THE JOURNAL OF PHYSICAL CHEMISTRY A

Article

pubs.acs.org/JPCA

Multiple Environment Single System Quantum Mechanical/Molecular Mechanical (MESS-QM/MM) Calculations. 1. Estimation of Polarization Energies

Alexander J. Sodt,[†] Ye Mei,[†,‡,§] Gerhard König,[†] Peng Tao,[∥] Ryan P. Steele,[⊥] Bernard R. Brooks,[†] and Yihan Shao*[,#]

[†]Laboratory of Computational Biology, National Heart, Lung and Blood Institute, National Institutes of Health, 5635 Fishers Lane, T-900 Suite, Rockville, Maryland 20852, United States

[‡]Center for Laser and Computational Biophysics, State Key Laboratory of Precision Spectroscopy, Department of Physics and Institute of Theoretical and Computational Science, East China Normal University, Shanghai 200062, China

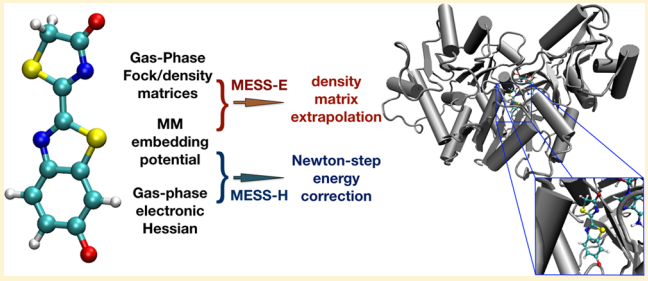
[§]NYU-ECNU Center for Computational Chemistry at NYU Shanghai, Shanghai 200062, China

[∥]Department of Chemistry, Southern Methodist University, 3215 Daniel Avenue, Dallas, Texas 75275, United States

[⊥]Department of Chemistry, University of Utah, 315 S 1400 E, Salt Lake City, Utah 84112, United States

[#]Q-Chem Inc., 6601 Owens Drive, Suite 105, Pleasanton, California 94588, United States

ABSTRACT: In combined quantum mechanical/molecular mechanical (QM/MM) free energy calculations, it is often advantageous to have a frozen geometry for the quantum mechanical (QM) region. For such multiple-environment single-system (MESS) cases, two schemes are proposed here for estimating the polarization energy: the first scheme, termed MESS-E, involves a Roothaan step extrapolation of the self-consistent field (SCF) energy; whereas the other scheme, termed MESS-H, employs a Newton−Raphson correction using an approximate inverse electronic Hessian of the QM region (which is constructed only once). Both schemes are extremely efficient, because the expensive Fock updates and SCF iterations in standard QM/MM calculations are completely avoided at each configuration. They produce reasonably accurate QM/MM polarization energies: MESS-E can predict the polarization energy within 0.25 kcal/mol in terms of the mean signed error for two of our test cases, solvated methanol and solvated β-alanine, using the M06-2X or ωB97X-D functionals; MESS-H can reproduce the polarization energy within 0.2 kcal/mol for these two cases and for the oxyluciferin−luciferase complex, if the approximate inverse electronic Hessians are constructed with sufficient accuracy.



I. INTRODUCTION

In the last two decades, combined quantum mechanical molecular mechanical (QM/MM) calculations have become increasingly popular in the computational study of molecular solvation, catalytic or enzymatic reactions,[1−7] and ligand binding.[8,9] In these calculations, typically the solute, the reactive region, or the active site is studied by quantum mechanics (QM) methods, whereas the environment (solvent molecules or the non-QM portion of the macromolecule) is described with molecular mechanics (MM) force fields.

In pure QM calculations, it is often sufficient to locate a few stationary points (local minima and saddle points) on the potential energy surface.[10,11] When it comes to QM/MM potential energy surfaces, such stationary points can still provide some useful insights into the reaction mechanisms or into the chemical nature of ligand−receptor binding. But only a small number of degrees of freedom (that are predominantly associated with QM atoms) connect these stationary points to each other. At a finite temperature, many other degrees of freedom are also accessible and can thus have non-negligible contributions to the free energy of interest, such as the solvation free energy, reaction free energy, or binding free energy. So it is essential to adequately sample all the relevant degrees of freedom.

The sampling of these QM and MM degrees of freedom can be achieved via QM/MM molecular dynamics simulations, which in practice fall into at least two general categories. In the first category, one performs conventional simulations using QM/MM forces throughout the trajectory, meaning that the QM/MM energy and gradient have to be computed at every

Special Issue: 25th Austin Symposium on Molecular Structure and Dynamics

Received: July 19, 2014
Revised: October 15, 2014
Published: October 16, 2014

dx.doi.org/10.1021/jp5072296 | J. Phys. Chem. A 2015, 119, 1511−1523

geometry step. Though such QM/MM MD simulations have proven to yield higher accuracy than classical MM simulations, they are, unfortunately, also orders of magnitude more expensive. For the three test cases in this work, for example, a single QM/MM MD step (using the B3LYP/6-31+G* level of theory for the QM region) would take 50 times longer than a MM step for solvated methanol, 400 times longer for solvated β-alanine, and 4500 times longer for oxyluciferin in its luciferase complex. For many applications, where an even larger QM region and/or a more sophisticated QM method are desired, a conventional QM/MM MD simulation remains quite expensive, even after much recent and continuing work on algorithmic developments to speed up QM calculations.[12−15]

In the second category, one performs "indirect" QM/MM simulations.[16−22] In these simulations, the conformational space is sampled using a MM force field (and thus with a higher computational speed). Then a subset of configurations along the MM trajectory are subjected to QM/MM energy/gradient evaluations, to account for the free energy differences between QM/MM and MM surfaces. This strategy works best if the QM/MM configurations are well represented by the MM ensemble, which requires considerable "overlap" between the QM/MM potential energy surface and the MM potential energy surface. In cases where the overlap is insufficient, the free energy can either fail to converge or converge to incorrect values. To alleviate this problem, several groups have proposed "freezing" some or all internal degrees of freedom within the QM region[16−18,23,24] during the MM simulations.

Such freezing of QM degrees of freedom has shown to improve the convergence of free energy calculations, but it might also lose accuracy due to the omission of any entropic contributions in the QM region.[25,26] This deficiency can potentially be mitigated with the introduction of three separate energy corrections: (a) an enthalpic correction which corresponds to calculating the shift of the energy minimum due to the constraints on the QM degrees of freedom; (b) a vibrational entropic correction based on a subsystem Hessian[27] (the same Hessian as in the enthalpic term); (c) a Jacobian term. These energy corrections are discussed in detail in separate publications.[22,28] In addition, systems with multiple rotational states can be addressed using the technique developed by Straatsma and McCammon.[29]

In this work, we shall focus on "indirect" QM/MM simulations where all internal degrees of freedom in the QM region are frozen in the MM simulations. In other words, the QM region simply becomes a rigid body. The subset of configurations coming out of this MM trajectory share a single QM geometry whereas the MM environment can be quite different. The main task then is to compute the QM/MM energy for such "multiple environment single system" (MESS) configurations. We shall show in this work that, when the QM method in use is Kohn−Sham density functional theory (DFT),[30−35] it is possible to estimate such MESS-QM/MM energies with a small computational cost and a reasonably high accuracy.

This article is structured as follows. In section 2, the underlying theory is formulated. The MESS-QM/MM energy is divided into three terms: (a) a gas-phase energy; (b) a first-order term, which describes the interaction of the gas-phase electron density and the MM electrostatic potential; (c) the QM/MM polarization energy, which corresponds to the second-order and higher terms. In the same section, two different schemes will be proposed for estimating the QM/MM polarization energy: MESS-E, which is based on a Roothaan step extrapolation, and MESS-H, which is based on a Newton−Raphson energy correction using an inverse electronic Hessian in a subspace representation. In section III, technical details will be provided about the implementation of these two schemes for MESS-QM/MM polarization energy estimation and about our test systems. In section IV, the results will be presented and discussed. Concluding remarking are made in section V.

II. THEORY

A. QM/MM Polarization Energy. *1. Gas-Phase Electronic Structure.* Let us consider a QM system that consists of N_{QM} atoms and is subjected to an external potential $v(\mathbf{r})$. Within the Kohn−Sham density functional theory using pure, hybrid or range-separated functionals, the total energy of such a system is given by a functional of the one-particle density matrix

$$E = E(\rho(\mathbf{r},\mathbf{r}')) \tag{1}$$

For pure functionals, this energy functional involves only the diagonal part of the one particle density matrix, namely the electron density $\rho(\mathbf{r})$. Hybrid functionals and range-separated functionals involve the off-diagonal elements as well through the Hartree−Fock exchange energy contributions.

From a set of orthonormal Kohn−Sham molecular orbitals (MO), ψ_p, which are often approximated as linear combinations of atom-centered basis functions, ϕ_μ,

$$\psi_p(\mathbf{r}) = \sum_\mu C_{\mu p}\phi_\mu(\mathbf{r}) \tag{2}$$

one can write the one-particle density matrix as

$$\rho(\mathbf{r},\mathbf{r}') = \sum_{i\in\mathrm{occ}} \psi_i(\mathbf{r})\,\psi_i(\mathbf{r}') = \sum_{\mu\nu} P^{\mu\nu}\phi_\mu(\mathbf{r})\,\phi_\nu(\mathbf{r}') \tag{3}$$

Here $P^{\mu\nu}$ is

$$P^{\mu\nu} = \sum_{i\in\mathrm{occ}} C_{\mu i}C_{\nu i} \tag{4}$$

and is often also referred to as the density matrix.

In a DFT calculation, one optimizes the MOs in eq 2 via a self-consistent field (SCF) iterative procedure to obtain the lowest energy for the system in eq 1. Within each iteration, from the current set of orthonormal MOs, one rotates the orbitals

$$\psi_i \to \psi_i + \sum_a \Theta_{ai}\psi_a \tag{5}$$

$$\psi_a \to \psi_a - \sum_i \Theta_{ai}\psi_i \tag{6}$$

where the unoccupied MOs, ψ_a, are mixed into the occupied ones, ψ_i. The actual free variables in the energy functional in eq 1 are thus the orbital rotations,

$$E = E(C,\{\Theta\}) \tag{7}$$

The iterative procedure continues until it reaches a set of optimized MOs that make the energy stationary with respect to the orbital rotations

$$\frac{\partial E}{\partial \Theta_{ai}} = 0 \tag{8}$$

In the gas phase, the external potential $v(\mathbf{r})$ only includes the nuclear attraction,

dx.doi.org/10.1021/jp5072296 | J. Phys. Chem. A 2015, 119, 1511−1523

$$v^{(0)}(\mathbf{r}) = \sum_{A=1}^{N_{QM}} \frac{Z_A}{|\mathbf{R}_A - \mathbf{r}|} \tag{9}$$

which is summed over all nuclei, A, with charges Z_A. This potential can be represented in the atomic basis

$$h_{\mu\nu}^{(0)} = \int \phi_\mu(\mathbf{r})\, v^{(0)}(\mathbf{r})\, \phi_\nu(\mathbf{r})\, d\mathbf{r} \tag{10}$$

The MO coefficients optimized under this gas phase external electrostatic potential will be denoted by $C^{(0)}$, and the corresponding gas-phase density matrix by

$$P^{\mu\nu,(0)} = \sum_{i\in occ} C_{\mu i}^{(0)} C_{\nu i}^{(0)} \tag{11}$$

and the Fock matrix, $F^{\mu\nu} = \partial E/\partial P^{\mu\nu}$, by $F^{(0)}$. The gas-phase Kohn−Sham energy is

$$E^{(0)} = E(C^{(0)},\{\Theta=0\}) \tag{12}$$

and the gas-phase electronic density is

$$\rho^{(0)}(\mathbf{r}) = \sum_{\mu\nu} P^{\mu\nu,(0)} \phi_\mu(\mathbf{r})\, \phi_\nu(\mathbf{r}) \tag{13}$$

2. MM Perturbation. In QM/MM calculations, the external potential includes both the nuclear attraction, $v^{(0)}(\mathbf{r})$ in eq 9, and an additional electrostatic potential due to a set of N_{MM} MM atoms,

$$v(\mathbf{r}) = v^{(0)}(\mathbf{r}) + \delta v(\mathbf{r}) \tag{14}$$

In the simplest cases (as in the examples in section IV), where the MM atoms are represented by point charges, $\{q_B, B = 1, ..., N_{MM}\}$, the additional potential is

$$\delta v(\mathbf{r}) = \sum_{B=1}^{N_{MM}} \frac{q_B}{|\mathbf{R}_B - \mathbf{r}|} \tag{15}$$

Its matrix representation is

$$\delta h_{\mu\nu} = \int \phi_\mu(\mathbf{r})\, \delta v(\mathbf{r})\, \phi_\nu(\mathbf{r})\, d\mathbf{r} \tag{16}$$

This external electrostatic potential from the MM atoms perturbs the energy of the system from $E^{(0)}$, its gas-phase value:

$$E = E^{(0)} + \delta E^{(1)} + \delta E^{(2)} + \cdots \tag{17}$$

where the first-order change is a simple sum of electronic and nuclear terms:

$$\begin{aligned}\delta E^{(1)} &= -\int \rho^{(0)}(\mathbf{r})\, \delta v(\mathbf{r})\, d\mathbf{r} + \sum_A Z_A \delta v(\mathbf{A}) \\ &= -\sum_{\mu\nu} P^{\mu\nu,(0)} \delta h_{\mu\nu} + \sum_A Z_A \delta v(\mathbf{A})\end{aligned} \tag{18}$$

where the minus signs reflect the negative charge from electrons. *To compute this first-order energy change, the dominant computational cost comes from the evaluation of $\delta h_{\mu\nu}$ in eq 16.*

The second-order change, $\delta E^{(2)}$, plus all higher order terms is the QM/MM polarization energy. It is directly caused by the polarization of the Kohn−Sham orbitals (and therefore the one-particle density matrix) by the MM electrostatic potential. Essentially, the Kohn−Sham orbitals will relax within the additional MM electrostatic potential in eq 15, stabilize the system, and lower the energy. In the next two subsections, we will describe two approximate schemes, MESS-E and MESS-H, for estimating $\delta E^{(2)}$, the second-order term of the QM/MM polarization energy.

B. Scheme MESS-E: Roothaan Step Extrapolation. Similar to the dual-basis projection technique for DFT and MP2 calculations,[36−38] the three extrapolations in density functional triple jumping,[39] and the Roothaan step correction in the absolutely localized molecule orbital (ALMO) based energy decomposition analysis,[40,41] one Roothaan step[42] is taken in the MESS-E scheme.

Here we write the perturbed Fock matrix as

$$F^{(1)} = F^{(0)} + \delta h \tag{19}$$

where $F^{(0)}$ is the gas-phase Fock matrix, and δh is the atomic basis representation of the electrostatic potential due to MM atoms (see eq 16). In a Roothaan step, the perturbed Fock matrix is diagonalized

$$F^{(1)}C^{(1)} = SC^{(1)}\epsilon^{(1)} \tag{20}$$

where S is the overlap matrix

$$S_{\mu\nu} = \int \phi_\mu(\mathbf{r})\, \phi_\nu(\mathbf{r})\, d\mathbf{r} \tag{21}$$

From the eigenvectors, which are the new set of MO coefficients, $C^{(1)}$, one constructs a new density matrix

$$P^{\mu\nu,(1)} = \sum_{i\in occ} C_{\mu i}^{(1)} C_{\nu i}^{(1)} \tag{22}$$

Then one can approximate the QM/MM polarization energy as a Roothaan step extrapolation

$$\delta E_{\text{MESS-E}}^{(2)} = F^{(1)}\cdot\delta P^{(1)} = F^{(0)}\cdot\delta P^{(1)} + \delta h\cdot\delta P^{(1)} \tag{23}$$

where $\delta P^{(1)}$ is the projected response in the density matrix

$$\delta P^{(1)} = P^{\mu\nu,(1)} - P^{\mu\nu,(0)} \tag{24}$$

The leading contributions to the MESS-E energy, eq 23, are clearly second-order: in the $F^{(0)}\cdot\delta P^{(1)}$ term, the gas-phase Fock matrix, $F^{(0)}$, can only couple to the occupied−occupied (oo) and virtual−virtual (vv) blocks of $\delta P^{(1)}$, both of which are quadratic in orbital response and thus in δh. In the $\delta h\cdot\delta P^{(1)}$ term, δh couples to the occupied-virtual (ov) and virtual-occupied (vo) blocks of $\delta P^{(1)}$, both of which are first order in orbital response (and thus in δh).

With the $F^{(0)}\cdot\delta P^{(1)}$ term, one removes a fractional number of electrons from occupied Kohn−Sham orbitals and puts them into (higher-energy) virtual orbitals. As a result, this term is always positive. To achieve a net stabilization of the QM system within the MM potential, the $\delta h\cdot\delta P^{(1)}$ term must be negative and have a larger absolute value. In our test cases, $\delta h\cdot\delta P^{(1)}$ was found to be roughly twice as large as $F^{(0)}\cdot\delta P^{(1)}$ (with an opposite sign).

It should be noted that, *in the computation of this second-order energy change for each MM environment, one avoids the expensive Fock builds and only performs a diagonalization of the perturbed Fock matrix* (eq 20) and a *few simple matrix operations* (eqs 19, 22 and 23).

C. Scheme MESS-H: Newton−Raphson Correction with an Approximate Hessian. As shown in eq 8, the gradient of the gas-phase Kohn−Sham energy with respect to orbital rotations vanishes at the set of optimized gas-phase Kohn−Sham orbitals. When the MM electrostatic potential in eq 15 is applied, the gradient is no longer zero:

dx.doi.org/10.1021/jp5072296 | J. Phys. Chem. A 2015, 119, 1511−1523

$$\frac{\partial E}{\partial \Theta_{ai}} = \delta h_{ai} \tag{25}$$

where δh_{ai} is the MO representation of the MM potential

$$\delta h_{ai} = \int \psi_a^{(0)}(\mathbf{r})\, \delta v(\mathbf{r})\, \psi_i^{(0)}(\mathbf{r})\, \mathrm{d}\mathbf{r} = \sum_{\mu\nu} C_{\mu a}^{(0)} \delta h_{\mu\nu} C_{\nu i}^{(0)} \tag{26}$$

If the electronic Hessian of the gas-phase energy, i.e., the set of second derivatives with respect to orbital rotations, $H_{ai,bj}^{(0)}$, is readily available, one can use its inverse, $H_{ai,bj}^{(0),-1}$, to approximate the QM/MM polarization energy with a Newton−Raphson energy correction

$$\begin{aligned} \delta E_{\text{MESS-H}}^{(2)} &= -\frac{1}{2} \delta h \cdot H^{(0),-1} \cdot \delta h \\ &= -\frac{1}{2} \sum_{ai,bj} \delta h_{ai} (H^{(0),-1})_{ai,bj} \delta h_{bj} \end{aligned} \tag{27}$$

which corresponds to a shift of the energy minimum with the following orbital response due to the external MM potential,

$$\delta \Theta_{ai} = -\sum_{bj} (H^{(0),-1})_{ai,bj} \delta h_{bj} \tag{28}$$

Here one only needs to compute the gas-phase electronic Hessian once and then store its inverse in memory or disk. In subsequent MESS-H calculations, for each MM environment, one simply reads in the inverse Hessian back from the storage and performs two simple matrix operations (eqs 26 and 27). We note that the inverse electronic Hessian is closely related to the linear response kernel in DFT, $(\delta^2 E)/(\delta v(\mathbf{r})\, \delta v(\mathbf{r}'))$, which has been studied analytically and numerically.[43−54]

One can and should avoid computing the exact gas-phase electronic Hessian, $H^{(0)}$, for three practical considerations: (a) it might be too expensive to compute the exact Hessian for systems with a large number of QM atoms and/or a large basis set, (b) it might be unnecessary to compute the exact value of the second-order term whereas higher-order terms are completely neglected, and most importantly, (c) *the commonly used stability analysis,[55] which usually involves using Davidson's iterative subspace method[56,57] to solve for the lowest eigenvalues of the electronic Hessian, provides a practical way to approximate the gas-phase electronic Hessian and therefore its inverse within a subspace representation.* Namely, if the Davidson procedure yields the lowest m eigenvectors, $HU_m = U_m \lambda_m$, we can approximate the inverse Hessian as $\sum_m (\lambda_m)^{-1} U_m U_m^{\mathrm{T}}$. Of course, this is based on the positive definiteness of a gas-phase electronic Hessian (assuming that the gas-phase Kohn−Sham energy is indeed optimized to its minimum), implying smaller and smaller contributions to the inverse Hessian from the eigenvectors as their eigenvalues increase.

III. COMPUTATIONAL DETAILS

The two schemes for estimating MESS QM/MM polarization energies, MESS-E and MESS-H, are implemented within the Q-Chem/CHARMM interface,[58] using version c38b2 of CHARMM[59] and a development version of Q-Chem 4.2.[60,61]

Three test cases employed in this work are shown in Figure 1:

- methanol solvated in a cubic box with 1000 water molecules
- β-alanine solvated in a cubic box with 1000 water molecules
- the oxyluciferin−luciferase complex[62−64]

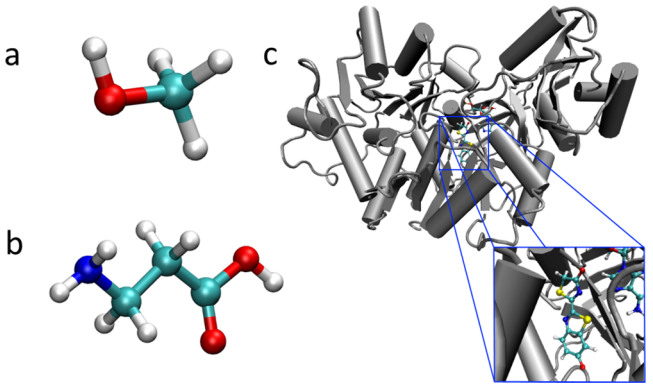


Figure 1. Three test systems in this work: (a) methanol and (b) β-alanine, both solvated in a cubic box of classical water molecules, and (c) oxyluciferin in its complex with luciferase. In the luciferase binding pocket, there is also an AMP molecule.

For the first two test cases, the solute molecules (methanol or β-alanine) are described with DFT methods (B3LYP,[65−67] M06-2X,[68] and ωB97X-D[69] functionals with 6-31+G* and 6-311++G** basis sets[70,71]) and water molecules with TIP3P model.[72] Constant-volume rigid body molecular dynamics simulations are performed for both systems at 298 K for 100 ps for an initial equilibration, and then for 1 ns during which configurations are collected at 1 ps intervals, leading to 1000 configurations for both systems.

For the oxyluciferin−luciferase complex, we started with the 2D1R protein data bank structure,[62] which includes 540 protein residues, the oxyluciferin substrate, and an AMP molecule also in the binding pocket. 626 water molecules are included to solvate the protein surface. The oxyluciferin substrate is also described with DFT methods, whereas the protein residues and AMP (10 307 atoms in total) are described with CHARMM c22 protein force field and CGenFF force field.[73] Constant-volume rigid body molecular dynamics simulations are performed for both systems at 298 K for 10 ps for an initial equilibration, and then for 100 ps. With one configuration collected every 1 ps, this leads to 100 configurations in total.

In all these calculations, the core region (methanol, β-alanine, or oxyluciferin) is constrained as a rigid body through the "CONS FIX" command in CHARMM, or can alternatively be enforced through the "SHAPE" command.[74] As a result, the core region retains the same geometry for all 100 or 1000 configurations used in subsequent MESS-E or MESS-H polarization energy estimations.

A "MESS" option is added to the Q-Chem interface in CHARMM, where "NROOTS n" is used to specify how many lowest eigenvalues/eigenvectors of the stability matrix (i.e., electronic Hessian) to be computed and used for constructing its inverse. The inverse electronic Hessian is constructed only once for the gas-phase density and saved to disk at the starting of the MESS-QM/MM calculation. For each configuration, we evaluate the MM potential for the QM atoms (eq 16), and then compute MESS-E and MESS-H estimations (the latter uses the inverse Hessian read in from disk) for the QM/MM polarization energy. To assess the accuracy of MESS-E and MESS-H energies, we also perform SCF calculations to fully converge DFT energies for each configuration. This is done

dx.doi.org/10.1021/jp5072296 | J. Phys. Chem. A 2015, 119, 1511−1523

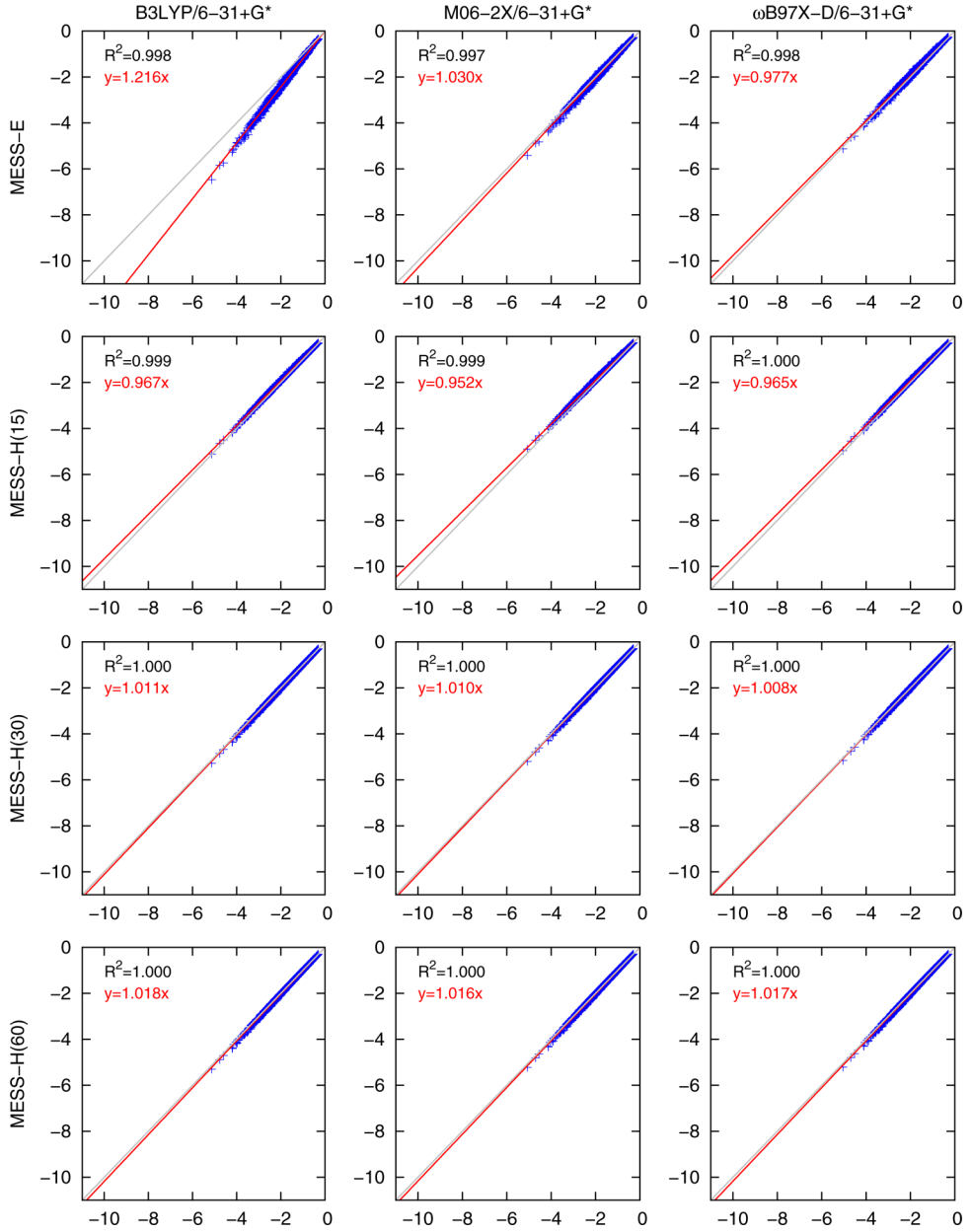


Figure 2. Estimated MESS-QM/MM polarization energies (*y*-axis) versus the exact values (*x*-axis) for 1000 configurations of a single methanol molecule solvated in a cubic box of 1000 water molecules. All energies are in kcal/mol. The MESS-H(15), MESS-H(30), and MESS-H(60) estimations started by solving iteratively for 15, 30, or 60 eigenvalues/eigenvectors of the electronic Hessian. B3LYP results are shown in the left four panels, M06-2X results in the middle, and ωB97X-D results in the right four panels. The gray lines correspond to a hypothetical perfect estimation. The red lines are linear fits of the estimated energy values against the exact values, with R^2 values and slopes listed at the corner of each panel.

only for the purpose of comparison, and it will be avoided in actual MESS-QM/MM calculations.

IV. RESULTS AND DISCUSSION

A. Solvated Methanol. The results for the solvated methanol systems are shown in Figure 2 and Table 1. The QM/MM polarization energy are confirmed to be all negative (i.e., stabilizing the system) and, with a range of −5 to 0 kcal/mol, its contribution to the total energy are thus non-negligible.

With all three functionals (B3LYP, M06-2X, and ωB97X-D), the QM/MM polarization energy estimated with the MESS-E scheme, as shown in Figure 2, correlates very well with the exact values, with R^2 values greater than 0.996. The fitted slope with the B3LYP calculations in Figure 2 is 1.216, suggesting that MESS-E tends to overestimate the polarization energy by roughly 20%, which corresponds to a 20.6% relative error (REL) and a −0.386 kcal/mol mean signed error (MSE) in Table 1. The maximum error (MAX) is 1.339 kcal/mol, which is slightly above the desired chemical accuracy of 1 kcal/mol. In contrast, the MESS-E scheme only overestimates the M06-2X values by 3% (slope 1.030, REL 3.3%, MSE −0.047 kcal/mol) and actually slightly overestimates the ωB97X-D values by 3−4% (slope 0.977, REL 3.8%, MSE 0.048 kcal/mol). It is very encouraging that, without performing even a single Fock build at each configuration, MESS-E can reproduce the M06-2X and ωB97X-D polarization energies with an average error less than 0.1 kcal/mol and a maximum error of 0.2−0.4 kcal/mol (0.358 kcal/mol with M06-2X and 0.209 kcal/mol with ωB97X-D).

dx.doi.org/10.1021/jp5072296 | J. Phys. Chem. A 2015, 119, 1511−1523

Table 1. Errors in the Estimated QM/MM Polarization Energy for a Methanol Molecule Solvated in a Cubic Box of 1000 Water Molecules[a]

		6-31+G* basis				6-311++G** basis			
functional	scheme	MSE	RMS	MAX	REL	MSE	RMS	MAX	REL
B3LYP	MESS-E	−0.386	0.432	1.339	20.6%	−0.420	0.470	1.459	21.2%
	MESS-H(15)	0.066	0.071	0.173	4.0%	0.101	0.107	0.234	5.7%
	MESS-H(30)	−0.016	0.033	0.176	1.0%	0.019	0.032	0.112	1.5%
	MESS-H(60)	−0.029	0.044	0.216	1.4%	−0.030	0.048	0.241	1.5%
M06-2X	MESS-E	−0.047	0.084	0.358	3.3%	−0.078	0.113	0.462	4.3%
	MESS-H(15)	0.090	0.097	0.230	5.2%	0.140	0.150	0.356	7.7%
	MESS-H(30)	−0.014	0.032	0.170	1.0%	0.028	0.039	0.118	1.9%
	MESS-H(60)	−0.025	0.041	0.205	1.3%	−0.026	0.043	0.218	1.3%
ωB97X-D	MESS-E	0.048	0.071	0.209	3.8%	0.048	0.075	0.223	3.7%
	MESS-H(15)	0.067	0.072	0.173	4.0%	0.095	0.103	0.247	5.4%
	MESS-H(30)	−0.010	0.030	0.151	1.0%	0.014	0.031	0.122	1.4%
	MESS-H(60)	−0.027	0.042	0.208	1.3%	−0.028	0.046	0.229	1.4%

[a]Mean signed errors (MSE), root mean square errors (RMS), maximum errors (MAX), all in kcal/mol, in the QM/MM polarization energies estimated with MESS-E and MESS-H schemes are listed, together with the average of the relative unsigned error (REL). For the MESS-H calculations, the number within the parentheses refers to the number of eigenvectors used in the construction of the inverse of the electronic Hessian. Averaged over 1000 configurations from a 1 ns trajectory.

When applying the MESS-H scheme to the solvated methanol, we first solved for 15, 30, or 60 eigenvectors of the electronic Hessian of the gas-phase molecule. We will refer to these calculations as MESS-H(15), MESS-H(30), and MESS-H(60) in subsequent discussions. As shown in Figure 2, the estimated MESS-H energies all correlate well with the exact values, with R^2 values greater than 0.998. MESS-H(15) slightly underestimates the polarization energy by 4−5%, with MSE values of 0.066 kcal/mol (B3LYP), 0.090 kcal/mol (M06-2X), and 0.067 kcal/mol (ωB97X-D). MESS-H(30) has a higher initial cost, because it requests twice as many eigenvectors for the gas-phase inverse Hessian construction, and it also leads to more accurate energies. MESS-H(30) energies are only 1% larger than the exact values, and the MSE values are −0.016 kcal/mol (B3LYP), −0.014 kcal/mol (M06-2X), and −0.010 kcal/mol (ωB97X-D). However, the results do not improve further upon requesting even more eigenvectors. As shown by the slopes in Figure 2 and the error values in Table 1, MESS-H(60) actually leads to slightly larger errors than MESS-H(30). Overall, all these MESS-H models led to highly accurate energies with all three functionals, with MSE errors smaller than 0.1 kcal/mol, and maximum errors of 0.15−0.23 kcal/mol.

The MESS-E and MESS-H results with the larger 6-311++G** basis set are also listed in Table 1, which shows roughly the same performance as the smaller 6-31+G* basis set. MESS-E energies are quite accurate in M06-2X and ωB97X-D calculations, with MSE errors of −0.078 kcal/mol (M06-2X) and 0.048 kcal/mol (ωB97X-D) and maximum errors of 0.462 kcal/mol (M06-2X) and 0.223 kcal/mol (ωB97X-D). The MESS-E errors are also much larger in B3LYP calculations. On the other hand, MESS-H(15), MESS-H(30), and MESS-H(60) all produce energies with MSE values of no more than 0.14 kcal/mol and maximum errors of no more than 0.36 kcal/mol. With both basis sets, the best performance occurs with MESS-H(30), where the number of requested eigenvectors is around twice the number of electrons—18 for methanol.

B. Solvated β-Alanine. The results for solvated β-alanine are shown in Figure 3 and Table 2. The QM/MM polarization energy is more significant for this system: −8 to −1 kcal/mol with the three functionals.

MESS-E energies correlate well with the exact values, with R^2 values of 0.989 (B3LYP) and 0.992 (M06-2X and ωB97X-D). In B3LYP calculations, MESS-E again significantly overestimates the polarization energy (MSE −0.982 kcal/mol with 6-31+G* basis and −1.041 kcal/mol with 6-311++G** basis; REL 24.5% or 24.6% with the two basis sets). Even more significant are the maximum errors, which are 3.046 kcal/mol with 6-31+G* and 3.224 kcal/mol with 6-311++G**, and thus these MESS-E values should not be used without rescaling. In M06-2X calculations, MESS-E energies are again relatively more accurate, with the MSE values with M06-2X functionals being −0.194 and −0.241 kcal/mol, respectively. Nevertheless, the maximum errors are still as large as 1.189 and 1.355 kcal/mol. The MESS-E energies are much more accurate in ωB97X-D calculations, with MSE values as small as 0.060 and 0.079 kcal/mol. But, with maximum errors of 0.601 and 0.612 kcal/mol, the MESS-E energy values should be used with caution, if not rescaled.

The MESS-H calculations all tend to underestimate the polarization energy here, and they produce increasingly more accurate results as more eigenvectors are requested. The MSE values are 0.484−0.711 kcal/mol with MESS-H(30) and are reduced to 0.162−0.336 kcal/mol with MESS-H(60), and further down to 0.049−0.174 kcal/mol with MESS-H(90). At the same time, the maximum errors are reduced from 1.088 to 1.418 kcal/mol with MESS-H(30) to 0.412−0.701 with MESS-H(60) and finally to 0.206−0.421 kcal/mol with MESS-H(90). The relative errors goes down from 12.8−17.4% and 4.4−8.4% and finally to 1.6−4.4%. The best performance here is observed with MESS-H(90), for which the number of requested eigenvectors is again about twice the number of electrons—48 for β-alanine.

C. Oxyluciferin−Luciferase Complex. As shown in Figure 4, the QM/MM polarization energies range from −5 to −2 kcal/mol with the three functionals and the 6-31+G* basis set. With slopes of 1.858, 1.364, and 1.184 for three functionals, the MESS-E scheme now even more significantly overestimates the polarization energy, when compared to two previous cases (methanol and β-alanine).

As show in Table 3, in B3LYP calculations, the MSE (over 100 configurations) is now as large as −2.731 kcal/mol (REL

dx.doi.org/10.1021/jp5072296 | J. Phys. Chem. A 2015, 119, 1511−1523

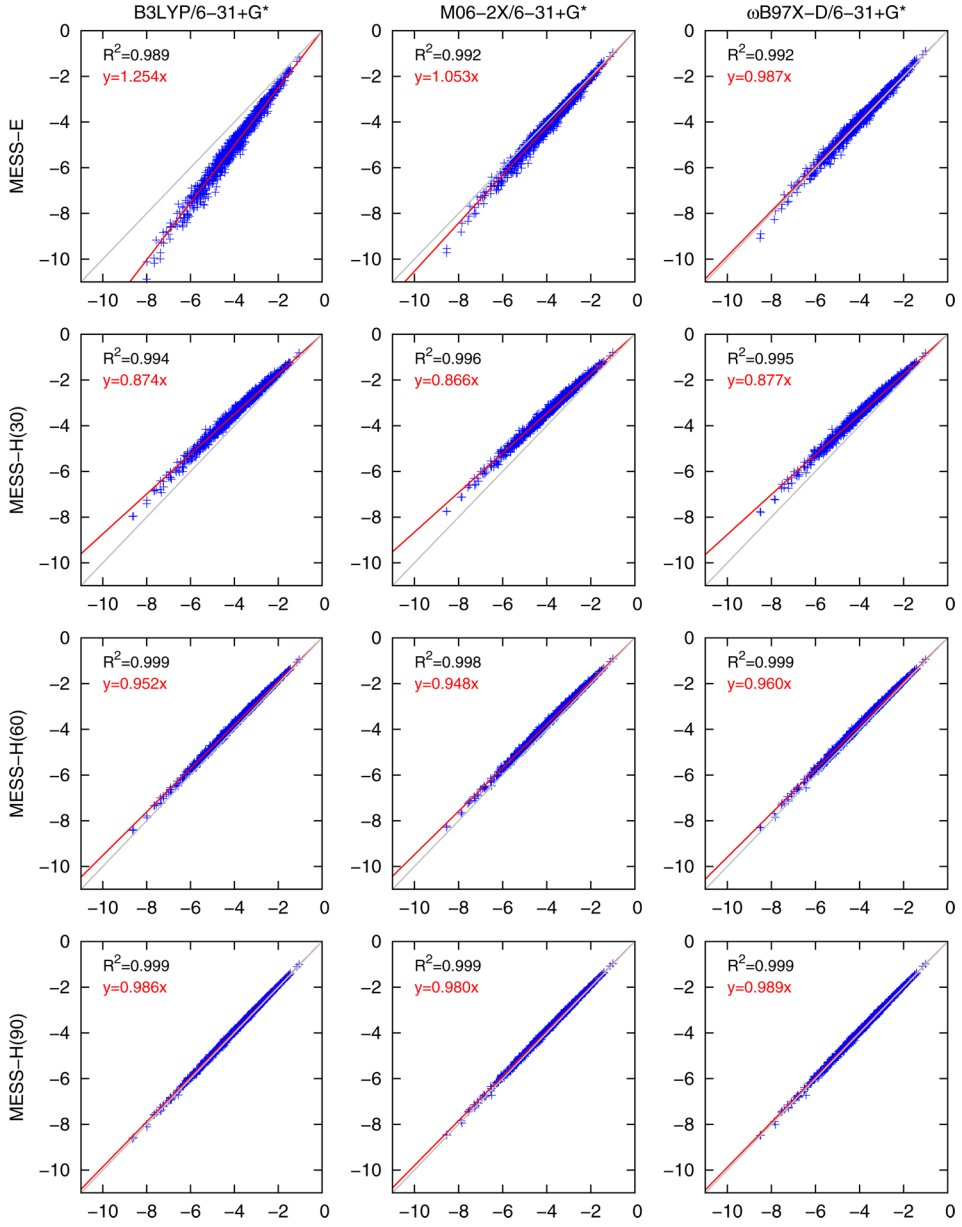


Figure 3. Estimated MESS-QM/MM polarization energies (y-axis) versus the exact values (x-axis) for 1000 configurations of a single β-alanine molecule solvated in a cubic box of 1000 water molecules. All energies are in kcal/mol. The MESS-H(30), MESS-H(60), and MESS-H(90) estimations started by solving iteratively for 30, 60, or 90 eigenvalues/eigenvectors of the electronic Hessian. B3LYP results are shown in the left four panels, M06-2X results in the middle, and ωB97X-D results in the right four panels. The gray lines correspond to a hypothetical perfect estimation. The red lines are linear fits of the estimated energy values against the exact values, with R^2 values and slopes listed at the corner of each panel.

82.7%), with a maximum error of 5.266 kcal/mol. This improves in M06-2X calculations, where the MSE is −1.114 kcal/mol, REL is 35.2%, and the maximum error is 2.129 kcal/mol. Even in ωB97X-D calcualtions, where MESS-E performs the best out of the three functionals, the MSE is still as large as −0.558 kcal/mol, and the maximum error is 1.219 kcal/mol, and the relative error is 17.6%. So for all three functionals, the MESS-E energy cannot be used without rescaling.

The MESS-H scheme has reliably led to more accurate energies with an increasing number of requested eigenvalues for the electronic Hessian. MESS-H(120) led to MSE values around 0.2 kcal/mol (0.251 kcal/mol (B3LYP), 0.222 kcal/mol (M06-2X), and 0.185 kcal/mol (ωB97X-D)) and maximum errors around 0.3 kcal/mol (0.341 kcal/mol (B3LYP), 0.296 kcal/mol (M06-2X), and 0.258 kcal/mol (ωB97X-D)). These errors are reduced roughly by one-third with MESS-H(180), and by one-half with MESS-H(240). Indeed, MESS-H(240) led to MSE values around 0.1 kcal/mol (0.136 kcal/mol (B3LYP), 0.104 kcal/mol (M06-2X), 0.077 kcal/mol (ωB97X-D)). Meanwhile, the MESS-H(240) maximum errors are all below 0.2 kcal/mol (0.190 kcal/mol (B3LYP), 0.161 kcal/mol (M06-2X), 0.129 kcal/mol (ωB97X-D)). Here, the number of requested eigenvectors, 240, is again roughly twice that of the number of electrons—128 for oxyluciferin.

D. Further Analysis of the MESS-E Scheme. So far, we have observed the following general trend for the errors in the MESS-E energies: (i) ωB97X-D < M06-2X ≪ B3LYP and (ii) methanol < β-alanine < oxyluciferin. On one end, MESS-E performs very well for solvated methanol, with MSE values less than 0.1 kcal/mol with M06-2X and ωB97X-D functionals. On

dx.doi.org/10.1021/jp5072296 | J. Phys. Chem. A 2015, 119, 1511−1523

Table 2. Errors in the Estimated QM/MM Polarization Energy for the β-Alanine Molecule Solvated in a Cubic Box of 1000 Water Molecules[a]

		6-31+G* basis				6-311++G** basis			
functional	scheme	MSE	RMS	MAX	REL	MSE	RMS	MAX	REL
B3LYP	MESS-E	−0.982	1.067	3.046	24.6%	−1.041	1.130	3.224	24.5%
	MESS-H(30)	0.509	0.534	1.159	13.3%	0.711	0.739	1.418	17.4%
	MESS-H(60)	0.199	0.210	0.425	5.3%	0.336	0.351	0.687	8.3%
	MESS-H(90)	0.062	0.075	0.206	1.8%	0.174	0.186	0.405	4.4%
M06-2X	MESS-E	−0.194	0.270	1.189	5.2%	−0.241	0.314	1.355	5.9%
	MESS-H(30)	0.528	0.551	1.048	14.0%	0.669	0.696	1.272	16.7%
	MESS-H(60)	0.206	0.218	0.462	5.5%	0.332	0.348	0.701	8.4%
	MESS-H(90)	0.081	0.094	0.244	2.3%	0.165	0.176	0.421	4.2%
ωB97X-D	MESS-E	0.060	0.161	0.601	3.6%	0.079	0.175	0.612	3.7%
	MESS-H(30)	0.484	0.508	1.088	12.8%	0.623	0.649	1.316	15.5%
	MESS-H(60)	0.162	0.174	0.412	4.4%	0.274	0.292	0.686	6.9%
	MESS-H(90)	0.049	0.070	0.306	1.6%	0.127	0.142	0.372	3.3%

[a]Mean signed errors (MSE), root mean square errors (RMS), maximum errors (MAX), all in kcal/mol, in the QM/MM polarization energies estimated with MESS-E and MESS-H schemes are listed, together with the average of the relative unsigned error (REL). For the MESS-H calculations, the number within the parentheses refers to the number of eigenvectors used in the construction of the inverse of the electronic Hessian. Averaged over 1000 configurations from a 1 ns trajectory.

the other end, it significantly overestimates the QM/MM polarization energy for oxyluciferin in luciferase: with B3LYP, the MSE is as large as −2.731 kcal/mol. Here we shall attempt to unravel the underlying reason for this rather uneven performance.

Because MESS-E almost always overestimates the polarization energy, the exception being ωB97X-D calculations on solvated methanol and β-alanine, let us introduce a tuning parameter, λ, into eq 19, which then becomes

$$F^{(\lambda)} = F^{(0)} + \lambda \delta h \tag{29}$$

By diagonalizing this Fock matrix,

$$F^{(\lambda)} C^{(\lambda)} = S C^{(\lambda)} \epsilon^{(\lambda)} \tag{30}$$

we obtain a set of MO coefficients, $C^{(\lambda)}$, which change continuously with parameter λ. From these MOs, we can compute the corresponding density matrix, $P^{(\lambda)}$, electron density, $\rho^{(\lambda)}$, and Kohn−Sham DFT energy,

$$E(\lambda) = E(C^{(\lambda)}) \tag{31}$$

Note that $E(\lambda)$ provides an upperbound to E(full), the actual DFT energy within a given MM potential, which corresponds to full converged MOs, $C^{(\text{full})}$, and its corresponding electron density, $\rho^{(\text{full})}$.

To assess the degree with which the occupied Kohn−Sham orbitals and thus the electron density respond to the applied MM potential, we computed the difference between $\rho^{(\lambda)}$ and the full-converged SCF density $\rho^{(\text{full})}$:

$$\Delta n = \int \left| \rho^{(\lambda)}(\mathbf{r}) - \rho^{(\text{full})}(\mathbf{r}) \right| \mathrm{d}\mathbf{r} \tag{32}$$

and plotted them and $E(\lambda)$ values against λ (ranging from 0 to 1.5) in Figure 5 for the first configuration of each test system with all three functionals. There, in the panels with energy curves, the energy values are relative to the full-converged DFT energy, so that the values at λ = 0 are (the absolute value of) the exact polarization energies.

It is clear in Figure 5 that, given a test system and a functional, the Δn curve and the $E(\lambda)$ curve reach minimum at the same λ values. And in all cases, the minimum occurs where λ is smaller than 1. This indicates that $\delta P^{(1)}$ always overestimates the response in the density matrix.

In ωB97X-D calculations, the energy reaches a minimum of 0.325 kcal/mol at λ = 0.9 for solvated methanol, 0.919 kcal/mol at λ = 0.8 for solvated β-alanine, and 0.497 kcal/mol at λ = 0.625 for the oxyluciferin−luciferase complex. In M06-2X calculations, the minimums are 0.296 kcal/mol at λ = 0.875 (methanol), 0.873 kcal/mol at λ = 0.775 (β-alanine), and 0.482 kcal/mol at λ = 0.55 (oxyluciferin). In B3LYP calculations, the minimums are 0.31 kcal/mol at λ = 0.725 (methanol), 1.004 kcal/mol at λ = 0.625 (β-alanine), and 0.613 kcal/mol at λ = 0.4 (oxyluciferin).

The total energy within the MESS-E scheme,

$$E_{\text{MESS-E}} = E^{(0)} + E^{(1)} + E^{(2)}_{\text{MESS-E}} \tag{33}$$

is a sum of gas-phase energy, $E^{(0)}$, the first-order energy change, $E^{(1)}$ (eq 18), and the MESS-E approximation to the polarization energy, $E^{(2)}_{\text{MESS-E}}$ (eq 23). Compared to $E(\lambda)$ at λ = 1 in Figure 5, the MESS-E total energy does not include (i) 2-electron integral contributions, $\delta P^{(1)} \cdot II \cdot \delta P^{(1)}$, and (ii) their DFT exchange−correlation counterparts. As a result, the MESS-E energies are no longer upper bounds to the full-converged DFT energy. Due to its dependence on $\delta P^{(1)}$, which overestimates the response in the density matrix, the MESS-E formula in eq 23 understandably tends to overestimate the QM/MM polarization energy. In Figure 6, the MSE in the MESS-E energies from Tables 1−3 are plotted against the λ values above. There is a clear trend: the smaller the λ value is at the minimum, the larger errors we find in the MESS-E energies.

V. CONCLUSIONS

In this work, we proposed two schemes, MESS-E and MESS-H, for a fast estimation of the MESS-QM/MM polarization energies and applied them to three test systems: solvated methanol, solvated β-alanine, and the luciferin−luciferase complex. The main findings are as follows:

- The MESS-E scheme in general overestimates the polarization energy, and the errors follow the trends ωB97X-D < M06-2X ≪ B3LYP and methanol < β-alanine < oxyluciferin. In the best cases, solvated

dx.doi.org/10.1021/jp5072296 | J. Phys. Chem. A 2015, 119, 1511−1523

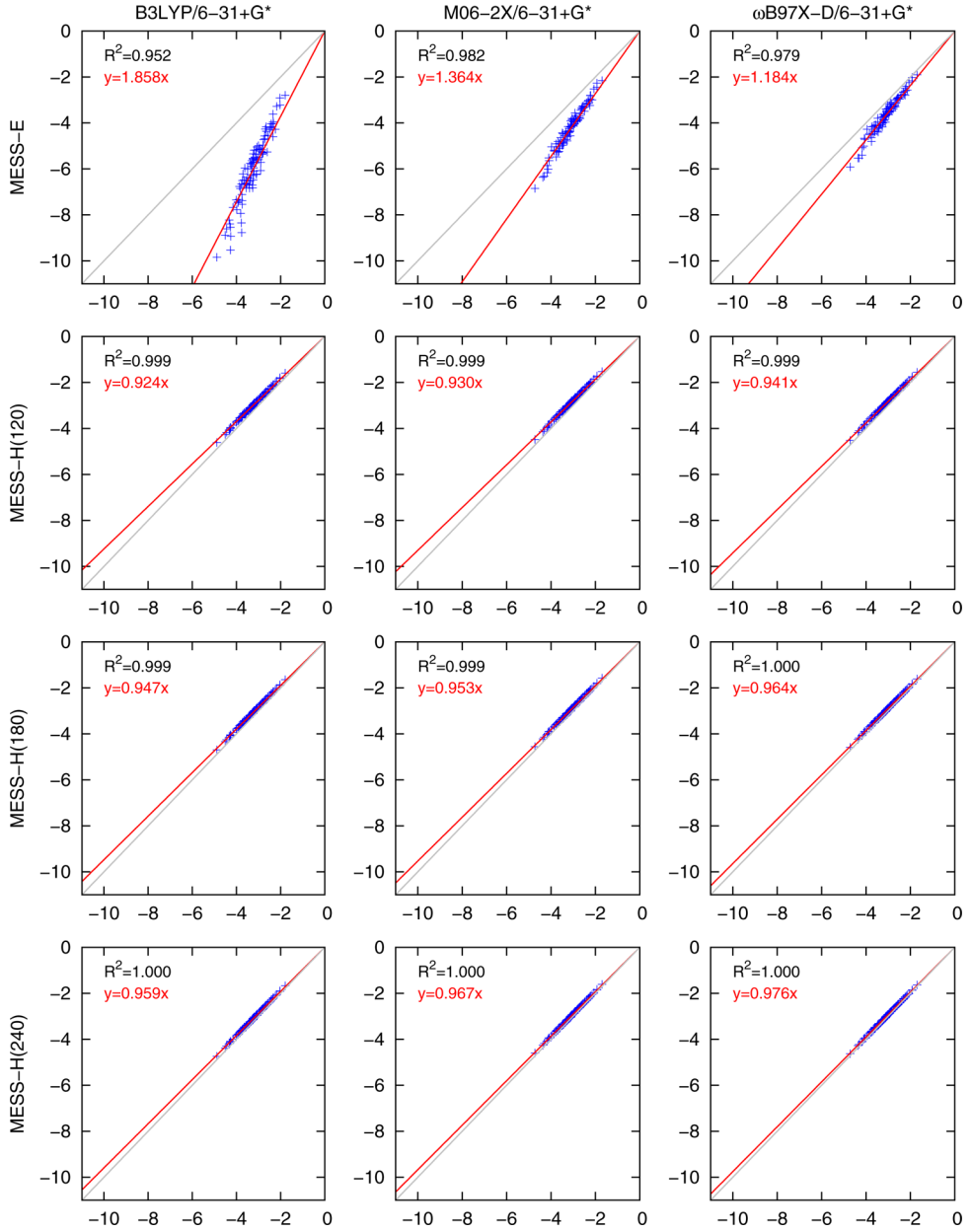


Figure 4. Estimated MESS-QM/MM polarization energies (y-axis) versus the exact values (x-axis) for 100 configurations of an oxyluciferin molecule embedded in luciferase. All energies are in kcal/mol. The MESS-H(120), MESS-H(180), and MESS-H(240) estimations started by solving iteratively for 120, 180, or 240 eigenvalues/eigenvectors of the electronic Hessian. B3LYP results are shown in the left four panels, M06-2X results in the middle, and ωB97X-D results in the right four panels. The gray lines correspond to a hypothetical perfect estimation. The red lines are linear fits of the estimated energy values against the exact values, with R^2 values and slopes listed at the corner of each panel.

methanol with the M06-2X or ωB97X-D functionals, for example, the MSE is smaller than 0.1 kcal/mol. In the worst cases, such as the luciferin−luciferase complex with the B3LYP functional, the MSE is as large as −2.731 kcal/mol. This significant difference in its performance is suggested to arise from the overestimation of the density matrix response in the MESS-E scheme, and a λ parameter was introduced to assess the degree of this overestimation.

- The MESS-H scheme can reliably estimate the polarization energy. The average errors can be around or even below 0.2 kcal/mol, if the number of eigenvectors used to construct the inverse electronic Hessian is at least about twice the number of electrons in the QM region.
- MESS-E and MESS-H schemes only require the evaluation of the MM electrostatic potential and completely avoid the SCF iterations at each configuration, especially the expensive updating of Coulomb, Hartree−Fock exchange, and DFT exchange−correlation matrices, and thus they can be performed with only a tiny fraction of the cost associated with standard QM/MM calculations.

This work is limited in a number of ways:

- Although we have demonstrated the accuracy of the two MESS schemes, especially MESS-H, for individual configurations of our test systems, it remains to be seen how much error these approximations cause to the overall free energies;

dx.doi.org/10.1021/jp5072296 | J. Phys. Chem. A 2015, 119, 1511−1523

Table 3. Errors in the Estimated QM/MM Polarization Energy for the Oxyluciferin−Luciferase Complex[a]

		6-31+G* basis			
functional	scheme	MSE	RMS	MAX	REL
B3LYP	MESS-E	−2.731	2.874	5.266	82.7%
	MESS-H(120)	0.251	0.254	0.341	7.9%
	MESS-H(180)	0.174	0.176	0.237	5.5%
	MESS-H(240)	0.136	0.137	0.190	4.3%
M06-2X	MESS-E	−1.114	1.169	2.129	35.2%
	MESS-H(120)	0.222	0.224	0.296	7.3%
	MESS-H(180)	0.149	0.150	0.208	4.9%
	MESS-H(240)	0.104	0.105	0.161	3.4%
ωB97X-D	MESS-E	−0.558	0.605	1.219	17.6%
	MESS-H(120)	0.185	0.187	0.258	6.1%
	MESS-H(180)	0.113	0.115	0.176	3.7%
	MESS-H(240)	0.077	0.079	0.129	2.5%

[a]The oxyluciferin substrate is the QM region, and the luciferase is described with the CHARMM force field. Mean signed errors (MSE), root mean square errors (RMS), maximum errors (MAX), all in kcal/mol, in the QM/MM polarization energies estimated with MESS-E and MESS-H schemes are listed, together with the average of the relative unsigned error (REL). Averaged over 100 configurations from a 100 ps trajectory.

- Though a λ parameter has been introduced to assess the overestimation of the density response in the MESS-E scheme, the dependence of the optimal λ values on the molecule, on the QM method and on the external MM environment needs to be explored in more detail before developing a MESS-E scheme with proper rescaling that is universally applicable and the least arbitrary. Specifically, there has been some clear indication that the optimal λ values as shown in Figure 5 are correlated with the HOMO−LUMO gaps. For example, B3LYP, which has the smallest fraction for the Hartree−Fock exact exchange (20%) out of the three functionals, leads to the smallest HOMO−LUMO gaps and then the smallest optimal λ values.
- In many cases, the gas-phase electron structure is not necessarily the best reference for the MESS-E and MESS-H calculations. Other references, a single configuration with external MM potential or an average over multiple configurations, should be tested.
- Although the QM/MM polarization energy can be estimated within a reasonable accuracy, one might want to estimate the QM/MM energy gradient as well, which is required to compute the enthalpic correction to the free energy due to the constraint force. It remains to be seen whether the QM/MM gradient is more sensitive to the error in Kohn−Sham orbitals than the energy.
- The MM electrostatic potential used in the solvated methanol or β-alanine calculations comes only from MM atoms in the central unit cell. Although their images in PBC simulations are not expected to significantly add to the polarization of the QM region, especially if a significantly large unit cell is used (as the case with this work), an extension of MESS-E and MESS-H schemes to QM/MM calculations with a periodic boundary condition[75−77] might be desirable.
- For a macromolecular system with tens of thousands of (if not more) MM atoms, as in our test case of the oxyluciferin−luciferase complex, the cost of evaluating the MM electrostatic potential in the QM region can be high. For example, for each MM configuration in the luciferase calculation with B3LYP/6-31+G* level of theory, it still takes about 8 s to evaluate the MM electrostatic potential on a single Intel Xeon E5420 2.5 GHz processor (whereas it would have required about 150 s to fully converge SCF within 7 cycles starting from

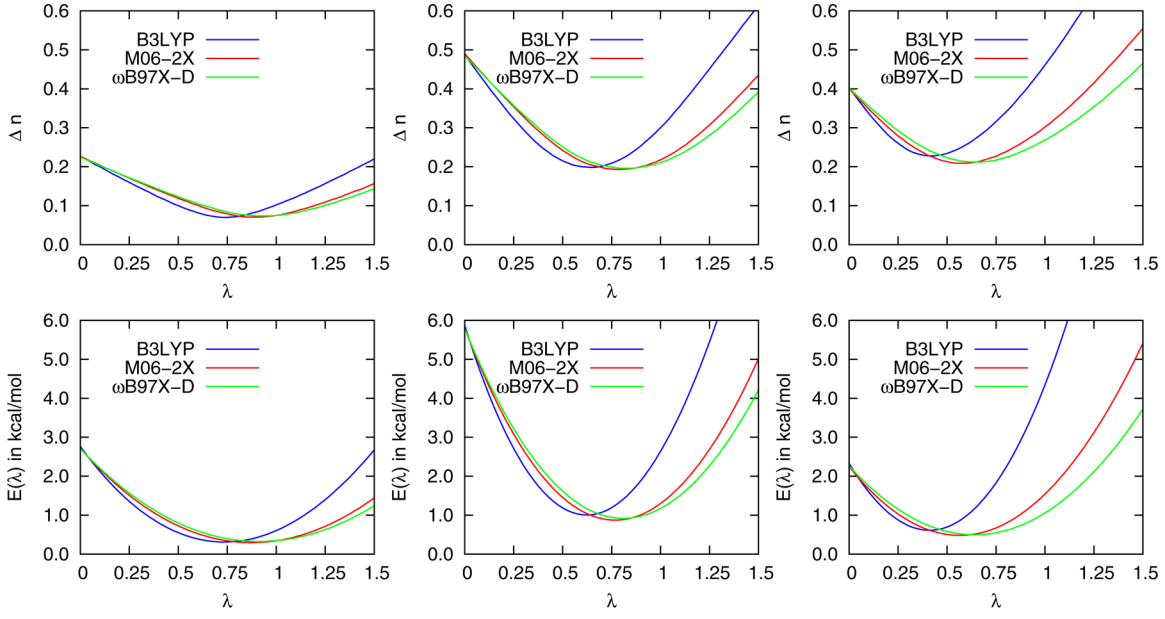


Figure 5. Deviation in the electron density and in DFT energies from the fully converged solutions versus the λ parameter. The deviation in the electron density is measured with Δn, which is defined in eq 32. The energy values are relative to the fully converged DFT energy, so that the values at $\lambda = 0$ are (the absolute value of) the exact polarization energies. The plotted energies contain extra 2-electron and exchange−correlation contributions that are not included in the MESS-E energies. The left panels correspond to results for solvated methanol, the middle panels for solvated β-alanine, and the right panels for the oxyluciferin−luciferase complex.

dx.doi.org/10.1021/jp5072296 | J. Phys. Chem. A 2015, 119, 1511−1523

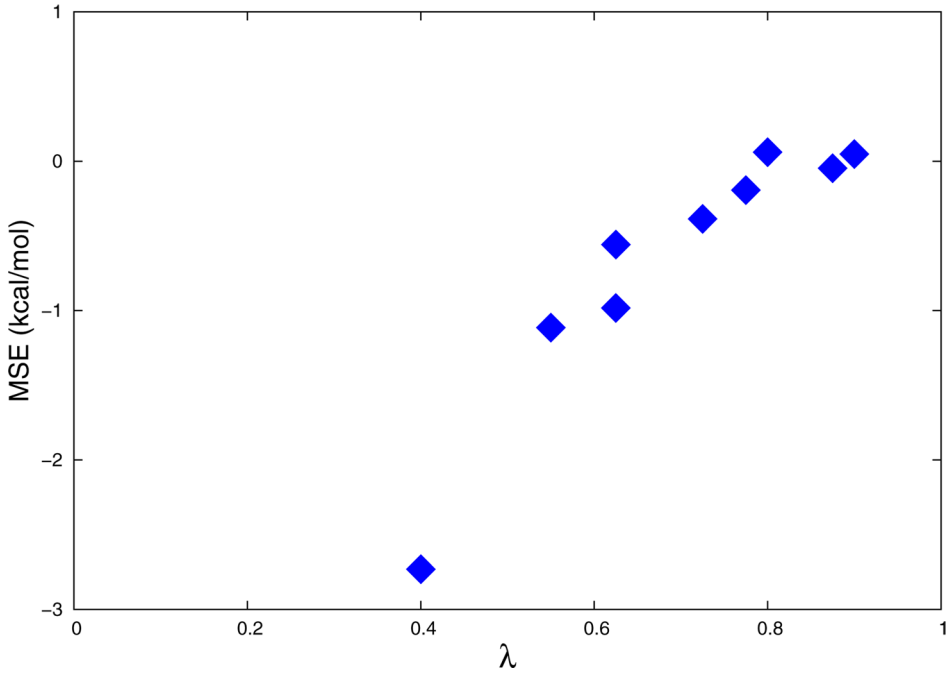


Figure 6. MSE in the MESS-E energies from Tables 1−3 versus the λ values at the minimums in Figure 5.

gas-phase molecular orbitals). Some techniques in the J-engine[78,79] and quantum Ewald mesh[80] methods can potentially be adapted to help speed up the evaluation of MM potential.

- At this moment, the MESS-E and MESS-H schemes are only applicable to Hartree−Fock, pure DFT, hybrid DFT, and range-separated DFT. One has yet to extend it to double-hybrid DFT or post-Hartree−Fock methods.
- We have limited our study to fixed point-charge force fields on the MM atoms, and therefore we have completely neglected the "back" polarization effects, where the QM nuclei/electrons also polarize the MM atoms.

Currently, we are actively pursuing work to address some of these limitations.

AUTHOR INFORMATION

Corresponding Author

*Y. Shao. Phone: 412-687-0695. E-mail: yihan@q-chem.com.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This research is partially supported by DOE grant No. DE-SC0011297 and NIH grant No. GM096678-02 (Y.S.), by the Intramural Research Program of the NIH, NHLBI (A.J.S., G.K., B.R.B.), and by the China Scholarship Council Grant No. 201208310344 (Y.M.). Y.S. thanks Drs. Andres Cisneros, Scott Perrin, Frank Pickard, Andy Simmonnet, Rick Veneble, and Lee Woodcock for helpful discussions. Y.S. also thanks Drs. Shushu Zhang, Chungen Liu and Shawn Chen for sharing their work on luciferase. Y.S. is very grateful to Prof. Peter Pulay for helpful discussions, especially some of their earlier work on the computation of MESS QM/MM polarization energies using generalized multipolar polarizabilities.[81,82] Computational resources and services used in this work were provided by the LoBoS cluster of the National Institutes of Health.

REFERENCES

(1) Gao, J. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; Wiley-VCH: New York, 1996; Vol. *17*; pp 119−185.
(2) Warshel, A. Computer Simulations of Enzyme Catalysis: Methods, Progress, and Insights. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 425−443.
(3) Friesner, R. A.; Guallar, V. Ab initio Quantum Chemical and Mixed Quantum Mechanics/Molecular Mechanics (QM/MM) Methods for Studying Enzymatic Catalysis. *Annu. Rev. Phys. Chem.* **2005**, *56*, 389−427.
(4) Senn, H. M.; Thiel, W. QM/MM Studies of Enzymes. *Curr. Opin. Chem. Biol.* **2007**, *11*, 182−187.
(5) Hu, H.; Yang, W. Free Energies of Chemical Reactions in Solution and in Enzymes with ab initio Quantum Mechanics/Molecular Mechanics Methods. *Annu. Rev. Phys. Chem.* **2008**, *59*, 573−601.
(6) Senn, H. M.; Thiel, W. QM/MM Methods for Biomolecular Systems. *Angew. Chem. Int.* **2009**, *48*, 1198−229.
(7) Lonsdale, R.; Ranaghan, K. E.; Mulholland, A. J. Computational Enzymology. *Chem. Commun.* **2010**, *46*, 2354−2372.
(8) Menikarachchi, L. C.; Gascón, J. A. QM/MM Approaches in Medicinal Chemistry Research. *Curr. Top. Med. Chem.* **2010**, *10*, 46−54.
(9) Lodola, A.; De Vivo, M. The Increasing Role of QM/MM in Drug Discovery. *Adv. Protein Chem. Struct. Biol.* **2012**, *87*, 337−362.
(10) Schlegel, H. B. Exploring Potential Energy Surfaces for Chemical Reactions: An Overview of Some Practical Methods. *J. Comput. Chem.* **2003**, *24*, 1514−1527.
(11) Schlegel, H. B. Geometry Optimization. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2011**, *1*, 790−809.
(12) Head-Gordon, M. Quantum Chemistry and Molecular Processes. *J. Phys. Chem.* **1996**, *100*, 13213−13225.
(13) Goedecker, S. Linear Scaling Electronic Structure Methods. *Rev. Mod. Phys.* **1999**, *71*, 1085−1123.
(14) Helgaker, T.; Klopper, W.; Tew, D. P. Quantitative Quantum Chemistry. *Mol. Phys.* **2008**, *106*, 2107−2143.
(15) Sherrill, C. D. Frontiers in Electronic Structure Theory. *J. Chem. Phys.* **2010**, *132*, 110902.
(16) Zhang, Y.; Liu, H.; Yang, W. Free Energy Calculation on Enzyme Reactions with an Efficient Iterative Procedure to Determine

Minimum Energy Paths on a Combined ab initio QM/MM Potential Energy Surface. *J. Chem. Phys.* **2000**, *112*, 3483−3492.
(17) Rod, T. H.; Ryde, U. Quantum Mechanical Free Energy Barrier for an Enzymatic Reaction. *Phys. Rev. Lett.* **2005**, *94*, 138302.
(18) Rod, T. H.; Ryde, U. Accurate QM/MM Free Energy Calculations of Enzyme Reactions: Methylation by Catechol O-methyltransferase. *J. Chem. Theory Comput.* **2005**, *1*, 1240−1251.
(19) Hu, L.; Soderhjelm, P.; Ryde, U. On the Convergence of QM/MM Energies. *J. Chem. Theory Comput.* **2011**, *7*, 761−777.
(20) Heimdal, J.; Ryde, U. Convergence of QM/MM Free-Energy Perturbations Based on Molecular-Mechanics or Semiempirical Simulations. *Phys. Chem. Chem. Phys.* **2012**, *14*, 12592−12604.
(21) König, G.; Hudson, P. S.; Boresch, S.; Woodcock, H. L. Multiscale Free Energy Simulations: An Efficient Method for Connecting Classical MD Simulations to QM or QM/MM Free Energies Using Non-Boltzmann Bennett Reweighting Schemes. *J. Chem. Theory Comput.* **2014**, *10*, 1406−1419.
(22) König, G.; Brooks, B. R. Correcting for the Free Energy Costs of Bond or Angle Constraints in Molecular Dynamics Simulations. *Biochim. Biophys. Acta* **2014**, DOI: 10.1016/j.bbagen.2014.09.001.
(23) Yamamoto, T. Variational and Perturbative Formulations of Quantum Mechanical/Molecular Mechanical Free Energy with Mean-Field Embedding and Its Analytical Gradients. *J. Chem. Phys.* **2008**, *129*, 244104.
(24) Galvan, I. F.; Martin, M. E.; Aguilar, M. A. A New Method to Locate Saddle Points for Reactions in Solution by Using the Free-Energy Gradient Method and the Mean Field Approximation. *J. Comput. Chem.* **2004**, 1227−1233.
(25) Bentzien, J.; Muller, R. P.; Florian, J.; Warshel, A. Hybrid ab initio Quantum Mechanics/Molecular Mechanics Calculations of Free Energy Surfaces for Enzymatic Reactions: The Nucleophilic Attack in Subtilisin. *J. Phys. Chem. B* **1998**, *102*, 2293−2301.
(26) Frushicheva, M. P.; Warshel, A. Towards Quantitative Computer-Aided Studies of Enzymatic Enantioselectivity: The Case of Candida Antarctica Lipase A. *ChemBioChem.* **2012**, *13*, 215−223.
(27) Woodcock, H. L.; Zheng, W.; Ghysels, A.; Shao, Y.; Kong, J.; Brooks, B. R. Vibrational Subsystem Analysis: A Method for Probing Free Energies and Correlations in the Harmonic Limit. *J. Chem. Phys.* **2008**, *129*, 214109.
(28) Tao, P.; Sodt, A. J.; Shao, Y.; König, G.; Brooks, B. R. Computing the Free Energy Along a Reaction Coordinate Using Rigid Body Dynamics. *J. Chem. Theory Comput.* **2014**, *10*, 4198−4207.
(29) Straatsma, T. P.; McCammon, J. A. Treatment of Rotational Isomers in Free Energy Evaluations. Analysis of the Evaluation of Free Energy Differences by Molecular Dynamics Simulations of Systems with Rotational Isomeric States. *J. Chem. Phys.* **1989**, *90*, 3300−3304.
(30) Hohenberg, P.; Kohn, W. Inhomegeneous Electron Gas. *Phys. Rev.* **1964**, *136*, B864−871.
(31) Kohn, W.; Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.* **1965**, *140*, A1133−A1138.
(32) Parr, R. G.; Yang, W. *Density-Functional Theory of Atoms and Molecules*; Oxford University Press: New York, 1989.
(33) Burke, K. Perspective on Density Functional Theory. *J. Chem. Phys.* **2012**, *136*, 150901.
(34) Cohen, A. J.; Mori-Sanchez, P.; Yang, W. Challenges for Density Functional Theory. *Chem. Rev.* **2012**, *112*, 289−320.
(35) Becke, A. D. Perspective: Fifty Years of Density-Functional Theory in Chemical Physics. *J. Chem. Phys.* **2014**, *140*, 18A301.
(36) Liang, W.; Head-Gordon, M. Approaching the Basis Set Limit in Density Functional Theory Calculations Using Dual Basis Sets without Diagonalization. *J. Phys. Chem. A* **2004**, *108*, 3206−3210.
(37) Steele, R. P.; Shao, Y.; DiStasio, R. A.; Head-Gordon, M. Dual-Basis Analytic Gradients. 1. Self-Consistent Field Theory. *J. Phys. Chem. A* **2006**, *110*, 13915−13922.
(38) Steele, R. P.; DiStasio, R. A.; Shao, Y.; Kong, J.; Head-Gordon, M. Dual-Basis Second-Order Möller-Plesset Perturbation Theory: A Reduced-Cost Reference for Correlation Calculations. *J. Chem. Phys.* **2006**, *125*, 074108.
(39) Deng, J.; Gilbert, A. T. B.; Gill, P. M. W. Density Functional Triple Jumping. *Phys. Chem. Chem. Phys.* **2010**, *12*, 10759−10765.
(40) Khaliullin, R. Z.; Head-Gordon, M.; Bell, A. T. An Efficient Self-Consistent Field Method for Large Systems of Weakly Interacting Components. *J. Chem. Phys.* **2006**, *124*, 204105.
(41) Khaliullin, R. Z.; Bell, A. T.; Head-Gordon, M. Analysis of Charge Transfer Effects in Molecular Complexes Based on Absolutely Localized Molecular Orbitals. *J. Chem. Phys.* **2008**, *128*, 184112.
(42) Roothaan, C. C. J. New Developments in Molecular Orbital Theory. *Rev. Mod. Phys.* **1951**, *23*, 69−89.
(43) Berkowitz, M.; Parr, R. G. Molecular Hardness and Softness, Local Hardness and Softness, Hardness and Softness Kernels, and Relations Among These Quantities. *J. Chem. Phys.* **1988**, *88*, 2554−2557.
(44) Cohen, M. H.; Ganduglia-Pirovano, M. V.; Kudrnovsk, J. Reactivity Kernels, the Normal Modes of Chemical Reactivity, and the Hardness and Softness Spectra. *J. Chem. Phys.* **1995**, *103*, 3543−3551.
(45) Senet, P. Nonlinear Electronic Responses, Fukui Functions and Hardnesses as Functionals of the Ground-State Electronic Density. *J. Chem. Phys.* **1996**, *105*, 6471−6489.
(46) Senet, P. Kohn-Sham Orbital Formulation of the Chemical Electronic Responses, including the Hardness. *J. Chem. Phys.* **1997**, *107*, 2516−2524.
(47) Morita, A.; Kato, S. Ab Initio Molecular Orbital Theory on Intramolecular Charge Polarization: Effect of Hydrogen Abstraction on the Charge Sensitivity of Aromatic and Nonaromatic Species. *J. Am. Chem. Soc.* **1997**, *119*, 4021−4032.
(48) Wang, C.; Zhao, D.; Yang, Z. Calculation of the Linear Response Function by the Atom Bond Electronegativity Equalization Method (ABEEM). *Chem. Phys. Lett.* **2000**, *330*, 132−138.
(49) Ayers, P. W.; Parr, R. G. Variational Principles for Describing Chemical Reactions. Reactivity Indices based on the External Potential. *J. Am. Chem. Soc.* **2001**, *123*, 2007−2017.
(50) Ayers, P. W. Strategies for Computing Chemcial Reactivity Indices. *Theor. Chem. Acc.* **2001**, *106*, 271−279.
(51) Morita, A.; Kato, S. The Charge Response Kernel with Modified Electrostatic Potential Charge Model. *J. Phys. Chem. A* **2002**, *106*, 3909−3916.
(52) Liu, S.; Li, T.; Ayers, P. W. Potentialphilicity and Potentialphobicity: Reactivity Indicators for External Potential Changes From Density Functional Reactivity Theory. *J. Chem. Phys.* **2009**, *131*, 114106.
(53) Boisdenghien, Z.; Alsenoy, C. V.; Proft, F. D.; Geerlings, P. Evaluating and Interpreting the Chemical Relevance of the Linear Response Kernel for Atoms. *J. Chem. Theory Comput.* **2013**, *9*, 1007−1015.
(54) Boisdenghien, Z.; Fias, S.; Van Alsenoy, C.; Proft, F. D.; Geerlings, P. Evaluating and Interpreting the Chemical Relevance of the Linear Response Kernel for Atom II: Open Shell. *Phys. Chem. Chem. Phys.* **2014**, *16*, 14614−14624.
(55) Seeger, R.; Pople, J. A. Self-Consistent Molecular Orbital Methods. XVIII. Constraints and Stability in HartreeFock Theory. *J. Chem. Phys.* **1977**, *66*, 3045−3050.
(56) Davidson, E. R. The Iterative Calculation of a Few of the Lowest Eigenvalues and Corresponding Eigenvectors of Large Real-Symmetric Matrices. *J. Comput. Phys.* **1975**, *17*, 87−94.
(57) Hirata, S.; Head-Gordon, M. Time-Dependent Density Functional Theory for Radicals: An Improved Description of Excited States with Substantial Double Excitation Character. *Chem. Phys. Lett.* **1999**, *302*, 375−382.
(58) Woodcock, H. L.; Hodošček, M.; Gilbert, A. T. B.; Gill, P. M. W.; Schaefer, H. F.; Brooks, B. R. Interfacing Q-Chem and CHARMM to Perform QM/MM Reaction Path Calculations. *J. Comput. Chem.* **2007**, *28*, 1485−1502.
(59) Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; et al. CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.* **2009**, *30*, 1564−1614.

(60) Shao, Y.; Fusti-Molnar, L.; Jung, Y.; Kussmann, J.; Ochsenfeld, C.; Brown, S. T.; Gilbert, A. T. B.; Slipchenko, L. V.; Levchenko, S. V.; O'Neill, D. P.; et al. Advances in Methods and Algorithms in a Modern Quantum Chemistry Program Package. *Phys. Chem. Chem. Phys.* **2006**, *8*, 3172−3191.
(61) Krylov, A. I.; Gill, P. M. W. Q-Chem: an Engine for Innovation. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2013**, *3*, 317−326.
(62) Nakatsu, T.; Ichiyama, S.; Hiratake, J.; Saldanha, A.; Kobashi, N.; Sakata, K.; Kato, H. Structural Basis for the Spectral Difference in Luciferase Bioluminescence. *Nature* **2006**, *440*, 372−376.
(63) da Silva, L. P.; Esteves Da Silva, J. C. Computational Studies of the Luciferase Light-Emitting Product: Oxyluciferin. *J. Chem. Theory Comput.* **2011**, *7*, 809−817.
(64) Song, C. I.; Rhee, Y. M. Dynamics on the Electronically Excited State Surface of the Bioluminescent Firefly Luciferase-Oxyluciferin system. *J. Am. Chem. Soc.* **2011**, *133*, 12040−12049.
(65) Becke, A. D. Density-Functional Exchange-Energy Approximation with Correct Asymptotic Behavior. *Phys. Rev. A* **1988**, *38*, 3098−3100.
(66) Becke, A. D. A New Mixing of Hartree-Fock and Local Density-Functional Theories. *J. Chem. Phys.* **1993**, *98*, 1372−1377.
(67) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Phys. Rev. B* **1988**, *37*, 785−789.
(68) Zhao, Y.; Truhlar, D. G. The M06 Suite of Density Functionals for Main Group Thermochemistry, Thermochemical Kinetics, Noncovalent Interactions, Excited States, and Transition Elements: Two New Functionals and Systematic Testing of Four M06-Class Functionals and 12 Other Function. *Theor. Chem. Acc.* **2007**, *120*, 215−241.
(69) Chai, J.-D.; Head-Gordon, M. Long-Range Corrected Hybrid Density Functionals with Damped Atom-Atom Dispersion Corrections. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615−6620.
(70) Hehre, W. J.; Ditchfield, R.; Pople, J. A. Self-Consistent Molecular Orbital Methods. XII. Further Extensions of Gaussian-Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules. *J. Chem. Phys.* **1972**, *56*, 2257−2261.
(71) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. Self-Consistent Molecular Orbital Methods. XX. A Basis Set for Correlated Wave Functions. *J. Chem. Phys.* **1980**, *72*, 650−654.
(72) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926−936.
(73) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; et al. CHARMM General Force Field: A Force Field for Drug-Like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields. *J. Comput. Chem.* **2009**, *31*, 671−690.
(74) Tao, P.; Wu, X.; Brooks, B. R. Maintain Rigid Structures in Verlet Based Cartesian Molecular Dynamics Simulations. *J. Chem. Phys.* **2012**, *137*, 134110.
(75) Nam, K.; Gao, J.; York, D. M. An Efficient Linear-Scaling Ewald Method for Long-Range Electrostatic Interactions in Combined QM/MM Calculations. *J. Chem. Theory Comput.* **2005**, *18*, 2−13.
(76) Walker, R. C.; Crowley, M. F.; Case, D. A. The Implementation of a Fast and Accurate QM/MM Potential Method in Amber. *J. Comput. Chem.* **2008**, *29*, 1019−1031.
(77) Holden, Z. C.; Richard, R. M.; Herbert, J. M. Periodic Boundary Conditions for QM/MM Calculations: Ewald Summation for Extended Gaussian Basis Sets. *J. Chem. Phys.* **2013**, *139*, 244108.
(78) Shao, Y.; Head-Gordon, M. An Improved J Matrix Engine for Density Functional Theory Calculations. *Chem. Phys. Lett.* **2000**, *323*, 425−433.
(79) Shao, Y.; White, C. A.; Head-Gordon, M. Efficient Evaluation of the Coulomb Force in Density-Functional Theory Calculations. *J. Chem. Phys.* **2001**, *114*, 6572−6577.
(80) Chang, C.-M.; Shao, Y.; Kong, J. Ewald Mesh Method for Quantum Mechanical Calculations. *J. Chem. Phys.* **2012**, *136*, 114112.
(81) Pulay, P.; Janowski, T. Efficient Calculation of the Energy of a Molecule in an Arbitrary Electric Field. *Int. J. Quantum Chem.* **2009**, *109*, 2113−2120.
(82) Janowski, T.; Wolinski, K.; Pulay, P. Ultrafast Quantum Mechanics/Molecular Mechanics Monte Carlo Simulations Using Generalized Multipole Polarizabilities. *Chem. Phys. Lett.* **2012**, *530*, 1−9.