Extension of the Effective Fragment Potential Method to Macromolecules

Pradeep Kumar Gurunathan,†,‡ Atanu Acharya,‡,§ Debashree Ghosh,‡,∥ Dmytro Kosenkov,†,∥ Ilya Kaliman,†,∥ Yihan Shao,⊥ Anna I. Krylov,‡,§ and Lyudmila V. Slipchenko*†

†Department of Chemistry, Purdue University, West Lafayette, Indiana 47907, United States
‡Department of Chemistry, University of Southern California, Los Angeles, California 90089-0482, United States
§Physical and Materials Chemistry Division, CSIR-National Chemical Laboratory, Pune 411008, Maharashtra, India
∥Department of Chemistry and Physics, Monmouth University, West Long Branch, New Jersey 07764, United States
⊥Q-Chem Inc., 6601 Owens Drive, Suite 105 Pleasanton, California 94588, United States

ABSTRACT: The effective fragment potential (EFP) approach, which can be described as a nonempirical polarizable force field, affords an accurate first-principles treatment of noncovalent interactions in extended systems. EFP can also describe the effect of the environment on the electronic properties (e.g., electronic excitation energies and ionization and electron-attachment energies) of a subsystem via the QM/EFP (quantum mechanics/EFP) polarizable embedding scheme. The original formulation of the method assumes that the system can be separated, without breaking covalent bonds, into closed-shell fragments, such as solvent and solute molecules. Here, we present an extension of the EFP method to macromolecules (mEFP). Several schemes for breaking a large molecule into small fragments described by EFP are presented and benchmarked. We focus on the electronic properties of molecules embedded into a protein environment and consider ionization, electron-attachment, and excitation energies (single-point calculations only). The model systems include chromophores of green and red fluorescent proteins surrounded by several nearby amino acid residues and phenolate bound to the T4 lysozyme. All mEFP schemes show robust performance and accurately reproduce the reference full QM calculations. For further applications of mEFP, we recommend either the scheme in which the peptide is cut along the Cα−C bond, giving rise to one fragment per amino acid, or the scheme with two cuts per amino acid, along the Cα−C and Cα−N bonds. While using these fragmentation schemes, the errors in solvatochromic shifts in electronic energy differences (excitation, ionization, electron detachment, or electron-attachment) do not exceed 0.1 eV. The largest error of QM/mEFP against QM/EFP (no fragmentation of the EFP part) is 0.06 eV (in most cases, the errors are 0.01−0.02 eV). The errors in the QM/molecular mechanics calculations with standard point charges can be as large as 0.3 eV.

1. INTRODUCTION

Electronic processes in complex environments are at the heart of numerous phenomena of fundamental and societal importance, such as catalysis, solar energy harvesting, and photovoltaics. Predictive computational modeling is instrumental for advancing our mechanistic understanding of the redox and photoinduced processes in a condensed phase; it requires a combination of quantum mechanical methods and an appropriate description of the environment (solvent, protein, molecular solids, etc.). The effect of the environment is multifaceted. First, it spatially confines the reacting species: for example, the protein matrix controls the structures of reaction centers in enzymes and restricts the range of motion of chromophores in photoactive proteins. Second, the environ-
ment serves as a thermal bath. Third, the environment often strongly perturbs the electronic structure of the system by local and long-range electric fields. Preferential stabilization of some electronic states relative to others leads to solvatochromism, solvent-induced shifts of electronic excitation energies, which can be as large as 1 eV. The effect on ionization/electron-attachment energies (quantities determining the redox potentials) is even more pronounced; shifts of several electronvolts in polar solvents and in proteins are rather common. Fourth, the environment itself can be perturbed by the solute: the changes in the electronic structure of an active center (e.g., ionization) induce both structural and electronic response of the solvent. In sum, even when the electronic processes are confined to a well-defined domain (such as solvated or protein-bound chromophores), electronic interactions between the solute and solvent need to be carefully accounted for.

Of course, full quantum mechanical treatment of the entire solvent—solute system would correctly describe these effects. However, such a brute-force approach is impractical due to steep computational scaling of electronic structure methods. For example, robust and reliable equation-of-motion coupled-cluster methods with single and double substitutions (EOM-CCSD) can be applied only to moderate-size systems (20–30 heavy atoms), even when combined with efficient parallelization and other algorithmic enhancements. Low-cost time-dependent density functional theory (TD-DFT) and scaled-opposite-spin configuration interaction singles with doubles correction (SOS-CIS(D)) methods can treat considerably larger systems (a hundred of heavy atoms), but not sufficiently large to model, for example, bulk solvation or an entire protein.

To overcome this hurdle, one can employ a more approximate description of the environment while treating the solute quantum mechanically. Several strategies of various degrees of sophistication have been developed toward this end. The simplest one is to describe the solvent by a polarizable continuum model. Multiple flavors of these methods exist, some have been shown to be capable of capturing solvatochromic effects and yielding reasonably accurate redox potentials in aqueous solutions. However, the description of specific solvent—solute interactions (such as hydrogen bonding) and inhomogeneous environments (such as proteins and interfaces) is beyond the reach of these models because of their implicit solvent nature. Thus, explicit solvent models, which are based on more detailed representation of the solvent, are needed to tackle proteins, interfaces, and complex molecular solids. The popular quantum mechanics/molecular mechanics (QM/MM) approach allows one to combine an arbitrary complex QM treatment of the solute (reaction center) with an atomistic description of the environment. The most common variants of QM/MM use standard (non-polarizable) force fields in which charges on the MM atoms are fixed. Thus, electronic interactions between the QM and MM parts are described as perturbation of the QM part by the (fixed) electrostatic field of the environment. The electronic response of the environment to the changes in the electronic structure of the solute is neglected, which may introduce undesirable errors in relative state energies. Moreover, the charges used in these force fields are known to be too large (to compensate for the lack of polarization); this can lead, for example, to overestimation of reorganization energies in electron-transfer calculations. Finally, the empirical nature of MM force-fields limits their predictive power.

Effective fragment potential (EFP) approach has been developed with an aim to address these limitations of QM/MM. EFP is based on a rigorous representation of different components of intermolecular interactions (electrostatics, polarization, dispersion, exchange-repulsion, and optional charge-transfer) based on perturbative expansion; it can be described as a nonempirical parameter-free polarizable force-field. In QM/EFP, the interactions between the QM and EFP subsystems include both electrostatics and mutual polarization. The extension of EFP to electronically excited and ionized states includes polarization response of the EFP environment to changes in the electronic distribution of the solute. The existing implementations allow one to combine EFP with many popular electronic structure methods including coupled-cluster (CC), EOM-CC, TD-DFT, CIS(D), and SOS-CIS(D).

EFP is similar to the polarizable embedding (PE) approach. The description of electrostatics and polarization is nearly identical in EFP and PE; however, EFP features more rigorous treatment of dispersion and exchange repulsion. For modeling intermolecular interactions, similar strategies based on multipolar expansion have been utilized. Recently, EFP was combined with the fragment molecular orbital (FMO) method giving rise to the EFMO method capable of describing both molecular aggregates and large molecules (however, the EFMO method has not yet been extended to excited states). The detailed reviews of various flavors of fragment-based methods can be found in refs and 46.

The original formulation of the EFP method assumes that the system can be separated, without breaking covalent bonds, into closed-shell rigid fragments, such as solvent and solute molecules. To extend EFP to macromolecules, one needs to figure out how to break large molecules into fragments that could then be described by EFP. A similar problem often arises in traditional QM/MM calculations when one needs to separate the QM part from the rest of the system by breaking covalent bonds. These broken bonds need to be saturated (or capped) for subsequent calculations. Various approaches including hydrogen atom caps and frozen molecular orbitals have been developed. In the context of fragmentation-based methods, schemes such as link atoms, the molecular tailoring approach, and molecular fractionation with conjugate caps have been used. Flexible EFP schemes have also been proposed.

In this work, we break the protein into amino acid fragments capped by hydrogen link atoms. We benchmark several schemes of breaking a macromolecule into effective fragments: breaking along (1) the peptide bond, (2) either the Cα−C or Cα−N bond, or (3) both the Cα−C and Cα−N bonds.

We focus on the electronic effects (such as excitation, ionization, and electron-attachment energies) of molecules embedded in the protein matrix. Here, our goal is to enable single-point calculations of electronic properties and energy differences rather than structure optimization. The test cases include excitation, ionization, and electron-attachment energies of common chromophores in the presence of a protein backbone.

The structure of the article is as follows. The next section presents a brief overview of the EFP scheme and then introduces different approaches for splitting macromolecules into effective fragments. Computational details are given in Section 3. Section 4 presents the results of benchmark
calculations and discusses relative merit of different fragmentation schemes. Our concluding remarks are given in Section 5.

2. METHODOLOGY

2.1. EFP Scheme: QM–EFP Interactions. The EFP method describes noncovalent interactions by using perturbation theory starting from the noninteracting (unperturbed) fragments. The total interaction energy \( E_{\text{EFP–QM}} \) between the effective fragments is broken down into four contributions: electrostatic \( (E_{\text{ele}}) \), polarization \( (E_{\text{pol}}) \), dispersion \( (E_{\text{disp}}) \), and exchange repulsion \( (E_{\text{exrep}}) \)

\[
E_{\text{EFP–QM}} = E_{\text{ele}} + E_{\text{pol}} + E_{\text{disp}} + E_{\text{exrep}}
\]

(1)

The interactions between the QM part and effective fragments are computed by the polarization embedding approach in which the Coulomb and polarization parts of the EFP subsystem contribute to the quantum Hamiltonian \( H \) via one-electron terms

\[
\hat{H} = \hat{H}_0 + \left\langle \sum_{\alpha} \hat{v}^\text{Coul}_{\alpha} + \hat{v}^\text{pol}_{\alpha} \right\rangle \sum_{\alpha} \hat{p}^\alpha q^\alpha
\]

(2)

where \( \hat{H}_0 \) is an unperturbed Hamiltonian of the QM part, \( \hat{v}^\text{Coul} \) and \( \hat{v}^\text{pol} \) are electrostatic and polarization perturbations, respectively, and \( \{p^\alpha\} \) and \( \{q^\alpha\} \) are the atomic orbitals in the QM part. The electrostatic and polarization terms are considered the most important ones, as far as the effect of the polarizable environment on the electronic properties of the QM solute are concerned. In the discussion below, we focus on these terms; the complete details of our EFP implementation including the treatment of dispersion and exchange-repulsion contributions can be found in refs 27–31. Note that eq 1 can be further augmented by including the charge-transfer term, which is important for strongly interacting fragments (i.e., ionic liquids); in our implementation, this term is omitted.

Both terms represent classic electrostatic interactions; \( \hat{v}^\text{Coul} \) is a Coulomb potential due to the fragments’ nuclear charges and their electron densities represented by the multipole expansion, whereas \( \hat{v}^\text{pol} \) describes an electrostatic field due to the induced dipoles (thus, it depends on the polarizabilities of the fragments). The induced dipoles are found by an iterative self-consistent procedure, such that the converged dipole is fully consistent with each other and with the electronic wave function of the QM part.

Neglecting the dispersion and exchange-repulsion contribution terms, the total ground-state (or, more precisely, reference-state) energy of the QM/EFP system is

\[
E_{\text{QM/EFP}} = \left\langle \Phi_{\text{ref}} | \hat{H}_0 + \hat{v}^\text{Coul}_{\text{ref}} + \hat{v}^\text{pol}_{\text{ref}} | \Phi_{\text{ref}} \right\rangle + E_{\text{Coul}}
\]

(3)

where \( \Phi_{\text{ref}} \) is the reference-state wave function, \( \hat{v}^\text{Coul}_{\text{ref}} \) and \( \hat{v}^\text{pol}_{\text{ref}} \) are the Coulomb and polarization EFP contributions to the Hamiltonian. The subscript ref indicates that the induced dipoles correspond to the electronic density of the reference state, that is, the ground electronic state in excited-state calculations, a closed-shell neutral state in the EOM-IP calculations of radical cations, and so forth. \( E_{\text{Coul}} \) is the electrostatic fragment–fragment interaction energy and \( E_{\text{pol},\text{ref}} \) is the self-consistent reference-state polarization energy of the QM/EFP system; it is computed using converged induced dipoles of the fragments and the fields due to the static fragment multipoles and the nuclei and electrons of the quantum region. Note that the polarization contributions appear in both the quantum Hamiltonian through \( \hat{v}^\text{pol} \) and the EFP energy as \( E_{\text{pol},\text{ref}} \) this is because self-consistency precludes the separation of the QM–EFP and EFP–EFP polarization contributions.

Electrostatics is the leading term in the total interaction energy in hydrogen-bonded and polar systems, \( E_{\text{Coul}} \). The effective fragments consist of charge–charge, charge–dipole, charge–quadrupole, charge–octupole, dipole–dipole, dipole–quadrupole, and quadrupole–quadrupole terms. At close separation between the fragments (or between a fragment and the QM region), the charge penetration may become significant; to correct classical multipoles for possible charge penetration, several types of damping functions can be used. Here, we employ an exponential damping in which the charge–charge interaction energy is damped using the following equation

\[
f_{\text{ch–ch}}(r) = 1 - \frac{b^2}{b^2 - a^2} \exp(-ar) - \frac{a^2}{a^2 - b^2} \exp(-br)
\]

(4)

where \( r \) is the distance between multipole points \( k \) and \( l \) and \( a \) and \( b \) are for the multipole points \( k \) and \( l \), respectively.

For QM–EFP interactions, the electrostatic potential of the molecule (QM part) at point \( x \) is expressed by the multipole expansion around \( K \) points located at the atomic centers and bond midpoints of the fragments

\[
V_{\text{Coul}}^K(x) = q_x T(r_{\text{nuc}}) - \sum_{\alpha} \frac{x_{\alpha}^K}{k} T_k(r_{\text{nuc}}) + \frac{1}{3} \sum_{a,b} \Theta_{a,b} T_{a,b}(r_{\text{nuc}})
\]

(5)

where \( q, \mu, \Theta, \) and \( \Omega \) are the net charge, dipole, quadrupole, and octupole located at \( K \) points, respectively, \( T \) are the electrostatic tensors of ranks 0–3. Interaction of the QM electronic density with multipole charges is augmented by a Gaussian-type damping function, such that eq 5 becomes

\[
V_{\text{Coul}}^K(x) = (q_{\text{nuc}} + q_{\text{ele}}(1 - \exp(-\alpha r_{\text{nuc}}^2))) T(r_{\text{nuc}}) - \sum_{\alpha} \frac{x_{\alpha}^K}{k} T_k(r_{\text{nuc}}) + \frac{1}{3} \sum_{a,b} \Theta_{a,b} T_{a,b}(r_{\text{nuc}})
\]

(6)

where \( q_{\text{nuc}} \) is the nuclear charge and \( q_{\text{ele}} \) is the electronic charge on multipole point \( k \), respectively. Thus, only the electronic charges are damped (smeared) by Gaussians. Damping parameters \( a, b, \) and \( \alpha \) in eqs 4 and 6 are determined by minimizing the difference between the electrostatic potentials from the damped multipole expansion and the electronic wave function in the parameter-generating calculation for each fragment.

Polarization is a many-body term, which is computed self-consistently because the induced dipoles of one fragment depend on the static electric field and induced dipoles of other fragments. The polarization energy of the QM–EFP system is computed as

DOI: 10.1021/acs.jpcb.6b04166
where $\mu^k$ and $\bar{\mu}^k$ are the induced and conjugated induced dipoles at distributed point $k$ (at the centers of the localized molecular orbitals (LMOs)); $F_{\text{mult},k}$ is the external field due to the static multipoles and nuclei of other fragments; $F_{\text{nuc},k}$ and $F_{\text{ai,k}}$ are the fields due to the nuclei and electrons of the QM region ($F_{\text{ai,k}}$ is computed using the reference-state density). The induced dipole at each polarizability point $k$ is given by

$$\mu^k = \sum_k \alpha^k (F_{\text{mult},k} + F_{\text{ind},k} + F_{\text{nuc},k} + F_{\text{ai,k}})$$

where $F_{\text{ind},k}$ is the field due to other induced dipoles. The many-body contribution is solved iteratively, as the induced dipole on a particular fragment depends on the values of the induced dipoles of all other fragments and the wave function of the QM region. The distributed polarizabilities ($\alpha^k$) are located at the centers of the LMOs.

To avoid polarization collapse when effective fragments approach each other (which is always the case in macromolecules; see below), the polarization damping functions are applied to the electric fields produced by multipoles and induced dipoles in eqs 7 and 8.

$$f_{\text{pol}} = 1 - \exp(-\sqrt{\alpha^k r_{kl}^2} (1 + \sqrt{\alpha^l r_{kl}^2}))$$

where $\alpha^k$ and $\alpha^l$ are the damping parameters on polarizability points $k$ and $l$, respectively. The default values of the polarization damping parameters are 0.6 for noncovalent neutral fragments and 0.1 for small anions and cations such as halide and alkali ions. Appropriate values of the polarization damping parameters for mEFP are explored in Section 4.

### 2.2. Breaking Macromolecules into Fragments

In its original formulation, the EFP method cannot be applied to macromolecules such as biopolymers, peptides, proteins, lipids, and DNA because it was designed as a rigid-fragment model for treating interactions between (small) molecules in clusters and liquids. The building block of EFP (the so-called fragment) is a molecule rather than an atom (as typical in classical force-fields). The EFP scheme does not include covalent interactions between the fragments.

To enable conformational flexibility of macromolecules while keeping the calculations affordable, the polymer chain should be split into small effective fragments, and the parameters for each of these fragments need to be generated, similar to that in other fragment-based methods. We generated the parameters for individual fragments using the following procedure. First, we saturate the dangling bonds of each fragment using a capping group, to obtain well-behaved closed-shell structures. In this work, we use hydrogens as the capping groups. Once the parameters for the capped closed-shell molecule are generated, we remove the capping groups and the associated parameters (multipoles and polarizabilities) from the fragment. This scheme enables calculation of the electrostatic and polarization energies in the macromolecule represented by effective fragments as well as electrostatic PE using QM/EFP. The specific details of the application of this scheme to polypeptides (such as positions of the cuts) are discussed below. Schemes for other macromolecules will be developed in future work, following similar strategies.

There are different ways to break macromolecules such as proteins and DNA into fragments, depending on the position of the cut between two covalently bound residues. For polypeptides, we consider the following cutting schemes (shown in Figure 1):

- cutting along the peptide link;
- cutting along the C$_\alpha$–C bonds;
- cutting along the C$_\alpha$–N bonds; and
- cutting along both the C$_\alpha$–C and C$_\alpha$–N bonds.

In the first three schemes, each fragment consists of a single amino acid, whereas the last one yields two fragments per amino acid, one fragment containing the peptide group and another containing the residue; see Figure 1.

The advantage of fractioning the protein along the peptide bond is that it yields “symmetric” fragments. However, in this scheme, highly polarized bonds are broken, which may lead to unphysical multipole and polarizability values. One may expect that the fragments obtained by cutting either the C$_\alpha$–C or C$_\alpha$–N bonds produce a more accurate representation of multipoles and polarizabilities near the cuts. Furthermore, smaller fragments and cutting both the C$_\alpha$–C and C$_\alpha$–N bonds will aid further extension of mEFP to dynamics, as this scheme ensures flexibility along the two most important degrees of freedom in proteins, dihedral $\phi$ and $\psi$ angles defining the conformation.

In all schemes, the capped fragments mimic the protein; however, the neighboring fragments have duplicated points (overlapping area) due to the hydrogen caps. The multipole expansion points and polarizability expansion points are extended on each isolated capped fragment by the standard procedure (see Section 3), and the multipoles and polarizabilities at the duplicated expansion points are removed. To maintain the net integer charge on each fragment, the monopole expansion of each cap is redistributed to the nearest atom. This method is called expand–remove–redistribute (ERR); see Figure 2). The polarizability points located through the “cut” bonds are removed to avoid over-polarization of the neighboring fragments.

We have implemented the mEFP approach in libepf, which is interfaced with the Q-Chem package.

![Figure 1. Various cutting schemes for polypeptides.](image-url)
3. COMPUTATIONAL DETAILS

The benchmark set was chosen with the focus on electronic properties. In particular, we consider the effect of the protein scaffold on the electronically excited states of chromophores and on their redox properties. First, we quantify the effect of the protein environment by comparing the properties of the bare chromophores (small QM) with the full QM calculations. Then, we test different fragmentation strategies and compare the results of QM/mEFP against full QM calculations and against QM/EFP (when no breaking along covalent bonds in the EFP part is performed). Finally, we consider QM/MM calculations with fixed point charges using standard force-fields.68,69

While setting up such benchmark calculations, one should keep in mind several important points. The QM/MM and QM/mEFP calculations can be compared with the full QM calculation only when the target electronic properties can be described as the electronic properties of the QM region perturbed by the environment. In the calculations of electronically excited states, this means that one should consider only the states in which both the initial and target molecular orbitals are confined within the QM part. In calculations of ionization or electron-attachment energies, the spin density of the target state needs to be localized in the QM region. Within Koopmans theorem, the highest occupied molecular orbital (HOMO) of the initial state is representative of the density of the unpaired electron; however, we found several examples where the shape of the HOMO is quite different from the spin density of the ionized/electron detached state. Because spin density represents the actual shape of the hole (or the unpaired electron), one needs to always consider spin density, whereas the shape of the HOMO is relevant only for assessing the validity of Koopmans’ description of ionization. In a similar vein, one should consider spin density while analyzing electron-attached states. In addition to the above consideration, one should also assess whether the states in question are bound or unbound with respect to electron ejection. This is particularly relevant for electron-attached states and excited states of anionic systems.

The following model systems were used in the benchmark calculations:

1. Green fluorescent protein (GFP) chromophore in its anionic (deprotonated) and neutral forms surrounded by four nearby amino acid residues (Figure 3).
2. The anionic form of the mPlum chromophore with nearby amino acids and one water molecule (Figure 3).

![Figure 3. Left: Protonated GFP chromophore and the VAL93, GLN94, GLU95, and ARG96 tetrapeptide strings; GLU95 is protonated. The structure of the model system with the anionic chromophore is the same. Right: Model mPlum system with a deprotonated extended chromophore. The protein is represented by the ARG88, VAL89, and MET90 tripeptide strings and one water molecule.]

Figure 4. Phenolate embedded into the apolar cavity consisting of four amino acids in the T4 lysozyme.
3. A phenolate molecule with four amino acids from the T4 lysozyme M102E/L99A mutant (Figure 4).

All relevant geometries and EFP parameters are provided in the Supporting Information (SI).

3.1. Model Systems.

3.1.1. GFP Model Systems. We constructed the model system as follows. We began from the X-ray structure of enhanced GFP (pdb id:1EMA). We added hydrogens following the protonation states determined in ref 70. Then, we optimized the structure using the CHARMM27 force field. The parameters of the chromophore are from ref 69. From the optimized structure, we extracted the model system comprising the chromophore and four other amino acid residues, VAL93, GLN94, GLU95, and ARG96. Because these residues constitute a single sequence, we capped only $\alpha$-carbons of VAL93 and ARG96 (these capping hydrogens were not included in the MM point charges in the QM/MM calculations). Also, two capping hydrogens were added to the chromophore.

We constructed two model systems, one with the anionic (deprotonated) chromophore and another with the neutral chromophore. In the model system with the anionic chromophore, GLU95 is deprotonated and ARG96 is protonated. The total charge of the model system is $-1$ (negatively charged chromophore and GLU95 and positively charged ARG96). The model structure is shown in Figure 3.

The model system with the neutral GFP chromophore was prepared following the same protocol, except that GLU95 was protonated (this is necessary for finite-cluster calculations to suppress the ionization and electronic excitation from the MO localized on GLU95). The total charge of the model system with the neutral GFP chromophore is $+1$.

In small QM, QM/MM, and QM/EFP calculations, only the chromophore constitutes the QM part. Vertical detachment and ionization energies (VDEs and VIEs, respectively) were calculated with $\omega$B97X-D/aug-cc-pVDZ.71 We note that using a range-separated functional is important for these charged...
systems. Vertical excitation energies were computed with SOS-CIS(D)/aug-cc-pVDZ. To investigate basis-set effects, we performed calculations also with a smaller basis set, 6-31G(d).

3.1.2. mPlum Model System. The mPlum model system features a larger anionic chromophore, so one can assess whether the effect of the environment depends on the extent of π-conjugation in the chromophore. The system was constructed from the PDB ID:2QLG structure of mPlum following the protocol from ref 72. The model system was extracted from the QM/MM-optimized structure (as described in ref 72). The model system comprises the extended conjugated anionic chromophore along with three amino acids (ARG88-VAL89-MET90) and one water molecule. The total charge of the system is 0. We rearranged the water molecule to facilitate hydrogen bonding between phenolate’s oxygen of the chromophore and water (this makes the radical-dianion state bound with respect to electron detachment). ARG89 and MET90 were capped with H-atoms at Cα.

In small QM, QM/MM, and QM/EFP calculations, only the chromophore constitutes the QM part. VDE and excitation energies were calculated using the same methods as in the GFP model systems, and the attachment (VEA) energies were computed using oB97X-D/aug-cc-pVDZ.

3.1.3. T4 Lysozyme Model System. The model system consists of the tyrosine residue in an apolar cavity in the T4 lysozyme.58 The following residues were retained from the crystal structure (3GOU): phenolate, ALA99, ILE100, GLU102 (protonated at the side chain), and VAL103. We choose phenolate rather than phenol to ensure the spin density to be localized in the QM region (phenolate). ALA99 and VAL103 were capped with hydrogens at the peptide bond (C–N). Hydrogen positions were optimized by oB97X-D/6-31+G(d).

All residues except for the phenolate are neutral; the total charge of the model system is −1. In small QM, QM/MM, and QM/EFP calculations, only the phenolate constitutes the QM part. We computed VDE using oB97X-D/aug-cc-pVDZ.

Figure 5a–e shows the spin densities of all open-shell cases under study with full QM treatment for each system. We see that the spin density is always localized on the small QM for respective systems. This justifies the comparison among small QM, QM/MM, QM/mEFP, and full QM.

3.2. QM/MM and QM/mEFP Calculations. The EFP parameters for the peptide residues were prepared following the procedure outlined in Section 2.2. In addition to the four fragmentation schemes described above, we also considered the superfragment scheme in which all covalently linked peptides constitute a single EFP fragment. That is, in the GFP and mPlum model systems, all residues form one superfragment; in the T4 lysozyme model system, there are two superfragments.

The EFP parameters for each isolated capped residue and superfragments have been generated using the MAKEFP job (RUNTYP = MAKEFP) of the GAMESS program.54 Only the parameters responsible for electrostatic and polarization terms (multipole expansion, damping parameters, and static polarizabilities) were computed at the Hartree–Fock level of theory and with the 6-31G(d) basis set. The 6-31G(d) basis set is usually recommended for computing polarization, dispersion, and exchange-repulsion terms.

Exponential electrostatic damping of charges (SCREEN2) is employed between the fragments;59 Gaussian-type damping of charges (SCREEN) is used to mitigate the charge-penetration errors in the QM–EFP interactions.59 Gaussian-type polarization damping controlled by a damping parameter is employed between the fragments but not between the QM and EFP regions.59 The effects of electrostatic screening and polarization damping on the QM/EFP energies are discussed below. The QM region is described at the same level of theory as in the full QM and QM/MM calculations.

3.2.1. QM/MM Calculations. In the QM/MM calculations, we tested three different schemes. Following the same strategy as in QM/mEFP, we do not include the capping H-atom in the MM part. We use CHARMM’s force-field parameters for the MM point charges.58 This creates an additional charge (e.g., +0.16 in the neutral GFP model system) in the MM part, making the total charge of the MM subsystem noninteger. There are several ways to handle this artifact of QM/MM. For example, we can add an extra −0.16 charge to the next atom (Cα in our case), or we can distribute that extra charge over the same residue. In QM/mEFP, the charge is added to the next carbon (see Figure 1). We compared these two protocols and the QM/MM calculations with noninteger charges for a protonated GFP model system:

1. QM/MM; protocol 1: charge distributed over all atoms of the entire residue.
2. QM/MM; protocol 2: charge added to the next atom.
3. QM/MM; protocol 3: no redistribution of charge; noninteger total MM charge.

Table 1 shows that all QM/MM calculations overestimate VIEs relative to full QM. Importantly, protocol 1 and protocol 2 produce very similar IEs, which indicates that IEs are not very sensitive to the details of charge redistribution, as long as the total charge is conserved. Because QM/mEFP uses a protocol similar to protocol 2, in all QM/MM calculations reported below, we employed protocol 2.

4. RESULTS

4.1. Electronic Properties of Model GFP Chromophores. Table 2 summarizes the results for the GFP model system with the anionic chromophore. The VDE of the bare chromophore is 2.70 eV. The environment strongly stabilizes the anionic chromophore leading to a significant blue shift (1.65 eV) in VDE. The VDE value in the full QM calculation is 4.35 eV. As illustrated in Figure 5, the spin density in the full QM calculations is localized on the chromophore, which means that QM/MM and QM/mEFP calculations with the QM part comprising the chromophore can be directly compared to the full QM and small QM calculations. The QM/MM calculations capture the effect of the stabilization of the anionic chromophore reasonably well, yielding VDE of 4.39 eV, which is blue-shifted relative to the full QM value by 0.04 eV. Similar to those for QM/MM, the QM/mEFP and QM/EFP VDEs and excitation energies are within 0.02–0.04 eV from the full QM results. Importantly, neither detachment nor...
excitation energies are sensitive to the choice of the mEFP fragmentation schemes.

The analysis of the excited-state calculations requires care. There are several types of excited states of chromophores in a condensed phase:4,75–77 (1) local excitations in which both the electron and hole are confined to the chromophore; (2) charge-transfer states in which the electron is transferred between the chromophore and another neighboring group;76,77 (3) charge-transfer-to-solvent states in which the electron is transferred from the chromophore to a cavity, forming states resembling the Koopmans onset.

In addition to these physical states, one should also be aware of pseudoocontinuum states corresponding to an electron on the surface). In addition to these physical states, one should also be aware of pseudocontinuum states corresponding to the electron detached from the system. In CIS calculations, these artificial states will appear above the Koopmans IE/DE when the basis is sufficiently diffuse80,81 they may spoil the description of the bright states that lie above the Koopmans onset.

When, in the QM/MM and QM/EFP calculations with the QM part comprising the chromophore alone, only local excitations corresponding to bound excited states can be described correctly. Thus, it is important to carefully analyze the character of the states in such calculations. Here, we focus on the bright \( \pi_2^* \) state of the chromophore, which can be easily identified by its large oscillator strength and by inspecting the target MO.

The Koopmans IE of the isolated anionic GFP chromophore is 2.86 eV (HOMO energy, HF/aug-cc-pVDZ). Thus, all CIS excited states above this energy are embedded in the electron-detachment continuum, which spoils their description.81,82 Full QM CIS and SOS-CIS(D) calculations yield two lowest states of different characters. One state is relatively dark and has charge-transfer character, whereas the other state is bright and corresponds to the local \( \pi_2^* \) excitation on the chromophore. The leading MOs for the bright states are shown in SI. We carefully checked that the nature of the HOMO and the target virtual orbital involved in the bright transition remains the same among different schemes of calculation. Table 2 shows the energies of the bright state. The energy of the second state, which cannot be correctly described in the QM/MM and QM/mEFP calculations with a small QM region, is given in the SI.

Table 3 summarizes the results for the GFP model system with the neutral chromophore. The VIE of the bare chromophore is 7.40 eV. The interaction with the neighboring residues leads to a blue shift of 2.5 eV (small QM vs full QM).

Relative to full QM, the QM/MM calculations overestimate the ionization energy by 0.05 eV. QM/mEFP with the double-fragmentation scheme performs well, matching the blue shift in VIE. We also note that QM/mEFP with the \( \text{C}_6^− \text{N} \) fragmentation scheme yields the largest deviation (+0.14 eV) relative to full QM.

In this system, the lowest excited state is the bright \( \pi_2^* \) state (the relevant orbitals are shown in the SI; as in the example before, they are of the same character in different calculations). The effect of the environment on the excitation energies is noticeable: the energy of the bright state is red-shifted by 0.24 eV in the full QM calculation relative to that in the bare chromophore (small QM). The QM/MM calculation overestimates the excitation energy by 0.02 eV. The QM/EFP superfragment calculation underestimates the excitation energy by 0.01 eV. Different QM/mEFP schemes show similar performance; they underestimate the full QM excitation energy by 0.01–0.02 eV.

Using the anionic and neutral forms of the GFP, we investigate the effects of electrostatic and polarization damping on electronic properties. The results for VDEs are collected in Table 4. In Tables 2 and 3 discussed above, the QM/mEFP values were computed using polarization damping between the EFP fragments with the default parameter value of 0.6 and electrostatic damping between the QM and EFP regions given in eq 6 with parameters precomputed as described in Section 3. In Table 4, the results with the default polarization and electrostatic damping are compared with the values obtained with (1) the reduced polarization damping parameter, which corresponds to a stronger damping of polarization energies between the fragments, (2) polarization completely turned off, which corresponds to electrostatic embedding, and (3) full polarization. In the default setting, we find that smaller values of this parameter, that is, stronger damping, are occasionally necessary for avoiding the polarization catastrophe. In particular, this often happens for strongly interacting ionic species.83 Although in the systems that we considered here, we did not encounter difficulties converging polarization energies with the default damping values, we explore the effect of polarization damping on IEs/DEs, in the case when the modification of damping parameter is required in more complex systems. Note that polarization damping is applied only to fragment–fragment polarization; the QM → EFP and EFP ← QM polarization interactions are not damped.

Table 2. Electronic Properties of the Anionic GFP Chromophore Surrounded by Four Nearby Amino Acid Residues

<table>
<thead>
<tr>
<th>calculation</th>
<th>VDE, eV</th>
<th>( E_{\text{def}} ) eV (( f_1 ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>small QM</td>
<td>2.70</td>
<td>2.76 (0.94)</td>
</tr>
<tr>
<td>full QM</td>
<td>4.35</td>
<td>2.56 (1.52)</td>
</tr>
<tr>
<td>QM/MM</td>
<td>4.39</td>
<td>2.57 (1.45)</td>
</tr>
<tr>
<td>QM/EFP: superfragment</td>
<td>4.39</td>
<td>2.60 (1.45)</td>
</tr>
<tr>
<td>QM/mEFP: C6–C</td>
<td>4.39</td>
<td>2.60 (1.45)</td>
</tr>
<tr>
<td>QM/mEFP: C6–N</td>
<td>4.39</td>
<td>2.60 (1.45)</td>
</tr>
<tr>
<td>QM/mEFP: C–N</td>
<td>4.38</td>
<td>2.60 (1.45)</td>
</tr>
<tr>
<td>QM/mEFP: C6–C and C6–N</td>
<td>4.38</td>
<td>2.60 (1.45)</td>
</tr>
</tbody>
</table>

VDEs computed using \( \omega \text{B97X-D/aug-cc-pVDZ} \). Excitation energies computed with SOS-CIS(D)/aug-cc-pVDZ; \( f_1 \) computed at the CIS level. The lowest bright state is shown (see the text).

Table 3. Electronic Properties of the Neutral GFP Chromophore Surrounded by Four Nearby Amino Acid Residues

<table>
<thead>
<tr>
<th>calculation</th>
<th>VIE, eV</th>
<th>( E_{\text{def}} ) eV (( f_1 ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>small QM</td>
<td>7.40</td>
<td>3.80 (1.05)</td>
</tr>
<tr>
<td>full QM</td>
<td>9.89</td>
<td>3.56 (1.09)</td>
</tr>
<tr>
<td>QM/MM</td>
<td>9.94</td>
<td>3.58 (1.08)</td>
</tr>
<tr>
<td>QM/EFP: superfragment</td>
<td>9.90</td>
<td>3.55 (1.03)</td>
</tr>
<tr>
<td>QM/mEFP: C6–C</td>
<td>9.91</td>
<td>3.55 (1.03)</td>
</tr>
<tr>
<td>QM/mEFP: C6–N</td>
<td>10.03</td>
<td>3.54 (1.04)</td>
</tr>
<tr>
<td>QM/mEFP: C–N</td>
<td>9.92</td>
<td>3.55 (1.03)</td>
</tr>
<tr>
<td>QM/mEFP: C6–C and C6–N</td>
<td>9.89</td>
<td>3.55 (1.04)</td>
</tr>
</tbody>
</table>

VIEs computed using \( \omega \text{B97X-D/aug-cc-pVDZ} \). Excitation energies computed with SOS-CIS(D)/aug-cc-pVDZ; \( f_1 \) computed at the CIS level. The lowest bright state is shown (see the text).
Comparing IEs computed with polarization turned off entirely or partially, we observe several interesting effects. The first observation is that polarization of the environment plays a more significant role in the anionic than in the neutral chromophore, as illustrated by the VIE/VDE differences between the default and polarization-off values in the superfragment calculations, which are 0.18 and 0.11 eV for the anionic and neutral chromophores, respectively. The second observation is that the polarization effect is smaller in fragmented (mEFP) than in the superfragment (EFP) calculations. For example, in the anionic form, the change in VDEs due to polarization decreases from 0.18 eV in QM/EFP to 0.06–0.08 eV in QM/mEFP, and similarly in the neutral form. This is an interesting observation, which suggests that neighboring fragments “depolarize” each other. This effect occurs in QM/mEFP but not in superfragment calculations. (In superfragment calculations, the QM and EFP subsystems polarize each other, but the EFP superfragment does not polarize itself.) Finally, a moderate decrease in polarization damping to 0.3 does not depolarize the environment well; VDE is overestimated by +0.04 eV relative to full QM/MM. QM/MM captures the effect of the environment well; VDE is overestimated by +0.04 eV relative to small QM. QM/MM captures the dissociation energy significantly, resulting in 2.4 eV blue shift in full QM/MM relative to small QM. QM/MM captures the effect of the environment well; VDE is overestimated by +0.04 eV relative to full QM. QM/mEFP also performs very well. We once again observe that the protein environment affects the electronic properties for the mPlum model system in Table 5. The mPlum chromophore features an extended π-system in which the conjugation extends beyond the imidazolinone ring and into the side chain. The extended π-system is responsible for red-shifted absorption relative to the GFP chromophore. The extended conjugation also leads to higher detachment energy of mPlum, as compared to deprotonated GFP (Table 2); the reason for that was described by Ghosh et al.84 In this system, the chromophore is also anionic and the environment strongly stabilizes its ground state. Thus, we observe similar trends in Tables 2 and 5.

We begin by considering the trends in VDE. From Table 5, we once again observe that the protein environment affects the detachment energy significantly, resulting in 2.4 eV blue shift in full QM/MM relative to small QM. QM/MM captures the effect of the environment well; VDE is overestimated by +0.04 eV relative to full QM. QM/mEFP also performs very well. We note that within the same fragmentation scheme, QM/mEFP

Table 4. VDE and VIE (in eV) of the Anionic (Top) and Neutral (Bottom) GFP Chromophores Surrounded by Four Nearby Amino Acid Residues as a Function of Electrostatic and Polarization Damping in the EFP Region

<table>
<thead>
<tr>
<th>calculation</th>
<th>pol damp = 0.6 (default)</th>
<th>pol damp = 0.3</th>
<th>pol damp = 0.1</th>
<th>pol off</th>
<th>elec damp off</th>
</tr>
</thead>
<tbody>
<tr>
<td>anionic GFP&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>QM/EFP: superfragment</td>
<td>3.93</td>
<td>3.94</td>
<td>3.87</td>
<td>3.87</td>
<td>3.75</td>
</tr>
<tr>
<td>QM/mEFP: Cα−C</td>
<td>3.94</td>
<td>3.94</td>
<td>3.87</td>
<td>3.87</td>
<td>3.90</td>
</tr>
<tr>
<td>QM/mEFP: Cα−N</td>
<td>3.92</td>
<td>3.92</td>
<td>3.86</td>
<td>3.86</td>
<td>3.87</td>
</tr>
<tr>
<td>QM/mEFP: C−N</td>
<td>3.93</td>
<td>3.93</td>
<td>3.85</td>
<td>3.85</td>
<td>3.89</td>
</tr>
<tr>
<td>QM/mEFP: Cα−C and Cα−N</td>
<td>3.93</td>
<td>3.93</td>
<td>3.87</td>
<td>3.87</td>
<td>3.88</td>
</tr>
<tr>
<td>neutral GFP&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>QM/mEFP: Cα−N</td>
<td>9.81</td>
<td>9.81</td>
<td>9.79</td>
<td>9.79</td>
<td>9.80</td>
</tr>
<tr>
<td>QM/mEFP: Cα−C and Cα−N</td>
<td>9.66</td>
<td>9.66</td>
<td>9.65</td>
<td>9.65</td>
<td>9.65</td>
</tr>
</tbody>
</table>

<sup>a</sup>ωB97X-D/6-31G(d) is used in all calculations. <sup>b</sup>Default polarization (with damp = 0.6) is used. <sup>c</sup>Full QM: VDE = 9.74 eV, QM/MM: VDE = 3.97 eV. <sup>d</sup>Full QM: VDE = 9.74 eV, QM/MM: VDE = 9.74 eV.

4.2. mPlum Model System. The electronic properties for the mPlum model system are collected in Table 5. The mPlum chromophore features an extended π-system in which the conjugation extends beyond the imidazolinone ring and into the side chain. The extended π-system is responsible for red-shifted absorption relative to the GFP chromophore. The extended conjugation also leads to higher detachment energy of mPlum, as compared to deprotonated GFP (Table 2); the reason for that was described by Ghosh et al.84 In this system, the chromophore is also anionic and the environment strongly stabilizes its ground state. Thus, we observe similar trends in Tables 2 and 5.

Table 5. Electronic Properties of the Anionic mPlum Chromophore Surrounded by Four Nearby Amino Acid Residues and Water<sup>a</sup>

<table>
<thead>
<tr>
<th>calculation</th>
<th>VDE, eV</th>
<th>E&lt;sub&gt;ie&lt;/sub&gt; eV (f&lt;sub&gt;i&lt;/sub&gt;)</th>
<th>VEA, eV</th>
</tr>
</thead>
<tbody>
<tr>
<td>small QM</td>
<td>3.20</td>
<td>2.30 (1.51)</td>
<td>2.03</td>
</tr>
<tr>
<td>full QM</td>
<td>5.57</td>
<td>2.65 (0.91)</td>
<td>0.21</td>
</tr>
<tr>
<td>QM/MM</td>
<td>5.61</td>
<td>2.53 (1.53)</td>
<td>−0.20</td>
</tr>
<tr>
<td>QM/EFP: superfragment</td>
<td>5.60</td>
<td>2.53 (1.54)</td>
<td>−0.22</td>
</tr>
<tr>
<td>QM/mEFP: Cα−C</td>
<td>5.63</td>
<td>2.52 (1.54)</td>
<td>−0.24</td>
</tr>
<tr>
<td>pol off</td>
<td>5.51</td>
<td>2.51 (1.54)</td>
<td>−0.08</td>
</tr>
<tr>
<td>QM/mEFP: Cα−N</td>
<td>5.61</td>
<td>2.53 (1.54)</td>
<td>−0.22</td>
</tr>
<tr>
<td>pol off</td>
<td>5.52</td>
<td>2.51 (1.54)</td>
<td>−0.07</td>
</tr>
<tr>
<td>QM/mEFP: C−N</td>
<td>5.59</td>
<td>2.52 (1.54)</td>
<td>−0.21</td>
</tr>
<tr>
<td>pol off</td>
<td>5.52</td>
<td>2.51 (1.54)</td>
<td>−0.08</td>
</tr>
<tr>
<td>QM/mEFP: Cα−C and Cα−N</td>
<td>5.61</td>
<td>2.53 (1.54)</td>
<td>−0.22</td>
</tr>
<tr>
<td>pol off</td>
<td>5.51</td>
<td>2.51 (1.54)</td>
<td>−0.07</td>
</tr>
</tbody>
</table>

<sup>a</sup>VDE and VEA computed with ωB97X-D/aug-cc-pVDZ. Excitation energies computed with SOS-CIS(D)/aug-cc-pVDZ; f<sub>i</sub> computed at the CIS level.
overestimates VDE when polarization is on (highest deviation +0.06 eV compared to full QM) and underestimates VDE with polarization switched off (the largest deviation from the full QM result is −0.06 eV).

In the bare chromophore (small QM), the lowest excited state is the bright excited state of the ππ* character, and \( E_{\text{ex}} = 2.30 \) eV. In the full QM calculations, we observe two blue-shifted excited states of similar character (the oscillator strength of the bright transition is distributed between the two). All QM/MM and QM/mEFP calculations yield only one bright ππ* state (which is the lowest excited state) carrying large oscillator strength. The excitation energy is rather insensitive to the fragmentation scheme (2.52−2.53 eV).

In this model system, we also consider electron-attached states. Although in the bare chromophore, such radical-dianion state is not bound electronically (i.e., it has positive electron affinity), it can be stabilized by interactions with the protein. In a recent study of KillerRed (which has the same chromophore as mPlum), such diion-radical states were found to be stable.\(^7\) Figures 5d and 6a−d show spin densities for the electron-attached states (the relevant MOs are shown in SI). As one can see from the data in Table 5, this state has a similar character across different calculation schemes. In Figure 6, inclusion of water in the QM part of small QM and QM/MM calculations does not affect the spin density, thereby confirming the effect of water to be purely electrostatic.

In this model system, the protein stabilizes the dianion state leading to a change of 2.24 eV in VEA between the small QM and full QM calculations. The QM/MM and QM/mEFP results are within 0.01 eV of the full QM value. The effect of polarization is more pronounced in the case of VEA: the effect of switching off polarization can be as large as 0.14 eV.

### 4.3. T4 Lysozyme Model System

Table 6 compares VDEs computed with small QM, QM/MM, full QM, QM/EFP, and QM/mEFP with different fragmentation schemes. In this system, we observe that QM/MM performs poorly and underestimates VDE by 0.33 eV, as compared to full QM. On the other hand, the QM/mEFP schemes perform really well, with the largest error being −0.08 eV relative to full QM. It is also noteworthy that in QM/EFP and QM/mEFP, polarization plays a significant role here: when polarization is turned off completely, the QM/EFP and QM/mEFP results change by 0.2−0.3 eV and become very similar to the QM/MM data. A significant contribution of polarization can be rationalized by the charge density of the phenolate being largely localized on O\(^−\), which leads to strong polarizing interactions with the GLU102 residue. Overall, the scheme with two fragments per amino acid (C\(_{\alpha}−C\) and C\(_{\alpha}−N\)) shows the

<table>
<thead>
<tr>
<th>system</th>
<th>VDE, eV</th>
<th>Koopmans IE (DFT)</th>
</tr>
</thead>
<tbody>
<tr>
<td>small QM</td>
<td>2.12</td>
<td>2.14</td>
</tr>
<tr>
<td>full QM</td>
<td>1.90</td>
<td>2.12</td>
</tr>
<tr>
<td>QM/MM</td>
<td>1.57</td>
<td>1.58</td>
</tr>
<tr>
<td>QM/mEFP: superfragment</td>
<td>1.85</td>
<td>2.04</td>
</tr>
<tr>
<td>pol off</td>
<td>1.57</td>
<td>1.58</td>
</tr>
<tr>
<td>QM/mEFP: C(_{\alpha}−C)</td>
<td>1.82</td>
<td>2.01</td>
</tr>
<tr>
<td>pol off</td>
<td>1.51</td>
<td>1.52</td>
</tr>
<tr>
<td>QM/mEFP: C(_{\alpha}−N)</td>
<td>1.92</td>
<td>2.12</td>
</tr>
<tr>
<td>pol off</td>
<td>1.64</td>
<td>1.63</td>
</tr>
<tr>
<td>QM/mEFP: C−N</td>
<td>1.84</td>
<td>2.01</td>
</tr>
<tr>
<td>pol off</td>
<td>1.61</td>
<td>1.61</td>
</tr>
<tr>
<td>QM/mEFP: C(<em>{\alpha}−C) and C(</em>{\alpha}−N)</td>
<td>1.90</td>
<td>2.10</td>
</tr>
<tr>
<td>pol off</td>
<td>1.60</td>
<td>1.61</td>
</tr>
</tbody>
</table>
The developed mEFP algorithm can be generalized to future studies.

Discrepancies between the full QM energies, provided that the electronic process is localized in other fragments. Four different fragmentation schemes have been tested; the most consistent results were obtained with the schemes in which the polypeptide is split either along the Cα−C bond or along both the Cα−C and Cα−N bonds.

In all systems considered here, QM/mEFP accurately reproduces excitation, ionization, and electron-attachment energies, provided that the electronic process is localized in the QM subsystem. Discrepancies between the full QM calculations and QM/mEFP calculations in all but one case do not exceed 0.1 eV. Polarization interactions have a noticeable effect on the electronic properties of biological chromophores. Turning off polarization in QM/mEFP deteriorates the accuracy and leads to additional errors of up to 0.3 eV. Short-range damping of electrostatic interactions between the QM and EFP subsystems, which corrects classical multipole expansion for charge-penetration energy, also brings the QM/mEFP results into closer agreement with the reference full QM values. In most cases, the errors of QM/mEFP against QM/EFP (no fragmentation of the EFP part) are 0.01–0.02 eV; the largest error is 0.06 eV.

The QM/mEFP approach provides a rigorous way to incorporate polarization embedding into studies of biological systems. The developed mEFP algorithm can be generalized to other polymers and flexible molecules, which will be exploited in future studies.

5. CONCLUSIONS

The extension of the EFP method to polypeptides, called mEFP, has been developed and validated by computing excitation, ionization, and electron-attachment energies of three biologically relevant systems, GFP/mPlum chromophores and phenolate in their natural surroundings. In the mEFP scheme, the polypeptide is split into smaller fragments, and the EFP parameters for each fragment are computed independently of other fragments. Four different fragmentation schemes have been tested; the most consistent results were obtained with the schemes in which the polypeptide is split either along the Cα−C bond or along both the Cα−C and Cα−N bonds.

The QM/mEFP approach provides a rigorous way to incorporate polarization embedding into studies of biological systems. The developed mEFP algorithm can be generalized to other polymers and flexible molecules, which will be exploited in future studies.

ASSOCIATED CONTENT

Supporting Information The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpcb.6b04166.

Computational details, relevant Cartesian geometries, molecular orbitals, and spin densities, as well as additional results PDF

AUTHOR INFORMATION

Corresponding Authors
*E-mail: krylov@usc.edu. Phone: +1(213)740-4929 (A.I.K.).
*E-mail: lsipchenko@purdue.edu. Phone: +1(765)494-5255 (L.V.S.).

Author Contributions
*A.A. and P.K.G. contributed equally.

Notes
The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work has been supported by the U.S. National Science Foundation (CHE-1566428 to A.I.K. and CHE-1465154 to L.V.S.). D.G. thanks CSIR XIIth Five Year Plan on Multiscale Modeling and DAE-BRNS for funding.

REFERENCES

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


(64) Other schemes exist such as the IUPAC definition with a the C=O bond or a definition splitting each amino acid into the backbone and the side-chain moiety. This fragmentation leads to asymmetrical fragments, which are not convenient for computing DMA distribution because they result in delocalized multipoles. However, the cut is between atoms of same chemical nature. Natural fragmentation solves the problem of asymmetry, but an extremely polar bond is cut.

(65) Other methods exist in the literature to guarantee a net integer charge: the expand-remove and scale (ERS) is similar to our method up to the removal step. The monopoles located on all caps are then scaled to reflect the net integer charge of the entire protein.\(^{56,57}\) Two other methods developed by the same group are the remove and expand (RE) method and the expand-collect-and-correct (ECC) method.\(^{56}\)


