JCTC Journal of Chemical Theory and Computation

pubs.acs.org/JCTC Article

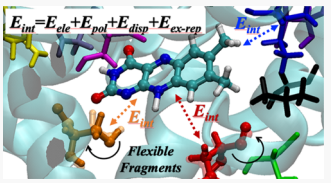
# Effective Fragment Potentials for Flexible Molecules: Transferability of Parameters and Amino Acid Database

Yongbin Kim, Yen Bui, Ruslan N. Tazhigulov, Ksenia B. Bravaya, and Lyudmila V. Slipchenko*

Cite This: *J. Chem. Theory Comput.* 2020, 16, 7735−7747 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

**ABSTRACT:** An accurate but efficient description of noncovalent interactions is a key to predictive modeling of biological and materials systems. The effective fragment potential (EFP) is an *ab initio*-based force field that provides a physically meaningful decomposition of noncovalent interactions of a molecular system into Coulomb, polarization, dispersion, and exchange-repulsion components. An EFP simulation protocol consists of two steps, preparing parameters for molecular fragments by a series of *ab initio* calculations on each individual fragment, and calculation of interaction energy and properties of a total molecular system based on the prepared parameters. As the fragment parameters (distributed multipoles, polarizabilities, localized wave function, *etc.*) depend on a fragment geometry, straightforward application of the EFP method requires recomputing parameters of each fragment if its geometry changes, for example, during thermal fluctuations of a molecular system. Thus, recomputing fragment parameters can easily become both computational and human bottlenecks and lead to a loss of efficiency of a simulation protocol. An alternative approach, in which fragment parameters are adjusted to different fragment geometries, referred to as "flexible EFP", is explored here. The parameter adjustment is based on translations and rotations of local coordinate frames associated with fragment atoms. The protocol is validated on extensive benchmark of amino acid dimers extracted from molecular dynamics snapshots of a cryptochrome protein. A parameter database for standard amino acids is developed to automate flexible EFP simulations in proteins. To demonstrate applicability of flexible EFP in large-scale protein simulations, binding energies and vertical electron ionization and electron attachment energies of a lumiflavin chromophore of the cryptochrome protein are computed. The results obtained with flexible EFP are in a close agreement with the standard EFP procedure but provide a significant reduction in computational cost.






## 1. INTRODUCTION

Molecular modeling became an essential tool for investigating and predicting properties of molecular systems. In biological and materials applications, where model molecular systems can contain thousands of atoms, a compromise between computational cost and accuracy of a chosen computational technique becomes a key for obtaining reliable answers. Chemistry of large molecular systems is often driven by noncovalent interactions. Correlated quantum mechanical (QM) and density functional theory methods were shown to provide sufficiently accurate description of noncovalent interactions, but modeling with such methods becomes unfeasible long before a model system can be considered realistic. On the other hand, molecular mechanics (MM) force fields (FF) can be routinely applied to systems containing millions of atoms, but the accuracy of MM simulations depends on good parameterization which makes them less reliable for molecules with unusual binding motifs. To address this challenge, a variety of computational techniques was introduced in recent years, including QM/MM models such as ONIOM,[1−3] advanced FFs,[4−9] fragmentation schemes,[10−16] and semi-empirical methods.[17−22] The effective fragment potential (EFP)[4,5,23−28] is a first principles-based FF that provides a physically meaningful description of noncovalent interactions by a fraction of computational cost of analogous QM methods such as the symmetry adapted perturbation theory (SAPT).[29−32]

The EFP method decomposes noncovalent interactions into Coulomb, polarization, dispersion, and exchange-repulsion terms described using precomputed parameters of isolated molecular fragments. Because of quantum-mechanically derived functional forms and parameters, EFP provides a physically meaningful insight into structural patterns and interaction energies of a chemical system. Previous benchmarks on S22 and S66 data sets demonstrate that EFP reliably describes noncovalent interactions of different types, with the accuracy approaching that of MP2 and exceeding that of classical FFs.[33] Additionally, EFP can be combined with QM models into polarizable QM/EFP schemes, with applications ranging from excitation and ionization properties of solvated chromophores to electron and energy transfer processes in biology.[4,23,25,27,34−41]

**Received:** July 21, 2020
**Published:** November 25, 2020




https://dx.doi.org/10.1021/acs.jctc.0c00758
*J. Chem. Theory Comput.* 2020, 16, 7735−7747

EFP parameters can be computed in an automated fashion from first-principles calculations for any chemical species. This makes EFP suitable for modeling a wide range of molecular systems, which might be challenging with standard FFs that require cumbersome parameterization of nonstandard molecules. On the other hand, the current EFP formalism assumes that an EFP fragment is a rigid molecule, such that one either uses the same parameters of the fragment in all successive computations (if structural changes of the molecule can be ignored) or needs to recompute parameters for each unique molecular geometry. Specifically, the latter scenario becomes unavoidable in biological simulations where both protein backbone and amino acid (AA) side chain residues possess high structural flexibility. However, re-computing parameters for different fragment geometries requires intensive computational resources, significantly increasing both computational and human costs of EFP simulations.

One approach that mitigates a problem of flexible fragments is updating fragment geometries and recomputing fragment parameters on-the-fly, as implemented in the effective fragment molecular orbital (EFMO) method.[42−45] In EFMO, which is a hybrid of the EFP and FMO (fragment molecular orbital) methods, interactions between neighboring fragments are computed quantum-mechanically with the FMO formalism, while the long-range interactions are obtained from the EFP formalism. The EFMO method has been shown to provide accurate energetics in water clusters, griffithsin−carbohydrate complex, and enzymatic catalysis.[46−50] Another possibility, which has been partly explored in our previous work,[9] is to adjust parameters computed at one fragment geometry to other geometries, without explicitly recomputing them. The goal of the present work is to refine and generalize this approach, and provide essential benchmarks. Specifically, we outline a procedure for adjusting parameters to an arbitrary fragment geometry by introducing local coordinate systems associated with fragment atoms and rotating and translating parameters accordingly. We validate this approach by considering test systems designed from molecular interactions occurring in a cryptochrome protein. Based on these benchmarks, we design a library of AA fragments that can be used for modeling noncovalent interactions in proteins.

## 2. THEORETICAL AND COMPUTATIONAL METHODS

**2.1. EFP Method.** The EFP method is a quantum-mechanical (QM) based model potential for describing intermolecular interactions in condensed phase systems. The basic premise of the EFP method is to represent a molecular system as a collection of rigid fragments. Noncovalent interactions between these fragments determine structure and binding patterns of the system.

The EFP energy ($E_{\mathrm{EFP-EFP}}$) of a molecular system consists of four dominant terms: electrostatic or Coulomb ($E_{\mathrm{elec}}$), polarization ($E_{\mathrm{pol}}$), dispersion ($E_{\mathrm{disp}}$), and exchange-repulsion ($E_{\mathrm{ex\text{-}rep}}$) energies. The charge-transfer term, which accounts for the resonance energy stabilization due to transfer of a charge between a pair of fragments, is omitted in this study.

$$E_{\mathrm{EFP-EFP}} = E_{\mathrm{elec}} + E_{\mathrm{pol}} + E_{\mathrm{disp}} + E_{\mathrm{ex\text{-}rep}} \tag{1}$$

In the presence of an *ab initio* subsystem, the interactions of the EFP fragments with the QM region are described by the polarization embedding approach; namely, fragments interact with the *ab initio* Hamiltonian through one-electron Coulomb and polarization terms

$$\hat{H}_{\mathrm{QM-EFP}} = \langle p|\hat{V}^{\mathrm{coul}} + \hat{V}^{\mathrm{pol}}|q\rangle p^{\dagger} q \tag{2}$$

where $p$ and $q$ are the atomic orbitals of the *ab initio* region. $\hat{V}^{\mathrm{Coul}}$ and $\hat{V}^{\mathrm{pol}}$ represent electrostatic and polarization perturbations to the quantum Hamiltonian, respectively.

The electrostatic term $E_{\mathrm{elec}}$ is computed-based on a multipole expansion obtained from the distributed multipole analysis by Stone et al.[51] The multipoles are centered at atoms and bond midpoints. The electrostatic interaction energy between EFP fragments $E_{\mathrm{elec}}$ consists of charge−charge, charge−dipole, charge−quadrupole, charge−octopole, dipole−dipole, dipole−quadrupole, and quadrupole−quadrupole terms. The electrostatic contribution to the *ab initio* Hamiltonian due to a multipole expansion point $k$ of a fragment can be written as

$$V_k^{\mathrm{Coul}}(x) = q_k T(r_{kx}) - \sum_a^{x,y,z} \mu_k{}^a T^a(r_{kx}) + \frac{1}{3}\sum_{a,b}^{x,y,z} \Theta_k{}^{ab} T^{ab}(r_{kx}) - \frac{1}{15}\sum_{a,b,c}^{x,y,z} \Omega_k{}^{abc} T^{abc}(r_{kx}) \tag{3}$$

where $q$, $\mu$, $\Theta$, and $\Omega$ are the charge, dipole, quadrupole, and octopole moments at the multipole point $k$. $T$, $T^a$, $T^{ab}$, and $T^{abc}$ are the electrostatic tensors of ranks zero to three, and $r_{kx}$ is the distance between the expansion point $k$ and the coordinate of an electron $x$ in the QM region. To account for the charge-penetration arising from the overlap of electron clouds between closely-spaced moieties, a Gaussian-type damping function is used for the first (charge) term in eq 3.[4,25]

The polarization energy $E_{\mathrm{pol}}$ is computed as an interaction of induced dipoles of a fragment with an electrostatic field produced by the other fragments and the *ab initio* region. Hence, polarization energy is calculated in a self-consistent manner because the induced dipole on one fragment is determined by the static multipoles as well as induced dipoles on the other fragments. In EFP, static anisotropic polarizability tensors $\alpha^k$ (2nd order tensors) of a fragment are assigned to the localized molecular orbital (LMO) centroids, where the induced dipoles are computed. Polarization energy (in the presence of the QM subsystem) is computed as

$$E_{\mathrm{pol}} = \frac{1}{2}\sum_a^{x,y,z}\sum_k [-\mu_a^k (F_a^{\mathrm{mult},k} + F_a^{\mathrm{nuc},k}) + \bar{\mu}_a^k F_a^{ai,k}] \tag{4}$$

where $\mu^k$ and $\bar{\mu}^k$ are the induced and conjugated induced dipoles on LMO centroid $k$. $F^{\mathrm{mult},k}$ is the electric field due to multipole moments on point $k$; $F^{\mathrm{nuc}}$ and $F^{ai}$ are electric fields due to nuclei and electrons of the QM subsystem, respectively. Induced dipoles at polarizability points $k$ are given by

$$\mu_\alpha^k = \sum_b^{x,y,z} a_{ab}^k (F^{\mathrm{mult},k} + F_b^{\mathrm{ind},k} + F_b^{\mathrm{nuc},k} + F_b^{ai,k}) \tag{5}$$

where $F^{\mathrm{ind},k}$ is the field due to induced dipoles of other fragments. Polarization perturbation to the *ab initio* Hamiltonian due to effective fragments is given as

$$V_k^{\mathrm{pol}}(x) = \frac{1}{2}\sum_a^{x,y,z} (\mu_a^k + \bar{\mu}_a^k) T_a(r_{kx}) \tag{6}$$

Dispersion energy, $E_{\mathrm{disp}}$, between EFP fragments accounts for the interaction between instantaneous dipoles and is formulated using dynamic polarizability tensors. Dispersion contribution is shown to be dominant in nonpolar molecules



https://dx.doi.org/10.1021/acs.jctc.0c00758
J. Chem. Theory Comput. 2020, 16, 7735−7747

and aromatic compounds. Similar to static polarizabilities, the dynamic polarizability tensors are computed and distributed on the LMO centroids of each fragment. Adamovic and Gordon adapted the formulation of $C_6$ dispersion coefficients by Amos et al. for EFP in 2005, in which the $C_6$ coefficients between a pair of distributed points $k$ and $j$ on different fragments are computed as integrals over imaginary frequencies $i\omega$ of a product of average distributed time-dependent polarizability tensors $\bar{\alpha}(i\omega)$[52,53]

$$C_6^{kj} = \int_0^\infty d\omega \bar{\alpha}^k(i\omega)\bar{\alpha}^j(i\omega) \tag{7}$$

The integral in the equation is computed using a quadrature, and dynamic polarizability tensors on each fragment are precomputed for a set of 12 imaginary frequencies using the time-dependent HF method. Using $C_6$ coefficients, the dispersion energy between EFP fragments can be expressed as

$$E_{disp} = -\frac{4}{3}\sum_{k\in A}\sum_{j\in B}\frac{f_{damp}C_6^{kj}}{R_{kj}^6} \tag{8}$$

where $f_{damp}$ corresponds to a damping function that accounts for interactions between fragments at close separation. The expression for dispersion energy between the QM and EFP region has been also derived and implemented.[54,55] The QM−EFP dispersion energy depends on the dynamic polarizability tensors at EFP fragments and electric field integrals and orbital energies of the QM region. While we do not consider the QM−EFP dispersion term in the present work, the proposed approach for obtaining EFP parameters at varying fragment geometries is applicable without any modifications to the QM−EFP dispersion term as well. Implementation of many-body dispersion that is shown to be important in molecular crystals, nonpolar complexes, as well as in biological systems, is an interesting future direction of the EFP development.[56−59]

The exchange-repulsion energy $E_{ex\text{-}rep}$ for each pair of fragments is calculated using fragments' localized wave functions, as a sum of exchange-repulsion interactions $E_{ex\text{-}rep}^{ij}$ between LMOs of each fragment[60,61]

$$E_{ex\text{-}rep} = \sum_{i\in A}\sum_{j\in B}E_{ex\text{-}rep}^{ij} \tag{9}$$

$$\begin{aligned} E_{ex\text{-}rep}^{ij} = {} & -4\sqrt{\frac{-2\ln|S_{ij}|}{\pi}}\frac{S_{ij}^2}{R_{ij}} \\ & - 2S_{ij}\left(\sum_{k\in A}F_{ik}^A S_{kj} + \sum_{l\in B}F_{jl}^B S_{li} - 2T_{ij}\right) \\ & + 2S_{ij}^2\left(-\sum_{J\in B}\frac{Z_J}{R_{iJ}} + 2\sum_{l\in B}\frac{1}{R_{il}} - \sum_{l\in A}\frac{Z_l}{R_{lj}}\right. \\ & \left. + 2\sum_{k\in A}\frac{1}{R_{kj}} - \frac{1}{R_{ij}}\right) \end{aligned} \tag{10}$$

where $i$, $j$, $k$, $l$ are the LMOs, $L$, $J$ are the nuclei, $S_{ij}$ and $T_{ij}$ are the overlap and kinetic energy integrals between $i$th and $j$th LMOs belonging to different fragments A and B, and $F$ is the Fock matrix elements of each fragment. The exchange-repulsion term is the most computationally expensive term in the EFP model, as it explicitly evaluates the overlap and kinetic energy integrals between the molecular orbital basis functions of fragments. Two distinct implementations of the QM−EFP exchange-repulsion term exist.[24,62,63] In the model from ref 63, the QM−EFP exchange-repulsion term is computed using overlap of *ab initio* wave function with repulsive Gaussian-like potentials positioned at the LMO centroids of each fragment. The parameters describing these interactions are scalar values (widths and heights of Gaussian functions). While the QM−EFP exchange-repulsion term is not considered in the present work (all presented QM/EFP calculations utilize a polarizable embedding model, with only electrostatic and polarization EFP contributions to the QM Hamiltonian, see eq 2), the developed mechanism for adjusting parameters to different fragment geometries can be easily extended to this term as well.

**2.2. Flexible EFP Scheme.** The EFP parameters consist of the following components: (i) atomic and bond midpoint coordinates, (ii) electrostatic multipoles (charges, dipoles, quadrupoles, and octopoles) and the electrostatic damping coefficients at atoms and bond midpoints, (iii) coordinates of the localized molecular orbital (LMO) centroids, (iv) static and dynamic polarizability tensors at the centroids of LMOs, (v) localized wave function described *via* atomic basis and molecular orbital coefficients, and the Fock matrix in the basis of localized molecular orbitals. The location of these parameters for the water molecule is shown in Figure 1. Multipoles, static and dynamic polarizability tensors, and localized molecular orbital wave functions are described as tensors of ranks zero to three.

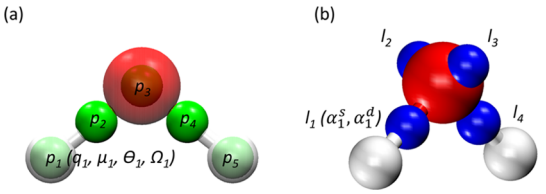



**Figure 1.** Distributed EFP parameters are shown on the example of a water molecule. (a) Multipole moments [charges ($q$), dipoles ($\mu$), quadrupoles ($\theta$), and octopoles ($\Omega$)] are distributed at atomic centers and bond midpoints $p$ shown with green spheres. (b) Static $\alpha^s$ and dynamic $\alpha^d$ polarizability tensors are placed at localized molecular orbital centroids (shown with blue spheres).

In this work, we explore whether it is feasible to "transfer" parameters computed at one fragment geometry to another geometry by translating and rotating the corresponding tensors according to the changes in fragment geometry.

We refer to a "flexible fragment" as a molecule that changes its internal geometry but not a chemical connectivity during a considered chemical process. In order to adjust precomputed potentials to a new geometry, we introduce local translational and rotational frames based on each triplet of neighboring atoms within a molecule. Here, we define the triplet of atoms as a group of any three atoms forming a valence angle, or, in other words, the three atoms out of which one is covalently bound to two others. Each triplet of atoms determines a coordinate plane. A rotation/translation matrix is computed for each pair of corresponding atom triplets in geometries A (original structure) and B (new structure). Then, a two-step rotation (R1 = R*R** = $\sum_{k=1}^{n} r_{ik}^{*} r_{kj}^{**}$ and R2 = R*R*** = $\sum_{k=1}^{n} r_{ik}^{*} r_{kj}^{***}$, where R*, R** and R*** are the rotation matrices, $i$ = 1,...,$m$, $j$ = 1,...,$m$, and $n$ = 3, $m$ = 3) brings all parameters positioned at the three atoms of structure A to the


https://dx.doi.org/10.1021/acs.jctc.0c00758
J. Chem. Theory Comput. 2020, 16, 7735−7747

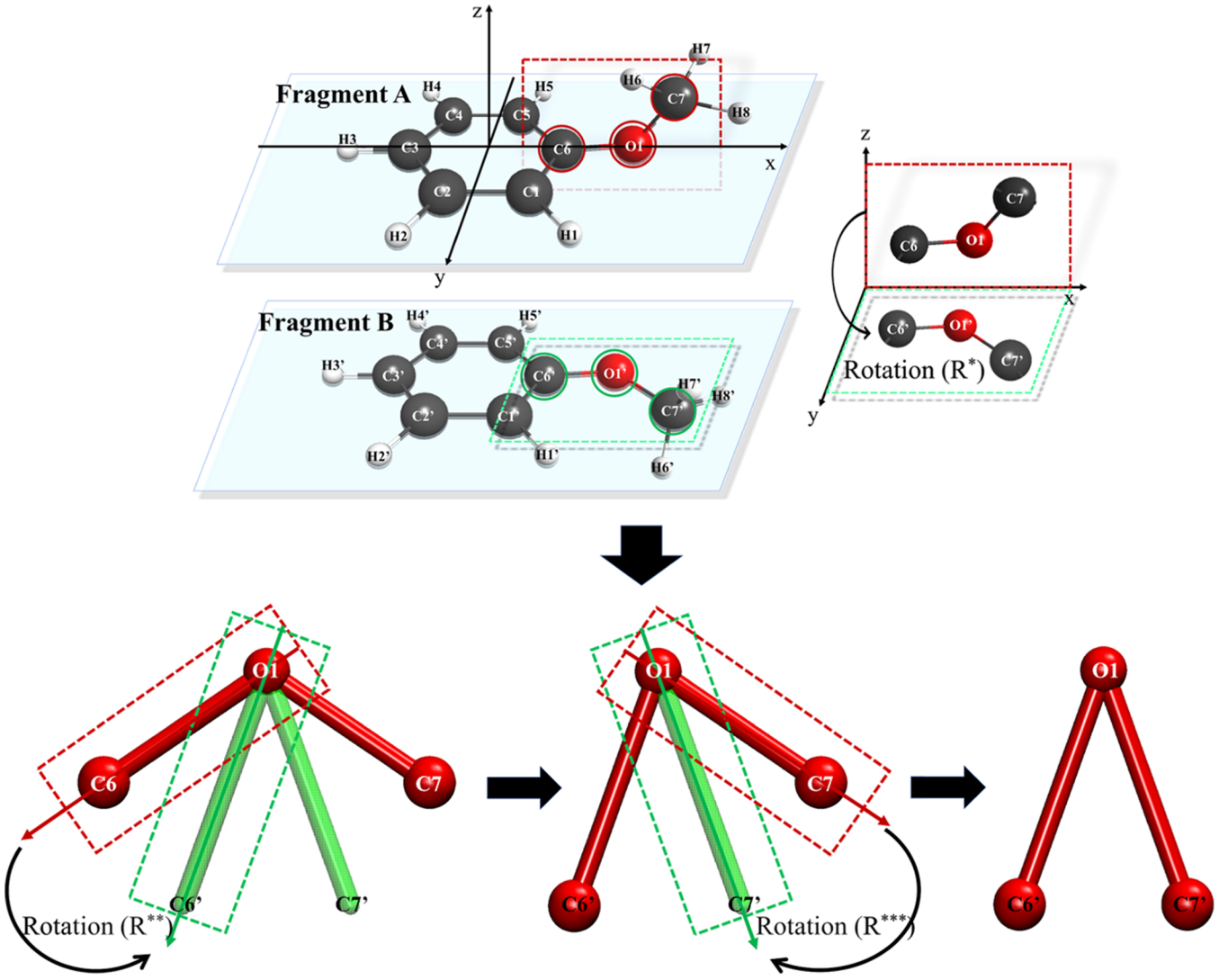



**Figure 2.** Translation of EFP parameters according to changes in fragment's geometry. Fragment A is a precomputed EFP fragment and B represents a fragment at a different geometry to which the parameters should be adjusted. (Upper panel) Rotation R* of a local plane defined by the triplet of atoms C6−O1−C7 between structures A and B. (Lower panel) Rotations R** and R*** bring positions and corresponding parameters of atoms C6 and C7 of fragment with structure A to the positions of atoms in structure B.

corresponding parameters in structure B, as shown in Figure 2. This three-atom local transformation procedure is repeated until parameters at all atoms of a fragment are shifted to a new geometry on an atom by atom basis. Atoms and corresponding potentials might undergo more than one translation/rotation. For example, in Figure 2, atom O1 undergoes two rotations, R1 = R*R** and R2 = R*R***. In this case, the resulting positions and corresponding parameters are averaged to produce the final potential. All parameters are averaged component by component, for example, the average position and the average dipole moment on an atom are found as the averages of $x$, $y$, and $z$ coordinates or dipole moment components produced by different rotations. Parameters at bond mid-points are shifted proportionally to changes in the corresponding bond length and rotated using the same rotation matrix as the atoms forming the bond. To adjust parameters positioned at LMO centroids, the following procedure is used. The LMOs are assigned to atom triplets based on the closest distance either to one of the bond mid-points of the triplet (this is typically the case for LMOs describing valence bonds) or to one of the atoms, in the case of lone-pair LMOs. Then, the LMOs are rotated and translated together with the atoms or bond mid-points of the corresponding atom triplet.

### 2.3. Construction of EFP Parameter Database EFPDB.

Twenty-five protein conformations were chosen from molecular dynamics (MD) trajectories of cryptochrome 1 from *Arabidopsis thaliana* (Cry1At, PDB: 1U3D[64]). A simulation protocol is described in detail in ref 65. Following the original implementation of the BioEFP fragmentation algorithm,[25] each of 25 protein configurations was fragmented into individual AA fragments along Cα−C bonds, resulting in 12125 BioEFP AA fragments.

However, we noticed in a preliminary work (data not shown) that the additional fragmentation of AAs into the backbone and residue groups improves the accuracy of the flexible EFP. Thus, in addition to the original BioEFP fragmentation, each AA fragment was further split along the Cα−Cβ bond into a backbone group (N−Cα−C═O) and a side-chain group. In this fragmentation, shown in Figure 3, a glycine is represented by a backbone fragment. The only exception from the above scheme is proline that is described as


https://dx.doi.org/10.1021/acs.jctc.0c00758
J. Chem. Theory Comput. 2020, 16, 7735−7747

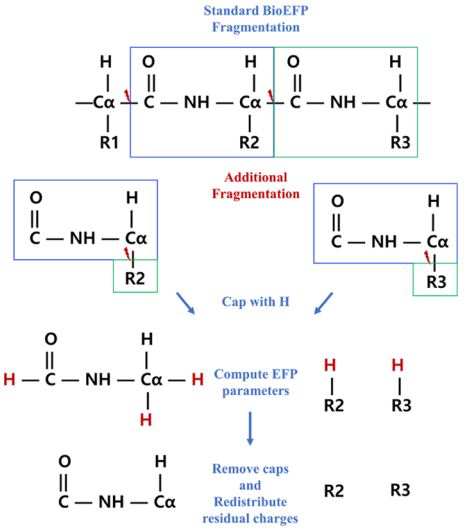



**Figure 3.** Fragmentation of a polypeptide chain into EFP fragments.

| | AA | # geom | | AA | # geom |
|---|---|---|---|---|---|
| | GLY | 7 | | PRO | 2 |
| | CYS | 3 | | THR | 9 |
| | MET | 22 | POLAR | ASN | 18 |
| ALIPHATIC | ILE | 27 | | GLN | 9 |
| | LEU | 27 | | SER | 3 |
| | VAL | 9 | CATIONIC | ARG | 26 |
| | ALA | 1 | | LYS | 54 |
| ARYL | TRP | 3 | | HIP | 1 |
| | HID | 4 | ANIONIC | ASP | 17 |
| | HIE | 1 | | GLU | 23 |
| | PHE | 3 | | | |
| | TYR | 8 | | **TOTAL** | **281** |

**Figure 4.** Unique structures of standard AA fragments. Peptide backbone group is represented by glycine (GLY). HID, HIE, and HIP are $\delta$-protonated, $\varepsilon$-protonated, and doubly protonated (positively charged) isomers of a histidine residue. EFP fragments are characterized as aryl (TRP, HID, HIE, PHE, and TYR), polar (THR, ASN, GLN, and SER), aliphatic (ILE, LEU, VAL, and ALA), cationic (ARG, LYS, and HIP), and anionic (ASP and GLU) residues.

a single fragment. Additionally, disulfide bridges are fragmented along the S−S bond.

To recombine the backbone and side-chain fragments in a complete AA moiety, parameters at a bond midpoint and LMO centroid between C$\alpha$ and C$\beta$ were excluded to ensure the stability of polarization self-consistent procedure (see Figure S1). So, compared to the original BioEFP scheme, the flexible EFP scheme has one less parameter point for each AA and is expected to show consistently smaller magnitudes of all energy terms than the original BioEFP. However, as demonstrated in dimer benchmarks below, the resulting discrepancies between two EFP schemes are minor in practice. In principle, parameters at the boundary points between fragments can be included if neighboring fragments do not polarize (or do not fully polarize) each other. Such models will be considered in our future work.

The 12125 AA backbone (glycine) and 12125 side-chain fragments obtained after fragmentation of 25 protein structures were protonated to cap cleavage sites and subjected to geometry optimizations at the HF/6-31G(d) level of theory using the GAMESS electronic structure package.[66] In total, 281 unique geometries corresponding to distinct local minima were isolated (see Figure 4). These (local-minima) structures constitute the AA EFP database.

EFP parameters for all unique AA structures (*i.e.*, 281 total fragments) were computed with a hybrid 6-31G(d)/6-31( , +)G(3df,2p) (for aryl residues) and 6-31G(d)/6-311++G-(3df,2p) (for other residues) basis using MAKEFP module of GAMESS. It was previously noted that the accuracy of EFP interaction energies improves significantly with the use of a hybrid basis set scheme for the parameter preparation step.[33,67,68] Our earlier studies indicate that electrostatic multipoles are more accurate (and nondivergent) when computed with a small basis set (6-31G(d)), while the rest of the EFP terms (polarization, dispersion, and exchange-repulsion) are more accurate when the parameters are computed in a larger basis set (6-311++G(3df,2p)).[33] Electrostatic exponential screening functions, Gaussian polarization screenings, and overlap-based dispersion screenings are employed to improve the description of corresponding EFP−EFP terms at short intermolecular separations and to avoid polarization catastrophe.[69] Gaussian screenings are utilized for short-range QM−EFP electrostatic interactions.[4,25]

As expected, AA residues possess high flexibility, with lysine producing the largest number of local minima (54) (Figure 4). A selection of a particular fragment from the database is based on the root mean square deviation (rmsd) between geometries of the target molecule and available database fragments. All fragment atoms including hydrogens were used in computing rmsd. Our preliminary investigations suggest that the accuracy of flexible EFP noncovalent interaction energies correlates with rmsd between initial and target structures (the dependence of absolute errors of the flexible EFP scheme on rmsd, shown on the example of interactions in lysine-benzoic acid dimer, is presented in Figure S2). Therefore, fragments with the smallest rmsd to the target structure are selected and subjected to rotation and shifting to match the target geometry. Figure 5 shows a spread in geometrical differences between fragments from the MD trajectory and the closest (in the rmsd sense) database fragments for each AA residue. Only for two residues, GLU and GLN, geometries of original fragments might deviate from geometries of database fragments by more than 0.5 Å. For other residues, typical rmsd differences are between 0.1 and 0.3 Å.

**2.4. Dimer Calculations.** To test the quality of the flexible EFP scheme, we considered a set of dimers, in which each monomer is composed of a peptide group (backbone) and an AA residue (side-chain), extracted from 25 Cry1At protein snapshots based on the distances between the centers of masses (COMs) of the AA residues. A threshold of 5 Å between the AA COMs resulted in 240 aryl−aryl, 533 aryl−aliphatic, 105 aryl−cationic, 49 aryl−anionic, 108 anionic−anionic, 305 anionic−cationic, and 388 polar−polar types of AA pairs (total 1728 pairs). Examples of considered dimers are shown in Figure 6. While the considered set does not include all possible interactions, we found it to be sufficiently large and diverse to be representative of protein noncovalent interactions.

The total interaction energies and individual energy components of these dimers are computed with flexible EFP, standard BioEFP, and the SAPT.



https://dx.doi.org/10.1021/acs.jctc.0c00758
J. Chem. Theory Comput. 2020, 16, 7735−7747

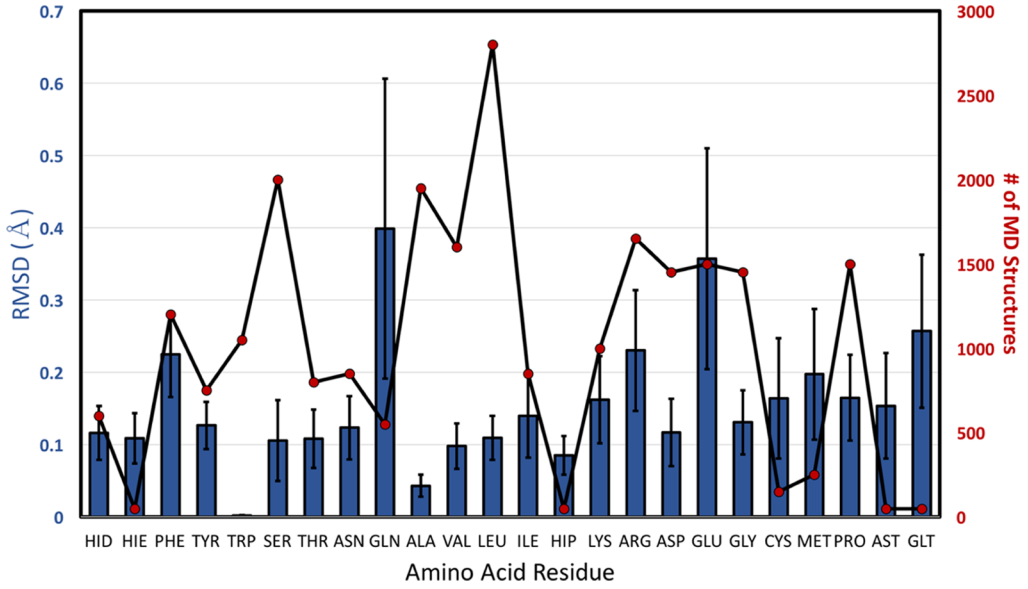



**Figure 5.** Average rmsd values (blue columns) and standard deviations of rmsd for each AA residue. The number of structures for each AA residue is shown with red dots.

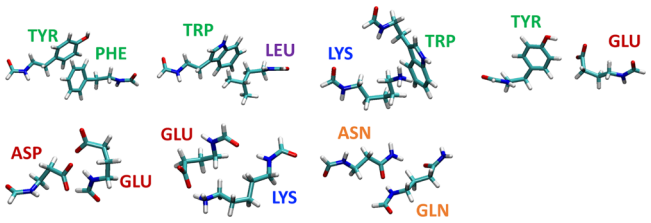



**Figure 6.** Representative structures of AA dimers.

*2.4.1. SAPT Calculations.* Sherrill and co-workers suggested the gold, silver, and bronze standards of SAPT methods based on the accuracy and computational cost of intermolecular interaction energy computations in several databases (S22, HBC6, NBC10, and HSG).[70] Taking into consideration a low computational cost and similarity of the description of noncovalent interactions to EFP [*i.e.*, no correlation (Hartree−Fock level) in describing fragments, second-order perturbation theory in describing interfragment interactions], we utilized sSAPT0/jun-cc-pVDZ (bronze standard) as a reference method.[70] SAPT interaction energies were computed with the PSI4 quantum chemistry package.[71]

*2.4.2. Standard BioEFP.* For performing standard EFP calculations, individual EFP parameters were generated for all monomers in the dimer complexes, totaling in 1728 × 2 = 3456 fragment parameters. Each monomer was represented by a single fragment containing a peptide backbone and AA residue groups. In other words, EFP parameters were explicitly computed for each fragment geometry following the original BioEFP procedure as reported in ref 25. Parameters were prepared in the hybrid basis sets: 6-31G(d)/6-31G( , +)G(3df,2p) for aryl compounds and 6-31G(d)/6-311+ +G(3df,2p) for the other monomers.

*2.4.3. Flexible EFP.* For flexible EFP calculations, the parameters of each monomer were obtained from the EFP parameter database discussed above. Each monomer (except glycine and proline) is constructed from the peptide backbone and side-chain fragments found in the database, using the smallest rmsd between the geometries of the target molecule and available database fragments as a selection criterion. Parameters of the selected database fragments are subjected to rotation and shifting to match the target geometry.

A comparison of workflows of standard BioEFP and flexible EFP calculations is presented in Figure 7.

All EFP interaction energies were computed with the EFPMD module of LibEFP software library.[72,73]

**2.5. Pairwise EFP.** Decomposition of the EFP energy of a many-fragment system into pairwise fragment−fragment interactions has been implemented in LibEFP software library.[72,73] Electrostatics, dispersion, and exchange-repulsion interactions are two-body, so the corresponding system energies are sums of all pairwise fragment−fragment interactions. Many-body polarization energy was decomposed into fragment−fragment interactions in the following way: (i)

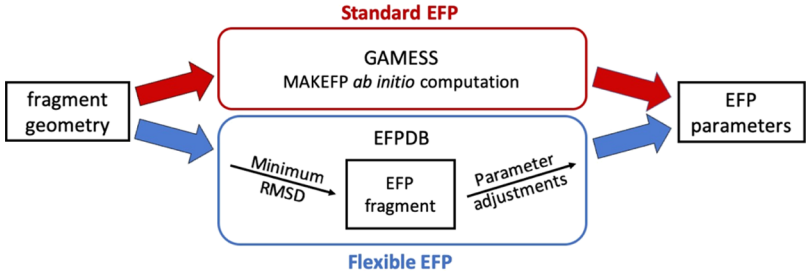



**Figure 7.** Workflow of standard (bio)EFP and flexible EFP calculations.



https://dx.doi.org/10.1021/acs.jctc.0c00758
J. Chem. Theory Comput. 2020, 16, 7735−7747

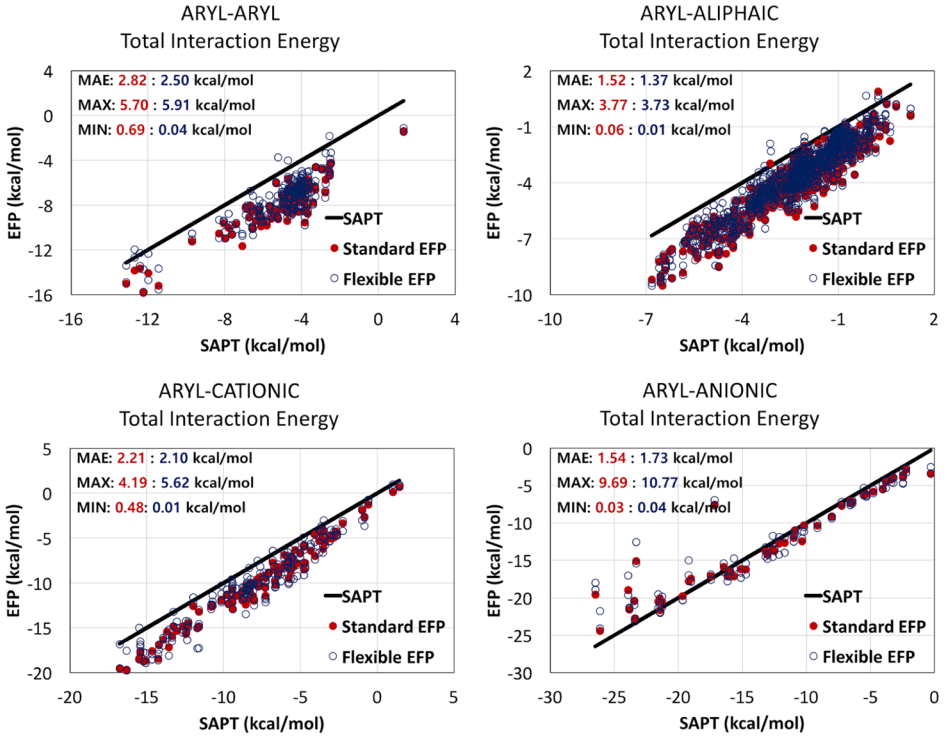



**Figure 8.** Total interaction energies in dimers containing aryl residues computed with standard EFP (red circles) and flexible EFP (blue circles) against sSAPT0/jun-cc-pVDZ.

induced dipoles are self-consistently converged for the whole system; (ii) using converged dipoles, pairwise polarization energies between fragments $i$ and $j$ are computed as

$$E_{ij} = -\frac{1}{2}\sum_{a}^{x,y,z}\left[\sum_{\mathrm{pp}\in i}\mu_{\mathrm{pp},a}F^{j}_{\mathrm{pp},a} + \sum_{\mathrm{pp}\in j}\mu_{\mathrm{pp},a}F^{i}_{\mathrm{pp},a}\right] \quad (11)$$

where $\mu_{pp}$ is the induced dipole moment at the polarizability point pp, $F_{pp}{}^{j}$ is the electric field at point pp due to all nuclei and static multipoles of fragment $j$. In this formulation, the total polarization energy is a sum of all pairwise energies, but each dimer energy implicitly incorporates many-body effects through the induced dipoles self-consistently converged for the whole system. Thus, the pairwise scheme can conveniently decompose the total noncovalent interaction energy into contributions of individual fragment pairs, or extract interactions of a particular fragment with the other fragments in the system, with potential applications in analysis of ligand-binding interactions.

**2.6. EFP and QM/EFP Calculations of Cryptochrome Cry1At Protein.** To demonstrate applicability of EFP to large-scale simulations, we performed a series of calculations on properties of cryptochrome Cry1At protein. First, using the pairwise energy decomposition, we computed the interaction energies of lumiflavin (a chromophore responsible for redox activity of Cry1At protein) with 483 AAs of the protein, excluding the interactions with terminal AAs, solvent molecules (water, $Na^+$ and $Mg^{2+}$), and cofactors (FAD and ATP). In the pairwise energy calculations, lumiflavin is represented as an EFP fragment, with the parameters prepared with hybrid 6-31G(d)/6-311++G(3df,2p) basis using the standard EFP protocol. Parameters for AAs were prepared as described in Section 2.4. Standard EFP and flexible EFP calculations were performed at 25 MD snapshots of the protein.

Then, ensemble-averaged values of vertical electron affinities ⟨VEA⟩ and vertical ionization energies ⟨VIE⟩ were computed based on the same 25 MD snapshots that were utilized for EFP-only benchmarks. For the exact comparison of ⟨VEA⟩ and ⟨VIE⟩ with the values from ref 65, the 6-31G(d) basis set was employed to compute both electrostatic and polarization parameters of all EFPDB fragments. Additionally, the octopoles of the electrostatic parameters were truncated. The fragment−fragment electrostatic interactions were damped using exponential screening functions, while the QM−fragment electrostatic interactions were damped with Gaussian-type screening.[25] Gaussian-type polarization damping was employed with POLAB values of 0.3 for AAs and 0.1 for ions to avoid overpolarization of neighboring AAs.[69] Water molecules were described with EFP potentials prepared with hybrid 6-31+G(d)/6-311G++(3df,2p) basis sets. To speed-up QM/EFP calculations, the multipole parameters of water fragment were truncated at the quadrupole moments and distributed to atom centers only, while polarizability was represented with a single polarizability tensor located at the water center of mass as described in ref 65. Following the flexible EFP protocol, the EFP parameters of the protein AAs at 25 MD structures were adjusted based on the fragments from the EFPDB. Terminal AAs were prepared with the standard EFP protocol. EFP parameters of solvent (water, $Na^+$, and $Mg^{2+}$) and cofactors (FAD and ATP) were taken from the ref 65.

The ensemble-averaged vertical electron affinity ⟨VEA⟩ is computed in the ensemble of the oxidized form of lumiflavin; the ensemble-averaged vertical ionization energy ⟨VIE⟩ is computed in the ensemble of the reduced form as


https://dx.doi.org/10.1021/acs.jctc.0c00758
J. Chem. Theory Comput. 2020, 16, 7735−7747

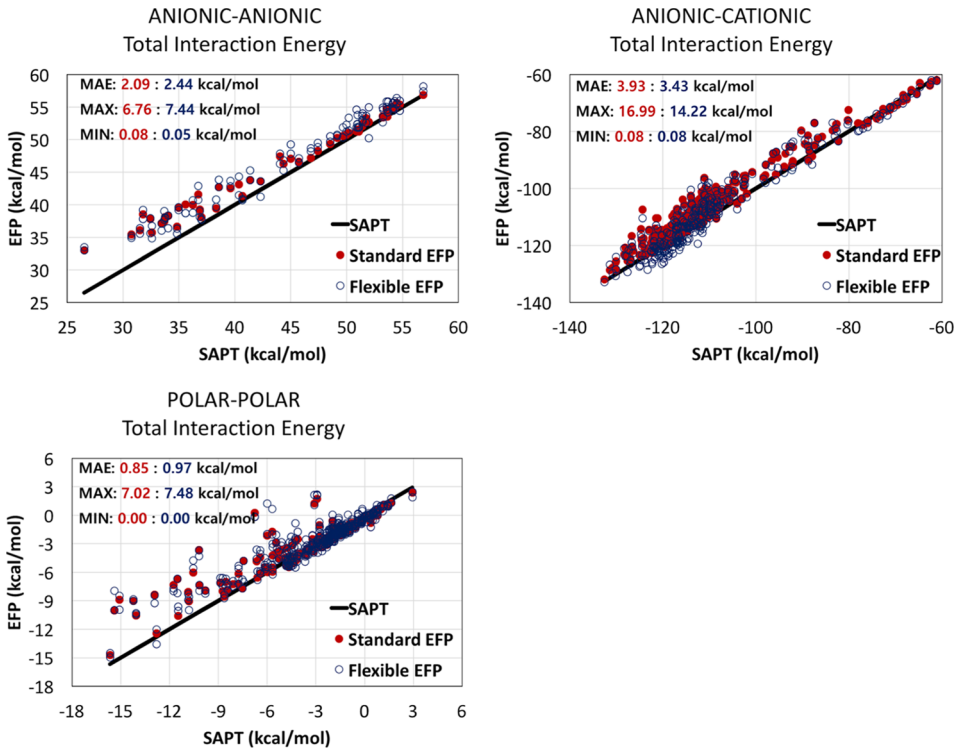



**Figure 9.** Total interaction energies in anionic−anionic, anionic−cationic, and polar−polar dimers computed with standard EFP (red circles) and flexible EFP (blue circles) against sSAPT0/jun-cc-pVDZ.

$$\langle \mathrm{VEA} \rangle = \langle E(\mathrm{ox}) - E(\mathrm{red}) \rangle_{\mathrm{Ox}}$$

$$\langle \mathrm{VIE} \rangle = \langle E(\mathrm{ox}) - E(\mathrm{red}) \rangle_{\mathrm{Red}}$$

where $E$(ox) and $E$(red) are the total QM/EFP energies for oxidized and semireduced states of lumiflavin. $\omega$B97X-D/6-31G(d) was used for describing the QM region.

## 3. RESULTS AND DISCUSSION

**3.1. Interaction Energies in AA Dimers.** Figures 8 and 9 provide a comparison of total interaction energies of original (standard) and flexible EFP with sSAPT0/jun-cc-pVDZ (referred to as "SAPT" in the following discussion) for aryl−aryl, aryl−aliphatic, aryl−cationic, aryl−anionic, anionic−anionic, anionic−cationic, and polar−polar dimers. Mean absolute errors (MAE), maximum errors (MAX), and minimum errors (MIN) of both EFP schemes against SAPT are also reported in Figures 8 and 9. Plots showing energy decomposition of these dimers are provided in the Supporting Information (Figures S3−S9). The presented graphs demonstrate that the total interaction energies computed with the flexible and standard EFP methods exhibit similar values and trends (*e.g.*, complexes with larger errors in standard EFP typically have similarly larger errors in flexible EFP). It is noteworthy that the two EFP schemes also exhibit similar values of energy components, meaning that shifting and rotating of the corresponding parameters does not introduce significant errors.

As follows from Figures 8 and 9, both standard and flexible EFP overestimate attractive interactions in aryl-containing complexes, while interactions in other analyzed complexes (polar−polar, cationic−anionic, and anionic−anionic) are overly repulsive. The reason can be deduced from analysis of the corresponding energy components. As seen in Figures S3−S9, the polarization component is underestimated by magnitude (*i.e.*, not sufficiently attractive) in most complexes because of missing a charge-transfer term and inaccuracies in the charge-penetration term, resulting in a systematic underestimation of the EFP binding energies. On the other hand, as follows from Figures S3−S6, both standard EFP and flexible EFP significantly underestimate the exchange-repulsion component in complexes containing aryl residues, while exchange-repulsion is accurately represented in other complexes (Figures S7−S9). This is due to the fact that the parameters for the exchange-repulsion term for aryls were prepared using 6-31( ,+)G(3df,2p) basis while 6-311++G-(3df,2p) basis was employed for the other compounds. As Figure S10 shows, employing 6-311++G(3df,2p) basis for the exchange-repulsion term in aryl compounds results in accurate exchange-repulsion energies. However, we noticed that rotation of diffuse s and p functions that compose exchange repulsion parameters resulted in large nonsystematic errors in the exchange-repulsion term of flexible EFP. Thus, we opted for using a smaller basis for aryl compounds when creating the parameter database, by a price of systematically underestimating exchange-repulsion energies (and somewhat overestimating total binding energies) in aryl-containing complexes. As a side note, the exchange-repulsion term is not included in a typical QM/EFP polarizable embedding model, such that inaccuracies in this component would not affect the performance of the flexible EFP scheme combined with the QM region, as discussed below.

**3.2. Binding Energy of Lumiflavin.** A power of EFP for applications in biological chemistry is its ability to compute pairwise interactions in a fully polarized system. In such a way, for example, binding energies of a ligand in a protein pocket can be efficiently evaluated. To demonstrate this functionality of EFP and validate flexible EFP for large-scale applications, we


https://dx.doi.org/10.1021/acs.jctc.0c00758
J. Chem. Theory Comput. 2020, 16, 7735−7747

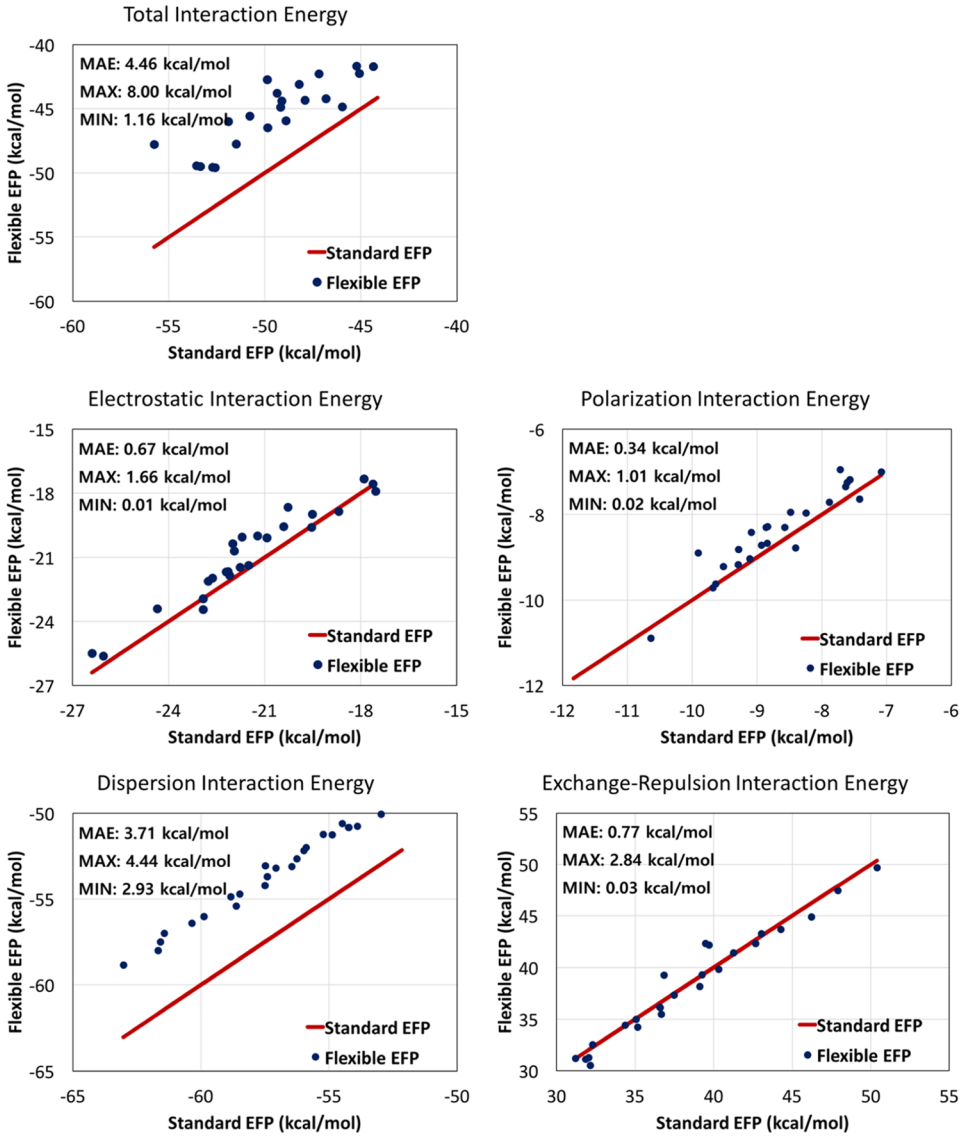



**Figure 10.** Comparison of the total interaction energies and energy components of the lumiflavin chromophore with the protein AAs, computed at 25 MD structural snapshots by the standard EFP and flexible EFP schemes.

computed the interaction energy of lumiflavin (a chromophore responsible for the redox activity of Cry1At protein) with 483 AAs of the protein, excluding interactions with terminal AAs, solvent molecules, and co-factors. Standard EFP and flexible EFP calculations are performed at 25 MD snapshots of the protein.

Comparisons of the standard EFP and flexible EFP total interaction energies and energy components between lumiflavin and surrounding AAs are presented in Figure 10. Energies at different geometrical snapshots are shown as separate points. As in the case of dimer systems, the electrostatic, polarization, and exchange-repulsion energies computed with the flexible EFP scheme are in excellent agreement with those obtained using the standard BioEFP method. However, the dispersion energies in the flexible EFP are significantly less attractive than those energies in the standard EFP. As a result, the MAE of the total interaction energies of flexible EFP against standard EFP is about 8.0 kcal/mol because of the poor agreement in the dispersion energies. However, it does not necessarily mean that the accuracy of flexible EFP is worse than that of standard BioEFP. For a better estimate of lumiflavin−protein binding energies, additional SAPT computations were carried out (see Figure S11). In these SAPT calculations, the lumiflavin−protein interaction energy was computed as a sum of interactions of all lumiflavin−AA dimers, where AAs were capped with H atoms at C and C$\alpha$ sites (*i.e.*, fragmentation was done analogously to the BioEFP scheme). While such SAPT calculations cannot fully reproduce the polarization component of the total interaction energy, electrostatic, dispersion, and exchange-repulsion terms do not incorporate strong many-body effects and should be well-reproduced by this additive scheme.[74] Interestingly, as shown in Figure S11, flexible EFP compares better to SAPT than standard EFP does, demonstrating almost perfect agreement in dispersion energies. The main difference in dispersion energies between the two EFP schemes is that there is no dynamic polarizability point on a border between the peptide group and AA residue fragments (*i.e.*, dynamic polarizability of the localized C$\alpha$−C$\beta$ bond) in flexible EFP, while such points are present in fragments in the standard EFP scheme. These boundary points containing static and dynamic polarizabilities are removed between all


https://dx.doi.org/10.1021/acs.jctc.0c00758
J. Chem. Theory Comput. 2020, 16, 7735−7747

Table 1. VEA and VIE of Lumiflavin in Cry1At Cryptochrome Protein, Computed with Standard and Flexible EFP Schemes[a]

| EFP scheme | VEA, eV | | VIE, eV | |
|---|---|---|---|---|
| | standard EFP | flexible EFP | standard EFP | flexible EFP |
| ensemble-averaged values ± standard deviations | 2.407 ± 0.262 | 2.378 ± 0.261 | 4.926 ± 0.214 | 4.882 ± 0.215 |
| average error | | 0.029 ± 0.038 | | 0.044 ± 0.037 |
| max error | | 0.130 | | 0.112 |
| min error | | 0.002 | | 0.000 |

[a]Ensemble-averaged values with standard deviations, as well as discrepancies between standard and flexible EFP schemes are provided.

covalently linked fragments to ensure stability of polarization self-consistent procedure. Indeed, both the standard and flexible EFP schemes miss such points between each pair of C and Cα atoms in a peptide chain. While the dynamic polarizability points at the boundary between fragments could be kept in one of the participating fragments (as the dispersion interactions do not have convergence issues), we leave exploring this possibility to future work.

Returning to the analysis of Figures 10 and S11, it is clear that the standard and flexible EFP schemes perform consistently between different geometrical snapshots, both for the energy components and for the total interaction energies. Additionally, with the exception of systematic and well-understood discrepancies in the dispersion component, the discrepancies between standard and flexible EFP are smaller than the errors of EFP against SAPT, justifying utilization of the flexible EFP model in large-scale biological applications.

### 3.3. Application of Flexible EFP in QM/EFP Modeling.

At last, we examine the applicability of the flexible EFP scheme in a framework of the polarizable embedding QM/EFP model. Here we examine two electronic properties, ensemble-averaged vertical electron affinity ⟨VEA⟩ and ensemble-averaged vertical ionization energy ⟨VIE⟩, which have been recently reported for lumiflavin in Cry1At protein.[65] In these calculations lumiflavin is treated quantum-mechanically, while the interactions between lumiflavin and protein are described with EFP electrostatic and polarization terms (eq 2). The overall setup of the system is identical to the one reported in ref 65 and is pictorially represented in Figure S12. We note that the goal of this benchmark is not to obtain the best description of lumiflavin within the protein but rather to demonstrate a close agreement between the standard and flexible EFP schemes.

As seen in Table 1, the average discrepancies between the standard and flexible EFP schemes are ∼0.029 eV (0.67 kcal/mol) and ∼0.042 eV (0.97 kcal/mol) for ⟨VEA⟩ and ⟨VIE⟩, respectively, which is almost an order of magnitude smaller than the uncertainties (standard deviations) in these values. Specific data on each MD snapshot, as well as a brief description of the computational protocol for computing ensemble-averaged VIE and VEA are provided in the Supporting Information (Table S1). Overall, this example demonstrates that the flexible EFP scheme is reliable in biological simulations and can be safely used for modeling electronic properties of biochromophores.

It should be emphasized that the flexible EFP scheme provides significant computational savings compared to standard EFP as the necessity to recompute parameters for hundreds of fragments at different geometries is avoided. For example, in the case of cryptochrome protein, using the flexible EFP protocol and the EFPDB database allowed to avoid generation of 485 EFP parameters for each snapshot which saved ∼20 and ∼200 CPU hours, for parameter calculations in 6-31G* and 6-311++G(3df,2p) bases, respectively. For comparison, a single EFP calculation of the binding energies in cryptochrome takes ∼0.2 CPU hour, while a typical QM/EFP calculation of VIE and VEA costs between 8 and 12 CPU hours on a 2.60 GHz processor. Thus, the flexible EFP scheme eliminates the most time-consuming step in both EFP and QM/EFP procedures. Additional critical benefit of using flexible EFP and EFPDB is a significant simplification of setup of EFP simulations and avoiding human errors in preparing, storing, and book-keeping parameters of multiple fragments. We expect the EFPDB parameter database is sufficiently diverse in covering an AA configurational space such that it can be used without or only with minor adjustments in modeling other proteins. We recommend although that rmsds between database fragments and AAs of a system of interest are used as a metric whether the database parameters are safe to use or new parameter sets have to be computed for ensuring the accuracy of calculations. The recommended rmsd cutoff for "safe" parameter rotations is 0.7 Å.

The scripts to perform flexible EFP calculations and the EFPDB database are stored on GitHub within the open source distribution of the LibEFP library at https://github.com/libefp2/libefp.

The flexible EFP protocol can be extended and generalized for other types of molecules and macromolecules. In our own work, we already applied the protocol to bacteriochlorophyll *a* and hexafluoro-*o*-xylene.[36,75] Extensions to other biological polymers such as lipids and DNA should be easily achievable. Other promising applications of the flexible EFP model are to synthetic organic polymers for which standard FF parameters might be less accurate, and to formulations of potential drug molecules.

## 4. CONCLUSIONS

In this paper, the concept and implementation of the flexible EFP model are reported. The flexible EFP formalism is based on connecting geometries of two chemically identical but structurally different fragments by a series of local coordinate rotations and translations. The distributed parameters associated with atoms, bond midpoints, and LMO centroids of one fragment are subjected to the corresponding reorientations, producing a new set of parameters that are suitable for describing the fragment in a new geometry. To utilize this approach in studies of proteins, we developed the EFPDB parameter database that contains EFP parameters of standard AAs. We validated the flexible EFP scheme by considering interaction energies in the AA dimers and showed that flexible EFP is at least as accurate as standard EFP with respect to sSAPT0/jun-cc-pVDZ. Applications of the flexible EFP scheme to binding energies and vertical electron attachment and ionization energies in the cryptochrome protein demonstrate substantial benefits of utilizing flexible


https://dx.doi.org/10.1021/acs.jctc.0c00758
J. Chem. Theory Comput. 2020, 16, 7735−7747

EFP in realistic simulations of biological systems. Although the computational cost of obtaining EFP parameters becomes negligible in the flexible EFP scheme, the predicted ensemble-averaged ⟨VEA⟩ and ⟨VIE⟩ deviate from those obtained with the standard EFP by only 0.029 and 0.042 eV, respectively. Hence, the flexible EFP approach represents a computationally efficient way of conducting large-scale EFP simulations without compromising accuracy.

Although the flexible EFP allows transferring the EFP parameters between different fragment geometries, the utilized model in which the distributed parameters are merely shifted and rotated but not otherwise adjusted to reflect changes in the electronic structure of a fragment at a new geometry might break in the case of significant geometrical rearrangements of the fragment. To address these concerns, the flexible EFP scheme might be further refined by accounting for changes in parameter magnitudes upon structural rearrangements of the fragment, for example, by expanding the parameter elements in a Taylor series around the equilibrium position. Future research will also explore enabling EFP-based dynamics simulations of proteins and other macromolecules by representing fragment internal degrees of freedom by quantum-mechanical, classical, or machine learned potentials.

## ■ ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.0c00758.

Interaction energy components in AA dimers, pairwise interaction energies between AAs and lumiflavin in the Cry1At protein, and vertical electron affinities and vertical ionization energies of lumiflavin in the Cry1At protein (PDF)

## ■ AUTHOR INFORMATION


### Corresponding Author

**Lyudmila V. Slipchenko** − *Department of Chemistry, Purdue University, West Lafayette, Indiana 47907, United States;* orcid.org/0000-0002-0445-2990; Email: lslipchenko@purdue.edu

### Authors

**Yongbin Kim** − *Department of Chemistry, Purdue University, West Lafayette, Indiana 47907, United States*

**Yen Bui** − *Department of Chemistry, Purdue University, West Lafayette, Indiana 47907, United States*

**Ruslan N. Tazhigulov** − *Department of Chemistry, Boston University, Boston, Massachusetts 02215, United States; Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, California 91125, United States;* orcid.org/0000-0002-0679-3078

**Ksenia B. Bravaya** − *Department of Chemistry, Boston University, Boston, Massachusetts 02215, United States;* orcid.org/0000-0002-5033-8240

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.0c00758


### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

Y.K., Y.B. and L.V.S. acknowledge support of the National Science Foundation (Grants CHE-1450088, CHE-1800505, and IIP-1640646). The research reported in this publication was supported in part by the National Institute of General Medical Sciences of the National Institutes of Health under award R43GM126804. This research was supported in part through computational resources provided by Boston University Shared Computing Cluster and Information Technology at Purdue, West Lafayette, Indiana.

## ■ REFERENCES


(1) Chung, L. W.; Sameera, W. M. C.; Ramozzi, R.; Page, A. J.; Hatanaka, M.; Petrova, G. P.; Harris, T. V.; Li, X.; Ke, Z.; Liu, F.; Li, H.-B.; Ding, L.; Morokuma, K. The ONIOM Method and Its Applications. *Chem. Rev.* **2015**, *115*, 5678−5796.

(2) Maseras, F.; Morokuma, K. IMOMM: A new integratedab initio + molecular mechanics geometry optimization scheme of equilibrium structures and transition states. *J. Comput. Chem.* **1995**, *16*, 1170−1179.

(3) Vreven, T.; Byun, K. S.; Komáromi, I.; Dapprich, S.; Montgomery, J. A.; Morokuma, K.; Frisch, M. J. Combining quantum mechanics methods with molecular mechanics methods in ONIOM. *J. Chem. Theory Comput.* **2006**, *2*, 815−826.

(4) Gordon, M. S.; Freitag, M. A.; Bandyopadhyay, P.; Jensen, J. H.; Kairys, V.; Stevens, W. J. The effective fragment potential method: A QM-based MM approach to modeling environmental effects in chemistry. *J. Phys. Chem. A* **2001**, *105*, 293−307.

(5) Gordon, M. S.; Slipchenko, L. V.; Li, H.; Jensen, J. H., Chapter 10 The Effective Fragment Potential: A General Method for Predicting Intermolecular Interactions. *Annual Reports in Computational Chemistry*; Elsevier, 2007; Vol. 3.

(6) Gresh, N.; Cisneros, G. A.; Darden, T. A.; Piquemal, J.-P. Anisotropic, Polarizable Molecular Mechanics Studies of Inter- and Intramolecular Interactions and Ligand−Macromolecule Complexes. A Bottom-Up Strategy. *J. Chem. Theory Comput.* **2007**, *3*, 1960−1986.

(7) Jensen, J. H.; Day, P. N.; Gordon, M. S.; Basch, H.; Cohen, D.; Garmer, D. R.; Kraus, M.; Stevens, W. J. *Modeling the Hydrogen Bond*; ACS Symposium Series; ACS Publications, 1994; p 139.

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

(9) Dubinets, N.; Slipchenko, L. V. Effective Fragment Potential Method for H-Bonding: How To Obtain Parameters for Nonrigid Fragments. *J. Phys. Chem. A* **2017**, *121*, 5301−5312.

(10) Babu, K.; Gadre, S. R. Ab initio quality one-electron properties of large molecules: Development and testing of molecular tailoring approach. *J. Comput. Chem.* **2003**, *24*, 484−495.

(11) Benson, S. W.; Cruickshank, F. R.; Golden, D. M.; Haugen, G. R.; O'Neal, H. E.; Rodgers, A. S.; Shaw, R.; Walsh, R. Additivity rules for the estimation of thermochemical properties. *Chem. Rev.* **1969**, *69*, 279−324.

(12) Collins, M. A.; Deev, V. A. Accuracy and efficiency of electronic energies from systematic molecular fragmentation. *J. Chem. Phys.* **2006**, *125*, 104104.

(13) Deev, V.; Collins, M. A. Approximateab initioenergies by systematic molecular fragmentation. *J. Chem. Phys.* **2005**, *122*, 154102.

(14) 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.

(15) Li, W.; Li, Y.; Lin, R.; Li, S. Generalized Energy-Based Fragmentation Approach for Localized Excited States of Large Systems. *J. Phys. Chem. A* **2016**, *120*, 9667−9677.

(16) Zhang, D. W.; Xiang, Y.; Zhang, J. Z. H. New Advance in Computational Chemistry: Full Quantum Mechanical ab Initio

Computation of Streptavidin−Biotin Interaction Energy. *J. Phys. Chem. B* **2003**, *107*, 12039−12041.
(17) Bikadi, Z.; Hazai, E. Application of the PM6 semi-empirical method to modeling proteins enhances docking accuracy of AutoDock. *J. Cheminf.* **2009**, *1*, 15.
(18) Dixon, S. L.; Merz, K. M. Semiempirical molecular orbital calculations with linear system size scaling. *J. Chem. Phys.* **1996**, *104*, 6643−6649.
(19) Dixon, S. L.; Merz, K. M. Fast, accurate semiempirical molecular orbital calculations for macromolecules. *J. Chem. Phys.* **1997**, *107*, 879−893.
(20) Nikitina, E.; Sulimov, V.; Zayets, V.; Zaitseva, N. Semiempirical calculations of binding enthalpy for protein-ligand complexes. *Int. J. Quantum Chem.* **2004**, *97*, 747−763.
(21) Stewart, J. J. P. Application of the PM6 method to modeling the solid state. *J. Mol. Model.* **2008**, *14*, 499−535.
(22) Stewart, J. J. P. Application of the PM6 method to modeling proteins. *J. Mol. Model.* **2009**, *15*, 765−805.
(23) Day, P. N.; Jensen, J. H.; Gordon, M. S.; Webb, S. P.; Stevens, W. J.; Krauss, M.; Garmer, D.; Basch, H.; Cohen, D. An effective fragment method for modeling solvent effects in quantum mechanical calculations. *J. Chem. Phys.* **1996**, *105*, 1968−1986.
(24) Gordon, M. S.; Smith, Q. A.; Xu, P.; Slipchenko, L. V. Accurate First Principles Model Potentials for Intermolecular Interactions. *Annu. Rev. Phys. Chem.* **2013**, *64*, 553−578.
(25) Gurunathan, P. K.; Acharya, A.; Ghosh, D.; Kosenkov, D.; Kaliman, I.; Shao, Y.; Krylov, A. I.; Slipchenko, L. V. Extension of the Effective Fragment Potential Method to Macromolecules. *J. Phys. Chem. B* **2016**, *120*, 6562−6574.
(26) Slipchenko, L. V.; Gurunathan, P. K., Effective fragment potential method: Past, present, and future. *Fragmentation: Toward Accurate Calculations on Complex Molecular Systems*; Wiley, 2017; pp 183−208.
(27) Tran, A. L.; Guidez, E. B. Quantum Mechanical Modeling of the Interactions between Noble Metal (Ag and Au) Nanoclusters and Water with the Effective Fragment Potential Method. *ACS Omega* **2020**, *5*, 7446−7455.
(28) Xu, P.; Guidez, E. B.; Bertoni, C.; Gordon, M. S. Perspective:Ab initioforce field methods derived from quantum mechanics. *J. Chem. Phys.* **2018**, *148*, 090901.
(29) Burns, L. A.; Marshall, M. S.; Sherrill, C. D. Appointing silver and bronze standards for noncovalent interactions: A comparison of spin-component-scaled (SCS), explicitly correlated (F12), and specialized wavefunction approaches. *J. Chem. Phys.* **2014**, *141*, 234111.
(30) Jeziorski, B.; Moszynski, R.; Ratkiewicz, A.; Rybak, S.; Szalewicz, K.; Williams, H. L. SAPT: A program for many-body symmetry-adapted perturbation theory calculations of intermolecular interaction energies. *Methods and Techniques in Computational Chemistry: METECC*; STEF, 1993; Vol. *94*, pp 79−129.
(31) Jeziorski, B.; Moszynski, R.; Szalewicz, K. Perturbation Theory Approach to Intermolecular Potential Energy Surfaces of van der Waals Complexes. *Chem. Rev.* **1994**, *94*, 1887−1930.
(32) Parker, T. M.; Burns, L. A.; Parrish, R. M.; Ryno, A. G.; Sherrill, C. D. Levels of symmetry adapted perturbation theory (SAPT). I. Efficiency and performance for interaction energies. *J. Chem. Phys.* **2014**, *140*, 094106.
(33) Flick, J. C.; Kosenkov, D.; Hohenstein, E. G.; Sherrill, C. D.; Slipchenko, L. V. Accurate Prediction of Noncovalent Interaction Energies with the Effective Fragment Potential Method: Comparison of Energy Components to Symmetry-Adapted Perturbation Theory for the S22 Test Set. *J. Chem. Theory Comput.* **2012**, *8*, 2835−2843.
(34) DeFusco, A.; Minezawa, N.; Slipchenko, L. V.; Zahariev, F.; Gordon, M. S. Modeling Solvent Effects on Electronic Excited States. *J. Phys. Chem. Lett.* **2011**, *2*, 2184−2192.
(35) Ghosh, D.; Isayev, O.; Slipchenko, L. V.; Krylov, A. I. Effect of Solvation on the Vertical Ionization Energy of Thymine: From Microhydration to Bulk. *J. Phys. Chem. A* **2011**, *115*, 6028−6038.
(36) Kim, Y.; Morozov, D.; Stadnytskyi, V.; Savikhin, S.; Slipchenko, L. V. Predictive First-Principles Modeling of a Photosynthetic Antenna Protein: The Fenna-Matthews-Olson Complex. *J. Phys. Chem. Lett.* **2020**, *11*, 1636−1643.
(37) Kosenkov, D.; Slipchenko, L. V. Solvent Effects on the Electronic Transitions ofp-Nitroaniline: A QM/EFP Study. *J. Phys. Chem. A* **2011**, *115*, 392−401.
(38) Slipchenko, L. V. Solvation of the Excited States of Chromophores in Polarizable Environment: Orbital Relaxation versus Polarization. *J. Phys. Chem. A* **2010**, *114*, 8824−8830.
(39) Bose, S.; Ghosh, D. An interaction energy driven biased sampling technique: A faster route to ionization spectra in condensed phase. *J. Comput. Chem.* **2017**, *38*, 2248−2257.
(40) Chakraborty, R.; Bose, S.; Ghosh, D. Effect of solvation on the ionization of guanine nucleotide: A hybrid QM/EFP study. *J. Comput. Chem.* **2017**, *38*, 2528−2537.
(41) Ho, J.; Shao, Y.; Kato, J. Do Better Quality Embedding Potentials Accelerate the Convergence of QM/MM Models? The Case of Solvated Acid Clusters. *Molecules* **2018**, *23*, 2466.
(42) Nagata, T.; Fedorov, D. G.; Kitaura, K.; Gordon, M. S. A combined effective fragment potential-fragment molecular orbital method. I. The energy expression and initial applications. *J. Chem. Phys.* **2009**, *131*, 024101.
(43) Nagata, T.; Fedorov, D. G.; Sawada, T.; Kitaura, K.; Gordon, M. S. A combined effective fragment potential-fragment molecular orbital method. II. Analytic gradient and application to the geometry optimization of solvated tetraglycine and chignolin. *J. Chem. Phys.* **2011**, *134*, 034110.
(44) Steinmann, C.; Fedorov, D. G.; Jensen, J. H. Effective Fragment Molecular Orbital Method: A Merger of the Effective Fragment Potential and Fragment Molecular Orbital Methods†. *J. Phys. Chem. A* **2010**, *114*, 8705−8712.
(45) Steinmann, C.; Fedorov, D. G.; Jensen, J. H. The Effective Fragment Molecular Orbital Method for Fragments Connected by Covalent Bonds. *PLoS One* **2012**, *7*, No. e41117.
(46) Nagata, T.; Fedorov, D. G.; Kitaura, K.; Gordon, M. S. A combined effective fragment potential-fragment molecular orbital method. I. The energy expression and initial applications. *J. Chem. Phys.* **2009**, *131*, 024101.
(47) Nagata, T.; Fedorov, D. G.; Sawada, T.; Kitaura, K. Analysis of Solute-Solvent Interactions in the Fragment Molecular Orbital Method Interfaced with Effective Fragment Potentials: Theory and Application to a Solvated Griffithsin-Carbohydrate Complex. *J. Phys. Chem. A* **2012**, *116*, 9088−9099.
(48) Pruitt, S. R.; Steinmann, C.; Jensen, J. H.; Gordon, M. S. Fully Integrated Effective Fragment Molecular Orbital Method. *J. Chem. Theory Comput.* **2013**, *9*, 2235−2249.
(49) Steinmann, C.; Fedorov, D. G.; Jensen, J. H. Mapping Enzymatic Catalysis Using the Effective Fragment Molecular Orbital Method: Towards all ab initio Biochemistry. *PLoS One* **2013**, *8*, No. e60602.
(50) Zahariev, F.; Gordon, M. S. Development of a combined quantum monte carlo-effective fragment molecular orbital method. *Mol. Phys.* **2019**, *117*, 1532−1540.
(51) Stone, A. J. *The Theory of Intermolecular Forces*; Oxford University Press: Oxford, 1996.
(52) Adamovic, I.; Gordon, M. S. Dynamic polarizability, dispersion coefficient C6and dispersion energy in the effective fragment potential method. *Mol. Phys.* **2005**, *103*, 379−387.
(53) Amos, R. D.; Handy, N. C.; Knowles, P. J.; Rice, J. E.; Stone, A. J. AB-initio prediction of properties of carbon dioxide, ammonia, and carbon dioxide...ammonia. *J. Phys. Chem.* **1985**, *89*, 2186−2192.
(54) Slipchenko, L. V.; Gordon, M. S.; Ruedenberg, K. Dispersion Interactions in QM/EFP. *J. Phys. Chem. A* **2017**, *121*, 9495−9507.
(55) Smith, Q. A.; Ruedenberg, K.; Gordon, M. S.; Slipchenko, L. V. The dispersion interaction between quantum mechanics and effective fragment potential molecules. *J. Chem. Phys.* **2012**, *136*, 244107.
(56) Alkan, M.; Xu, P.; Gordon, M. S. Many-Body Dispersion in Molecular Clusters. *J. Phys. Chem. A* **2019**, *123*, 8406−8416.


https://dx.doi.org/10.1021/acs.jctc.0c00758
*J. Chem. Theory Comput.* 2020, 16, 7735−7747

(57) Yu, K.; Schmidt, J. R. Many-body effects are essential in a physically motivated CO2 force field. *J. Chem. Phys.* **2012**, *136*, 034503.
(58) Marom, N.; DiStasio, R. A., Jr.; Atalla, V.; Levchenko, S.; Reilly, A. M.; Chelikowsky, J. R.; Leiserowitz, L.; Tkatchenko, A. Many-Body Dispersion Interactions in Molecular Crystal Polymorphism. *Angew. Chem., Int. Ed.* **2013**, *52*, 6629−6632.
(59) Stöhr, M.; Tkatchenko, A. Quantum mechanics of proteins in explicit water: The role of plasmon-like solute-solvent interactions. *Sci. Adv.* **2019**, *5*, No. eaax0024.
(60) Jensen, J.; Gordon, M. An approximate formula for the intermolecular Pauli repulsion between closed shell molecules. *Mol. Phys.* **1996**, *89*, 1313−1325.
(61) Jensen, J. H.; Gordon, M. S. An approximate formula for the intermolecular Pauli repulsion between closed shell molecules. II. Application to the effective fragment potential method. *J. Chem. Phys.* **1998**, *108*, 4772−4782.
(62) Kemp, D. D.; Rintelman, J. M.; Gordon, M. S.; Jensen, J. H. Exchange repulsion between effective fragment potentials and ab initio molecules. *Theor. Chem. Acc.* **2010**, *125*, 481−491.
(63) Rojas, C. I. V.; Fine, J.; Slipchenko, L. V. Exchange-repulsion energy in QM/EFP. *J. Chem. Phys.* **2018**, *149*, 094103.
(64) Brautigam, C. A.; Smith, B. S.; Ma, Z.; Palnitkar, M.; Tomchick, D. R.; Machius, M.; Deisenhofer, J. Structure of the photolyase-like domain of cryptochrome 1 from Arabidopsis thaliana. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 12142−12147.
(65) Tazhigulov, R. N.; Gurunathan, P. K.; Kim, Y.; Slipchenko, L. V.; Bravaya, K. B. Polarizable embedding for simulating redox potentials of biomolecules. *Phys. Chem. Chem. Phys.* **2019**, *21*, 11642−11650.
(66) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. General atomic and molecular electronic structure system. *J. Comput. Chem.* **1993**, *14*, 1347−1363.
(67) Hands, M. D.; Slipchenko, L. V. Intermolecular Interactions in Complex Liquids: Effective Fragment Potential Investigation of Water-tert-Butanol Mixtures. *J. Phys. Chem. B* **2012**, *116*, 2775−2786.
(68) Rankin, B. M.; Hands, M. D.; Wilcox, D. S.; Fega, K. R.; Slipchenko, L. V.; Ben-Amotz, D. Interactions between halide anions and a molecular hydrophobic interface. *Faraday Discuss.* **2013**, *160*, 255−270.
(69) Slipchenko, L. V.; Gordon, M. S. Damping functions in the effective fragment potential method. *Mol. Phys.* **2009**, *107*, 999−1016.
(70) Parker, T. M.; Burns, L. A.; Parrish, R. M.; Ryno, A. G.; Sherrill, C. D., Levels of symmetry adapted perturbation theory (SAPT). I. Efficiency and performance for interaction energies. *J. Chem. Phys.* **2014**, *140*. DOI: 10.1063/1.4867135
(71) Turney, J. M.; Simmonett, A. C.; Parrish, R. M.; Hohenstein, E. G.; Evangelista, F. A.; Fermann, J. T.; Mintz, B. J.; Burns, L. A.; Wilke, J. J.; Abrams, M. L.; Russ, N. J.; Leininger, M. L.; Janssen, C. L.; Seidl, E. T.; Allen, W. D.; Schaefer, H. F.; King, R. A.; Valeev, E. F.; Sherrill, C. D.; Crawford, T. D. PSI4: an open-source ab initio electronic structure program. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012**, *2*, 556−565.
(72) Kaliman, I. A.; Slipchenko, L. V. LIBEFP: A new parallel implementation of the effective fragment potential method as a portable software library. *J. Comput. Chem.* **2013**, *34*, 2284−2292.
(73) Kaliman, I. A.; Slipchenko, L. V. Hybrid MPI/OpenMP parallelization of the effective fragment potential method in thelibefpsoftware library. *J. Comput. Chem.* **2015**, *36*, 129−135.
(74) Parrish, R. M.; Sitkoff, D. F.; Cheney, D. L.; Sherrill, C. D. The Surprising Importance of Peptide Bond Contacts in Drug-Protein Interactions. *Chem.—Eur. J.* **2017**, *23*, 7887−7890.
(75) Herbers, S.; Fritz, S. M.; Mishra, P.; Kim, Y.; Slipchenko, L.; Zwier, T. S. The unusual symmetry of hexafluoro-o-xylene-A microwave spectroscopy and computational study. *J. Chem. Phys.* **2020**, *152*, 064302.


https://dx.doi.org/10.1021/acs.jctc.0c00758
*J. Chem. Theory Comput.* 2020, 16, 7735−7747