JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Accuracy of Protein Embedding Potentials: An Analysis in Terms of Electrostatic Potentials

Jógvan Magnus Haugaard Olsen,*[,†,‡] Nanna Holmgaard List,[‡] Kasper Kristensen,[¶] and Jacob Kongsted[‡]

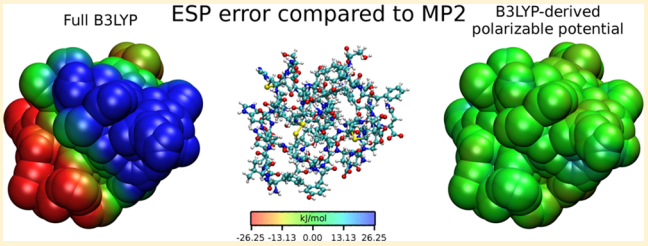
[†]Laboratory of Computational Chemistry and Biochemistry, École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland

[‡]Department of Physics, Chemistry and Pharmacy, University of Southern Denmark, DK-5230 Odense M, Denmark

[¶]qLEAP Center for Theoretical Chemistry, Department of Chemistry, Aarhus University, DK-8000 Aarhus, Denmark

Supporting Information

ABSTRACT: Quantum-mechanical embedding methods have in recent years gained significant interest and may now be applied to predict a wide range of molecular properties calculated at different levels of theory. To reach a high level of accuracy in embedding methods, both the electronic structure model of the active region and the embedding potential need to be of sufficiently high quality. In fact, failures in quantum mechanics/molecular mechanics (QM/MM)-based embedding methods have often been associated with the QM/MM methodology itself; however, in many cases the reason for such failures is due to the use of an inaccurate embedding potential. In this paper, we investigate in detail the quality of the electronic component of embedding potentials designed for calculations on protein biostructures. We show that very accurate explicitly polarizable embedding potentials may be efficiently designed using fragmentation strategies combined with single-fragment ab initio calculations. In fact, due to the self-interaction error in Kohn−Sham density functional theory (KS-DFT), use of large full-structure quantum-mechanical calculations based on conventional (hybrid) functionals leads to less accurate embedding potentials than fragment-based approaches. We also find that standard protein force fields yield poor embedding potentials, and it is therefore not advisable to use such force fields in general QM/MM-type calculations of molecular properties other than energies and structures.



1. INTRODUCTION

Accurate modeling of molecular response properties and excited states in large and complex systems, such as biomolecules or solute−solvent systems, requires the use of methods based on quantum mechanics. However, the size of such systems is often prohibitively large for conventional quantum-mechanical methods to be used, and while state-of-the-art linear-scaling techniques allow full quantum-mechanical calculations on relatively large structures,[1,2] they require a large amount of computing resources. Furthermore, limitations in commonly used density functional theory (DFT) methods could pose a serious problem when used on large systems.[2] However, in many cases it is in fact not necessary to treat the full system at the quantum-mechanical level, for example, molecular properties of a solute in a solvent or excitation energies of a chromophore embedded inside a protein. In such cases it can be advantageous to use combined quantum mechanics and molecular mechanics (QM/MM) methods[3] which are now standard for exploring energetics and structures of large and complex molecular systems.[4,5]

The underlying assumption in QM/MM is that the major part of the total system is adequately described using classical force fields. Methods based on quantum mechanics are then used for the parts where it is necessary to explicitly take into account the electronic degrees of freedom. For instance when modeling a chemical reaction or to achieve better accuracy for molecules/fragments that are not well described by a classical force field. Most traditional QM/MM implementations and applications use standard molecular mechanics (MM) force fields,[4,5] such as AMBER,[6−8] CHARMM,[9,10] or OPLS,[11−13] which makes sense when exploring energetics and structures because they have been parametrized exactly for such purposes. However, when modeling electronic response properties or excitation processes, use of traditional force fields is often not sufficiently accurate and therefore requires a large part of the system to be treated at the quantum-mechanical level.[1,14] For such calculations it is paramount to treat the effects from the classical region explicitly at the electronic level, that is, by including them in the Hamiltonian. In the electrostatic embedding scheme, the environment is included as a static embedding potential that polarizes the electron density, whereas in polarized embedding schemes the classical subsystem is also polarizable thus allowing mutual polarization between the quantum and classical regions. One of the advantages of polarized embedding is the ability to describe differential environmental polarization effects between ground and excited states of the quantum region. This effect becomes

Received: November 12, 2014
Published: March 23, 2015

DOI: 10.1021/acs.jctc.5b00078
J. Chem. Theory Comput. 2015, 11, 1832−1842

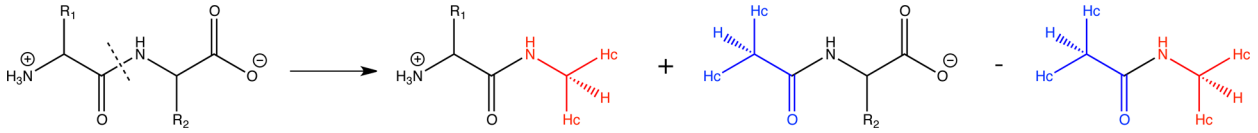
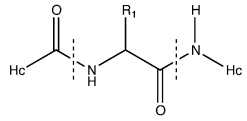
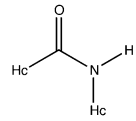
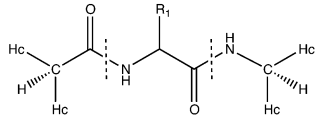
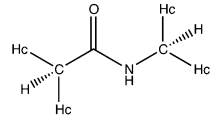


Figure 1. Illustration of the MFCC fragmentation procedure.

Table 1. Capped Fragments and Concaps According to Capping Level

level	capped fragment	concap
1		
2		

important when there is a significant rearrangement of the electronic structure upon electronic excitation.[15−18] The polarizable embedding (PE) model developed by some of the present authors[15,19−24] is a computational scheme that focuses on the calculation of molecular response properties using polarizable embedding potentials. Recently, this model was implemented in the PE library using an efficient atomic orbital (AO) density-based formalism.[25,26]

The aim of this study is to conduct an investigation of protein embedding potentials used within the PE model and other similar QM/MM-type embedding models. This is done by a direct comparison between electrostatic potentials (ESPs) calculated from the embedding potentials to a fully quantum-mechanical ESP. In this way, we have a simple and direct measure of the quality of the potentials as they are used in an actual embedding calculation because it is the ESP that enters the hybrid Hamiltonian. It is a very stringent test compared to evaluating calculated properties, such as excitation energies or response properties, because any such comparison will be system- and property-dependent, whereas an accurate reproduction of the ESP ensures accurate predictions of all molecular properties within the limitations of the embedding strategy. The protein embedding potentials used in this work are derived using the principles of the molecular fractionation with conjugate caps (MFCC) method[27] as defined by Söderhjelm and Ryde[28] for distributed properties. We examine two types of ab initio-derived potentials that are based on either distributed multipole moments or ESP-fitted charges, and, in both cases, we include distributed isotropic or anisotropic dipole−dipole polarizabilities to account for explicit polarization effects. Furthermore, we use charges from standard AMBER force fields as representatives of embedding potentials that are commonly used in QM/MM applications.[4,5]

2. METHODS

2.1. Embedding Potentials. We use the MFCC scheme developed by Zhang and Zhang[27] to fragment proteins and derive the embedding potential using the procedure reported by Söderhjelm and Ryde.[28] In the MFCC approach, a protein is decomposed into amino acid-based fragments that are capped with groups built from the neighboring amino acids. This satisfies the valency of the broken bonds and also mimics the effects from the covalently bonded neighbors. In addition to the capped amino acid fragments, two neighboring capping groups are merged to form a so-called conjugate cap fragment (concap). An illustration of the fragmentation procedure for a dipeptide is shown in Figure 1. The parts shown in red and blue are the capping groups (i.e., *N*-methyl in red and acetyl in blue). The hydrogens denoted "Hc" mark where heavier atoms have been replaced by hydrogens to terminate a fragment. The positions of the terminating hydrogens are adjusted so that the direction of the bond is kept intact, but the bond length is scaled to typical bond lengths (i.e., 1.09 Å for CH, 1.01 Å for NH, and 1.35 Å for SH).

The MFCC method was originally used to calculate protein−ligand interaction energies[27] where the procedure is straightforward: the total interaction energy is simply the sum of the capped fragment−ligand interaction energies minus the sum of the concap−ligand interaction energies, i.e., the concaps are used to remove the double counting from using overlapping fragments. The same principle is used to derive the embedding potential.[28] Thus, a property P^a, e.g., a multipole or polarizability, centered on atom a is calculated as

$$P^a = \sum_{f=1}^{F} P_f^a - \sum_{c=1}^{C} P_c^a \tag{1}$$

where P_f^a is the property on atom a in the f'th capped amino acid fragment, and P_c^a is the property on atom a in the c'th concap. The number of capped amino acid fragments and concaps that contain atom a is F and C, respectively. Terminating hydrogens ("Hc" in Figure 1) are considered equivalent to the atoms that they replace when using eq 1. The procedure thus consists of the following four steps:

1. Decompose the protein or peptide into capped amino acid fragments.
2. Merge neighboring caps to form the concaps.
3. Calculate the atom-centered properties of all fragments created in step 1 and 2.
4. Use the atom-centered properties from step 3 and apply eq 1 for each atom in the original protein or peptide to obtain the final embedding potential parameters.

We built fragments with capping groups of two different sizes to examine the effects that the size of the capping group has on

DOI: 10.1021/acs.jctc.5b00078
J. Chem. Theory Comput. 2015, 11, 1832−1842

the quality of the embedding potential. To describe the fragmentation type we define a capping level as shown in Table 1.

The polarizable sites, i.e., atoms that have a polarizability attached to them, also have an exclusion list associated that keeps track of which other sites are allowed to polarize it. The exclusion rule is simple;[28] a site is not allowed to be polarized by other sites if they have been in the same capped amino acid fragment. The reason is that the atom-centered multipoles and polarizabilities of a given fragment correspond to the quantum-mechanically optimized charge distribution of that fragment, and it is therefore reasonable to exclude polarization between those sites.

In this work we used the local properties (LoProp) method[29] implemented in Molcas[30] to calculate the atom-centered multipole moments and polarizabilities and four conventional charge-fitting schemes to obtain ESP-fitted charges, namely the restrained electrostatic potential method (RESP),[31] the Merz−Singh−Kollmann scheme (MK),[32,33] the charges from electrostatic potentials using a grid method (CHELPG),[34] and the Hu-Lu-Yang scheme (HLY).[35] The MK, CHELPG, and HLY charges were calculated using Gaussian 09,[36] while the RESP charges were fitted using AmberTools' Antechamber module[37,38] using an ESP from Gaussian 09. The MK, CHELPG, and HLY charges were constrained to reproduce the molecular dipole moment. We used 10 layers and 17 points per unit area for the fitting of RESP, MK, and HLY charges, while for CHELPG we used default settings. The B3LYP[39−42] exchange-correlation (XC) functional was used in all cases, together with various basis sets depending on the analysis performed. For a given embedding potential we always use the same method and basis set for all parameters.

2.2. Electrostatic Potentials. To evaluate the quality of the embedding potentials we benchmark the ESPs generated by the embedding potentials against quantum-mechanical ESPs. For the amino acid test systems we analyze the ESPs from static multipole moments and ESP-fitted charges. The polarization is already accounted for because the embedding potential is derived from a single quantum-mechanical calculation. In the case of the dipeptides and insulin there are also contributions from induced dipoles that result from electric fields originating both from the permanent multipoles and other induced dipoles.

In the following, we use a multi-index notation[43] that allows compact equations describing the electrostatics using multipole moment expansions. In this notation we define a multi-index as a 3-tuple, $k = (k_x,k_y,k_z)$, where the individual indices are associated with the respective Cartesian coordinates. The multi-index norm is defined as $|k| = k_x + k_y + k_z$, the factorial is $k! = k_x!\,k_y!k_z!$, and the power is $\mathbf{r}^k = x^{k_x}y^{k_y}z^{k_z}$. The partial derivative operator is defined as $\partial^k_{\mathbf{r}} = (\partial^{k_x+k_y+k_z})/(\partial x^{k_x}\partial y^{k_y}\partial z^{k_z})$. Summations involving the norm are over $(|k|+1)(|k|+2)/2$ multi-indices for each $|k|$, e.g., $\sum_{|k|=0}^{1}k = (0,0,0) + (1,0,0) + (0,1,0) + (0,0,1)$.

The ESP, produced by the embedding potential, at $\mathbf{R}_i$ is calculated as

$$V^{\mathrm{emb}}_{\mathrm{esp}}(\mathbf{R}_i) = \sum_{s=1}^{S}\sum_{|k|=0}^{K}\frac{(-1)^{|k|}}{k!}T^{(k)}_{si}M^{(k)}_s \tag{2}$$

where $M^{(k)}_s$ is the k'th component of the $|k|$'th-order multipole moment situated at coordinate $\mathbf{R}_s$, and $T^{(k)}_{si}$ is the k'th component of an interaction tensor defined as

$$T^{(k)}_{si} = \partial^k_{\mathbf{R}_i}\frac{1}{|\mathbf{R}_i - \mathbf{R}_s|} \tag{3}$$

The first summation in eq 2 is over S distributed multipole moments, which includes both permanent and induced moments, and the second summation goes up to the maximum expansion order K.

The distributed polarizabilities give rise to induced dipoles in the presence of an electric field according to

$$\bar{\boldsymbol{\mu}}_s = \boldsymbol{\alpha}_s\mathbf{F}(\mathbf{R}_s) \tag{4}$$

where $\bar{\boldsymbol{\mu}}_s$ is an induced dipole, $\boldsymbol{\alpha}_s$ is a dipole−dipole polarizability, both located at $\mathbf{R}_s$, and $\mathbf{F}(\mathbf{R}_s)$ is the total electric field at $\mathbf{R}_s$. The induced dipoles depend on the field from all other induced dipoles, and eq 4 is therefore solved self-consistently.

The quantum-mechanical ESP is given by

$$V^{\mathrm{qm}}_{\mathrm{esp}}(\mathbf{R}_i) = \sum_{\mu\nu}v^{\mathrm{AO}}_{\mu\nu}(\mathbf{R}_i)D^{\mathrm{AO}}_{\mu\nu} + \sum_{m=1}^{M}\frac{Z_m}{|\mathbf{R}_i - \mathbf{R}_m|} \tag{5}$$

where Z_m is the nuclear charge of the m'th nucleus out of M nuclei, and indices μ and ν refer to AOs. The integrals $v^{\mathrm{AO}}_{\mu\nu}(\mathbf{R}_i)$ are defined by

$$v^{\mathrm{AO}}_{\mu\nu}(\mathbf{R}_i) = -\int\frac{\chi^*_\mu(\mathbf{r})\chi_\nu(\mathbf{r})}{|\mathbf{R}_i - \mathbf{r}|}\,\mathrm{d}\mathbf{r} \tag{6}$$

and $D^{\mathrm{AO}}_{\mu\nu}$ is an element of the AO density matrix.

The classical and quantum-mechanical ESPs were calculated on a cubic grid (4 points per bohr) using the Dalton quantum chemistry suite[44] employing the PE library[26] to calculate classical values (eq 2) and the Dalton[45] and LSDalton programs[46] for the quantum-mechanical reference values (eq 6). For the amino acid and dipeptide quantum-mechanical reference ESPs we used B3LYP, while B3LYP, CAM-B3LYP, or the divide-expand-consolidate (DEC) MP2 method[47] was used for the ESP of insulin. In all cases we used a basis set corresponding to the one used to calculate the embedding potential that is being compared. To facilitate the comparison between classical and quantum-mechanical ESPs, we calculate root-mean-square deviations (RMSDs) using values on a molecular surface. The surfaces are defined by the vdW atomic radii multiplied by a factor. The RMSDs are calculated using points on a given surface and within a threshold distance defined by an increment added to the factor that defines the surface. In the following we use surfaces defined by the vdW radii multiplied by a factor of 2.0 and include all points within factors 2.0 ± 0.01 from the surface unless otherwise specified. Surface plots were created using the VMD Molecular Graphics Viewer.[48,49]

2.3. Structures. The amino acids and dipeptides include acetyl (ACE) and *N*-methyl (NME) capping groups, except AspArg, which has a positive N-terminal and ProPhe, which has a negative O-terminal. The geometries were optimized at the B3LYP/6-31+G*[50−53] level of theory using Gaussian 09. The insulin structure was obtained from the work by Jakobsen et al.[2] (PDBID: 1MSO[54]).

3. RESULTS AND DISCUSSION

In the following we present and discuss the results obtained from the analysis of the embedding potentials. We start with a preliminary analysis on smaller systems, i.e., a set of amino acids

DOI: 10.1021/acs.jctc.5b00078
J. Chem. Theory Comput. 2015, 11, 1832−1842

of different types (Asp, Lys, Phe, and Cys) and a series of dipeptides (AspArg, ArgVal, ValTyr, TyrIle, IleHis, HisPro, and ProPhe). This allows us to evaluate the errors that are introduced by the use of classical parameters, i.e., the multipole moments, ESP-fitted charges and polarizabilities, and by the fragmentation procedure. Therefore, the same method and basis set is used to derive the embedding potentials and to calculate the reference quantum-mechanical ESPs. In the second part we examine the protein embedding potentials, using insulin as the test system, and benchmark them against a high-level quantum-mechanical ESP. Here we use MP2 as the reference because a direct comparison to KS-DFT is not suitable due to the severity of the self-interaction errors for this system.[2] This allows us to directly evaluate the performance of the embedding potentials as they would be used in a real application. In this part we also compare standard molecular mechanics force fields that are commonly used in QM/MM applications.

Table 2. Basis Set Dependence of Amino Acid ESPs Produced by B3LYP Relative to B3LYP/aug-cc-pVQZ

	RMSD[a,b] (kJ/mol)		
	6-31+G*	aug-cc-pVDZ	aug-cc-pVTZ
Asp	1.8	0.22	0.063
Lys	1.5	0.15	0.025
Phe	0.97	0.21	0.032
Cys	1.2	0.29	0.081

[a]Reported as energies from the interaction with a unit point charge. [b]Evaluated on a molecular surface defined by twice the vdW atomic radii.

We use an M*x*P*y*-*z* notation to indicate the type of the LoProp embedding potential, where *x* gives the order of the multipole moment expansion, *y* indicates the polarizability type used (*y* = 1: isotropic, *y* = 2 anisotropic), and *z* is the capping level (see Table 1). The P*y* part is omitted in case no polarizabilities are present. The M*x* part is replaced with RESP, MK, CHELPG, or HLY, respectively, when using ESP-fitted charges instead of permanent multipole moments.

3.1. Preliminary Analysis. Starting first with the amino acids, we examine the basis set dependence of the ESPs by comparing the ESP values, calculated using the B3LYP XC functional and the 6-31+G*, aug-cc-pVDZ, and aug-cc-pVTZ[55−57] basis sets, to B3LYP/aug-cc-pVQZ reference values. The results of the basis set analysis are shown in Table 2 as RMSDs in terms of energies from the interaction with a unit point charge. Overall, the errors are relatively small, especially when compared to the ranges of the ESP values, which for the B3LYP/aug-cc-pVQZ reference are Asp −391.17 to −156.51 kJ/mol; Lys 98.42 to 387.23 kJ/mol; Phe −71.51 to 43.88 kJ/mol; Cys −65.62 to 44.18 kJ/mol. The size of the basis set errors can be a useful reference in the following when evaluating the errors introduced by the embedding potentials. We find that aug-cc-pVDZ represents a good compromise between computational requirements and accuracy.

To examine the convergence with respect to the order of the multipole moment expansion, we evaluate the LoProp embedding potentials by comparing the ESP created by the multipole moments to the quantum-mechanical equivalent, i.e., using the same XC functional and basis set throughout. To obtain reasonable multipole moments it is important to consider which polarization functions are available in a given basis set, since *s* orbitals on an atom can only describe a charge on that atom, *p* orbitals only a dipole, and so on. Therefore, we used the aug-cc-pVQZ basis set for this part of the study, since it contains polarization functions up to *g*-type orbitals which allows us to go up to hexadecapoles, i.e., a fourth-order expansion. The importance of higher-order angular momentum functions for the multipole moments is seen in Figure S1 in the

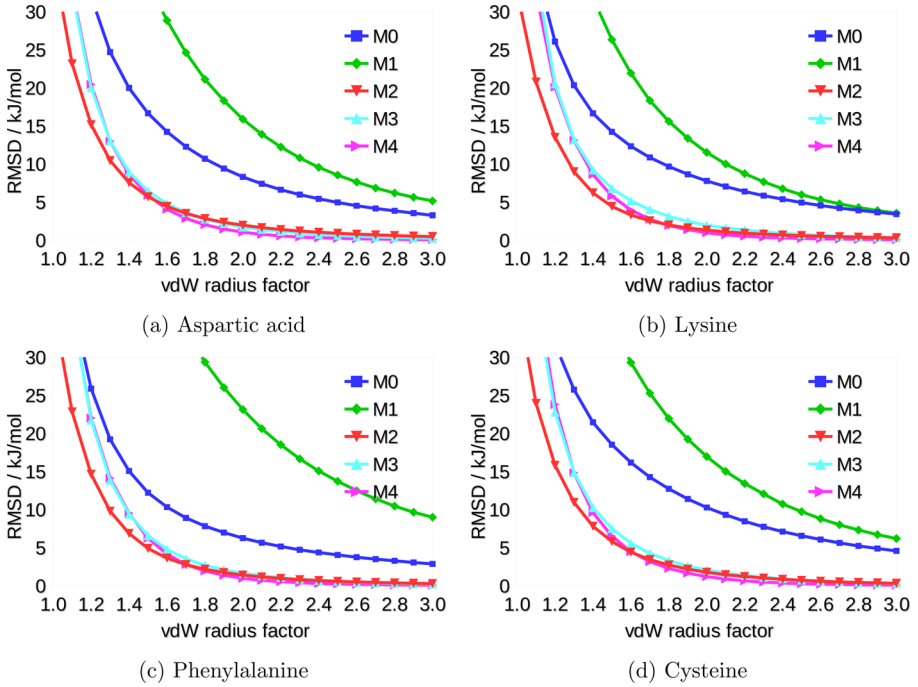


Figure 2. Distance dependence of amino acid ESPs produced by LoProp potentials relative to B3LYP. The values are reported as energies from the interaction with a unit point charge.

DOI: 10.1021/acs.jctc.5b00078
J. Chem. Theory Comput. 2015, 11, 1832−1842

Supporting Information (SI), which shows the convergence of the multipole moments with respect to basis set size. Nevertheless, the error that is introduced by the multipole expansion is consistent across different basis sets (see Table S1 in the SI) which means that conclusions based on aug-cc-pVQZ are also valid for other basis sets.

In Figure 2 we show the RMSDs of the amino acid ESPs with respect to distance from the molecule and order of the multipole moment expansion. First of all we note that the convergence behavior is maybe not as one would expect from a multipole expansion. Normally, we would expect an improved ESP with respect to increasing order of the expansion. However, in the LoProp scheme, the charges-only potential (M0) generally produces much better ESPs than the potential that also includes dipoles (M1). We observe this spurious convergence behavior across all amino acids and have observed the same behavior in previous studies on solvent embedding potentials.[15,58] Including quadrupoles (M2) yields a significant improvement compared to the M0 potential at all distances, while going to higher orders has very little effect in comparison, in accordance with previous observations.[59] Also noteworthy is the fact that the M2 and higher-order potentials perform very well even quite close to the molecule. The errors from around a factor of 2 of the vdW radii (see Table 3) are comparable to the

Table 3. Accuracy of Amino Acid ESPs Produced by LoProp Potentials Relative to B3LYP

	RMSD[a,b] (kJ/mol)				
	M0	M1	M2	M3	M4
Asp	8.3	15.8	1.9	1.7	1.0
Lys	7.8	11.4	1.3	2.0	1.0
Phe	6.3	22.7	1.4	1.6	1.0
Cys	10.2	16.9	1.8	2.1	1.2

[a]Reported as energies from the interaction with a unit point charge. [b]Evaluated on a molecular surface defined by twice the vdW atomic radii.

basis set error from using 6-31+G* as compared to aug-cc-pVQZ in full quantum-mechanical calculations (see Table 2). However, the errors increase rapidly as we approach the actual vdW surface. These large errors close to the nuclei can be attributed to charge penetration errors[60] and the divergence of the multipole expansion.[61] How far these errors extend from the nuclei depends partly on the distribution algorithm used for atom−atom midpoint sites and partly on the diffuseness of the basis set as shown in Figure S3 in the SI. In relation to intermolecular interactions we do not expect to have a significant amount of charge close to the vdW surface and therefore do not expect these errors to have a large impact in, for instance, embedding calculations.

Finally, we also benchmark the ESP-fitted charge potentials. For this analysis we used the aug-cc-pVDZ basis set and only consider a single molecular surface which is the one defined by twice the vdW radii. However, we note that the behavior of the ESP-fitted charge potentials with respect to distance from the molecule is very similar and resembles also the LoProp potentials (see Figure S4 in the SI). We also include the LoProp M2 potential (based on aug-cc-pVDZ to make the comparisons consistent). The RMSDs are presented in Table 4. Overall, the ESP-fitted charge potentials have very similar performances that are comparable to those of the M2 potential. The small variations between the different ESP-fitted charge

Table 4. Accuracy of Amino Acid ESPs Produced by ESP-Fitted Charge Potentials Relative to B3LYP

	RMSD[a,b](kJ/mol)				
	RESP	MK	CHELPG	HLY	M2
Asp	2.6	2.1	2.7	2.0	2.0
Lys	2.4	2.0	2.3	1.9	1.2
Phe	2.5	1.7	3.4	1.4	1.3
Cys	3.4	3.0	3.3	2.9	1.7

[a]Reported as energies from the interaction with a unit point charge. [b]Evaluated on a molecular surface defined by twice the vdW atomic radii.

potentials can be attributed to differences in the fitting procedures. For instance, the RESP procedure constrains charges on chemically equivalent atoms to be equal which removes some degrees of freedom compared to the others. Interestingly, the cysteine RMSDs are around 3 kJ/mol for all of the ESP-fitted charge potentials, while this is not the case for the M2 potential. The reason could be that there are anisotropies that are described better by the higher-order multipole moments than by the fitted charges.

To conclude, based on the analysis performed on the amino acids, we find that the M2 potential, which includes multipoles up to quadrupoles, is a very good compromise between computational complexity and performance. It yields significant improvement compared to lower-order potentials and is essentially converged with respect to the order of the multipole moment expansion. All the ESP-fitted charge potentials produce ESPs that are comparable in quality to the M2 potential, and they are therefore an attractive alternative to the LoProp potentials due to their simplicity. However, it comes at a cost because we also find that the multipole expansion is more robust, consistently yielding lower RMSDs.

3.2. Dipeptide Potentials. Analyzing the dipeptide ESPs allows us to examine the effects from the fragmentation procedure more closely. We only consider M2-based potentials since these were shown to be optimal. In addition, we include RESP-based potentials as a representative of ESP-fitted charge potentials.

In Table 5 we present the RMSDs of the dipeptide ESPs calculated from the M2-based potentials relative to the quantum-mechanical reference. We include potentials with isotropic (M2P1) or anisotropic (M2P2) polarizabilities and compare to potentials without polarizabilities. In addition, we also include potentials that have been derived using different

Table 5. Accuracy of Dipeptide ESPs Produced by LoProp Potentials Relative to B3LYP

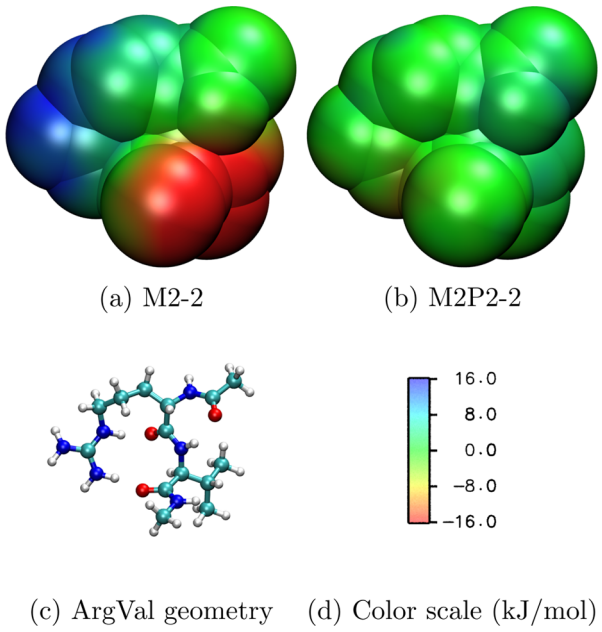
	RMSD[a,b] (kJ/mol)			
	M2-2	M2P2-2	M2P1-2	M2P2-1
AspArg	3.3	1.6	1.7	2.5
ArgVal	10.4	2.7	3.2	3.3
ValTyr	1.8	1.8	1.8	3.2
TyrIle	2.6	1.9	1.7	3.3
IleHis	2.7	1.6	1.7	2.8
HisPro	2.5	4.2	4.2	9.9
ProPhe	5.5	2.6	2.8	4.2

[a]Reported as energies from the interaction with a unit point charge. [b]Evaluated on a molecular surface defined by twice the vdW atomic radii.

DOI: 10.1021/acs.jctc.5b00078
J. Chem. Theory Comput. 2015, 11, 1832−1842

capping levels (see Table 1). Looking first at the M2-2 potential, which does not contain polarizabilities, we see that in most cases the RMSD is roughly 2−3 kJ/mol which is only slightly higher than the amino acid RMSDs (i.e., compared to M2 in Table 4). This additional error is due to the fragmentation scheme. There are also cases where the RMSD is significantly higher (10.4 kJ/mol for ArgVal and 5.5 kJ/mol for ProPhe) which shows that the fragmentation can also introduce a larger error in some cases even for small systems.

Including anisotropic polarizabilities (M2P2-2) generally lowers the RMSDs to a level comparable to the RMSDs seen for the single amino acids. The RMSDs in the ProPhe and ArgVal cases, which are high without polarization, are down to below 3 kJ/mol. This is not as low as some of the other dipeptides (<2 kJ/mol) but clearly shows that the polarizabilities can recover a large part of the error that is introduced with the fragmentation procedure. Especially, the ArgVal case demonstrates this nicely as seen in Figure 3 which shows the



(a) M2-2 (b) M2P2-2

(c) ArgVal geometry (d) Color scale (kJ/mol)

Figure 3. Difference between ArgVal ESP produced by (a) M2-2 and (b) M2P2-2 relative to B3LYP. Surfaces are defined by twice the vdW atomic radii.

ESP differences on a surface using M2-2 and M2P2-2 potentials relative to the full quantum-mechanical ESP. Here we see that the ESP around the arginine side-chain is too positive, while the ESP surrounding the valine is too negative when no polarization is included (Figure 3a). Figure 3b shows the ESP when polarization is included, and here we see that the problematic areas are now essentially gone.

There is also one case where the polarization worsens the ESP, namely HisPro where the RMSD increases from 2.5 to 4.2 kJ/mol. Suspecting overpolarization effects, we performed test calculations using Tholes exponential damping scheme[62,63] which damps the electric fields between close-lying sites. By adjusting the amount of damping we were able to lower the RMSD to 2.1 kJ/mol, which supports our suspicion; however, including the same amount of damping for the other dipeptides resulted in higher RMSDs, so we did not pursue this path further.

Using isotropic polarizabilities instead of full anisotropic polarizability tensors does not change the RMSDs significantly. The largest difference is seen for ArgVal where there is a modest increase of 0.5 kJ/mol. Using smaller capping groups, i.e., capping level one compared to level two (see Table 1), increases the RMSDs by roughly 1 kJ/mol. This is not a severe deterioration, but in most cases it is not an improvement compared to not including polarization. HisPro is an exception again, where the RMSD more than doubles to 9.9 kJ/mol, which also points toward overpolarization effects in this particular dipeptide.

The RMSDs of the RESP-based potentials are shown in Table 6 including also the RMSDs from M2-based potentials

Table 6. Accuracy of Dipeptide ESPs Produced by RESP-Based Potentials Relative to B3LYP

	RMSD[a,b] (kJ/mol)			
	RESP-2	RESPP1-2	M2-2	M2P2-2
AspArg	3.9	3.0	3.3	1.6
ArgVal	11.2	5.5	10.4	2.7
ValTyr	4.0	4.2	1.8	1.8
TyrIle	4.0	5.5	2.6	1.9
IleHis	3.7	3.2	2.7	1.6
HisPro	3.9	6.9	2.5	4.2
ProPhe	7.6	3.6	5.5	2.6

[a]Reported as energies from the interaction with a unit point charge. [b]Evaluated on a molecular surface defined by twice the vdW atomic radii.

for comparison (RMSDs of the other ESP-fitted charge potentials are shown in Table S2 in the SI). Without polarization the RMSDs are about 4 kJ/mol which is slightly higher than the RMSDs from the M2-2 potentials. This is not unexpected since the same was observed for the amino acids. The trends when including polarizabilities are not as clear as they are in the case of the LoProp potentials where we observe improvements in most of the dipeptides. There is a small gain for AspArg and IleHis, and there is still a nice improvement in the ArgVal and ProPhe cases. However, there are three dipeptides, HisPro, TyrIle and ValTyr, where the RMSDs increase due to the polarization. A possible explanation for this discrepancy between multipole- and ESP-fitted charge potentials is that each higher-order multipole explicitly models anisotropy effects, whereas the charge-fitting schemes must rely on several point charges in order to describe anisotropies.

From the analysis on the dipeptides we find that the polarizabilities can recover some of the errors that are introduced by the fragmentation procedure. This is quite clear for the LoProp potentials but less obvious for the ESP-fitted charge potentials. Furthermore, we also find that using isotropic polarizabilities instead of the anisotropic tensors does not result in significant loss in quality. Using smaller capping groups, however, does give rise to a small penalty. It is important to keep in mind that we here only consider polarization between two neighboring amino acids, whereas in a larger system there would also be large many-body polarization effects. This will be investigated in the following.

3.3. Protein Embedding Potentials. Here we present the results from calculations using a whole protein exemplified by the case of insulin. The ESP calculated from the embedding potentials is benchmarked against a full-structure MP2/cc-pVDZ ESP. The ab initio-derived embedding potentials are

DOI: 10.1021/acs.jctc.5b00078
J. Chem. Theory Comput. 2015, 11, 1832−1842

based on B3LYP/cc-pVDZ calculations so that the same basis set is used throughout. In Table 7 we show the RMSDs of the insulin ESPs on a molecular surface defined by twice the vdW atomic radii, and in Figure 4 we show the RMSDs as a function of distance from the protein. It is immediately clear that the B3LYP ESP has very large errors, especially considering that the (MP2) ESP ranges from −498 to 278 kJ/mol. This is basically what Jakobsen et al.[2] observed in their study. These large errors are attributed to the self-interaction error in KS-DFT that leads to fractional electron transfer in anions.[64,65] This can be especially problematic for proteins[2] where fractional electrons can be artificially transferred from anionic to cationic sites (see Figure 5a). We also see that these errors can partly be corrected by using the long-range-corrected CAM-B3LYP functional (see Figure 5b).

Table 7. Accuracy of the Insulin ESPs Produced by M2- and RESP-Based Potentials, AMBER Force Fields, B3LYP, and CAM-B3LYP Relative to MP2

	RMSD[a,b] (kJ/mol)
M2-2	24.3
M2P2-2	4.7
M2P1-2	6.0
M2P2-1	6.4
RESP-2	25.3
RESPP1-2	8.2
ff94	33.5
ff03	36.1
GAFF-like	28.7
B3LYP	43.9
CAMB3LYP	7.7

[a]Reported as energies from the interaction with a unit point charge. [b]Evaluated on a molecular surface defined by twice the vdW atomic radii.

The M2P2-2 potential produces a significantly improved ESP with relatively small RMSDs of around 5 kJ/mol, except very close to the protein where the RMSD increases due to charge penetration errors[60] and nonconvergent multipole expansions.[61] Even though the M2P2-2 embedding potential has been calculated using B3LYP, we avoid the fractional electron transfer issue by fragmenting the protein into amino acid components and thus force the electrons to stay localized on each fragment (see Figure 5c). Interestingly, the M2P2-2 potential produces a slightly better ESP than the full CAM-B3LYP ESP even at distances quite close to the molecule (until around 1.5 times the vdW radii). The reason is that CAM-B3LYP also suffers slightly from the fractional electron transfer issues (see Figure 5b). These results clearly demonstrate the excellent performance that can be achieved using the M2P2-2 potential and also stress the importance of modeling explicit polarization effects. The M2-2 RMSD, calculated at twice the vdW radii, decreases from 24.3 kJ/mol to just 4.7 kJ/mol for M2P2-2, which is remarkable considering it is 3 kJ/mol less than the RMSD of the full-structure CAM-B3LYP ESP. Using isotropic polarizabilities increases the RMSD slightly as does the use of smaller capping groups. The ESP from the RESP-2 potential has errors that are comparable to M2-2, as one might expect based on the results from the preliminary analysis. However, where the dipeptide analysis showed mixed results when polarization was included, we here see an improvement that is on par with the M2-based potentials.

The results presented here show that the ab initio-derived potentials are highly suitable for embedding calculations because they will allow use of smaller quantum regions while still retaining high accuracy. Furthermore, they can also be used directly for the evaluation of ESPs. Alternatively one could also use a fragment-based quantum-mechanical method.[66] However, polarization effects can be very important, as demonstrated here, and they are computationally expensive to obtain in full from such fragment-based methods, since it requires an iterative procedure that involves many quantum mechanical calculations on each fragment. Using instead classical polarization, as in our case, requires only one quantum-mechanical calculation on each fragment and one self-consistent determination of the induced dipoles which is comparatively cheap.

We also included two different variants of AMBER force fields in this benchmark, namely ff94 and ff03. Note that the ff94 charges are also used in many of the newer AMBER force fields such as ff99 and variants. In addition, we also calculated general AMBER force field (GAFF)[67] charges, meaning that they are based on HF/6-31G* RESP calculations but otherwise derived using the same procedure as the other ab initio-derived potentials. It may be argued that the comparison is not entirely straightforward since the ff94 and GAFF-like charges implicitly include some solvent polarization, which is not present in the calculations performed here. The electrostatic component of the ff03 force field, however, has been constructed so that the implicit polarization mimics that of a protein which makes it a

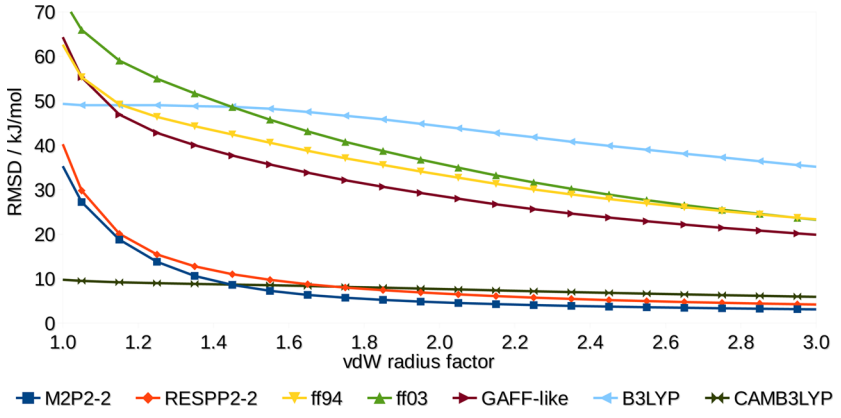


Figure 4. Distance dependence of the insulin ESPs produced by the LoProp potentials, AMBER force fields, B3LYP, and CAM-B3LYP at different distances from the protein relative to MP2. The values are reported as interaction energies with a unit point charge.

DOI: 10.1021/acs.jctc.5b00078
J. Chem. Theory Comput. 2015, 11, 1832−1842

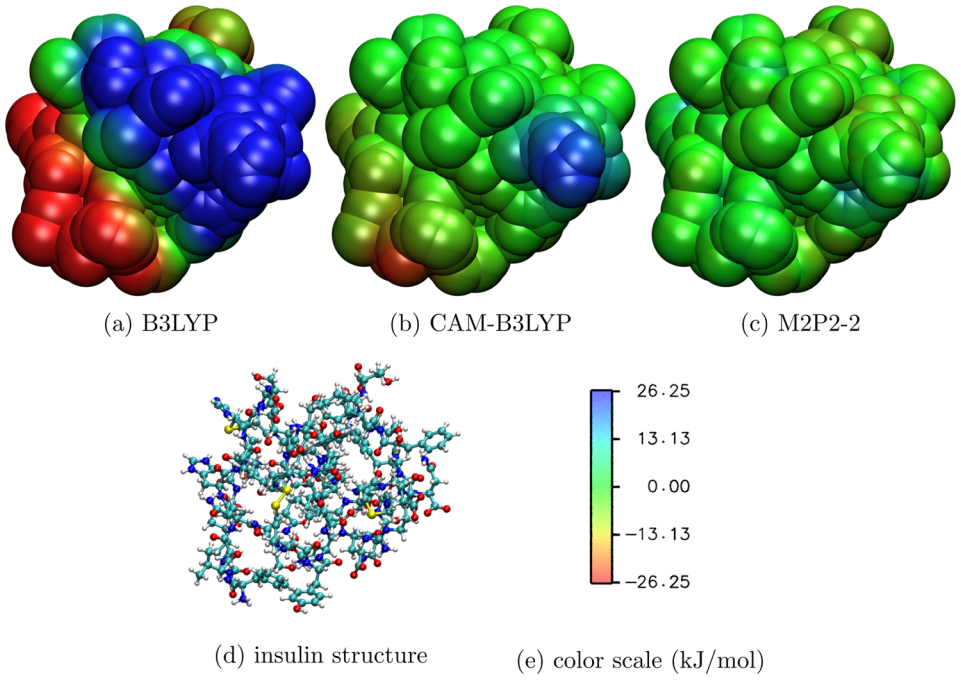


Figure 5. Difference between insulin ESPs produced by (a) B3LYP, (b) CAM-B3LYP, and (c) M2P2-2 relative to MP2. The surface is defined by twice the vdW atomic radii.

more rigorous comparison. All things considered, we find that the comparisons presented here are useful because AMBER and other standard molecular mechanics force fields are routinely used for electrostatic embedding QM/MM calculations of molecular response properties (see e.g., ref 5 and references therein). For such applications it is essential that the electrostatic component is of good quality, although the importance of the environment obviously varies. Our results show that the AMBER force fields produce an ESP which is quite poor compared to the MP2 reference. They are roughly equivalent to the nonpolarizable versions of the M2-2 and RESP-2 potentials showing that implicit polarization is not flexible enough to yield a decent quality ESP. These results explain why some studies conclude that very large quantum regions are needed to get converged QM/MM electronic properties.[1,14] The reason is not inherent in the QM/MM formalism but simply that use of low quality embedding potentials is not suitable for such calculations. The results presented here clearly show that what is really needed is a better embedding potential that can reproduce the full quantum-mechanical ESP also at short distances. This has also been demonstrated in recent studies. In an investigation of solvent effects on nuclear magnetic shielding constants by Steinmann et al. it was shown that the size of the quantum region can be substantially reduced when a high-quality embedding potential is used compared to standard solvent potentials.[24] In addition, Schwabe et al. demonstrated that electronic excitations in green fluorescent protein (GFP) can be accurately reproduced using polarizable embedding potentials when compared to full quantum-mechanical calculations.[68]

4. SUMMARY AND OUTLOOK

We have analyzed the performance of polarizable protein potentials designed for QM/MM-type embedding calculations of molecular response properties. In particular, we benchmark the electrostatic potentials calculated from ab initio-derived embedding potentials against suitable quantum-mechanical references using amino acids, dipeptides, and insulin model systems. The potentials were derived using the molecular fractionation with conjugate caps method[27] as defined by Söderhjelm and Ryde.[28] Two types of potentials were examined, one based on distributed multipole moments[29] and the other based on ESP-fitted atomic charges.[31,32,34,35] In both cases we included either anisotropic or isotropic polarizabilities.

Starting from the amino acids, we find that the accuracy that can be obtained is of the same order as the basis set error from using 6-31+G* compared to aug-cc-pVQZ in full quantum-mechanical calculations. The errors close to the van der Waals surface can be quite large due to charge penetration errors and divergences in the multipole expansion; however, we do not expect this error to be significant at distances relevant for most intermolecular interactions. We find that the most optimal embedding potential, with respect to performance and complexity, is obtained by including multipoles up to quadrupoles. The different ESP-fitting schemes yield very similar results that are comparable to the multipole-based potential. However, the multipole-based potentials have the lowest error and the most consistent behavior in all cases considered here. The analysis of the dipeptides shows that large errors can be introduced by the fragmentation procedure; however, these errors can be significantly reduced by including explicit polarization.

For the analysis of the full protein embedding potentials we use insulin as a case study. The insulin ESPs calculated from the embedding potentials are compared to a high-level MP2 reference. Jakobsen et al.[2] have shown that the insulin ESP produced by a full-structure B3LYP calculation, as well as other common XC functionals, contains substantial errors due to issues with fractional electron transfer.[64,65] We demonstrate that an accurate ESP, compared to the MP2 reference, can be obtained from polarizable embedding potentials derived using a fragmentation procedure using the B3LYP functional for the

DOI: 10.1021/acs.jctc.5b00078
J. Chem. Theory Comput. 2015, 11, 1832−1842

individual fragments. It is critical to include explicit polarization in order to obtain high accuracy. We find that the ESPs calculated from these embedding potentials rivals the fully quantum-mechanical ESP from CAM-B3LYP even quite close to the van der Waals surface. This makes the potentials highly suitable for QM/MM embedding calculations of electronic properties since small quantum regions can be used. Some recent studies report that very large quantum regions are needed for QM/MM property calculations;[1,14] however, their conclusions are based on the use of standard molecular mechanics force fields that are not suitable for such calculations. The reason is that the ESP produced by the partial charges from such force fields is quite poor. This is mainly because polarization is important and cannot be modeled satisfactorily by including it implicitly. Interestingly, we find that using isotropic polarizabilities instead of the anisotropic does not lead to significant deterioration. This, and the fact that the ESP-fitted charges produce reasonably good ESPs, is promising for the possibility of making a transferable high-quality embedding potential, based on atom-centered charges and isotropic polarizabilities, for QM/MM calculations of molecular response properties.

■ ASSOCIATED CONTENT

Supporting Information

Figure S1 shows the convergence of the distributed multipole moments of Asp with respect to basis-set size (atom labeling is given in Figure S2). Figure S3 displays the effect that the diffuseness of the basis set has on the errors in the MK potential for Asp. The distance dependence of the amino acid ESPs produced by ESP-fitted charge potentials relative to B3LYP are given in Figure S4. Table S1 reports RMSDs that show the basis set dependence of the errors in the amino acid ESPs produced by the LoProp M2 potential compared to B3LYP. Table S2 provides the RMSDs between the amino-acid ESPs produced by the ESP-fitted charge potentials also compared to B3LYP. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: magnus.olsen@epfl.ch.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

J.M.H.O. acknowledges financial support from the Danish Council for Independent Research (DFF) through the Sapere Aude research career program. J.K. thanks the Danish Council for Independent Research (the Sapere Aude program), the Villum Foundation, the Lundbeck Foundation, and the Danish e-Infrastructure Cooperation (DeIC) for financial support. The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013)/ERC Grant Agreement No. 291371.

■ REFERENCES

(1) Flaig, D.; Beer, M.; Ochsenfeld, C. Convergence of Electronic Structure with the Size of the QM Region: Example of QM/MM NMR Shieldings. *J. Chem. Theory Comput.* **2012**, *8*, 2260−2271.

(2) 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.

(3) 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.

(4) Lin, H.; Truhlar, D. QM/MM: what have we learned, where are we, and where do we go from here? *Theor. Chem. Acc.* **2007**, *117*, 185−199.

(5) Senn, H. M.; Thiel, W. QM/MM Methods for Biomolecular Systems. *Angew. Chem., Int. Ed.* **2009**, *48*, 1198−1229.

(6) 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.

(7) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* **2006**, *65*, 712−725.

(8) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* **2003**, *24*, 1999−2012.

(9) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.

(10) Mackerell, A. D.; Feig, M.; Brooks, C. L. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* **2004**, *25*, 1400−1415.

(11) Jorgensen, W. L.; Tirado-Rives, J. The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.* **1988**, *110*, 1657−1666.

(12) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225−11236.

(13) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides. *J. Phys. Chem. B* **2001**, *105*, 6474−6487.

(14) Isborn, C. M.; Götz, A. W.; Clark, M. A.; Walker, R. C.; Martínez, T. J. Electronic Absorption Spectra from MM and ab Initio QM/MM Molecular Dynamics: Environmental Effects on the Absorption Spectrum of Photoactive Yellow Protein. *J. Chem. Theory Comput.* **2012**, *8*, 5092−5106.

(15) Olsen, J. M.; Aidas, K.; Kongsted, J. Excited States in Solution through Polarizable Embedding. *J. Chem. Theory Comput.* **2010**, *6*, 3721−3734.

(16) Sneskov, K.; Schwabe, T.; Christiansen, O.; Kongsted, J. Scrutinizing the effects of polarization in QM/MM excited state calculations. *Phys. Chem. Chem. Phys.* **2011**, *13*, 18551−18560.

(17) List, N. H.; Olsen, J. M. H.; Jensen, H. J. Aa.; Steindal, A. H.; Kongsted, J. Molecular-Level Insight into the Spectral Tuning Mechanism of the DsRed Chromophore. *J. Phys. Chem. Lett.* **2012**, *3*, 3513−3521.

(18) Beerepoot, M. T. P.; Steindal, A. H.; Kongsted, J.; Brandsdal, B. O.; Frediani, L.; Ruud, K.; Olsen, J. M. H. A polarizable embedding DFT study of one-photon absorption in fluorescent proteins. *Phys. Chem. Chem. Phys.* **2013**, *15*, 4735−4743.

DOI: 10.1021/acs.jctc.5b00078
J. Chem. Theory Comput. 2015, 11, 1832−1842

(19) Olsen, J. M. H.; Kongsted, J. Molecular Properties through Polarizable Embedding. *Adv. Quantum Chem.* **2011**, *61*, 107−143.
(20) Sneskov, K.; Schwabe, T.; Kongsted, J.; Christiansen, O. The polarizable embedding coupled cluster method. *J. Chem. Phys.* **2011**, *134*, 104108.
(21) Schwabe, T.; Sneskov, K.; Olsen, J. M. H.; Kongsted, J.; Christiansen, O.; Hättig, C. PERI-CC2: A Polarizable Embedded RI-CC2 Method. *J. Chem. Theory Comput.* **2012**, *8*, 3274−3283.
(22) Hedegård, E. D.; List, N. H.; Jensen, H. J. Aa.; Kongsted, J. The multi-configuration self-consistent field method within a polarizable embedded framework. *J. Chem. Phys.* **2013**, *139*, 044101.
(23) Pedersen, M. N.; Hedegård, E. D.; Olsen, J. M. H.; Kauczor, J.; Norman, P.; Kongsted, J. Damped Response Theory in Combination with Polarizable Environments: The Polarizable Embedding Complex Polarization Propagator Method. *J. Chem. Theory Comput.* **2014**, *10*, 1164−1171.
(24) Steinmann, C.; Olsen, J. M. H.; Kongsted, J. Nuclear Magnetic Shielding Constants from Quantum Mechanical/Molecular Mechanical Calculations Using Polarizable Embedding: Role of the Embedding Potential. *J. Chem. Theory Comput.* **2014**, *10*, 981−988.
(25) Olsen, J. M. H. Development of Quantum Chemical Methods towards Rationalization and Optimal Design of Photoactive Proteins. Ph.D. Thesis, University of Southern Denmark, Odense, Denmark, 2012; DOI:10.6084/m9.figshare.156852.
(26) Olsen, J. M. H. PElib: The Polarizable Embedding library (version 1.0.8). 2014.
(27) 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.
(28) 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.
(29) Gagliardi, L.; Lindh, R.; Karlström, G. Local properties of quantum chemical systems: The LoProp approach. *J. Chem. Phys.* **2004**, *121*, 4494−4500.
(30) Aquilante, F.; De Vico, L.; Ferré, N.; Ghigo, G.; Malmqvist, P.-A.; Neogrády, P.; Pedersen, T. B.; Pitoňák, M.; Reiher, M.; Roos, B. O.; Serrano-Andrés, L.; Urban, M.; Veryazov, V.; Lindh, R. MOLCAS 7: The Next Generation. *J. Comput. Chem.* **2010**, *31*, 224−247.
(31) Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.* **1993**, *97*, 10269−10280.
(32) Singh, U. C.; Kollman, P. A. An approach to computing electrostatic charges for molecules. *J. Comput. Chem.* **1984**, *5*, 129−145.
(33) Besler, B. H.; Merz, K. M.; Kollman, P. A. Atomic charges derived from semiempirical methods. *J. Comput. Chem.* **1990**, *11*, 431−439.
(34) Breneman, C. M.; Wiberg, K. B. Determining atom-centered monopoles from molecular electrostatic potentials. The need for high sampling density in formamide conformational analysis. *J. Comput. Chem.* **1990**, *11*, 361−373.
(35) Hu, H.; Lu, Z.; Yang, W. Fitting Molecular Electrostatic Potentials from Quantum Mechanical Calculations. *J. Chem. Theory Comput.* **2007**, *3*, 1004−1013.
(36) 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 A.02*; Gaussian, Inc.: Wallingford, CT, 2009.
(37) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graphics Modell.* **2006**, *25*, 247260.
(38) Case, D.; Darden, T.; Cheatham, T. E.; Simmerling, I.; Wang, C.; Duke, J.; Luo, R.; Walker, R.; Zhang, W.; Merz, K.; Roberts, B.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossvai, I.; Wong, K.; Paesani, F.; Vanicek, J.; Liu, J.; Wu, X.; Brozell, S.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M.-J.; Cui, G.; Roe, D.; Mathews, D.; Seetin, M.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. *AMBER 11*; University of California: San Francisco, 2010.
(39) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648−5652.
(40) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **1988**, *37*, 785−789.
(41) Vosko, S. H.; Wilk, L.; Nusair, M. Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Can. J. Phys.* **1980**, *58*, 1200−1211.
(42) 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.
(43) Raymond, X. S. *Elementary Introduction to the Theory of Pseudodifferential Operators*; CRC Press: Boca Raton, FL, 1991.
(44) 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.; Jansik, 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.; Salek, 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.
(45) Dalton, a molecular electronic structure program, Release Dalton2013, 2013. See http://daltonprogram.org/ (accessed Mar 27, 2015).
(46) LSDalton, a linear scaling molecular electronic structure program, Release LSDalton2013, 2013. See http://daltonprogram.org/ (accessed Mar 27, 2015).
(47) 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.
(48) Humphrey, W.; Dalke, A.; Schulten, K. VMD - Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14*, 33−38.
(49) Stone, J. An Efficient Library for Parallel Ray Tracing and Animation. M.Sc. Thesis, Computer Science Department, University of Missouri—Rolla, 1998.
(50) 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.

DOI: 10.1021/acs.jctc.5b00078
J. Chem. Theory Comput. 2015, 11, 1832−1842

(51) Hariharan, P.; Pople, J. The influence of polarization functions on molecular orbital hydrogenation energies. *Theor. Chim. Acta* **1973**, *28*, 213−222.
(52) Francl, M. M.; Pietro, W. J.; Hehre, W. J.; Binkley, J. S.; Gordon, M. S.; DeFrees, D. J.; Pople, J. A. Self-consistent molecular orbital methods. XXIII. A polarization-type basis set for second-row elements. *J. Chem. Phys.* **1982**, *77*, 3654−3665.
(53) Clark, T.; Chandrasekhar, J.; Spitznagel, G. W.; Schleyer, P. V. R. Efficient diffuse function-augmented basis sets for anion calculations. III. The 3-21+G basis set for first-row elements, LiF. *J. Comput. Chem.* **1983**, *4*, 294−301.
(54) Smith, G. D.; Pangborn, W. A.; Blessing, R. H. The structure of T_6 human insulin at 1.0 Å resolution. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2003**, *59*, 474−482.
(55) 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.
(56) 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.
(57) Woon, D. E.; Dunning, T. H. Gaussian basis sets for use in correlated molecular calculations. III. The atoms aluminum through argon. *J. Chem. Phys.* **1993**, *98*, 1358−1371.
(58) Schwabe, T.; Olsen, J. M. H.; Sneskov, K.; Kongsted, J.; Christiansen, O. Solvation Effects on Electronic Transitions: Exploring the Performance of Advanced Solvent Potentials in Polarizable Embedding Calculations. *J. Chem. Theory Comput.* **2011**, *7*, 2209−2217.
(59) Söderhjelm, P.; Krogh, J. W.; Karlström, G.; Ryde, U.; Lindh, R. Accuracy of distributed multipoles and polarizabilities: Comparison between the LoProp and MpProp models. *J. Comput. Chem.* **2007**, *28*, 1083−1090.
(60) Stone, A. J. *The Theory of Intermolecular Forces*, 2nd ed.; Oxford University Press: UK, 2013; Chapter 8.1.
(61) Stone, A. J. *The Theory of Intermolecular Forces*, 2nd ed.; Oxford University Press, UK, 2013; Chapter 7.1−7.3.
(62) Thole, B. Molecular polarizabilities calculated with a modified dipole interaction. *Chem. Phys.* **1981**, *59*, 341−350.
(63) van Duijnen, P. T.; Swart, M. Molecular and Atomic Polarizabilities: Thole's Model Revisited. *J. Phys. Chem. A* **1998**, *102*, 2399−2407.
(64) Vydrov, O. A.; Scuseria, G. E. Assessment of a long-range corrected hybrid functional. *J. Chem. Phys.* **2006**, *125*, 234109.
(65) Jensen, F. Describing Anions by Density Functional Theory: Fractional Electron Affinity. *J. Chem. Theory Comput.* **2010**, *6*, 2726−2735.
(66) Gordon, M. S.; Fedorov, D. G.; Pruitt, S. R.; Slipchenko, L. V. Fragmentation Methods: A Route to Accurate Calculations on Large Systems. *Chem. Rev.* **2012**, *112*, 632−672.
(67) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157−1174.
(68) Schwabe, T.; Beerepoot, M. T. P.; Olsen, J. M. H.; Kongsted, J. Analysis of computational models for an accurate study of electronic excitations in GFP. *Phys. Chem. Chem. Phys.* **2015**, *17*, 2582−2588.

DOI: 10.1021/acs.jctc.5b00078
J. Chem. Theory Comput. 2015, 11, 1832−1842