JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

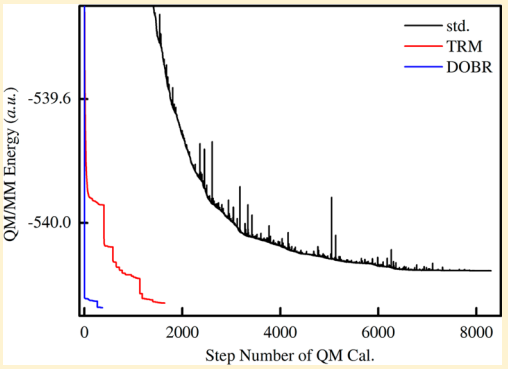
High-Efficiency Microiterative Optimization in QM/MM Simulations of Large Flexible Systems

Yan Zhang,* Peng Xie, Xiaohu He, and Keli Han*

State Key Laboratory of Molecular Reaction Dynamics, Dalian Institute of Chemical Physics, Chinese Academy of Science, Zhongshan Road 457, Dalian 116023, China

Supporting Information

ABSTRACT: We present here a double-optimizations-of-buffer-region (DOBR) microiterative scheme for high-efficiency energy minimizations of large, flexible systems in combined quantum-mechanical/molecular-mechanical (QM/MM) calculations. In the DOBR scheme, an entire system is divided into three regions: the core, buffer, and outer regions. The core region includes QM atoms and the MM atoms within a cutoff distance R_1 to the QM atoms (denoted by MM_1 atoms), and the buffer region consists of MM atoms within another cutoff distance R_2 to MM_1 atoms. Each DOBR microcycle involves two steps: First, QM atoms are assigned electrostatic-potential (ESP) charges, and the buffer and outer regions are optimized at the MM level with the core region kept frozen. Second, the core and buffer regions are optimized at the QM/MM level using the electrostatic embedding with the outer region kept frozen. The two steps are repeated until two optimizations converge at one structure. The DOBR scheme was tested in the optimizations of nucleobases solvated in water spheres of 30 Å radius, where the initial geometries were extracted from the trajectories of classical molecular dynamics simulations, and the cutoff distances R_1 and R_2 were set to 5.0 and 4.0 Å, respectively. For comparisons, the optimizations were also carried out by a "standard" scheme without microiteration and by the two-region microiterative (TRM) method. We found that the averaged number of QM calculations for the DOBR scheme is only ∼1% of that of the standard scheme and ∼6% of the TRM approach. The promising results indicate that the DOBR scheme could significantly increase the efficiency of geometry optimizations for large, flexible systems in QM/MM calculations.



I. INTRODUCTION

A combined quantum-mechanical and molecular-mechanical (QM/MM) approach has become a relatively mature technology after the development over the past decades, which is successfully applied in the simulations for large complex systems.[1−7] In QM/MM computations, the full system is divided into the active region treated at the quantum mechanical (QM) level and the environmental region computed at the molecular mechanics (MM) level. Thus, the QM/MM approach has excellent features of the QM high accuracy and the MM low computational cost. In QM/MM calculations, there are unsaturated dangly bonds in the QM region when the boundary between the QM and MM regions crosses the covalent bonds. Several special models are reported to treat the unsaturated bonds.[2,3,7] The electrostatic interactions between the QM and MM regions are very important in QM/MM simulations. Based on the treatment manner of the electrostatic interactions, the QM/MM method can be classified as mechanical-embedding (ME), electrostatic-embedding (EE), and polarized-embedding (PE).[3,8]

The electrostatic interactions between QM and MM atoms in ME QM/MM simulations are computed at the MM level.[8−12] In contrast, the MM charges are included in the QM calculation of the EE QM/MM calculations. The electrostatic interaction is handled at the QM level, and the QM region would be polarized by MM charges.[8,12−17] Several PE models of recent development allow the polarization of the MM region by the QM region. The numerical values or positions of the charges involved in QM calculations would be changed due to the effect of QM region polarizations.[18−23] In spite of their higher accurateness, the QM/MM PE simulations have an obvious shortcoming which is the expensive computational cost. Thus, the EE treatment is still popular in current QM/MM simulations considering the balance of computational accuracy and cost.

Geometry optimization plays a significant role in various QM/MM investigations. Computational effort of QM/MM optimization is primarily determined by the number of optimized active atoms and the QM region. Direct QM/MM optimization would be impossible for the system in which the optimized active region is very large, and the QM region is computed at a high level. Hence, the full system can be divided into two parts which are core and outer regions, as shown in Figure S1 of the Supporting Information. The core region must be larger than or equal to the QM part. Based on the

Received: May 26, 2016
Published: August 9, 2016

DOI: 10.1021/acs.jctc.6b00547
J. Chem. Theory Comput. 2016, 12, 4632−4643

partitioning design, two microiterative schemes were suggested to optimize the geometry of the large system in QM/MM simulations.[24−37] One is that the outer region is fully optimized after the optimization of each step for the core region is performed.[24−26,28−31,34,36,37] The other is that the full optimization of the outer region is performed after the core region is optimized.[27,32,33,35] In the optimization of the core region for the two microiterative schemes, the gradients on the core atoms are calculated at the QM/MM level. The second-order methods are usually used in the optimization of the core and outer regions due to their rapid convergence.

The microiterative schemes would dramatically reduce the step number of the QM calculations comparing with the direct QM/MM optimization. The consequence would decrease the computational cost for QM/MM geometry optimizations in which QM calculations are at a very high level. In the optimization of the outer region of two microiterative schemes, the gradients on the outer atoms are computed at the MM level. For QM/MM ME simulations, the MM optimization of the outer region could be directly performed by using the exact gradients.[24,34] Theoretically, QM calculations should be carried to calculate the gradients on the atoms of the outer region in the QM/MM EE calculations. It is impossible in the actual optimization of the outer region. Thus, QM density is frozen in the optimization of the outer region. For example, the QM electrostatic potential (ESP) charges are fitted and used in the calculation of the gradients on the atoms of the outer region. The approximative gradients are different from QM/MM EE simulations when the core region is smaller (e.g., the QM part). Several models were suggested to correct the inaccurate gradients in the optimization of the outer region, which promote the efficiency of the geometry optimizations.[30,34,36,37]

The above microiterative schemes of the two-region partition have been extensively employed in many QM/MM computations including the optimization of transition structure[38,39] and the determination of reaction pathways[40−42] and conical intersection.[43] However, the earlier work found that the coupling between the core and outer regions is very important in the microiterative optimizations of the two-region partition. Their total steps of QM calculations are not evidently related to the size of the core region.[35] In addition, the previous investigations revealed that the structures of direct standard QM/MM and microiterative optimizations using the very high converged criteria (e.g., 0.000045 au) are very close.[34] The step number of QM calculations for the microiterative schemes is 1/10−1/2 of that for the standard direct QM/MM optimization.[34,36]

A computational effort would be quite huge in QM/MM EE optimizations of large flexible systems when QM parts are very large and computed at a high level. Even if we use the above two microiterative schemes, the optimizations of such structures are still not feasible. In this work, we present a double-optimizations-of-buffer-region (DOBR) microiterative scheme to optimize the geometries of energy local minima for complex systems in QM/MM EE simulations. The DOBR scheme was used to optimize the geometries of adenine, cytosine, and guanine in water. The performance shows that the DOBR scheme has a high efficiency. The DOBR microiterative scheme would be described in detail below. Section II gives the model of the DOBR scheme, and the computational details of test calculations are described in section III. The results and discussions are shown in section IV. Section V presents the conclusions.

II. METHODOLOGY

The QM/MM method is usually applied in the computations of large systems including 10,000−100,000 atoms.[7] Full QM/MM Hamiltonian has two principal expressions which are subtractive and additive methods.[3] Although the subtractive approach is simple, it requires the complete MM parameters of the QM part, which could not be obtained in some investigations. Here, we use an additive scheme, and the QM/MM total energy can be defined as[3,36]

$$E_{\mathrm{QM/MM}} = E_{\mathrm{QM}}(X_{\mathrm{QM}}) + E_{\mathrm{QM-MM}}(X_{\mathrm{QM}}, X_{\mathrm{MM}}) + E_{\mathrm{MM}}(X_{\mathrm{MM}}) \quad (1)$$

where $E_{\mathrm{QM}}(X_{\mathrm{QM}})$ is the energy of the QM part and the function of the coordinates X_{QM} of QM atoms. $E_{\mathrm{QM-MM}}(X_{\mathrm{QM}}, X_{\mathrm{MM}})$ as the function of the coordinates of all atoms in a whole system is the interaction energy between the QM and MM parts. In this work, the electrostatic interaction involved in $E_{\mathrm{QM-MM}}$ is treated by EE. $E_{\mathrm{MM}}(X_{\mathrm{MM}})$ is the energy of the MM parts and the function of the coordinates X_{MM} of MM atoms.

In our DOBR microiterative scheme, an entire system is divided into three parts which are core, buffer, and outer regions, as shown in Figure 1. The core region consists of QM

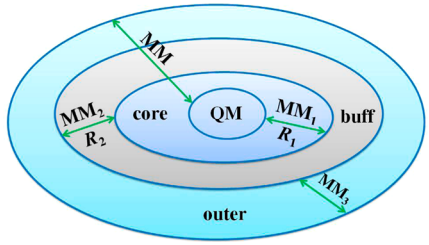


Figure 1. Partitioning of the system for the DOBR microiterative scheme.

atoms and the MM atoms within a cutoff distance R_1 to QM atoms which are denoted by MM_1. The buffer region includes the MM atoms within a cutoff distance R_2 to MM_1 atoms which are marked by MM_2. The outer region comprises the MM atoms outside the core and buffer regions designated by MM_3. If the cutoff distance R_1 is enough large, the interactions between the QM atoms and the MM_2 and MM_3 atoms only contain the nonbonding electrostatic and van der Waals energies, and the ESP charges on QM atoms to represent the QM electronic density can be applied to compute the electrostatic interaction between QM atoms and MM atoms in the buffer and outer regions. The standard direct QM/MM energy minimization of the full system is actually to solve the nonlinear equation set

$$\begin{cases} \dfrac{\partial E_{\mathrm{QM/MM}}}{\partial X_{\mathrm{QM}}} = \dfrac{\partial E_{\mathrm{QM}}}{\partial X_{\mathrm{QM}}} + \dfrac{\partial E_{\mathrm{QM-MM}}}{\partial X_{\mathrm{QM}}} = 0 \\ \dfrac{\partial E_{\mathrm{QM/MM}}}{\partial X_{\mathrm{MM_1}}} = \dfrac{\partial E_{\mathrm{QM-MM}}}{\partial X_{\mathrm{MM_1}}} + \dfrac{\partial E_{\mathrm{MM}}}{\partial X_{\mathrm{MM_1}}} = 0 \\ \dfrac{\partial E_{\mathrm{QM/MM}}}{\partial X_{\mathrm{MM_2}}} = \dfrac{\partial E_{\mathrm{QM-MM}}}{\partial X_{\mathrm{MM_2}}} + \dfrac{\partial E_{\mathrm{MM}}}{\partial X_{\mathrm{MM_2}}} = 0 \\ \dfrac{\partial E_{\mathrm{QM/MM}}}{\partial X_{\mathrm{MM_3}}} = \dfrac{\partial E_{\mathrm{QM-MM}}}{\partial X_{\mathrm{MM_3}}} + \dfrac{\partial E_{\mathrm{MM}}}{\partial X_{\mathrm{MM_3}}} = 0 \end{cases} \quad (2)$$

DOI: 10.1021/acs.jctc.6b00547
J. Chem. Theory Comput. 2016, 12, 4632−4643

There are many first-order and second-order methods to determine the solution of the equation set. The conventional first-order approaches are the steepest descent and conjugated gradient optimization.[44] A second-order method such as Newton−Raphson converges faster than first-order optimization because it uses the Hessian matrix of the full system which can be given by

$$H = \begin{bmatrix} H_{c,c} & H_{c,MM_2} & H_{c,MM_3} \\ H_{MM_2,c} & H_{MM_2,MM_2} & H_{MM_2,MM_3} \\ H_{MM_3,c} & H_{MM_3,MM_2} & H_{MM_3,MM_3} \end{bmatrix} \quad (3)$$

$H_{c,c}$, H_{MM_2,MM_2}, and H_{MM_3,MM_3} respectively, are the diagonal elements of the Hessian matrix for the core−core, buffer−buffer, and outer−outer interactions. The coupling matrix element between the interaction of the core and buffer region is denoted by $H_{MM_2,c}$ (H_{c,MM_2}). $H_{MM_3,c}$ (H_{c,MM_3}) is the core-outer coupling Hessian matrix element, and the buffer-outer coupling is H_{MM_3,MM_2} (H_{MM_2,MM_3}).

In the optimization of a large system, the equation set 2 will be constituted of several tens of thousands or more equations. Even though second-order methods are used in the energy minimization, the number of the optimized cycle will be extremely enormous in the direct standard QM/MM optimization. The computational cost would be very expensive if the QM region is larger and computed at a high level. Although the previous microiterative schemes could reduce the step number of QM calculations, the computational effort to optimize such a system still seems to be too high. Considering the partition of three regions in the present DOBR scheme, the equation set 2 can be composed into

$$\begin{cases} \dfrac{\partial E_{QM/MM}}{\partial X_{QM}} = \dfrac{\partial E_{QM}}{\partial X_{QM}} + \dfrac{\partial E_{QM-MM}}{\partial X_{QM}} = 0 \\ \dfrac{\partial E_{QM/MM}}{\partial X_{MM_1}} = \dfrac{\partial E_{QM-MM}}{\partial X_{MM_1}} + \dfrac{\partial E_{MM}}{\partial X_{MM_1}} = 0 \\ \dfrac{\partial E_{QM/MM}}{\partial X_{MM_2}} = \dfrac{\partial E_{QM-MM}}{\partial X_{MM_2}} + \dfrac{\partial E_{MM}}{\partial X_{MM_2}} = 0 \end{cases} \quad (4)$$

and

$$\begin{cases} \dfrac{\partial E_{QM/MM}}{\partial X_{MM_2}} = \dfrac{\partial E_{QM-MM}}{\partial X_{MM_2}} + \dfrac{\partial E_{MM}}{\partial X_{MM_2}} = 0 \\ \dfrac{\partial E_{QM/MM}}{\partial X_{MM_3}} = \dfrac{\partial E_{QM-MM}}{\partial X_{MM_3}} + \dfrac{\partial E_{MM}}{\partial X_{MM_3}} = 0 \end{cases} \quad (5)$$

The solution of equation set 2 can be achieved by the DOBR microiterative scheme, which iteratively solves the equation sets 4 and 5. The flowchart of the DOBR scheme is shown in Figure 2. In the DOBR scheme, the optimization of the structure of the core (QM and MM_1 atoms) and buffer (MM_2 atoms) regions corresponds to solve the equation set 4 and is performed at the QM/MM EE level with the outer region frozen. The solution of the equation set 5 is the geometry of the buffer and outer (MM_3 atoms) regions optimized at the MM level. In the MM optimizations, the atoms in the core region are frozen, and the ESP charges on QM atoms are used to compute the electrostatic interactions between the QM atoms and the MM atoms in the buffer and outer regions. The two optimizations in the microcycle of the DOBR scheme are independently performed by the different levels. The buffer region is optimized twice in one microcycle. One is at the QM/MM EE level and the other is at the MM level. The cutoff distances R_1 and R_2 are the key in the DOBR scheme. The entire system is optimized by the QM/MM EE method at the end of the DOBR procedure with the purpose of checking the convergence of the structure for the full system. Normally the optimization would converge within several steps.

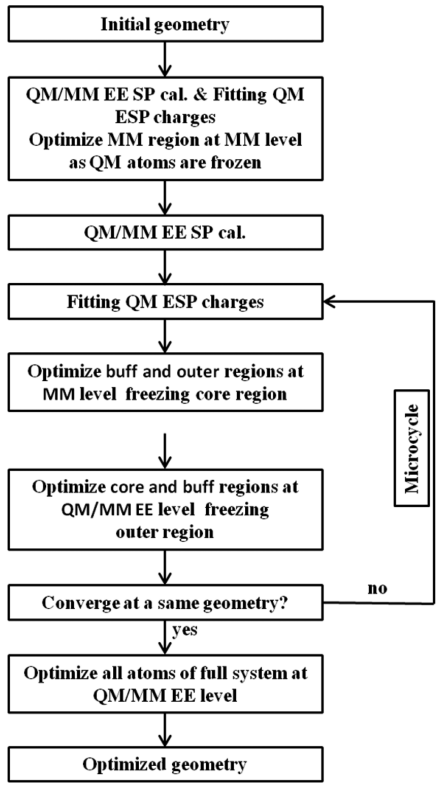


Figure 2. Procedure to perform the DOBR microiterative scheme.

III. COMPUTATIONAL DETAILS

The large systems to test the DOBR optimization in the work are adenine, cytosine, and guanine, which dissolve in the water sphere with 30.0 Å radius, and contain 10842, 10846, and 10840 atoms, respectively. For three systems, molecule dynamics (MD) simulations were carried out to get the initial geometries of a series of the snapshots. In the MD simulations, NVT simulations with the fixed nucleobases were first done for 20 ps at 300 K to equilibrate the water of full systems. At the pre-equilibrated configurations, the systems were heated to 300 K in 20 ps and then equilibrated at 300 K for 2000 ps. In the last 1000 ps of the equilibration, the structures at intervals of 50 ps were extracted to obtain the initial geometries of the optimizations. Thus, there are 20 snapshots for one large system. We separately denote the snapshots by S1050−S2000 (see Table S1 of the Supporting Information).

We perform a QM/MM EE single point (SP) calculation at a given geometry of the system. Then, ESP charges on QM atoms are fitted by the Kollman parameters,[45] and the MM SP gradients of the full system are calculated by applying the ESP charges on the QM atoms. Based on the SP gradients calculated by the QM/MM EE and MM approaches, we compute the

DOI: 10.1021/acs.jctc.6b00547
J. Chem. Theory Comput. 2016, 12, 4632−4643

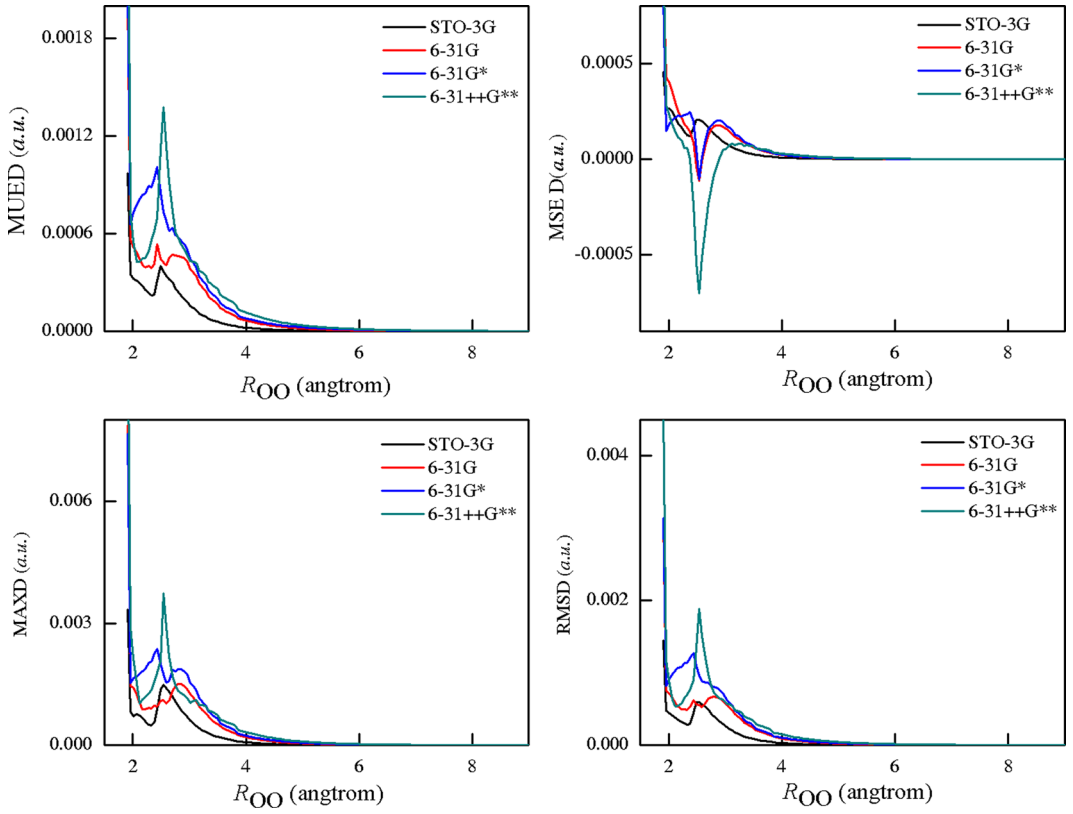


Figure 3. MUED, MSED, MAXED, and RMSD of the MM atoms for the system including two water molecules at the B3LYP level.

gradient differences of full or part MM atoms for the two methods. Finally, mean unsigned error of the gradient deviations (MUED), mean signed error of the gradient deviations (MSED), maximum of the absolute values of the gradient deviations (MAXD), and root-mean-square of the gradient deviations (RMSD) are evaluated to examine the cutoff distance R_1 in the DOBR scheme. The model including two water molecules is a simpler system. In the QM/MM simulations of the model, one H_2O was treated by QM (remarked as WAT1), and the other was computed by the TIP3P force field[46] (remarked as WAT2). The QM calculations used the HF[47,48] method and the B3LYP functional[49−51] of Density Functional Theory (DFT) with the STO-3G,[52,53] 6-31G,[54−56] 6-31G*,[54−57] and 6-31++G**[54−58] basis sets. Moreover, the four snapshots for adenine in water, which include the neutral S1500 and S1900 snapshots and cationic S1250 and S1800 snapshots, were also studied to determine the cutoff distance R_1. In the QM/MM SP simulations for four snapshots, adenine was computed at the B3LYP/6-31+G*[54−58] level, and MM water molecules were calculated by the TIP3P water model.

In the simulations of the nucleobases, the distance between the nucleobases and a water molecule is defined by the minimal value among all lengths between the O atom of a water molecule and each atom of nucleobases. We use the distance to describe the magnitude of the QM region, the MM_1 part, and the buffer region. The neutral and cationic geometries of 20 snapshots for adenine in water were optimized by the standard direct QM/MM EE, two region microiterative, and DOBR microiterative approaches to compare the step number of QM calculations in the optimizations. Furthermore, the DOBR scheme was used to optimize the geometries of the neutral species, anion and cation for 20 snapshots of cytosine and guanine in water. In these optimizations, the DFT approach using the B3LYP functional with the 6-31+G* basis set was applied to compute QM regions, which only include the atoms of the nucleobases, and MM water molecules were treated by the TIP3P water model. Furthermore, we also optimized the geometries of the neutral, anionic, and cationic species for the cytosine S1050 snapshot and the guanine S1750 snapshot using the larger QM regions. The QM parts of the systems were computed at the B3LYP/6-31+G* level, and the TIP3P water model was used to calculate the MM part.

In addition, we also computed two gas enzyme systems which are the chain A of the catalytically active C-terminal domain (A3B-CTD; PDB ID code 2NBQ) for human APOBEC3B of cytidine deaminases[59] and a natural mutation for α-esterase 7 of *Luculia cuprina* (NMAE; PDB ID code 5C8V)[60] to test the DOBR microiterative scheme. The QM region in the QM/MM EE computations of two enzymes was treated at the B3LYP/6-31G level, and the CHARMM31[61] force field was applied in the MM calculations. There are 22 atoms in the QM part of A3B-CTD (totally 3208 atoms), and one of them is the Zn^{2+} ion. The NMAE (totally 9042 atoms) QM part excludes the metallic atom and is also comprised of 22 atoms. We calculated the MSED, MUED, RMSD, and MAXD of different MM regions for two systems and optimized their geometries using the DOBR and two region microiterative approaches. In the present investigation, the QM/MM calculations used the ChemShell 3.5 package.[62] The system setups and MD simulations were carried out by employing the CHARMM.[61] The TURBOMOLE 6.4 program[63] was used for QM calculations and fitting QM ESP charges. All optimizations used the DL-FIND optimizer[64] which is implemented in ChemShell.

DOI: 10.1021/acs.jctc.6b00547
J. Chem. Theory Comput. 2016, 12, 4632−4643

Table 1. MAXD Computed by Using the Different MM Regions for Four Snapshots of Adenine in Water Including Two Neutral (S1500 and S1900 Snapshots) and Two Cationic (S1250 and S1800 Snapshots) Systems[a]

		MM	2.4 Å	3.0 Å	4.0 Å	5.0 Å	6.0 Å	9.0 Å
neutral	S1500	4.0×10^{-3}	4.0×10^{-3}	1.1×10^{-3}	2.4×10^{-4}	2.4×10^{-4}	1.0×10^{-4}	3.9×10^{-5}
	S1900	4.2×10^{-3}	4.2×10^{-3}	1.1×10^{-3}	2.6×10^{-4}	8.0×10^{-5}	6.3×10^{-5}	2.9×10^{-5}
cation	S1250	3.5×10^{-3}	3.5×10^{-3}	1.3×10^{-3}	1.7×10^{-4}	6.3×10^{-5}	3.4×10^{-5}	8.5×10^{-6}
	S1800	2.3×10^{-3}	2.3×10^{-3}	1.6×10^{-3}	1.5×10^{-4}	6.7×10^{-5}	6.0×10^{-5}	2.0×10^{-5}

[a]The different MM regions are the full MM part and MM water molecules excluding the water molecules within 2.4, 3.0, 4.0, 5.0, 6.0, and 9.0 Å around adenine, respectively. Adenine is treated by the QM level. MAXD in au.

Table 2. MAXD and RMSD of Two Gas Enzymes Computed by Using Different MM Regions[a]

		MM	4.0 Å	5.0 Å	6.0 Å	7.0 Å	9.0 Å
RMSD	A3B-CTD	7.0×10^{-05}	5.6×10^{-05}	5.4×10^{-05}	5.3×10^{-05}	5.2×10^{-05}	5.1×10^{-05}
	NMAE	2.1×10^{-05}	1.5×10^{-05}	1.5×10^{-05}	1.4×10^{-05}	1.3×10^{-05}	1.2×10^{-05}
MAXD	A3B-CTD	5.9×10^{-03}	9.0×10^{-04}	7.7×10^{-04}	5.6×10^{-04}	4.3×10^{-04}	2.8×10^{-04}
	NMAE	1.4×10^{-02}	6.7×10^{-04}	6.7×10^{-04}	3.3×10^{-04}	2.5×10^{-04}	1.3×10^{-04}

[a]The different MM parts are full MM atoms and MM atoms which the group distances to the QM region respectively are larger than 4.0, 5.0, 6.0, 7.0, and 9.0 Å. MAXD and RMSD are in au.

IV. RESULTS AND DISCUSSIONS

Sizes of the MM_1 Part and the Buffer Region. In the DOBR scheme, the ESP charges on QM atoms are used to optimize the geometry of the buffer and outer regions to promote the computational efficiency. The ESP charges should ensure that the gradients on the buffer and outer atoms obtained by the MM calculations agree well with those computed by the QM/MM EE method. Thus, the MM optimization of the buffer and outer regions would be performed on the potential energy surface (PES) of the QM/MM EE computations. Observing the partition of the DOBR scheme in Figure 1, it could be found that the MM_1 cutoff distance R_1 should be large enough to meet the above requirement.

We first studied the two-water model to determine the cutoff distance R_1. Along the distance between the O atoms of two water molecules, the PES scans, which the step is 0.1 au, were carried out by the constrained optimizations of full QM. At each optimized geometry, the SP gradients were calculated by the QM/MM EE and MM methods. Figure 3 shows MUED, MSED, MAXD, and RMSD on the WAT2 atoms as the functions of the distance between the O atoms of two water molecules. The figure presents the results of the B3LYP calculations with four basis sets. It is clear that the gradients on the WAT2 atoms computed by the QM/MM EE method are significantly different from the results of MM calculations as two water molecules are close. At the larger distance (e.g., >6.0 Å), the computed MUED, MSED, MAXD, and RMSD are very small and almost zero. It means that the gradients on the WAT2 atoms obtained by the QM/MM EE and MM simulations are well consistent. The results computed by the HF theory with the four basis sets have the same characters, as can be seen in Figure S2 of the Supporting Information.

In the structural optimization using the DL-FIND optimizer, the default converged criteria are 0.00045 au for the maximum gradient component and 0.00030 au for the root-mean-square of the gradients. The two converged standards are also used as the defaults in others computational codes. We consider that the MAXD and RMSD on the MM atoms in the buffer and outer regions should be much smaller than the converged standards of the structural optimization (e.g., <2/3 of the criteria). In the situation, the MM gradients on the buffer and outer atoms would be in good agreement with the QM/MM EE computation. It can be observed from Figure 3 that MAXD and RMSD on the WAT2 atoms would satisfy the requirement as the O−O distance being larger than 4.2 Å. The calculations by the HF theory have the same results (see Figure S2 of the Supporting Information).

In the study of the large systems, we first got the structures of the four snapshots (S1500, S1900, S1250, and S1800) for adenine in water from the MD simulations. At the four geometries, we performed the QM/MM EE and MM SP calculations. MUED, MSED, MAXD, and RMSD on the MM atoms in seven different regions, which are the full MM part and MM water molecules excluding the water molecules within 2.4, 3.0, 4.0, 5.0, 6.0, and 9.0 Å around adenine, respectively, were subsequently computed to determine the R_1 value of the DOBR core region. The atom numbers of the selected MM regions are very huge for large systems. Thus, RMSD, MSED, and MUED of four snapshots are extremely small because they are the averaged quantities. Especially, RMSD of seven different MM regions are far less than 0.0002 au. RMSD, MSED, and MUED of four snapshots are respectively listed in Tables S2, S3, and S4 of the Supporting Information. The more important MAXD of the several MM regions for adenine in water are given in Table 1. The results in the table indicate that the gradients on the atoms within 3.0 Å around adenine computed by the MM method are distinctly inconsistent with the QM/MM EE simulations. MAXDs of four snapshots are less than 0.0003 au for the water atoms outside 4.0 Å around adenine. It implies that R_1 in the DOBR optimizations of four snapshots should be at least 4.0 Å.

Moreover, MSED, MUED, RMSD, and MAXD of A3B-CTD and NMAE enzymes were calculated at the geometries optimized by MM (tolerance = 0.0025 au). Here, we define the group distance which is the minimal value among all lengths between each atom in one MM group and QM atom. It was used to establish the MM region, in which MSED, MUED, RMSD, and MAXD were computed. Table 2 presents RMSD and MAXD of A3B-CTD and NMAE enzymes calculated by employing different MM atoms. The several MM parts include the full MM region and MM atoms for which the group distances are larger than 4.0, 5.0, 6.0, 7.0, and 9.0 Å, respectively. The corresponding MSED and MUSED are

DOI: 10.1021/acs.jctc.6b00547
J. Chem. Theory Comput. 2016, 12, 4632−4643

Table 3. Total Step Number of QM Computations (N_{QM}) and Microcycle (N_{mic}) in the TRM Optimizations for Four Snapshots of Adenine in Water Including Two Neutral (S1500 and S1900 Snapshots) and Two Cationic (S1250 and S1800 Snapshots) Systems Using Four Different Core Regions (5.0, 6.0 9.0, and 12.0 Å R_c)

		5.0 Å		6.0 Å		9.0 Å		12.0 Å	
		N_{QM}	N_{mic}	N_{QM}	N_{mic}	N_{QM}	N_{mic}	N_{QM}	N_{mic}
neutral	S1500	813	18	975	27	1167	20	2417	23
	S1900	850	22	884	19	2168	34	3104	37
cation	S1250	638	14	970	19	2185	37	3351	33
	S1800	7377	19	1447	27	1525	27	2053	23

shown in Table S5 of the Supporting Information. RMSDs in Table 2 are very small, which are similar to the calculations of four snapshots of adenine. In contrast, the MAXD decay of two enzymes is slow, and at the 7.0 Å group distance their MAXDs are smaller than the criterion of the structural optimization (0.00045 au). Especially, NMAE MAXD is less than 0.0003 au at the distance.

The above results of the two-water model and large systems including four snapshots of adenine and two enzymes indicate that the ESP charges on QM atoms can only be used to calculate the gradients on MM atoms for which the distances to QM atoms are larger. Then, the gradients computed by the MM method will agree well with the QM/MM EE SP calculations. According to the considerations, the cutoff distance R_1 should be larger than 5.0 Å in the DOBR optimization of nucleobases and 7.0 Å for two enzymes. Finally, it could be deduced that the cutoff distance R_1 in the DOBR scheme should be at least 5.0 Å. One should notice that 5.0 Å R_1 is not exact and fixed. The R_1 value of the DOBR scheme is approximate and varies with the computational systems.

In the DOBR scheme, the interactions between the core and buffer regions are accurately computed by the QM/MM EE approach, and the coupling matrix element $H_{MM_2,c}$ (H_{c,MM_2}) between two regions in eq 3 is included in the QM/MM EE optimization of the core and buffer regions. The interactions of the buffer and outer regions are calculated at the MM level, and the coupling matrix element H_{MM_3,MM_2} (H_{MM_2,MM_3}) is computed in the MM optimization of the buffer and outer regions. The DOBR scheme neglects the interactions between core and outer regions and the corresponding coupling $H_{MM_3,c}$ (H_{c,MM_3}) of the two regions which is related to compute the step size in a second-order optimizer.[44] At a larger cutoff distance R_2 of the buffer region, there are only the electrostatic and van der Waals interactions between the core and outer regions. R_2 determines the magnitude of the interaction energies which are used to calculate the coupling matrix element $H_{MM_3,c}$ (H_{c,MM_3}) of eq 3. It is possible that an appropriate cutoff distance R_2 could cause the coupling matrix element $H_{MM_3,c}$ (H_{c,MM_3}) to be very small.

There are many different formulas to compute the electrostatic and van der Waals energies according to the difference of the force field. However, the value of coupling matrix element could have the same order of $Q_cQ_{MM_3}/R^3_{c,MM_3}$. R_{c,MM_3} is the distance between the atoms of core and outer regions, which the minimal value is R_2. Q_c and Q_{MM_3} respectively are the atomic charges of core and outer regions. As R_2 is equal to 4.0 Å, the $1/R^3_{c,MM_3}$ maximal value is 0.0023 au. The value is close to the default criterion 0.0018 au of the maximal step in the geometry optimization. Due to the parameters of the atomic charge, the absolute value of $Q_cQ_{MM_3}/R^3_{c,MM_3}$ could be much smaller than 0.0023 au. Considering the coupling matrix elements $H_{MM_3,c}$ (H_{c,MM_3}) to be an off-diagonal element, its effect could be much less than 0.0018 au even though the coupling is related to the step size of a second-order optimizer. The result implies that the cutoff distance R_2 of the buffer region should at least be 4.0 Å.

Microiterative Scheme of the Two-Region Partition. For the comparison, the investigation also presents the two-region microiterative (TRM) scheme, which is similar to the previous literature,[27,32,33,35] as shown in Figure S1 of the Supporting Information. The core region of the TRM scheme includes QM atoms and MM atoms within a cutoff distance R_c to the QM atoms. The procedure to carry out the TRM scheme is illustrated by Figure S3 of the Supporting Information. ESP charges on the QM atoms are applied in the MM optimization of the outer region. Although the optimizations of core and outer regions converge at the same geometry, the structure could not be a local minimum of the full system on the QM/MM EE PES. Thus, a direct QM/MM EE optimization of the full system is performed at the end of the TRM scheme. This is the major difference from the former studies.

The cutoff distance R_c could affect the step number of QM calculations in the TRM optimization. In order to study the relation between them, the TRM scheme was used to optimize the geometries of the neutral S1500 and S1900 snapshots and the cationic S1250 and S1800 snapshots for adenine in water. For four snapshots, adenine was calculated by the DFT method employing the B3LYP functional with the 6-31+G* basis set in the QM/MM EE calculations, and MM atoms were computed by the TIP3P water model. Four different R_c values including 5.0, 6.0, 9.0, and 12.0 Å were used in the TRM optimizations of each snapshot. The atom numbers of core regions are listed in Table S6 of the Supporting Information.

Table 3 gives the numbers of QM computational steps (N_{QM}) and the microcycle (N_{mic}) in the TRM optimizations. The converged criteria of the maximum gradient component were the default 0.00045 au in the TRM optimizations. Except for the S1800 snapshot with 5.0 Å R_c, the direct QM/MM optimizations of the entire system at the end of the TRM procedure converged after one cycle for others examples. In the simulations of S1800 snapshot with 5.0 Å R_c, the QM/MM and MM optimizations in the microcycle, in which total 758 QM calculations were performed, reached the convergence at a common geometry after 19 cycles. The direct QM/MM optimization at the end of the TRM procedure had 6619 iterations. It seems that the default standard of the maximum gradient component in the microcycle is too high. Alternatively, the criterion could be reduced to be 0.00015 au. Using the criterion, the structure optimized after 19 microcycles was as the initial geometry. We carried out a new TRM optimization for the S1800 snapshot. The minima of the QM/MM EE and

DOI: 10.1021/acs.jctc.6b00547
J. Chem. Theory Comput. 2016, 12, 4632−4643

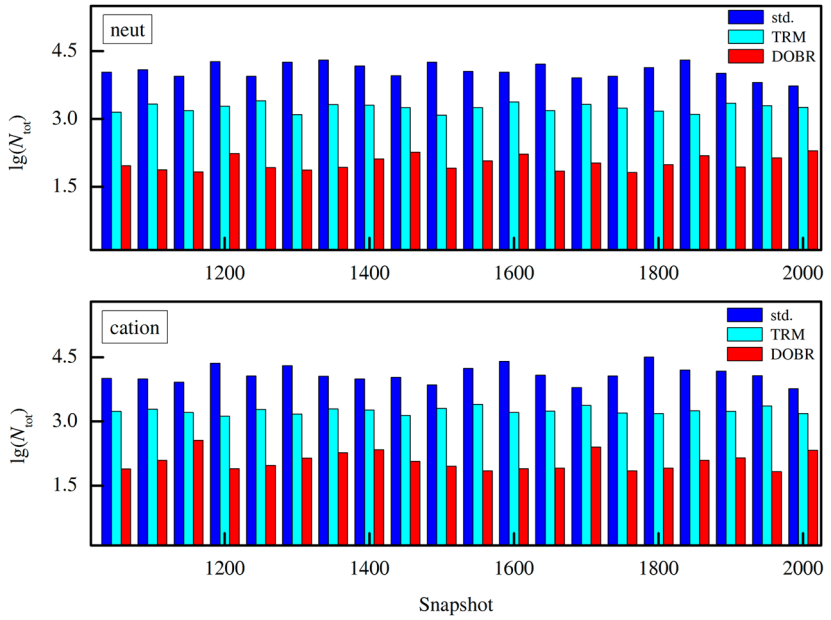


Figure 4. Total step numbers of QM computations (N_{tot}) in the standard direct (std.), two-region microiterative (TRM), and double-optimizations-of-buffer-region (DOBR) microiterative optimizations of the neutral (20 snapshots) and cationic (20 snapshots) adenine systems.

MM energies at a same structure were again obtained after 17 microcycles in which there are 442 QM computations. The subsequent QM/MM EE optimization of the full system has only one cycle. In general, the default criterion 0.00045 au of the maximum gradient component is sufficient for most TRM optimizations.

The results in Table 3 indicate that the step number of QM computations in the TRM optimization increases with the core region enlarging. Thus, the computational cost of the 5.0 Å R_c is cheaper in the TRM optimizations. The microcycle number has not a direct relation to R_c and there are normally dozens of microcycles in one TRM optimization. Comparing with the atom numbers of core regions in Table S6 of the Supporting Information, we found that the growth of the QM step number straightly comes from the addition of the atom number in the core region. The ratio of the QM step number to the atom number in the core region decrease as the cutoff distance R_c increases. Its value is about 6 at the 5.0 Å R_c and about 2 at the 12.0 Å R_c.

Efficiency of the DOBR Microiterative Scheme. Most computational cost of the QM/MM EE calculations at a very high QM level is produced by QM simulation. Thus, step numbers of QM computations in the optimizations using the standard direct QM/MM, TRM, and DOBR methods are applied to study their efficiencies. We optimized the geometries of total 40 snapshots for adenine in water using the three approaches. Among the structures, each of neutral and cationic systems has 20 snapshots. According to the previous discussions, the 5.0 Å R_1 and 4.0 Å R_2 are used in the DOBR optimizations. For comparison, the R_c of the TRM scheme is taken as 9.0 Å, which is corresponding to the MM_1 part and the buffer region of the DOBR scheme. In the treatment, the atom number in the core region of the TRM scheme will be equal to the sum of the core and buffer atoms of the DOBR scheme. Table S1 of the Supporting Information lists the atom number of the MM_1 part, the buffer region, and the outer region of the DOBR scheme.

Figure 4 describes the total step numbers (N_{tot}) of QM calculations in the standard direct QM/MM, TRM, and DOBR optimizations of 40 snapshots for adenine in water. The maximal step number of QM computations for the DOBR method is 363 at the cationic S1150 snapshot. In contrast, the minimal step numbers of QM calculations in the TRM and standard direct QM/MM optimizations respectively are 1222 at the neutral S1500 snapshot and 5376 at the neutral S2000 snapshot. Clearly, the maximal value of the DOBR scheme is much smaller than the TRM minimum and appears little comparing with the minimal number in the standard direct QM/MM optimization. The averaged step number of QM simulations in the DOBR scheme is 122 for total 40 snapshots. The averaged amounts for the TRM and standard direct QM/MM optimizations of 40 snapshots are respectively 1798 and 13144.

We computed the ratios between the step numbers of QM computations (*Rat*) in the optimizations using the three approaches to study the efficiency improvement of the DOBR scheme. They respectively are the standard direct QM/MM optimization to the TRM scheme (std./TRM), the standard direct QM/MM optimization to the DOBR scheme (std./DOBR), and the TRM approach to the DOBR scheme (TRM/DOBR). The three ratios of 40 snapshots for adenine in water are given in Figure 5. Observing the figure, it can be found that the std./TRM maximum is 21 at the cationic S1800 snapshot and its most values are lower than 10. The std./TRM average for 40 snapshots is 8, which is in agreement with the earlier studies.[34,36] The std./DOBR maximum is 394 at the cationic S1800 snapshot and its minimal values is 23 at the cationic S1150 snapshot. The results imply that the step number of QM calculations for the DOBR scheme is 1/394−1/23 of the standard direct QM/MM optimization. The std./DOBR average for 40 snapshots is 134. Obviously, the DOBR scheme dramatically improves the computational efficiency in contrast to the standard direct QM/MM optimization. In addition, the TRM/DOBR maximal value is 36 and locates at the cationic S1550 snapshot. Most TRM/DOBRs are between 10 and 30

DOI: 10.1021/acs.jctc.6b00547
J. Chem. Theory Comput. 2016, 12, 4632−4643

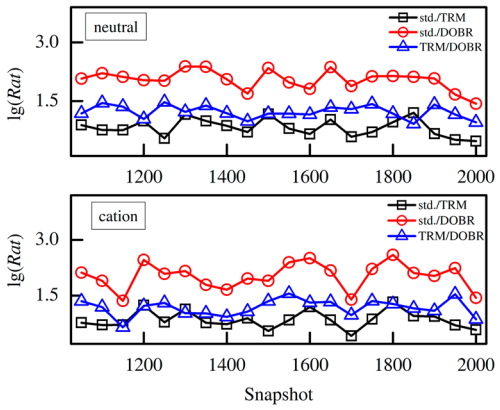


Figure 5. Ratios of total QM computational steps (*Rat*) in standard direct, TRM, and DOBR optimizations for the neutral (20 snapshots) and cationic (20 snapshots) adenine systems. std./TRM corresponds to standard direct optimization to the TRM scheme, and std./DOBR denotes standard direct optimization to the DOBR microiterative scheme. The TRM approach to the DOBR microiterative scheme is given by TRM/DOBR.

and the TRM/DOBR average is 17. The DOBR scheme seems to promote the computational efficiency by at least 1 order of magnitude comparing with the TRM optimization.

We analyzed the microcycle number in the TRM and DOBR optimizations to study the main contribution of the DOBR treatment to the efficiency promotion. Figure 6 displays the

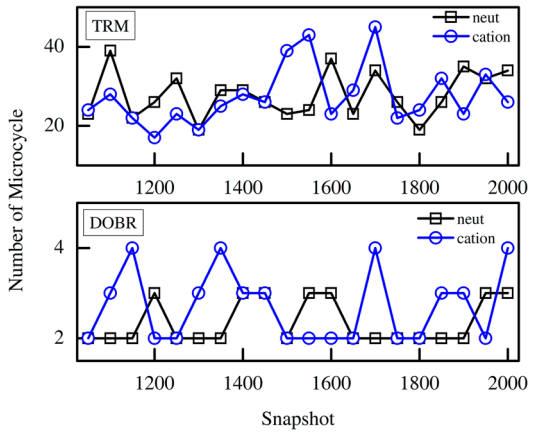


Figure 6. Microcycle number in TRM and DOBR microiterative optimizations of the neutral (20 snapshots) and cationic (20 snapshots) adenine systems.

microcycle number of 40 snapshots. For the TRM scheme, all microcycle numbers of 40 snapshots are larger than 15, and its most values are between 20 and 40. The averaged microcycle number is 28 for 40 snapshots. In the DOBR optimizations of 40 snapshots, all microcycle numbers are less than 5, and the average of the microcycle number is 3. We compared the averaged microcycle number of the TRM and DOBR schemes. It can be found that the averaged microcycle number of the TRM scheme is about 9 times of that for the DOBR optimization. It is evident that the DOB scheme can significantly decrease the microcycle number in the optimization.

DOBR Microiterative Optimizations of Additional Systems. We also optimized the structures of cytosine and guanine in water using the DOBR scheme to further explore its efficiency. Figure 7 presents step numbers of QM computations in the DOBR optimizations of 120 geometries, in which QM

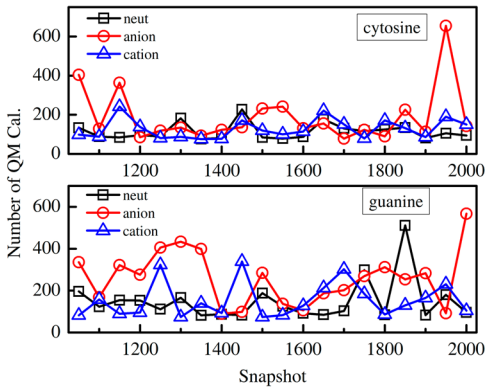


Figure 7. Step numbers of QM computations in the DOBR microiterative optimization of the neutral (20 snapshots), anionic (20 snapshots), and cationic (20 snapshots) systems for cytosine and guanine in water.

regions of the QM/MM EE computations only contain the nucleobases atoms. For each of the neutral, anionic, and cationic species of cytosine and guanine in water, we optimized the structures of 20 snapshots applying the DOBR scheme. R_1 and R_2 in the DOBR scheme respectively are 5.0 and 4.0 Å. The corresponding atom numbers in the MM_1 part and the buffer region are given in Table S1 of the Supporting Information.

In the DOBR optimizations of the 120 snapshots, there are several hundreds of QM computational steps for each snapshot. Except for the S1950 snapshot of anionic cytosine, the microcycle numbers of others 119 snapshots are less than 5. The DOBR optimization of the anionic cytosine S1950 snapshot requires 654 QM calculations and 7 microcycles. The averaged step numbers of QM calculations for the neutral, cationic, and anionic cytosine systems are 113, 128, and 188, respectively. For guanine systems, the averaged values are slight larger than those of cytosine, which are 150 for neutral species, 155 for cation, and 261 for anion. However, the DOBR scheme can be successfully applied in the optimizations of 120 snapshots and exhibits a powerful performance.

Moreover, we utilized the DOBR scheme to optimize the geometries of cytosine S1050 and guanine S1750 snapshots with the larger QM regions in the QM/MM calculations. In the computations of guanine S1750 snapshot, the QM part consists of a guanine molecule and water molecules which the distance to guanine are within 4.0 Å and contains 108 atoms. For cytosine S1050 snapshot, the QM region is composed of cytosine atoms and water molecules of the distances 4.2 Å to cytosine and includes 100 atoms. Figure 8 shows the QM/MM EE SP energy of each step in the DOBR optimizations for the two snapshots. 4.8 Å R_1 and 4.0 Å R_2 were used in the DOBR optimizations of the cytosine S1050 snapshot. For the guanine S1750 snapshot, 5.0 Å R_1 and 4.0 Å R_2 were employed. It can be observed from the figure that the QM/MM EE energies have several distinct vertical decreases. The features are produced by the MM optimizations of the buffer and outer regions in the microcycle. The DOBR optimizations would converge within several hundreds of QM computations. The results imply that the DOBR scheme can have a high efficiency in the QM/MM EE simulation using a larger QM region.

DOI: 10.1021/acs.jctc.6b00547
J. Chem. Theory Comput. 2016, 12, 4632−4643

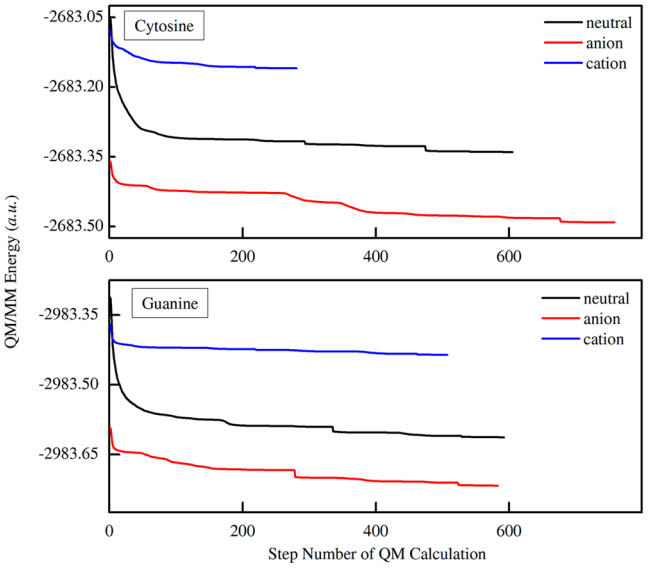


Figure 8. QM/MM energies of the two snapshots including cytosine S1050 and guanine S1750 in the DOBR microiterative optimization.

For the optimizations of the rigid structures, the efficiency promotion of the DOBR optimization could be smaller in contrast to the TRM scheme. We optimized the structure of two gas enzymes (A3B-CTD and NMAE) by using the DOBR and TRM schemes. The initial geometries of the QM/MM optimizations were obtained by relaxing the crystal structures at the MM level. 11.0 Å R_c was employed in TRM optimizations and for the DOBR scheme R_1 and R_2 were respectively 7.0 and 4.0 Å. Figure 9 shows the QM/MM SP energies of each step in

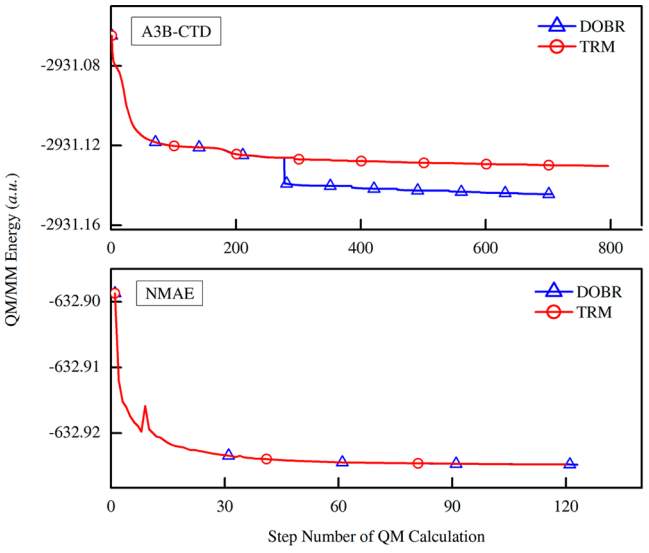


Figure 9. QM/MM energies of A3B-CTD and NMAE two gas enzymes in the DOBR and TRM microiterative optimizations.

the DOBR and TRM optimizations for the two enzymes. For A3B-CTD enzyme, the DOBR scheme slightly improves the efficiency comparing with the TRM optimization. The numbers of QM computational step in the DOBR and TRM optimizations for NMAE enzyme are very close. The result indicates that the DOBR scheme could be reduced to the TRM scheme in the situation of very small coupling between the buffer and outer regions. However, it could be expected that the DOBR scheme would have a high efficiency in the optimizations of the enzymes solvated in water.

Overall Assessment. The structures of nucleobases in water are flexible. QM/MM PESs of these systems are very flat and there are many local minima on them. The DOBR, TRM, and standard optimizations walk along the different paths and could reach the different minima. Figure S4 of the Supporting Information displays the final optimized QM/MM energies obtained by the DOBR, TRM, and standard methods for 40 snapshots of adenine in water. It can be found in the figure that each approach has the lowest QM/MM energies for about 1/3 of 40 snapshots. In the structural optimization of two gas enzymes, the difference of QM/MM energy optimized by the DOBR and TRM schemes is 0.39 kcal/mol for A3B-CTD and 0.0011 kcal/mol for NMAE using the standard converged criterion (tolerance = 0.00045 au). When the criterion is tight (tolerance = 0.000045 au), the A3B-CTD difference would reduced to be 0.0013 kcal/mol. The results imply that the DOBR and TRM schemes can converge to a same minimum of the QM/MM energy (See Figure 9) in the optimizations of rigid structures.

Using the DOBR scheme, one does not modify the source code of the optimizer because of the MM_1 presence. The DOBR scheme can easily be carried out by an external script and employing the different optimizer. Nevertheless, the MM_1 part and the buffer region of the DOBR scheme determine that the model can only be applied to optimize the geometry of the very large system (e.g., $MM_1 + MM_2 + MM_3 > 12$ Å). The MM_1 part of the core region increases the number of active atoms in the QM/MM EE optimization of core and buffer regions and the step number of QM computations. The approaches suggested to correct the atom gradients of the outer region in the microiterative optimization[30,34,36,37] can be applied to cut off the MM_1 part of the DOBR scheme. Using the methods, the additional efforts have to be paid for the modification of the source code of the optimizer. After cutting the MM_1 part, the step number of QM calculations in the DOBR scheme probably is further reduced by about a half according to the results in Table 3. In the QM/MM ME calculations, the DOBR scheme can be directly used to optimize the structures without the MM_1 part. In the DOBR optimizations of nucleobases in water, the 5.0 Å R_1, 4.0 Å R_2, and default converged criteria are enough to optimize the structure of the local minima of the QM/MM EE energies. For the enzyme optimization, R_1 of the DOBR scheme could be at least 7.0 Å. The DOBR scheme is an appropriate approach to optimize the geometries of local minima of the QM/MM EE energy for the large flexible systems and specially scan the PES of the QM/MM EE calculations.

V. CONCLUSIONS

In this work, we describe a DOBR microiterative scheme to optimize the geometry of the energy local minima for the large flexible system in the QM/MM calculations using electrostatic embedding. In the approach, an entire system is divided three regions which are the core, buffer, and outer regions. The core region is larger than the QM part of QM/MM simulation and consists of the QM and MM_1 atoms. In the procedure to perform the DOBR scheme, the buffer and outer regions are optimized at the MM level by fixing the core region and the ESP charges on QM atoms are used to compute the interaction between QM atoms and the buffer and outer regions. Subsequently, the structural optimization of the core and

DOI: 10.1021/acs.jctc.6b00547
J. Chem. Theory Comput. 2016, 12, 4632−4643

buffer regions is carried out by applying the QM/MM method and freezing the atoms of the outer region. The two optimizations are iteratively performed until they reach the converged criteria at a common geometry. Hence, the buffer region would be optimized twice in each microcycle of the DOBR scheme. One is the optimization at the MM level and the other is at the QM/MM level. We add a QM/MM optimization of the full system at the end of the procedure to check the structure.

The cutoff distances R_1 and R_2 are the keys in the DOBR scheme. We determine the cutoff distances R_1 according to MSED, MUED, MAXD, and RMSD because the gradients of MM atoms are very important in the MM optimization. The four quantities reveal the gradient deviations on the MM atoms between the QM/MM SP calculations and the MM SP simulations in which QM atoms were assigned ESP charges. We analyzed the four quantities of a two-water system, four snapshots for adenine in water (S1500, S1900, S1250, and S1800), and two gas enzymes (A3B-CTD and NMAE). We focus on RMSD and MAXD which are related to the converged criteria of the geometry optimization and compare the two functions with the default converged standard of the structural optimization. The results indicate that the cutoff distance R_1 is at least 5.0 Å for the nucleobases in water and 7.0 Å for the gas enzymes. Furthermore, we studied the Hessian matrix in the DOBR optimization based on the formulas of the force field. The numerical analysis of the off-diagonal Hessian matrix elements neglected by the DOBR scheme suggests that at least 4.0 Å R_2 should be applied in the DOBR optimization.

The 40 geometries for adenine in water, which contain the neutral systems of 20 snapshots and cation of 20 snapshots, were optimized by the standard QM/MM approach, the TRM method, and the DOBR scheme. The performances of three methods were measured by their total step numbers of QM calculations in the optimizations. Generally, the step number of QM computation for the DOBR scheme is very small in comparison with the standard direct and TRM optimizations. On average, the step number of QM computation for the DOBR scheme is ∼1% (1/134) of the QM step amount for the standard direct QM/MM optimization and ∼6% (1/17) for the TRM approach. Comparing with the former work of two-region microiterative schemes,[34,36] the DOBR scheme reduces the computational cost by about ten times. The DOBR scheme is also used to optimize the structures of the cytosine and guanine in water. All DOBR optimizations will converge within 800 QM calculations. These show that the DOBR approach also has a high efficiency in the tests of the large flexible systems. Moreover, the DOBR and TRM schemes were used to optimize the geometries of two gas enzymes. The results indicate that the efficiency improvement of the DOBR scheme is not prominent in the optimizations of rigid structures.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.6b00547.

TRM partitioning (Figure S1). MUSED, MSED, MAXD, and RMSD at the HF level (Figure S2). TRM procedure (Figure S3). QM/MM optimized energies (Figure S4). Atom number of DOBR different regions (Table S1). RMSD of four snapshots (Table S2). MSED of four snapshots (Table S3). MUED of four snapshots (Table S4). MSED and MUED of two enzymes (Table S5). Atom number of TRM core regions for four snapshots (Table S6) (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

*E-mail: zhangyanhg@dicp.ac.cn.

*E-mail: klhan@dicp.ac.cn.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

Y.Z. thanks Prof. Hai Lin for helpful discussions. This work was supported by a grant from the National Natural Science Foundation of China (21403225).

■ REFERENCES

(1) Senn, H. M.; Thiel, W. QM/MM Methods for Biological Systems. *Top. Curr. Chem.* **2007**, *268*, 173−290.

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

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

(4) Shaik, S.; Cohen, S.; Wang, Y.; Chen, H.; Kumar, D.; Thiel, W. P450 Enzymes: Their Structure, Reactivity, and Selectivity-Modeled by QM/MM Calculations. *Chem. Rev.* **2010**, *110*, 949−1017.

(5) Acevedo, O.; Jorgensen, W. L. Advances in Quantum and Molecular Mechanical(QM/MM) Simulations for Organic and Enzymatic Reactions. *Acc. Chem. Res.* **2010**, *43*, 142−151.

(6) 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.

(7) Brunk, E.; Rothlisberger, U. Mixed Quantum Mechanical/Molecular Mechanical Molecular Dynamics Simulations of Biological Systems in Ground and Electronically Excited States. *Chem. Rev.* **2015**, *115*, 6217−6263.

(8) Bakowies, D.; Thiel, W. Hybrid models for combined quantum mechanical and molecular mechanical approaches. *J. Phys. Chem.* **1996**, *100*, 10580−10594.

(9) Weiner, S. J.; Singh, U. C.; Kollman, P. A. Simulation of Formamide Hydrolysis by Hydroxide Ion in the Gas Phase and in Aqueous Solution. *J. Am. Chem. Soc.* **1985**, *107*, 2219−2229.

(10) Maseras, F.; Morokuma, K. IMOMM: A New Integrated *Ab Initio* + Molecular Mechanics Geometry Optimization Scheme of Equilibrium Structures and Transition States. *J. Comput. Chem.* **1995**, *16*, 1170−1179.

(11) Svensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K. ONIOM: A Multilayered Integrated MO + MM Method for Geometry Optimizations and Single Point Energy Predictions. A Test for Diels-Alder Reactions and $Pt(P(t\text{-}Bu)_3)_2 + H_2$ Oxidative Addition. *J. Phys. Chem.* **1996**, *100*, 19357−19363.

(12) Vreven, T.; Byun, K. S.; Komáromi, I.; Dapprich, S.; Montgomery, J. A., Jr.; Morokuma, K.; Frisch, M. J. Combining Quantum Mechanics Methods with Molecular Mechanics Methods in ONIOM. *J. Chem. Theory Comput.* **2006**, *2*, 815−826.

(13) Singh, U. C.; Kollman, P. A. A Combined *Ab Initio* Quantum Mechanical and Molecular Mechanical Method for Carrying out Simulations on Complex Molecular Systems: Applications to the $CH_3C1 + C1^-$ Exchange Reaction and Gas Phase Protonation of Polyethers. *J. Comput. Chem.* **1986**, *7*, 718−730.

(14) de Vries, A. H.; Sherwood, P.; Collins, S. J.; Rigby, A. M.; Rigutto, M.; Kramer, G. J. Zeolite Structure and Reactivity by Combined Quantum-Chemical-Classical Calculations. *J. Phys. Chem. B* **1999**, *103*, 6133−6141.

DOI: 10.1021/acs.jctc.6b00547
J. Chem. Theory Comput. 2016, 12, 4632−4643

(15) Eichinger, M.; Tavan, P.; Hutter, J.; Parrinello, M. A hybrid method for solutes in complex solvents: Density functional theory combined with empirical force fields. *J. Chem. Phys.* **1999**, *110*, 10452−10467.
(16) Das, D.; Eurenius, K. P.; Billings, E. M.; Sherwood, P.; Chatfield, D. C.; Hodošček, M.; Brooks, B. R. Optimization of quantum mechanical molecular mechanical partitioning schemes: Gaussian delocalization of molecular mechanical charges and the double link atom method. *J. Chem. Phys.* **2002**, *117*, 10534−10547.
(17) Lin, H.; Truhlar, D. G. Redistributed Charge and Dipole Schemes for Combined Quantum Mechanical and Molecular Mechanical Calculations. *J. Phys. Chem. A* **2005**, *109*, 3991−4004.
(18) Gao, J. L. Energy Components of Aqueous Solution: Insight from Hybrid QM/MM Simulations Using a Polarizable Solvent Model. *J. Comput. Chem.* **1997**, *18*, 1061−1071.
(19) Gascon, J. A.; Leung, S. S. F.; Batista, E. R.; Batista, V. S. A Self-Consistent Space-Domain Decomposition Method for QM/MM Computations of Protein Electrostatic Potentials. *J. Chem. Theory Comput.* **2006**, *2*, 175−186.
(20) Zhang, Y.; Lin, H.; Truhlar, D. G. Self-Consistent Polarization of the Boundary in the Redistributed Charge and Dipole Scheme for Combined Quantum-Mechanical and Molecular-Mechanical Calculations. *J. Chem. Theory Comput.* **2007**, *3*, 1378−1398.
(21) Biswas, P. K.; Gogonea, V. A Polarizable Force-Field Model for Quantum-Mechanical-Molecular-Mechanical Hamiltonian Using Expansion of Point Charges into Orbitals. *J. Chem. Phys.* **2008**, *129*, 154108.
(22) Lu, Z.; Zhang, Y. Interfacing *Ab Initio* Quantum Mechanical Method with Classical Drude Osillator Polarizable Model for Molecular Dynamics Simulation of Chemical Reactions. *J. Chem. Theory Comput.* **2008**, *4*, 1237−1248.
(23) Boulanger, E.; Thiel, W. Solvent Boundary Potentials for Hybrid QM/MM Computations Using Classical Drude Oscillators: A Fully Polarizable Model. *J. Chem. Theory Comput.* **2012**, *8*, 4527−4538.
(24) Maseras, F.; Morokuma, K. IMOMM: A New Integrated *Ab Initio* + Molecular Mechanics Geometry Optimization Scheme of Equilibrium Structures and Transition States. *J. Comput. Chem.* **1995**, *16*, 1170−1179.
(25) Ryde, U. The Coordination of the Catalytic Zinc Ion in Alcohol Dehydrogenase Studied by Combined Quantum-Chemical and Molecular Mechanics Calculations. *J. Comput.-Aided Mol. Des.* **1996**, *10*, 153−164.
(26) Turner, A. J.; Moliner, V.; Williams, I. H. Transition-state Structural Refinement with GRACE and CHARMM: Flexible QM/MM Modelling for Lactate Dehydrogenase. *Phys. Chem. Chem. Phys.* **1999**, *1*, 1323−1331.
(27) 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.
(28) Sierka, M.; Sauer, J. Finding Transition Structures in Extended Systems: A Strategy Based On a Combined Quantum Mechanics−Empirical Valence Bond Approach. *J. Chem. Phys.* **2000**, *112*, 6983−3496.
(29) Hall, R. J.; Hindle, S. A.; Burton, N. A.; Hillier, I. H. Aspects of Hybrid QM/MM Calculations: The Treatment of the QM/MM Interface Region and Geometry Optimization with an Application to Chorismate Mutase. *J. Comput. Chem.* **2000**, *21*, 1433−1441.
(30) Murphy, R. B.; Philipp, D. M.; Friesner, R. A. A Mixed Quantum Mechanics/Molecular Mechanics (QM/MM) Method for Large-Scale Modeling of Chemistry in Protein Environments. *J. Comput. Chem.* **2000**, *21*, 1442−1457.
(31) Billeter, S. R.; Turner, A. J.; Thiel, W. Linear Scaling Geometry Optimisation and Transition State Search in Hybrid Delocalised Internal Coordinates. *Phys. Chem. Chem. Phys.* **2000**, *2*, 2177−2186.
(32) Hayashi, S.; Ohmine, I. Proton Transfer in Bacteriorhodopsin: Structure, Excitation, IR Spectra, and Potential Energy Surface Analyses by an ab Initio QM/MM Method. *J. Phys. Chem. B* **2000**, *104*, 10678−10691.
(33) Prat-Resina, X.; González-Lafont, À.; Lluch, J. M. How Important is the Refinement of Transition State Structures in Enzymatic Reactions? *J. Mol. Struct.: THEOCHEM* **2003**, *632*, 297−307.
(34) Vreven, T.; Morokuma, K.; Farkas, Ö.; Schlegel, H. B.; Frisch, M. J. Geometry Optimization with QM/MM, ONIOM, and Other Combined Methods. I. Microiterations and Constraints. *J. Comput. Chem.* **2003**, *24*, 760−769.
(35) Prat-resina, X.; Bofill, J. M.; González-lafont, Á.; Lluch, J. M. Geometry Optimization and Transition State Search in Enzymes: Different Options in the Microiterative Method. *Int. J. Quantum Chem.* **2004**, *98*, 367−377.
(36) Kästner, J.; Thiel, S.; Senn, H. M.; Sherwood, P.; Thiel, W. Exploiting QM/MM Capabilities in Geometry Optimization: A Microiterative Approach Using Electrostatic Embedding. *J. Chem. Theory Comput.* **2007**, *3*, 1064−1072.
(37) Rokob, T. A.; Rulíšek, L. Curvature Correction for Microiterative Optimizations with QM/MM Electronic Embedding. *J. Comput. Chem.* **2012**, *33*, 1197−1206.
(38) Lundberg, M.; Kawatsu, T.; Vreven, T.; Frisch, M. J.; Morokuma, K. Transition States in a Protein Environment − ONIOM QM:MM Modeling of Isopenicillin N Synthesis. *J. Chem. Theory Comput.* **2009**, *5*, 222−234.
(39) Maeda, S.; Ohno, K.; Morokuma, K. An Automated and Systematic Transition Structure Explorer in Large Flexible Molecular Systems Based on Combined Global Reaction Route Mapping and Microiteration Methods. *J. Chem. Theory Comput.* **2009**, *5*, 2734−2743.
(40) Maeda, S.; Abe, E.; Hatanaka, M.; Taketsugu, T.; Morokuma, K. Exploring Potential Energy Surfaces of Large Systems with Artificial Force Induced Reaction Method in Combination with ONIOM and Microiteration. *J. Chem. Theory Comput.* **2012**, *8*, 5058−5063.
(41) Polyak, I.; Boulanger, E.; Sen, K.; Thiel, W. A Microiterative Intrinsic Reaction Coordinate Method for Large QM/MM Systems. *Phys. Chem. Chem. Phys.* **2013**, *15*, 14188−14195.
(42) Jung, J.; Re, S.; Sugita, Y.; Ten-no, S. Improved Constrained Optimization Method for Reaction-Path Determination in the Generalized Hybrid Orbital Quantum Mechanical/Molecular Mechanical Calculations. *J. Chem. Phys.* **2013**, *138*, 044106.
(43) Ruiz-Barragan, S.; Morokuma, K.; Blancafort, L. Conical Intersection Optimization Using Composed Steps Inside the ONIOM(QM:MM) Scheme: CASSCF:UFF Implementation with Microiterations. *J. Chem. Theory Comput.* **2015**, *11*, 1585−1594.
(44) Fensen, K. In *Introduction to Computational Chemistry*, 2nd ed.; John Wiley & Sons Ltd.: Chichester, West Sussex, England, 2006; Chapter 12, pp 383−389.
(45) Singh, U. C.; Kollman, P. A. An Approach to Computing Electrostatic Charges for Molecules. *J. Comput. Chem.* **1984**, *5*, 129−145.
(46) 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*, 926−935.
(47) Roothaan, C. C. J. New Developments in Molecular Orbital Theory. *Rev. Mod. Phys.* **1951**, *23*, 69−89.
(48) Eichkorn, K.; Treutler, O.; Öhm, H.; Häser, M.; Ahlrichs, R. Auxiliary Basis Sets to Approximate Coulomb Potentials (erratum, 1995, 242, 283). *Chem. Phys. Lett.* **1995**, *242*, 652−660.
(49) 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: Condens. Matter Mater. Phys.* **1988**, *37*, 785−789.
(50) Becke, A. D. Density-Functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98*, 5648−5652.
(51) Ahlrichs, R.; Furche, F.; Grimme, S. Comment on "Assessment of Exchange Correlation Functional. *Chem. Phys. Lett.* **2000**, *325*, 317−321.
(52) Hehre, W. J.; Stewart, R. F.; Pople, J. A. Self-Consistent Molecular Orbital Methods. 1. Use of Gaussian Expansions of Slater-Type Atomic Orbitals. *J. Chem. Phys.* **1969**, *51*, 2657−2664.

DOI: 10.1021/acs.jctc.6b00547
J. Chem. Theory Comput. 2016, 12, 4632−4643

(53) Collins, J. B.; Schleyer, P. v. R.; Binkley, J. S.; Pople, J. A. Self-Consistent Molecular Orbital Methods. 17. Geometries and Binding Energies of Eecond-Row Molecules. A comparison of Three Basis Sets. *J. Chem. Phys.* **1976**, *64*, 5142−5151.
(54) Ditchfield, R.; Hehre, W. J.; Pople, J. A. Self-Consistent Molecular Orbital Methods. IX. Extended Gaussian-Type Basis for Molecular-Orbital Studies of Organic Molecules. *J. Chem. Phys.* **1971**, *54*, 724−728.
(55) 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.
(56) Francl, M. M.; Pietro, W. J.; Hehre, W. J.; Binkley, J. S.; DeFrees, D. J.; Pople, J. A.; Gordon, M. S. Self-Consistent Molecular Orbital Methods. 23. A Polarization-Type Basis Set for 2nd-Row Elements. *J. Chem. Phys.* **1982**, *77*, 3654−3665.
(57) Frisch, M. J.; Pople, J. A.; Binkley, J. S. Self-Consistent Molecular Orbital Methods. 25. Supplementary Functions for Gaussian Basis Sets. *J. Chem. Phys.* **1984**, *80*, 3265−3269.
(58) Clark, T.; Chandrasekhar, J.; Spitznagel, G. W.; Schleyer, P. v. R. Efficient Diffuse function-Augmented Basis-Sets for Anion Calculations. 3. The 3-21+G Basis Set for 1st-Row Elements, Li-F. *J. Comput. Chem.* **1983**, *4*, 294−301.
(59) Byeon, I.-J. L.; Byeon, C.-H.; Wu, T.; Mitra, M.; Singer, D.; Levin, J. G.; Gronenborn, A. M. Nuclear Magnetic Resonance Structure of the APOBEC3B Catalytic Domain: Structural Basis for Substrate Binding and DNA Deaminase Activity. *Biochemistry* **2016**, *55*, 2944−2959.
(60) Mabbitt, P. D.; Correy, G. J.; Meirelles, T.; Fraser, N. J.; Coote, M. L.; Jackson, C. J. Conformational Disorganization within the Active Site of a Recently Evolved Organophosphate Hydrolase Limits Its Catalytic Efficiency. *Biochemistry* **2016**, *55*, 1408−1417.
(61) Brooks, B. R.; Brooks, C. L., III; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: The Biomolecular simulation Program. *J. Comput. Chem.* **2009**, *30*, 1545−1615.
(62) Sherwood, P.; de Vries, A. H.; Guest, M. F.; Schreckenbach, G.; Catlow, C. R. A.; French, S. A.; Sokol, A. A.; Bromley, S. T.; Thiel, W.; Turner, A. J.; Billeter, S.; Terstegen, F.; Thiel, S.; Kendrick, J.; Rogers, S. C.; Casci, J.; Watson, M.; King, F.; Karlsen, E.; Sjøvoll, M.; Fahmi, A.; Schäfer, A.; Lennartz, C. QUASI: A General Purpose Implementation of the QM/MM Approach and Its Application to Problems in Catalysis. *J. Mol. Struct.: THEOCHEM* **2003**, *632*, 1−28.
(63) TURBOMOLE V6.4 2012, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989−2007, TURBOMOLE GmbH, since 2007. Available from http://www.turbomole.com (accessed Aug 16, 2016).
(64) Kästner, J.; Carr, J. M.; Keal, T. W.; Thiel, W.; Wander, A.; Sherwood, P. DL-FIND: An Open-Source Geometry Optimizer for Atomistic Simulations. *J. Phys. Chem. A* **2009**, *113*, 11856−11865.

DOI: 10.1021/acs.jctc.6b00547
J. Chem. Theory Comput. 2016, 12, 4632−4643