K.; Fias, S.; Vandenbrande, S.; Berrocal, D.; Ayers, P. W. HORTON 2.1.0. 2017.

(38) Saff, E. B.; Kuijlaars, A. B. Distributing many points on a sphere. *Math. Intell.* **1997**, *19*, 5–11.

(39) Reinholdt, P.; Kjellgren, E. PropertyFit. 2019; `https://github.com/peter-reinholdt/propertyfit`.

(40) Kraft, D. A software package for sequential quadratic programming. *Forschungsbericht-Deutsche Forschungs- und Versuchsanstalt fur Luft- und Raumfahrt* **1988**,

(41) Virtanen, P.; Gommers, R.; Oliphant, T. E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; van der Walt, S. J.; Brett, M.; Wilson, J.; Jarrod Millman, K.; Mayorov, N.; Nelson, A. R. J.; Jones, E.; Kern, R.; Larson, E.; Carey, C.; Polat, İ.; Feng, Y.; Moore, E. W.; VanderPlas, J.; Laxalde, D.; Perktold, J.; Cimrman, R.; Henriksen, I.; Quintero, E. A.; Harris, C. R.; Archibald, A. M.; Ribeiro, A. H.; Pedregosa, F.; van Mulbregt, P.; Contributors, S. . . SciPy 1.0–Fundamental Algorithms for Scientific Computing in Python. *arXiv e-prints* **2019**, arXiv:1907.10121.

(42) 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.; Hettema, 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.; Pawlowski, 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.

(43) Dalton, a molecular electronic structure program, Release Dalton2018.1 (2019), see `http://www.daltonprogram.org`.

(44) Vahtras, O. LoProp for Dalton (version 1.0). 2014; `https://doi.org/10.5281/zenodo.13276`.

(45) Olsen, J. M. H. PyFraME: Python tools for Fragment-based Multiscale Embedding (version 0.2.0). 2018; `https://doi.org/10.5281/zenodo.1443314`.

(46) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.

(47) Ren, P.; Ponder, J. W. Polarizable Atomic Multipole Water Model for Molecular Mechanics Simulation. *J. Phys. Chem. B.* **2003**, *107*, 5933–5947.

(48) Grossfield, A.; Ren, P.; Ponder, J. W. Ion Solvation Thermodynamics from Simulation with a Polarizable Force Field. *J. Am. Chem. Soc.* **2003**, *125*, 15671–15682.

(49) Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio, R. A.; Head-Gordon, M.; Clark, G. N. I.; Johnson, M. E.; Head-Gordon, T. Current Status of the AMOEBA Polarizable Force Field. *J. Phys. Chem. B.* **2010**, *114*, 2549–2564.

(50) Loco, D.; Polack, E.; Caprasecca, S.; Lagardère, L.; Lipparini, F.; Piquemal, J.-P.; Mennucci, B. A QM/MM approach using the AMOEBA polarizable embedding: from ground state energies to electronic excitations. *J. Chem. Theory Comput.* **2016**, *12*, 3654–3661.

(51) Loco, D.; Lagardère, L.; Caprasecca, S.; Lipparini, F.; Mennucci, B.; Piquemal, J.-P. Hybrid QM/MM molecular dynamics with AMOEBA polarizable embedding. *J. Chem. Theory Comput.* **2017**, *13*, 4025–4033.

(52) Loco, D.; Lagardère, L.; Cisneros, G. A.; Scalmani, G.; Frisch, M.; Lipparini, F.; Mennucci, B.; Piquemal, J.-P. Towards large scale hybrid QM/MM dynamics of complex systems with advanced point dipole polarizable embeddings. *Chem. Sci.* **2019**, *10*, 7200–7211.

(53) Rackers, J. A.; Wang, Z.; Lu, C.; Laury, M. L.; Lagardère, L.; Schnieders, M. J.; Piquemal, J.-P.; Ren, P.; Ponder, J. W. Tinker 8: Software Tools for Molecular Design. *J. Chem. Theory Comput.* **2018**, *14*, 5273–5289.

(54) Schrödinger Release 2018-2: Maestro, Schrödinger, LLC, New York, NY, 2018.

(55) Watts, K. S.; Dalal, P.; Murphy, R. B.; Sherman, W.; Friesner, R. A.; Shelley, J. C. ConfGen: A Conformational Search Method for Efficient Generation of Bioactive Conformers. *J. Chem. Inf. Model.* **2010**, *50*, 534–546.

(56) Schrödinger Release 2018-2: ConfGen, Schrödinger, LLC, New York, NY, 2018.

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

(58) Hehre, W. J.; Ditchfield, R.; Pople, J. A. Self—Consistent Molecular Orbital Methods. XII. Further Extensions of Gaussian—Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules. *J. Chem. Phys.* **1972**, *56*, 2257–2261.

(59) Bochevarov, A. D.; Harder, E.; Hughes, T. F.; Greenwood, J. R.; Braden, D. A.; Philipp, D. M.; Rinaldo, D.; Halls, M. D.; Zhang, J.; Friesner, R. A. Jaguar: A high-performance quantum chemistry software program with strengths in life and materials sciences. *Int. J. Quantum Chem.* **2013**, *113*, 2110–2142.

(60) Schrödinger Release 2018-2: Jaguar, Schrödinger, LLC, New York, NY, 2018.

(61) Weigend, F.; Hser, M.; Patzelt, H.; Ahlrichs, R. RI-MP2: optimized auxiliary basis sets and demonstration of efficiency. *Chem. Phys. Lett.* **1998**, *294*, 143–152.

(62) Neese, F.; Wennmohs, F.; Hansen, A.; Becker, U. Efficient, approximate and parallel Hartree–Fock and hybrid DFT calculations. A chain-of-spheres algorithm for the Hartree–Fock exchange. *Chem. Phys.* **2009**, *356*, 98–109.

(63) Weigend, F. Accurate Coulomb-fitting basis sets for H to Rn. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1057–1065.

(64) Riplinger, C.; Pinski, P.; Becker, U.; Valeev, E. F.; Neese, F. Sparse maps—A systematic infrastructure for reduced-scaling electronic structure methods. II. Linear scaling

domain based pair natural orbital coupled cluster theory. *J. Chem. Phys.* **2016**, *144*, 024109.

(65) Kulik, H. J.; Luehr, N.; Ufimtsev, I. S.; Martinez, T. J. Ab initio quantum chemistry for protein structures. *J. Phys. Chem. B* **2012**, *116*, 12501–12509.

(66) Heyd, J.; Scuseria, G. E.; Ernzerhof, M. Hybrid functionals based on a screened Coulomb potential. *J. Chem. Phys.* **2003**, *118*, 8207–8215.

(67) Izmaylov, A. F.; Scuseria, G. E.; Frisch, M. J. Efficient evaluation of short-range Hartree-Fock exchange in large molecules and periodic systems. *J. Chem. Phys.* **2006**, *125*, 104103.

(68) Henderson, T. M.; Izmaylov, A. F.; Scalmani, G.; Scuseria, G. E. Can short-range hybrids describe long-range-dependent properties? *J. Chem. Phys.* **2009**, *131*, 044108.

(69) 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.

(70) 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.

(71) Nåbo, L. J.; Olsen, J. M. H.; Martínez, T. J.; Kongsted, J. The Quality of the Embedding Potential Is Decisive for Minimal Quantum Region Size in Embedding Calculations: The Case of the Green Fluorescent Protein. *J. Chem. Theory Comput.* **2017**, *13*, 6230–6236.

(72) Sneskov, K.; Olsen, J. M. H.; Schwabe, T.; Hättig, C.; Christiansen, O.; Kongsted, J.

Computational screening of one- and two-photon spectrally tuned channelrhodopsin mutants. *Phys. Chem. Chem. Phys.* **2013**, *15*, 7567.

(73) Kato, H.; Ishitani, R.; Nureki, O. Crystal Structure of the Closed State of Channelrhodopsin. 2012; `https://doi.org/10.2210/pdb3ug9/pdb`.

(74) Scheurer, M.; Herbst, M. F.; Reinholdt, P.; Olsen, J. M. H.; Dreuw, A.; Kongsted, J. Polarizable Embedding Combined with the Algebraic Diagrammatic Construction: Tackling Excited States in Biomolecular Systems. *J. Chem. Theory Comput.* **2018**, *14*, 4870–4883.

(75) 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.

(76) Jensen, F. Unifying general and segmented contracted basis sets. Segmented polarization consistent basis sets. *J. Chem. Theory Comput.* **2014**, *10*, 1074–1085.

(77) 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.

(78) Olsen, J. M. H.; List, N. H.; Steinmann, C.; Steindal, A. H.; Nørby, M. S.; Reinholdt, P. PElib: The Polarizable Embedding library (version 1.3.7). 2018; `https://doi.org/10.5281/zenodo.1434124`.

(79) Gao, B.; Thorvaldsen, A. J.; Ruud, K. GEN1INT: A unified procedure for the evaluation of one-electron integrals over Gaussian basis functions and their geometric derivatives. *Int. J. Quantum Chem.* **2011**, *111*, 858–872.

(80) Gao, B. Gen1Int (version 0.2.1). 2012; `http://gitlab.com/bingao/gen1int`.

(81) Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L.-P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D.; Wiewiora, R. P.; Brooks, B. R.; Pande, V. S. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLOS Comput. Biol.* **2017**, *13*, 1–17.

(82) Wang, J.; Cieplak, P.; Li, J.; Hou, T.; Luo, R.; Duan, Y. Development of polarizable models for molecular mechanical calculations I: parameterization of atomic polarizability. *J. Phys. Chem. B* **2011**, *115*, 3091–3099.

Graphical TOC Entry