

QM/MM Reweighting Free Energy SCF for Geometry Optimization on Extensive Free Energy Surface of Enzymatic Reaction

Takahiro Kosugi and Shigehiko Hayashi*

Department of Chemistry, Graduate School of Science, Kyoto University, Kyoto 606-8502, Japan

 Supporting Information

ABSTRACT: We developed a quantum mechanical/molecular mechanical (QM/MM) free energy geometry optimization method by which the geometry of a quantum chemically treated (QM) molecule is optimized on a free energy surface defined with thermal distribution of the surrounding molecular environment obtained by molecular dynamics simulation with a molecular mechanics (MM) force field. The method called QM/MM reweighting free energy self-consistent field combines a mean field theory of QM/MM free energy geometry optimization developed by Yamamoto (Yamamoto, T. *J. Chem. Phys.* **2008**, *129*, 244104) with a reweighting scheme for updating the MM distribution introduced by Hu et al. (Hu, H., et al. *J. Chem. Phys.* **2008**, *128*, 034105) and features high computational efficiency suitable for exploring the reaction free energy surface of extensive protein conformational space. The computational efficiency with improved treatment of a long-range electrostatic (ES) interaction using the Ewald summation technique permits one to take into account global conformational relaxation of an entire protein of an enzyme in the free energy geometry optimization of its reaction center. We applied the method to an enzymatic reaction of a substrate complex of psychrophilic α -amylase from Antarctic bacterium *Pseudoalteromonas haloplanktis* and succeeded in geometry optimizations of the reactant and the product of the catalytic reaction that involve large conformational changes of protein loops adjacent to the reaction center on time scales reaching sub-microseconds. We found that the adjacent loops in the reactant and the product form in different conformations and produce catalytic ES potentials on the reaction center.

INTRODUCTION

Catalytic reaction mechanisms in enzymes have been interesting and important topics in chemistry and biology.^{1,2} Theoretically, a widely used technique to study enzymatic reactions is a combined quantum mechanical/molecular mechanical (QM/MM) method.^{3–7} In this technique, the enzymatic reaction center is described quantum mechanically, while the surrounding environment of biomacromolecules such as proteins is treated by using a molecular mechanics (MM) force field. The method allows one to take into account complex environmental effects of enzymes for catalysis very efficiently, providing powerful means for studying enzymatic chemical reactions.

Despite the success of the QM/MM approach, however, the method suffers from a difficulty due to the high computational cost of the included QM part. Because of the time-consuming QM calculation, the potential energy profile in only one conformation of a protein is examined in a conventional QM/MM procedure, whereas the chemical reaction proceeds in thermal fluctuation. Thus its energetics are characterized in terms of free energy determined by the thermal average of conformations. In particular, such an approach based on the potential energy surface cannot be applied to enzymatic reaction processes where the chemical steps are correlated with protein conformational changes.

Although, in principle, statistical samples of protein conformations can be collected from a molecular dynamics (MD) simulation with a QM/MM Hamiltonian, the high computational cost of the QM/MM method limits severely the sampling calculation, leading to a poor statistical convergence of the sampling. In particular, a very long trajectory calculation for the sampling is

required for protein systems since the structural relaxation of the protein is known to be very slow.^{8–10} Such a slow relaxation motion of the protein is also suggested to be coupled with a local event of reaction.¹¹ The relaxation time could be tens of picoseconds,¹¹ which is more than 2 orders of magnitude longer than that of water solvent,¹² indicating that a much longer MD calculation for protein than for water solution is necessary for obtaining a properly converged statistical sampling.

It should be noted that use of an adequate QM method in the QM/MM treatment is also crucial for accurate evaluation of the reaction profile. Since catalytic reaction centers are often highly polar, extended QM molecular regions with large basis functions are necessary for a description of the complex electronic nature including polarization and charge transfer. An increase in the computational cost of the QM method for the complex catalytic reaction centers therefore introduces a serious dilemma between the accuracies of the QM description and the statistical sampling.

Several QM/MM methodologies that can take into account thermal fluctuation of a protein conformation for the examination of an enzymatic reaction profile have been developed. MD simulations with QM/MM Hamiltonians have been carried to evaluate reaction free energy surfaces along reaction coordinates. In such straightforward approaches, to improve the convergence of statistical samplings, QM/MM Hamiltonians are approximated with semiempirical methods^{13–18} or empirical potential energy functions^{19–21} of which parameters were determined by more accurate QM or QM/MM methods.

Received: August 20, 2011

Published: November 09, 2011

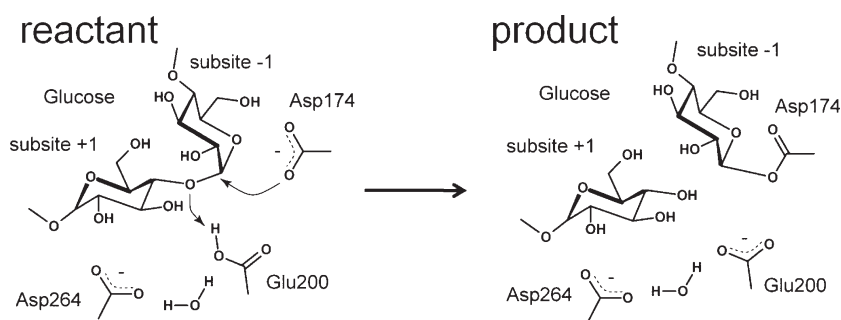


Figure 1. Reaction scheme of cleavage of the α -glycosidic bond in α -amylase catalysis studied in the present study. The product state corresponds to an intermediate of the overall α -amylase catalytic reaction proposed previously.³⁶ Numbers, +1 and -1, indicate the subsites of the substrate.

The other approaches, i.e., QM/MM free energy geometry optimizations,^{22–30} have also been developed extensively. In these approaches, reaction paths are determined by geometry optimizations of enzymatic reaction groups represented by QM/MM Hamiltonians on free energy surfaces defined by statistical samples of the surrounding protein conformations described by MM force fields. Although the approaches omit explicit sampling of the conformations of QM molecules, the omission separates the time-consuming QM calculation and the MM conformational sampling one, enabling one to utilize highly accurate ab initio QM methodologies directly for the description of large QM molecules and to enrich statistical samples of the surrounding protein conformations. However, the convergence of the statistical sampling for enzymes has not been well assessed since the methodologies developed so far are still not efficient enough to obtain sufficient statistical samples for protein systems with very slow conformational relaxation.

In the present study, we developed an efficient QM/MM free energy geometry optimization method that combines a method based on a mean field approximation developed by Yamamoto³⁰ with a reweighting update scheme for the statistical ensemble of a protein conformation introduced by Yang and co-workers.²⁵ Unlike others, the present method, called the QM/MM reweighting free energy self-consistent field (QM/MM-RWFE-SCF), determines a fully variationally electronic wave function of the QM molecules in the surrounding molecular field represented in a mean field manner, leading to a great efficiency in the computational procedure and an improvement of convergence behavior of the geometry optimization necessary for examination of the protein systems. We also incorporated the Ewald summation technique in evaluation of the QM/MM electrostatic (ES) interaction for a periodic boundary condition (PBC) system. In addition to improved accuracy of the QM/MM ES interaction with full electrostatics, the Ewald method provides a consistent description of the ES interaction of the QM/MM method with that of existing highly sophisticated MD program packages, enabling one to obtain efficiently sufficient statistical samples of the protein conformations by the latter.

The QM/MM-RWFE-SCF method was applied to an enzymatic reaction of psychrophilic α -amylase from the Antarctic bacterium *Pseudoalteromonas haloplanktis*. This enzyme and its homologues exhibit a notable temperature dependence for the enzymatic activity, which is suggested to originate from the difference in protein structural flexibility.^{31–35} For the first step to understanding the role of the protein structural flexibility in the enzymatic catalysis, we determined free energetically optimal structures of the active site in a catalytic reaction step schematically

depicted in Figure 1. The reaction, cleavage of the α -glycosidic bond, is the first step of the overall catalytic reaction process proposed previously,³⁶ and thus the state labeled as the product in the present study corresponds to an intermediate state of the overall catalysis.

The high efficiency provided by the present method drastically increases the sizes of the enzyme systems and protein conformational samples treated; we identified the optimized structures of the QM catalytic site described by an ab initio QM method with more than 600 basis functions on free energy surfaces of the conformational samples of the MM protein consisting of more than 68 000 atoms obtained by MD simulations for tens of nanoseconds. In particular, we succeeded in optimizing a structure with the MM conformational samples obtained by a MD trajectory for ~ 90 ns, which involves large and slow conformational changes at a loop adjacent to the catalytic reaction site. The calculations demonstrate critical importance of the sufficient conformational sampling by a long MD simulation.

THEORY

A free energy functional of the QM Born–Oppenheimer (BO) electronic wave function, $\Psi(\mathbf{r}; \mathbf{R}, \mathbf{X})$, is introduced³⁰ as

$$F[\Psi] = -\beta^{-1} \ln \int d\mathbf{R} d\mathbf{X} \exp(-\beta E[\Psi(\mathbf{r}; \mathbf{R}, \mathbf{X}); \mathbf{R}, \mathbf{X}]) \quad (1)$$

where \mathbf{r} , \mathbf{R} , and \mathbf{X} are coordinates of electrons, QM atoms, and MM atoms, respectively, and $E[\Psi(\mathbf{r}; \mathbf{R}, \mathbf{X}); \mathbf{R}, \mathbf{X}]$ is the expectation value of the total energy of the QM/MM system:

$$E[\Psi(\mathbf{r}; \mathbf{R}, \mathbf{X}); \mathbf{R}, \mathbf{X}] = \langle \Psi(\mathbf{r}; \mathbf{R}, \mathbf{X}) | \hat{H} | \Psi(\mathbf{r}; \mathbf{R}, \mathbf{X}) \rangle_{\mathbf{r}} \quad (2)$$

\hat{H} is the total Hamiltonian,

$$\hat{H}(\mathbf{r}, \mathbf{R}, \mathbf{X}) = \hat{H}^0(\mathbf{r}, \mathbf{R}) + \hat{H}^{\text{QM-MM}}(\mathbf{r}, \mathbf{R}, \mathbf{X}) + E^{\text{MM}}(\mathbf{X}) \quad (3)$$

where $\hat{H}^0(\mathbf{r}, \mathbf{R})$ is the gas electronic Hamiltonian of the QM molecules, $E^{\text{MM}}(\mathbf{X})$ is the energy function of the MM ones, and $\hat{H}^{\text{QM-MM}}(\mathbf{r}, \mathbf{R}, \mathbf{X})$ represents the interaction between the QM and MM parts. The QM–MM interaction is usually given by a sum of the ES nonbonding interaction, $\hat{H}_{\text{ES}}^{\text{QM-MM}}(\mathbf{r}, \mathbf{R}, \mathbf{X})$, and the interaction other than ES, $E_{\text{non-ES}}^{\text{QM-MM}}(\mathbf{R}, \mathbf{X})$, such as the Lennard-Jones nonbonding one and the intramolecular bonding potentials:

$$\hat{H}^{\text{QM-MM}}(\mathbf{r}, \mathbf{R}, \mathbf{X}) = \hat{H}_{\text{ES}}^{\text{QM-MM}}(\mathbf{r}, \mathbf{R}, \mathbf{X}) + E_{\text{non-ES}}^{\text{QM-MM}}(\mathbf{R}, \mathbf{X}) \quad (4)$$

Note that the QM electrons interact with effective point charges of the MM atoms through the ES interaction, whereas the non-ES interaction is independent of the QM electronic coordinates and is described solely by MM force field parameters. In a standard QM/MM scheme, the ES interaction is given as

$$\hat{H}_{\text{ES}}^{\text{QM-MM}}(\mathbf{r}, \mathbf{X}) = \sum_i^{N_{\text{elec}}} \sum_{\alpha}^{N_{\text{MM}}} \frac{Q_{\alpha}}{|\mathbf{r}_i - \mathbf{X}_{\alpha}|} \quad (5)$$

where N_{elec} and N_{MM} are the numbers of electrons and MM atoms, respectively, and Q_{α} represents the effective charges of the MM atoms.

Taking variation of the free energy functional with respect to the QM wave function, i.e., $\delta F[\Psi]/\delta \Psi = 0$, with a constraint for the normalization of the QM wave function, $\langle \Psi | \Psi \rangle = 1$, provides the variational condition which the optimal QM wave function, $\tilde{\Psi}$, has to satisfy as

$$\int d\mathbf{R} d\mathbf{X} [\hat{H}|\tilde{\Psi}(\mathbf{r}; \mathbf{R}, \mathbf{X})\rangle - E[\tilde{\Psi}(\mathbf{r}; \mathbf{R}, \mathbf{X}); \mathbf{R}, \mathbf{X}]|\tilde{\Psi}(\mathbf{r}; \mathbf{R}, \mathbf{X})\rangle] \times \exp(-\beta E[\tilde{\Psi}(\mathbf{r}; \mathbf{R}, \mathbf{X}); \mathbf{R}, \mathbf{X}]) = 0 \quad (6)$$

The condition indicates that the optimal free energy is given with the BO wave functions and their energies variationally determined at conformations of \mathbf{R} and \mathbf{X} in the Boltzmann distribution of the total energies, $E[\tilde{\Psi}(\mathbf{r}; \mathbf{R}, \mathbf{X}); \mathbf{R}, \mathbf{X}]$. A sampling procedure that fulfills the condition is a straightforward QM/MM MD simulation where the time evolution of the trajectory is computed with the QM/MM wave function and its energy variationally determined at each step of the trajectory calculation. Unfortunately, such a straightforward QM/MM MD simulation, especially based on the ab initio QM method, is computationally impractical for a QM/MM system with a large QM region. As mentioned above, the large QM size is a crucial factor for accurate description of the QM/MM system. Furthermore, the high computational cost limits the sampling time of the trajectory, leading to a poor convergence of the statistical sampling for the thermal distribution.

To reduce the sampling cost, thermal distribution of the QM coordinates, \mathbf{R} , is omitted, and instead, the free energy functional is optimized with respect to \mathbf{R} . This approximate procedure is the so-called free energy geometry optimization.^{22,23} For this purpose, the free energy surface, which is an explicit function of \mathbf{R} , is defined as

$$F[\Psi; \mathbf{R}] = -\beta^{-1} \ln \int d\mathbf{X} \exp(-\beta E[\Psi(\mathbf{r}; \mathbf{R}, \mathbf{X}); \mathbf{R}, \mathbf{X}]) \quad (7)$$

Gradients on the free energy surface are then derived as the averaged energy gradients

$$\frac{\partial F[\tilde{\Psi}; \mathbf{R}]}{\partial \mathbf{R}} = \left\langle \frac{\partial E[\tilde{\Psi}(\mathbf{r}; \mathbf{R}, \mathbf{X}); \mathbf{R}, \mathbf{X}]}{\partial \mathbf{R}} \right\rangle_{\mathbf{X}, E[\tilde{\Psi}(\mathbf{r}; \mathbf{R}, \mathbf{X}); \mathbf{R}, \mathbf{X}]} \quad (8)$$

where $\langle \dots \rangle_{\mathbf{X}, E[\tilde{\Psi}(\mathbf{r}; \mathbf{R}, \mathbf{X}); \mathbf{R}, \mathbf{X}]}$ indicates the thermal average over \mathbf{X} with its thermal distribution obtained for the energy, $E[\tilde{\Psi}(\mathbf{r}; \mathbf{R}, \mathbf{X}); \mathbf{R}, \mathbf{X}]$. The procedure also requires calculations of the optimal QM/MM energy and its forces on \mathbf{R} and \mathbf{X} at every MM conformation so that the computational cost is still very demanding.

The computational cost of the free energy geometry optimization procedure is drastically reduced by introducing the mean field approximation for the electronic wave function where the

explicit dependence on \mathbf{X} is neglected, i.e., $\Psi(\mathbf{r}; \mathbf{R}, \mathbf{X}) \rightarrow \Psi_{\text{MF}}(\mathbf{r}; \mathbf{R})$. The mean field free energy functional is then written as

$$F[\Psi_{\text{MF}}; \mathbf{R}] = E^0[\Psi_{\text{MF}}(\mathbf{r}; \mathbf{R}); \mathbf{R}] - \beta^{-1} \ln \int d\mathbf{X} \exp(-\beta[E^{\text{QM-MM}}[\Psi_{\text{MF}}(\mathbf{r}; \mathbf{R}); \mathbf{R}, \mathbf{X}] + E^{\text{MM}}(\mathbf{X})]) \quad (9)$$

where E^0 and $E^{\text{QM-MM}}$ are expectation values of the gas Hamiltonian, $\hat{H}^0(\mathbf{r}; \mathbf{R})$, and the QM–MM interaction, $\hat{H}^{\text{QM-MM}}(\mathbf{r}; \mathbf{R}, \mathbf{X})$, respectively. A Hartree–Fock equation that determines the optimal mean field wave function and energy can be derived by taking variation of the free energy functional with respect to molecular orbitals. The Fock operator is expressed as

$$\hat{f}_{\text{ES}}^{\text{QM/MM}}(\mathbf{r}; \mathbf{R}) = \hat{f}^0(\mathbf{r}; \mathbf{R}) + \langle \hat{f}_{\text{ES}}^{\text{QM-MM}}(\mathbf{r}; \mathbf{R}, \mathbf{X}) \rangle_{\mathbf{X}, E[\Psi_{\text{MF}}(\mathbf{r}; \mathbf{R}); \mathbf{R}, \mathbf{X}]} \quad (10)$$

where $\hat{f}^0(\mathbf{r}; \mathbf{R})$ is the gas Fock operator and $\hat{f}_{\text{ES}}^{\text{QM-MM}}(\mathbf{r}; \mathbf{R}, \mathbf{X})$ represents the ES interaction of an electron with the MM effective point charges. The Fock term corresponding to the Hamiltonian of eq 5 is expressed as

$$\hat{f}_{\text{ES}}^{\text{QM-MM}}(\mathbf{r}, \mathbf{X}) = \sum_{\alpha}^{N_{\text{MM}}} \frac{Q_{\alpha}}{|\mathbf{r} - \mathbf{X}_{\alpha}|} \quad (11)$$

As eq 10 indicates, the mean field treatment allows one to determine the optimal wave function by solving only one variational problem with the averaged operator, which is much less time-consuming than QM/MM MD simulations where the variational solution has to be obtained at every MM conformation. However, for the standard QM–MM ES interaction operator, the treatment requires calculation of one-electron integrals at every MM conformation, which is still computationally demanding. We therefore introduced the charge operator, $\hat{q}_A(\mathbf{r}, \mathbf{R})$, the expectation value of which gives an effective point charge of a QM atom:

$$q_A(\mathbf{d}, \mathbf{R}) = \langle \Psi_{\text{MF}}(\mathbf{r}, \mathbf{R}) | \hat{q}_A(\mathbf{r}, \mathbf{R}) | \Psi_{\text{MF}}(\mathbf{r}, \mathbf{R}) \rangle_{\mathbf{r}} \quad (12)$$

where \mathbf{d} is the one-electron density matrix. In the present study, the RESP charge operator³⁷ was employed for the charge operator. The QM–MM ES interaction energy is then expressed as

$$E_{\text{ES}}^{\text{QM-MM}}(\mathbf{d}; \mathbf{R}, \mathbf{X}) = \sum_A^{N_{\text{QM}}} q_A(\mathbf{d}; \mathbf{R}) V_A(\mathbf{R}, \mathbf{X}) \quad (13)$$

where $V_A(\mathbf{R}, \mathbf{X})$ is the ES potential of the MM atoms on the QM atom A and N_{QM} is the number of atoms in the QM region. Details on the RESP charge operator and the ES potential are found in ref 37. The averaged Fock operator term of the ES interaction is given as³⁰

$$\langle \hat{f}_{\text{ES}}^{\text{QM-MM}}(\mathbf{r}; \mathbf{R}, \mathbf{X}) \rangle_{\mathbf{X}, E[\Psi_{\text{MF}}(\mathbf{r}; \mathbf{R}); \mathbf{R}, \mathbf{X}]} = \sum_A^{N_{\text{QM}}} \hat{q}_A(\mathbf{r}; \mathbf{R}) \langle V_A(\mathbf{R}, \mathbf{X}) \rangle_{\mathbf{X}, E[q(\mathbf{d}, \mathbf{R}); \mathbf{R}, \mathbf{X}]} \quad (14)$$

One can see in the equation that the ES potentials are independent of the coordinates of electrons so that the ensemble average is taken only with the classical variables \mathbf{R} and \mathbf{X} . Hence, the ensemble average does not include evaluation of the one-electron integrals, which drastically reduces the computational cost.

The ensemble average in eq 14, however, still suffers from a computational difficulty. For the ensemble average over \mathbf{X} , a classical MD simulation of \mathbf{X} is carried with a given $q_A(\mathbf{d}, \mathbf{R})$ and \mathbf{R}

of the QM atoms. Hence, the classical MD simulation has to be performed at every step of SCF and geometry optimization cycles where \mathbf{d} and \mathbf{R} are updated, respectively. Such frequent MD samplings require a very demanding computational cost since tight convergence of the ensemble for each MD sampling needs to be imposed for stable convergence of the SCF and the geometry optimization. In a computational scheme developed by Yamamoto,³⁰ a micro-iteration procedure which avoids the frequent MD samplings is introduced. During the micro-iteration of SCF and geometry optimization, the ensemble of \mathbf{X} is unchanged; even \mathbf{d} and \mathbf{R} are updated. Then a macro-iteration for the update of the ensemble of \mathbf{X} with \mathbf{d} and \mathbf{R} converged at the micro-iteration is carried out in order to achieve the global convergence. However, as seen below, the procedure may slow the global convergence because the update of \mathbf{d} and \mathbf{R} does not proceed on the proper free energy surface defined by the functional of eq 9 at most steps of the micro-iteration.

In the QM/MM-RWFE-SCF method developed in the present study, we have incorporated a reweighting update of the ensemble of \mathbf{X} employed in QM/MM-MFEP methods by Yang and co-workers²⁵ into the MF free energy functional formalism by Yamamoto.³⁰ By introducing a reference QM system with an electronic wave function, $\Psi_{\text{MF,ref}}(\mathbf{r}_{\text{ref}}; \mathbf{R}_{\text{ref}})$, and a geometry, \mathbf{R}_{ref} , the free energy difference from the reference value, $\Delta F[\Psi_{\text{MF}}; \mathbf{R}] = F[\Psi_{\text{MF}}; \mathbf{R}] - F[\Psi_{\text{MF,ref}}; \mathbf{R}_{\text{ref}}]$, is rewritten as

$$\begin{aligned} \Delta F[\Psi_{\text{MF}}; \Psi_{\text{MF,ref}}; \mathbf{R}; \mathbf{R}_{\text{ref}}] \\ = E^0[\Psi_{\text{MF}}(\mathbf{r}; \mathbf{R}); \mathbf{R}] - E^0[\Psi_{\text{MF,ref}}(\mathbf{r}_{\text{ref}}; \mathbf{R}_{\text{ref}}); \mathbf{R}_{\text{ref}}] \\ - \beta^{-1} \ln \langle \exp(-\beta \Delta E^{\text{QM-MM}}(\mathbf{d}, \mathbf{d}_{\text{ref}}; \mathbf{R}, \mathbf{R}_{\text{ref}}; \mathbf{X})) \rangle_{\mathbf{X}, E[\mathbf{q}(\mathbf{d}_{\text{ref}}, \mathbf{R}_{\text{ref}}); \mathbf{R}_{\text{ref}}, \mathbf{X}]} \end{aligned} \quad (15)$$

where \mathbf{d}_{ref} is the one electron density of the reference electronic wave function and $\Delta E^{\text{QM-MM}}(\mathbf{d}, \mathbf{d}_{\text{ref}}; \mathbf{R}, \mathbf{R}_{\text{ref}}; \mathbf{X})$ is the QM–MM interaction energy difference from the reference value. The reference density and coordinates are better chosen to be similar to the optimal ones in order to attain smaller QM–MM interaction energy differences which provide quicker convergence of the statistical sampling, although they can be arbitrary in principle. In the present study, we adopted the ones obtained by conventional static QM/MM calculations for the initial values. As seen below, the reference density and coordinates are renewed in a sequential sampling. The ensemble average over \mathbf{X} is taken with MD samples obtained for the reference QM system, and thus the update of the ensemble of \mathbf{X} at each step of SCF and geometry optimization cycles is not necessary.

The ensemble average of the ES potential in the Fock operator of eq 14 is also rewritten with \mathbf{d}_{ref} and \mathbf{R}_{ref} as

$$\begin{aligned} \langle V_A(\mathbf{R}, \mathbf{X}) \rangle_{\mathbf{X}, E[\mathbf{q}(\mathbf{d}, \mathbf{R}); \mathbf{R}, \mathbf{X}]} \\ = \left[\int d\mathbf{X} V_A(\mathbf{R}, \mathbf{X}) \exp[-\beta(\Delta E^{\text{QM-MM}}(\mathbf{d}, \mathbf{d}_{\text{ref}}; \mathbf{R}, \mathbf{R}_{\text{ref}}; \mathbf{X}))] \right. \\ \left. \times \exp[-\beta(E^{\text{QM-MM}}(\mathbf{d}_{\text{ref}}; \mathbf{R}_{\text{ref}}; \mathbf{X}) + E^{\text{MM}}(\mathbf{X}))] \right] \\ \left/ \left[\int d\mathbf{X} \exp[-\beta(\Delta E^{\text{QM-MM}}(\mathbf{d}, \mathbf{d}_{\text{ref}}; \mathbf{R}, \mathbf{R}_{\text{ref}}; \mathbf{X}))] \right. \right. \\ \left. \left. \times \exp[-\beta(E^{\text{QM-MM}}(\mathbf{d}_{\text{ref}}; \mathbf{R}_{\text{ref}}; \mathbf{X}) + E^{\text{MM}}(\mathbf{X}))] \right] \right. \\ = \langle V_A(\mathbf{R}, \mathbf{X}) \omega(\mathbf{d}, \mathbf{d}_{\text{ref}}; \mathbf{R}, \mathbf{R}_{\text{ref}}; \mathbf{X}) \rangle_{\mathbf{X}, E[\mathbf{q}(\mathbf{d}_{\text{ref}}, \mathbf{R}_{\text{ref}}); \mathbf{R}_{\text{ref}}, \mathbf{X}]} \end{aligned} \quad (16)$$

where $\omega(\mathbf{d}, \mathbf{d}_{\text{ref}}; \mathbf{R}, \mathbf{R}_{\text{ref}}; \mathbf{X})$ is the reweighting factor:

$$\begin{aligned} \omega(\mathbf{d}, \mathbf{d}_{\text{ref}}; \mathbf{R}, \mathbf{R}_{\text{ref}}; \mathbf{X}) \\ = \frac{\exp(-\beta \Delta E^{\text{QM-MM}}(\mathbf{d}, \mathbf{d}_{\text{ref}}; \mathbf{R}, \mathbf{R}_{\text{ref}}; \mathbf{X}))}{\langle \exp(-\beta \Delta E^{\text{QM-MM}}(\mathbf{d}, \mathbf{d}_{\text{ref}}; \mathbf{R}, \mathbf{R}_{\text{ref}}; \mathbf{X})) \rangle_{\mathbf{X}, E[\mathbf{q}(\mathbf{d}_{\text{ref}}, \mathbf{R}_{\text{ref}}); \mathbf{R}_{\text{ref}}, \mathbf{X}]} \end{aligned} \quad (17)$$

Because of the reweighting factor, the Fock operator of the ES interaction between the QM and MM regions depends on the electron density, unlike conventional QM/MM methods, so that the matrix element is calculated at each step of the SCF cycle. Extension of the present scheme to other variational methods such as density functional theory (DFT) and complete active space SCF is straightforward.

The reweighting treatment is also applied straightforwardly to the calculation of the mean force given by eq 8. Expression of the derivative of the RESP charge operator given by eq 12 with respect to \mathbf{R} is found in ref 37. It is noteworthy that one does not need to carry out coupled perturbed calculations to evaluate explicitly derivatives of linear-combination-of-atomic-orbitals (LCAO) coefficients of the molecular orbitals with respect to \mathbf{R} . The molecular orbitals are determined variationally for the free energy functional, eq 15, by solving the consistent Fock or Kohn–Sham (KS) equation with the operator, eq 10, with eqs 14 and 16. Hence, the free energy gradient terms including derivatives of the LCAO coefficients can be replaced by terms with the orbital energies and derivatives of the overlap integrals as in conventional energy gradient techniques. The reweighting scheme also allows one to calculate efficiently the Hessian matrix on the free energy surface through the finite differential method, since the same MM conformational samples can be used for the free energy gradient calculations of the slightly displaced QM geometries. The efficient scheme for the calculation of the Hessian matrix enables one to determine a transition state structure and to evaluate zero point energy and vibrational entropy. Those calculations are now ongoing and will be reported elsewhere.

As described above, the reweighting method permits one to update the ensemble of \mathbf{X} without reevaluating it with an MD trajectory calculation. However, the reweighting update of the ensemble of \mathbf{X} is valid only if the ensemble of \mathbf{X} covers a large configurational space of \mathbf{X} . Thus, for limited samples of \mathbf{X} obtained by an MD simulation, the changes of \mathbf{d} and \mathbf{R} from their references, \mathbf{d}_{ref} and \mathbf{R}_{ref} , need to stay small in SCF and geometry optimization; if the deviation of \mathbf{d} and \mathbf{R} from their references is large, only a few samples come to possess dominantly large reweighting factors, eq 17, because of the exponential nature of weighting, leading to a poor ensemble average over \mathbf{X} . It is therefore necessary to perform the sequential sampling²⁵ where an MD trajectory calculation is iteratively carried out at each end of the geometry optimization in order to update the ensemble of \mathbf{X} for renewed \mathbf{d}_{ref} and \mathbf{R}_{ref} until the deviations of \mathbf{d} and \mathbf{R} from the references stay small in the following geometry optimization and consequently a sufficiently large number of the samples come to contribute to the average over \mathbf{X} . Since the behavior of the averaging with the reweighting factors depends strongly on the sample size, careful assessment of the convergence of the sampling is necessary.

In order to evaluate the ES interaction between the QM and MM regions accurately and efficiently, the Ewald method was

implemented in the QM/MM-RWFE-SCF calculation. The ES interaction energy is expressed as

$$V = V^{\text{real}} + V^{\text{rec}} + V^{\text{corr}} \quad (18)$$

where V^{real} and V^{rec} are interaction energies in real and reciprocal spaces, respectively

$$V^{\text{real}} = \sum_n \sum_i^{N_{\text{QM}}} \sum_j^{N_{\text{MM}}} \hat{q}_i q_j \frac{\text{erfc}(\alpha |\mathbf{R}_i - \mathbf{X}_j + \mathbf{L}n|)}{|\mathbf{R}_i - \mathbf{X}_j + \mathbf{L}n|} \quad (19)$$

$$V^{\text{rec}} = \frac{4\pi}{V} \sum_{\mathbf{G} \neq 0} \frac{\exp(-|\mathbf{G}|^2/4\alpha^2)}{|\mathbf{G}|^2} \left\{ \sum_i^{N_{\text{QM}}} \hat{q}_i \cos(\mathbf{G} \cdot \mathbf{R}_i) W_{\cos}(\mathbf{G}) + \sum_i^{N_{\text{QM}}} \hat{q}_i \sin(\mathbf{G} \cdot \mathbf{R}_i) W_{\sin}(\mathbf{G}) \right\} \quad (20)$$

$$W_{\cos}(\mathbf{G}) = \sum_j^{N_{\text{MM}}} q_j \cos(\mathbf{G} \cdot \mathbf{X}_j) \quad (21)$$

$$W_{\sin}(\mathbf{G}) = \sum_j^{N_{\text{MM}}} q_j \sin(\mathbf{G} \cdot \mathbf{X}_j)$$

and V^{corr} is a correction which subtracts self-interaction within the QM–MM boundary region from the term in reciprocal space, eq 20

$$V^{\text{corr}} = - \sum_{(i,j)}^{N_{\text{QM-MM}}} \hat{q}_i q_j \frac{\text{erf}(\alpha |\mathbf{R}_i - \mathbf{X}_j|)}{|\mathbf{R}_i - \mathbf{X}_j|} \quad (22)$$

N_{QM} and N_{MM} are the numbers of atoms in the QM and MM regions, respectively, and $N_{\text{QM-MM}}$ is that of atom pairs included in bonding interactions at the boundary between the QM and MM regions. \hat{q} and q represent the RESP charge operator of the QM atoms and the atomic charge of the MM force field, respectively. α is the screening parameter. \mathbf{L} indicates the box length vector, and V is the box volume. \mathbf{G} is the reciprocal space vector

$$\mathbf{G} = 2\pi \begin{pmatrix} k_x/L_x \\ k_y/L_y \\ k_z/L_z \end{pmatrix} \quad (23)$$

where k_x , k_y , and k_z are integer numbers. Since the RESP charge operator is used for the QM–MM ES interaction, computation of the Ewald interaction terms is straightforward.

For computational simplicity, interaction between the QM regions in different image cells is neglected. Because distance between the nearest neighbor QM regions is large and alignment of the periodic QM images is homogeneous, the QM images add only a constant ES field with negligibly small deviation, as shown in Figures S2 and S3 in the Supporting Information. As the summations for the MM part, $W_{\cos}(\mathbf{G})$ and $W_{\sin}(\mathbf{G})$, do not include the QM atom, these terms are calculated only once at the first step of free energy optimization upon update of the MM ensemble. We used a small screening parameter, $\alpha = 0.13149 \text{ \AA}^{-1}$, and a large cutoff distance of 20 Å for the QM–MM interaction in real space, eq 19, compared with those employed in MD simulations. These parameters give a quick convergence of the time-consuming reciprocal space summation with respect to \mathbf{G} in

eq 20 and thus lead to drastic reduction of the overall computational cost, despite a moderate increase of cost of the QM–MM interaction in real space, eq 19, which is only proportional to the number of atoms in the cutoff range. According to the convergence behavior shown in Figure S4 in the Supporting Information, k_x , k_y , and k_z are set to be from -10 to $+10$, respectively. The present parameters of the Ewald summation were confirmed to give essentially the same description of the ES interaction as that by a particle mesh Ewald (PME) method³⁸ with default parameters of Amber9 program packages;³⁹ the discrepancy of the ES interaction energies for a conformation was evaluated to be within $1.0 \times 10^{-3} \%$.

It is noted that most of recent sophisticated MD programs which execute trajectory calculations very efficiently utilize Ewald-based techniques such as the PME method. The equivalent description of the ES interaction in the QM/MM-RWFE-SCF calculation to that in the MD programs permits one to employ externally the efficient existing MD programs for the conformational sampling of \mathbf{X} in the QM/MM-RWFE-SCF procedure.

The protocol of free energy optimization with the QM/MM RWFE-SCF method is summarized as follows:

- (1) Initial reference coordinates, \mathbf{R}_{ref} and a density matrix, \mathbf{d}_{ref} of the QM region are given. In this study, results of conventional QM/MM calculation for a cluster system were used as the initial values.
- (2) To obtain an ensemble of \mathbf{X} , a MD simulation is performed with the fixed \mathbf{R}_{ref} and \mathbf{q}_{ref} , which is the RESP charges derived from \mathbf{d}_{ref} .
- (3) The optimal wave function is calculated by the SCF procedure with eq 10. $W_{\cos}(\mathbf{G})$ and $W_{\sin}(\mathbf{G})$ in eq 20 are calculated at the first step of SCF iteration when the ensemble of \mathbf{X} is updated in the sequential sampling procedure and stored throughout a geometry optimization cycle. At each SCF iteration step, the ensemble of \mathbf{X} is updated by eq 16 as the density matrix, i.e., the QM charges, changes and accordingly the ensemble average of the ES potential in the Fock or KS operator of eq 14 is reevaluated.
- (4) Using the optimized wave function, the averaged force is calculated with eq 8 and the QM geometry is updated in a standard geometry optimization manner. The reweighted distribution is used for the calculation of the averaged forces of the QM/MM interaction. If the force is smaller than a convergence criterion, the optimization procedure goes to step 5. Otherwise, the optimization procedure returns to step 3 after the QM geometry is updated.
- (5) If the number of the effective MM samples for the reweighting average is not sufficient, a MD trajectory calculation for the sequential sampling is carried out to update the ensemble of \mathbf{X} as discussed above. The QM reference coordinates and densities, \mathbf{R}_{ref} and \mathbf{d}_{ref} are updated with those optimized at the preceding cycle of steps 3 and 4, and then the procedure returns to step 2 to obtain a new ensemble of \mathbf{X} . If displacements from the reference values upon the following two successive sequential samplings are small and the reweighting averages are taken with sufficient numbers of the MM samples, the optimization procedure is finished.

As seen above, the QM/MM-RWFE-SCF geometry optimization determines the stationary structure on the free energy surface of a given MM ensemble obtained at the end of the

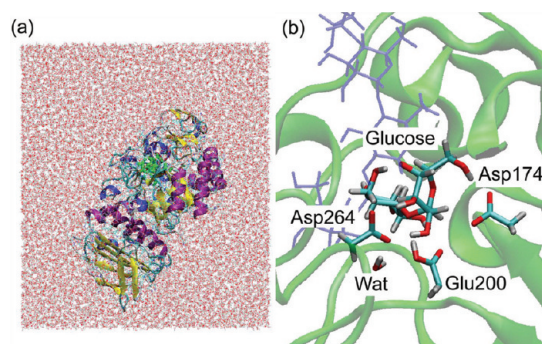


Figure 2. (a) Total simulation system of the QM/MM-RWFE-SCF calculation. The QM region is drawn in green licorice representation. (b) A close-up view of the QM region of the QM/MM-RWFE-SCF calculation. The QM region is drawn in licorice representation.

sequential sampling. For the examination of the reaction free energy profile, the free energy minimum structures of the states in the reaction process such as the reactant and product are therefore determined with different MM ensembles. Hence, the QM/MM-RWFE-SCF calculation cannot solely evaluate the free energy differences between the states. Nevertheless, one can calculate the free energy differences by standard techniques such as the free energy perturbation using the coordinates and the charges of the QM molecule determined by the QM/MM-RWFE-SCF geometry optimizations. The calculation of the reaction free energy profile is now ongoing and will be reported elsewhere.

COMPUTATIONAL DETAIL

The initial protein structure of α -amylase was taken from PDB 1G94.⁴⁰ The structure is of the protein with a saccharide substrate analog compound which occupies subsites -4 to $+3$. We replaced the substrate analog with an amylose, which consists of six α -glucoses in subsites -4 to $+2$. We determined the charge state of acidic residues located at the active site based on a proposed reaction mechanism of α -amylases,³⁶ while standard charge states were assumed for other acidic and basic residues. Hydrogen atoms were added by the LEaP module of AMBER9.³⁹ AMBER ff99 and GLYCAM04 parameter sets⁴¹ were utilized for the protein and glucoses, respectively. For water molecules, the TIP3P model⁴² was used. The protein system was immersed in a water box of $80 \text{ \AA} \times 100 \text{ \AA} \times 90 \text{ \AA}$ (Figure 2) in PBC. To neutralize the system, 16 sodium ions were put in the box. The AMBER9 software suite³⁹ was used for all MD simulations. Long-range ES interactions were treated with the PME method. Short range nonbonded interactions were cut off at 10 \AA .

The QM region employed in the QM/MM calculations is depicted in Figure 2. It consists of two glucoses at the subsites -1 and $+1$, the side chains of two aspartic acids, Asp174 and Asp264, and the side chain from the C_γ atom to the end of Glu200. We used the DFT method with the B3LYP functional for the QM region. The 6-31G* basis set was employed, except for the carboxyl groups of Asp and Glu, where the 6-31+G* basis set was employed. A link atom approach was used to terminate properly chemical bonds at the QM–MM boundaries.³⁷ The convergence criterion of the geometry optimization was set to be 1.0×10^{-3} Hartree/Bohr for the largest component of the gradient for the QM region. The QM/MM-RWFE-SCF method was implemented in the GAMESS program package.⁴³

In order to obtain the initial reference coordinates and electronic density, R_{ref} and d_{ref} , conventional QM/MM calculations based on the potential energy surface³⁷ for a sphere cluster system were first carried out. The initial system of the QM/MM calculations was prepared as follows. A 500-ps equilibrium MD simulation with the reactant substrate in a constant-NPT (300 K, 1 atm) ensemble was carried out. The sphere cluster that contains all atoms at less than 40 \AA from the C atom of the scissile bond in the reactant state was taken from the last structure of the MD simulation. Then, an equilibrium 400-ps MD simulation at 300 K was carried out for the sphere cluster system. The residues in the region more than 32 \AA away from the C atom of the scissile bond were fixed in order to keep the shape of the sphere cluster. In the last 200-ps, the temperature was lowered gradually. Finally, an energy minimization was performed until the RMS energy gradient was below 1.0×10^{-4} kcal/mol/ \AA (8.4×10^{-8} Hartree/Bohr). The last structure was used as the initial reactant structure for the QM/MM potential energy geometry optimization. The convergence criterion of the geometry optimization for the MM region was set to be 1.0×10^{-4} Hartree/Bohr for the RMS gradient. We also determined the reference structures and charges of the product state through the QM/MM potential energy geometry optimization from an initial structure obtained by cleaving the α -glycosidic bond of the reactant QM molecule. For the sphere cluster system, no cutoff of the nonbonded interaction was applied throughout the MD and QM/MM calculations.

MD simulations for obtaining the MM ensemble in the sequential sampling of the QM/MM-RWFE-SCF geometry optimization were performed for the PBC system with the fixed reference coordinates and charges of the QM region. A 3-ns trajectory was calculated at each iteration step of the sequential sampling, and the first 1-ns and the last 2-ns trajectories were employed for equilibration and sampling of the MM ensemble, respectively. The MM conformational samples were taken at every 100 fs, and thus the MM ensemble at each step of the sequential sampling was comprised of 20 000 configurations. For the first step of the iteration of the sequential sampling, the MD simulation was carried out from the initial structure where the sphere cluster optimized by the QM/MM calculation described above was inserted into the PBC system (a water box of $80 \text{ \AA} \times 100 \text{ \AA} \times 90 \text{ \AA}$). The total number of atoms in this box is 68 533. The first MD simulation was performed by the SANDAR module under constant-NPT (1 atm, 283 K) conditions for the QM region fixed with the SHAKE algorithm. For MD simulations in the iteration of the sequential sampling after the first step, the trajectories were calculated by the PMEMD module under constant-NVT (283 K) conditions for the space fixed QM region, which drastically accelerates the computation of the trajectory.

RESULTS AND DISCUSSION

Long-Range ES Interaction of the QM Catalytic Site with the Protein. It is well-known that long-range ES interaction is important for protein systems where charge distributions are very inhomogeneous.⁴⁴ Nevertheless, because of computational difficulty, sphere-shaped cluster QM/MM systems were often employed, and long-range ES interactions beyond the cluster sizes were neglected. In order to access the importance of the long-range ES interaction, we compared the ES potentials on the QM sites from the MM surroundings of the cluster system with that of the PBC system with the Ewald method (PBC-Ewald). Figure 3 shows differences in the ES potentials between the cluster systems

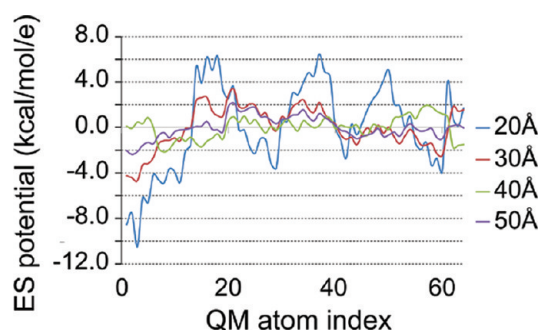


Figure 3. Comparison of ES potentials of the MM region acting on the QM atoms obtained by the Ewald method and cutoff ones. Deviations of ES potentials in the reactant calculated by the residue-based cutoff methods with cutoff distances of 20, 30, 40, and 50 Å from those by the Ewald method are shown. The cutoff distance is defined as the distance from C(40), i.e., the C atom of the scissile bond. The QM atom index is given in Figure S1 in the Supporting Information.

with different cutoff distances and the PBC-Ewald system. The ES potentials for the smaller cluster systems deviate largely from that for the PBC-Ewald system. As the sphere cluster size increases, the ES potentials of the cluster systems converge gradually to those of the PBC-Ewald one, and finally the difference between those of the cluster systems with a cutoff of 50 Å and of the PBC-Ewald system falls within ± 2.0 kcal/mol. This indicates clearly that the long-range ES interaction between the QM and MM regions is important, and the Ewald method provides an accurate description of the interaction.

It should be noted that, although a large cluster system with a cutoff of 50 Å could give an accuracy of the ES interaction comparable with the PBC-Ewald system, the MD sampling simulation in the QM/MM-RWFE-SCF procedure for the large cluster system is much more computationally demanding than that for the PBC-Ewald system with the PME method; the increase of its computational cost for the former is proportional to N^2 (N is the number of atoms treated), whereas that for the latter is proportional to $N \log N$. On an Intel Xeon E5450 computer cluster with 32 cores, the PMEMD module with the PME method for the PBC system executes already much faster than the SANDER module for the sphere cluster system of 30 Å. Furthermore, the cluster system requires a fixed or restrained sphere shell boundary to keep its shape and thus arrests global flexibility of the protein, which may play a role in the catalytic reaction. The PBC-Ewald system is therefore advantageous over the cluster system in terms of both computational accuracy and efficiency.

Convergence Behavior of MM Conformational Distribution Sampled by MD Simulation. As described above, the electronic wave function and the geometry of the QM region are determined under mean fields of the MM conformational distributions sampled by MD simulations. Thus, statistical convergence of the MM distribution is a crucial factor for the stable calculation. We therefore examined carefully the convergence behaviors of the MM distributions. In the QM/MM-RWFE-SCF geometry optimization, convergences of the MM distributions at two different iterations are necessary, i.e., convergence of the geometry optimization in each iteration of the sequential sampling and that of the macro-iteration of the sequential sampling.

First, we assessed the former convergence. In each step of the sequential sampling iteration, the geometry is optimized with a

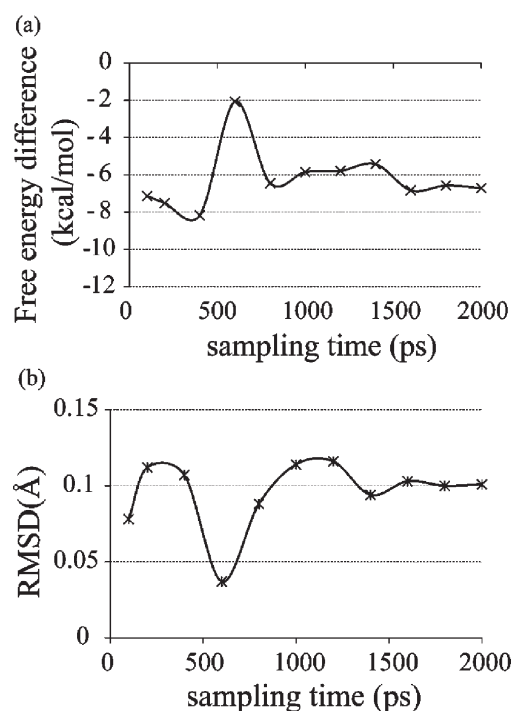


Figure 4. Convergence behaviors of the QM/MM free energy optimization with a single MM conformational distribution in the first step of the sequential sampling iteration along its cumulative sampling time by a MD simulation. Changes of free energy difference given by eq 15 (a) and RMSDs from the initial structure (b) are plotted.

single set of the MM samples. Figure 4 indicates convergences of free energy and root-mean-square deviation (RMSD) of the QM geometry optimized at the first step of the sequential sampling iteration with respect to cumulative MD sampling time for the MM distribution. For the present system, both the free energy and the QM geometry do not converge before 1.5 ns, which is much longer than simulation time of other previous studies of free energy geometry optimization,^{23,25,29,30} i.e., several hundreds of picoseconds or less. It was found that the optimized QM geometries with MM distributions of different cumulative MD sampling times before 1.5 ns differ significantly from one another, which leads to an unstable search for the geometry optimization, whereas after 1.5 ns the geometry optimizations with different MM distributions converge to a single QM geometry. We therefore employed MD trajectories for 2.0 ns for each step of the sequential sampling iteration in this study.

Second, convergence of the overall iteration of the sequential sampling was examined. As described above, the reweighting scheme with limited MM samples suffers from the difficulty of poor averaging when the deviations of \mathbf{d} and \mathbf{R} from their reference values become large. The difficulty can be seen in Figure S5, which shows a histogram of the energy difference, $\Delta E^{\text{QM-MM}}(\mathbf{d}, \mathbf{d}_{\text{ref}}; \mathbf{R}, \mathbf{R}_{\text{ref}}; \mathbf{X})$, in eqs 15 and 17 for the MM samples at the optimized geometry in the first step of the sequential sampling iteration and its average over the reweighted MM distribution

$$\begin{aligned} & \langle \Delta E^{\text{QM-MM}} \rangle_{\mathbf{X}, E[\mathbf{q}(\mathbf{d}, \mathbf{R}); \mathbf{R}, \mathbf{X}]} \\ &= \langle \Delta E^{\text{QM-MM}}(\mathbf{d}, \mathbf{d}_{\text{ref}}; \mathbf{R}, \mathbf{R}_{\text{ref}}; \mathbf{X}) \omega(\mathbf{d}, \mathbf{d}_{\text{ref}}; \mathbf{R}, \mathbf{R}_{\text{ref}}; \mathbf{X}) \rangle_{\mathbf{X}, E[\mathbf{q}(\mathbf{d}_{\text{ref}}, \mathbf{R}_{\text{ref}}); \mathbf{R}_{\text{ref}}, \mathbf{X}]} \end{aligned} \quad (24)$$

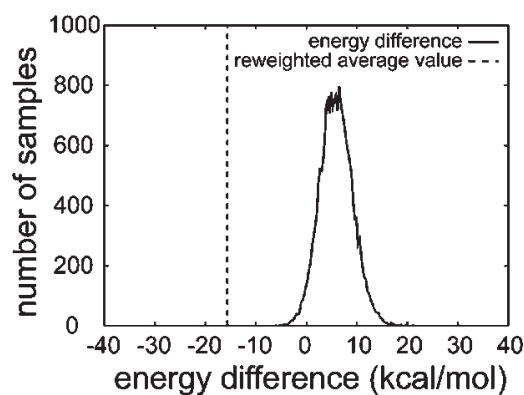


Figure 5. Histogram of the QM–MM interaction energy differences of MM conformational samples, $\Delta E^{\text{QM-MM}}(\mathbf{d}, \mathbf{d}_{\text{ref}}, \mathbf{R}, \mathbf{R}_{\text{ref}}, \mathbf{X})$, and its reweighted average, eq 24, at the end of the geometry optimization cycle in the first step of the sequential sampling.

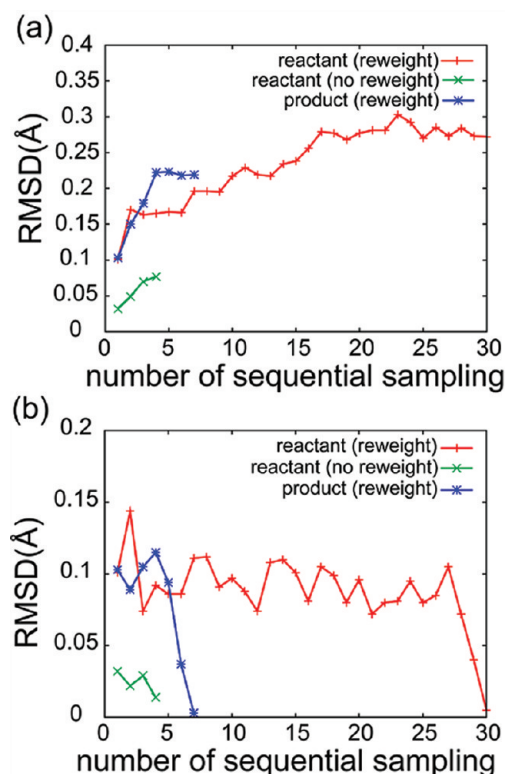


Figure 6. Changes of the QM geometries during the sequential samplings in the reactant and the product. RMSDs from the initial QM coordinates (a) and RMSDs from the QM coordinates optimized at the preceding steps of the sequential samplings (b) are plotted. Red and blue lines indicate RMSDs in the reactant and the product, respectively. Green lines indicate RMSDs of free energy geometry optimizations without the reweighting of the MM distribution in the reactant.

The reweighted average value is far out of the distribution as one MM conformation which exhibits a large energy difference gives a dominantly large reweighting factor. Although the MM conformation is found to provide a strongly stabilizing QM–MM interaction, the statistical averaging over the reweighted distribution represented only by the single MM conformation is no longer valid. This ill behavior of the reweighted average is due to

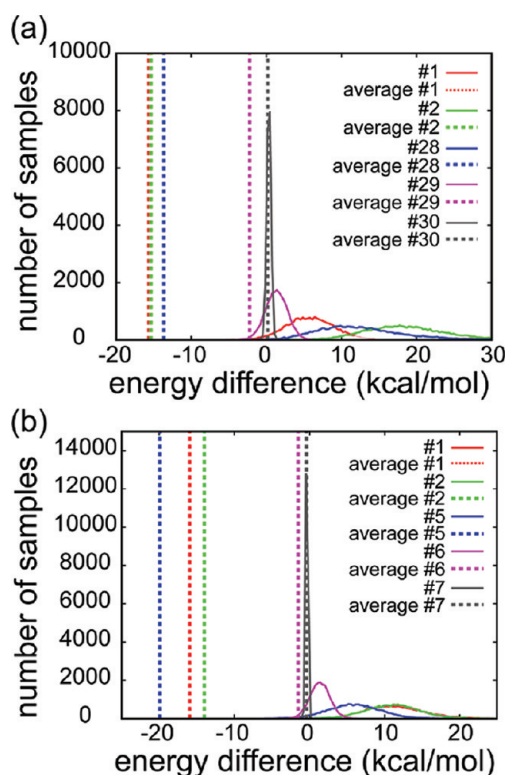


Figure 7. Changes of histogram of the QM–MM interaction energy differences of MM conformational samples, $\Delta E^{\text{QM-MM}}(\mathbf{d}, \mathbf{d}_{\text{ref}}, \mathbf{R}, \mathbf{R}_{\text{ref}}, \mathbf{X})$, and its reweighted average, eq 24, during the sequential samplings in the reactant (a) and the product (b). The histogram and the average are calculated with QM coordinates and charges at the end of the geometry optimization in each step of the sequential samplings. Histograms and reweighted averages of the first two and the last three steps of the sequential samplings are shown.

limited MM samples which fail to cover the MM distributions for the updated \mathbf{d} and \mathbf{R} during the geometry optimization. Naturally, the fewer MM samples obtained by the shorter MD trajectories tested above were observed to lead to ill behavior as well. The sequential sampling which redistributes the MM samples therefore needs to be continued until the ill behavior of the reweighted average disappears. The high computational efficiency featured in the present method is therefore a prerequisite for obtaining sufficient MM samples that avoid the ill behavior of the reweighting scheme.

Figure 6 depicts RMSD changes of the QM coordinates along the sequential sampling iterations. The QM geometries undergo changes from the initial ones as the sequential samplings proceed. Then, the geometry optimizations converged at 30 and 7 steps of the sequential samplings for the reactant and product states, respectively. Since the MD trajectory at each step of the sequential sampling is calculated for 3 ns (1 ns for equilibration and 2 ns for the sampling of the MM distribution), the MD simulations were carried out for 90 and 21 ns in total for the reactant and the product. The reason for the very long MD simulation time required for the convergence in the reactant state is discussed later. As seen in RMSDs from the previous optimized QM coordinates (Figure 6b), the movements of the QM coordinates become small in the last two steps of the sequential samplings, indicating clear convergence behavior of the MM distribution. Figure 7 shows the histograms of the energy difference and their

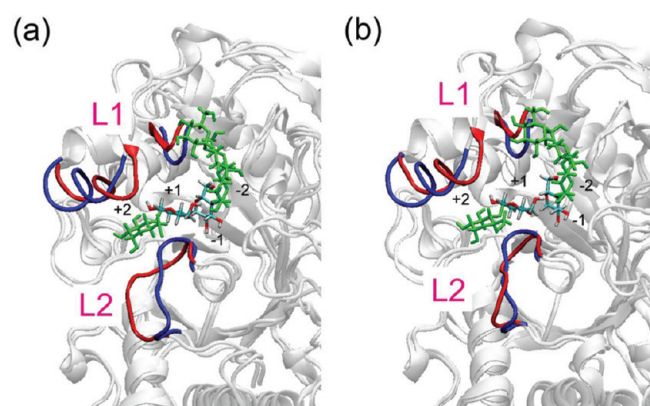


Figure 8. Conformational changes of the protein loops, L1 and L2, adjacent to the substrate binding site observed in the free energy geometry optimizations. The loops undergo large conformational transitions from the initial MM structure obtained by the QM/MM potential energy geometry optimization (blue) to the final structure of the free energy geometry optimization (red) in the reactant (a) and the product (b). Reweighted average structures are shown for the free energetically optimized ones. The L1 and L2 loops correspond to regions 5 and 9, respectively, defined in a previous paper.³⁵ The substrates are depicted in licorice representation, and their QM and MM regions are drawn in colors based on the atom type and in green, respectively. Numbers, -2 to $+2$, are the subsite indices of the substrate.

averaged values at the early and final steps of the sequential sampling iterations. For both of the reactant and product states, the reweighted averages exhibit the ill behavior seen above before the last two steps of the sequential samplings. On the other hand, in the last two steps, the reweighted averages stay in well-distributed regions of the histograms. The averaging with the reweighted distribution therefore becomes statistically valid with the MM distribution at the end of the sequential sampling.

Finally, the validity of the mean field approximation for the electronic wave function is assessed with the MM distribution of the last sampling. We compared the QM charges and the QM/MM energy (the QM energy plus the QM–MM interaction energy) obtained in the mean ES field to those evaluated without the mean field approximation. The latter are obtained through a reweighted average of the quantities which are determined by a series of QM/MM calculations for individual MM configurations in the distribution. For comparison, we limited the number of MM configuration for the ensemble averages to 2000 out of 20 000 of the MM distribution since the series of QM/MM calculations for the average evaluation without the mean field approximation is very time-consuming. We recalculated the mean field quantities using a QM/MM-RWFE-SCF calculation with the small MM ensemble and confirmed that the reweighted average of the QM–MM interaction energy differences stays in a well-distributed region of their histogram for the small ensemble.

Figure S5 (Supporting Information) compares the QM charges determined with and without the mean field approximation. The mean field charges are in accord with those without the mean field approximation. The maximum value and the standard deviation of the error are 7.07×10^{-4} and 2.19×10^{-4} , respectively. The error of the QM/MM energy, 0.48 kcal/mol, is also reasonably small. The mean field approximation therefore provides a reasonable description for the complex protein system as well as for simple solutions systems reported previously.^{30,45,46}

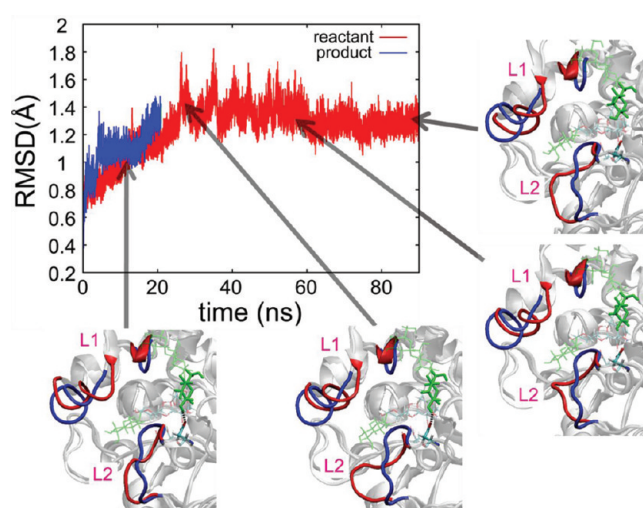


Figure 9. Time evolution of the protein conformational changes during the sequential samplings. RMSDs of the MM structures from the initial ones in the reactant and the product are shown. Representative snapshots of the structure during the sequential sampling in the reactant are also depicted. L1 and L2 protein loops drawn in blue and red are the initial structure and the snapshot structures, respectively. The substrate and Asp264 are drawn in licorice representation, and their QM and MM regions are depicted in colors based on the atom type and in green, respectively. A dashed line indicates a hydrogen bond between glucose (-2) and Asp264 drawn in thick licorice representation, which undergoes dissociation around 60 ns.

Large Conformational Changes of Protein Found in the QM/MM-RWFE-SCF Geometry Optimization. As seen above, the QM/MM-RWFE-SCF geometry optimizations converged after the sequential sampling with the MD simulations for 90 and 21 ns in total for the reactant and the product, respectively, which are much longer than those in previous studies.^{23,25,29,30} During the MD simulations, the protein structures around the binding site underwent large conformational changes, as shown in Figure 8. A loop, L1, approaches the substrate and forms interaction with it in both of the reactant and product states in a similar fashion. Another loop, L2, interacting with the substrate from the other side of L1 also changes its conformation in the reactant state, whereas it does not exhibit large changes in the product state. Consequently, the L2 loop forms an extensive interaction with glucose($+2$) (the number in parentheses is the index of the subsite) in the reactant state, which is absent in the product structure after the geometry optimization as well as in the initial structure of the optimization.

Figure 9 depicts the time evolution of the conformational changes. The movement of the L1 loop almost completes within 10 ns in both the reactant and the product. On the other hand, the L2 loop keeps fluctuating up to 60 ns in the reactant state. At ~ 60 ns, a hydrogen bond between a carboxylate of Asp264 in the QM region and a hydroxyl group of glucose (-2) in the MM one breaks. Then, the L2 loop forms the extensive interaction with glucose($+2$) and becomes stable after breakage of the hydrogen bond. Finally, the optimization of the QM coordinates converges in equilibration for a few tens of nanoseconds. In the case of the product state, the convergence is attained much earlier because the hydrogen bond breaks at a few nanoseconds and the movement of the L2 loop is smaller.

The observations above demonstrate that the present method is capable of determining the optimal QM geometry on an

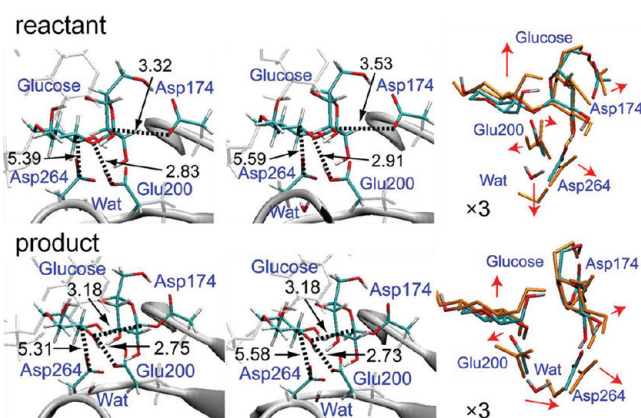


Figure 10. Changes of the QM geometries obtained with the QM/MM free energy optimizations. The left and middle panels depict the initial QM structures determined by the QM/MM potential energy geometry optimization and the free energetically optimized QM ones, respectively. Distances between Asp174:O(6) and glucose:C(41) for the reactant, glucose:O(40) and glucose:C(41) for the product, Glu200:O(12) and glucose:O(40), and Asp264:C(17) and glucose:C(30) are shown (see Figure S1, Supporting Information, for the atom index). The right panels illustrate overall changes of the QM geometries. Structures drawn in colors based on the atom type and in orange indicate the initial ones and the final ones where the structural changes are magnified by a factor of 3 for emphasis.

extensive free energy surface of the MM distribution by following slow conformational relaxation of the protein on a submicrosecond scale. In the present case, one reason for the conformational changes at L1 and L2 in the relaxation are presumably due to a distorted initial conformation by crystal packing of the X-ray crystallographic model since those loops which were found to be flexible³⁵ are in contact with an adjacent protein in the crystal packing. The present free energy geometry optimization method therefore removes properly the possible large distortion of the protein conformation that is hardly detectable by shorter MD simulations.

Furthermore, more notably, the submicrosecond MD searches from the similar initial protein conformations for the reactant and the product found the remarkably different conformations of the L2 loop depending on the catalytic reaction states in the QM region (Figure 8), which cannot be found with an MD conformational search on a less than nanoseconds scale. The observation may imply that the present free energy geometry optimization identified successfully the large and slow conformational change of the protein that couples with the catalytic reaction step and thus plays a role in the enzymatic catalysis, although further examination is necessary for the proposal because of the possibility that those conformations are trapped in local free energy minima.

Optimized Structures of the Catalytic Site. Figure 10 shows comparison between the QM structures optimized by the present method and those by a QM/MM method based on the potential energy surface. In the reactant state, the intramolecular distances between the substrate and three carboxyl groups in the catalytic site, i.e., Asp174, Glu200, and Asp264, increase on the free energy surface (see also Figure S6 in the Supporting Information). The thermal fluctuation of the MM region taken into account therefore relaxes the interaction for the substrate binding. Two negatively charged carboxylates among them, which repel each other, may induce expansion of the binding pocket in thermal fluctuation and thus enhance the relaxation of the interaction.

In the product state, the distance between the substrate and Asp264 increases as well as in the reactant state because of the dissociation of a hydrogen bond described above (Figure 9). On the other hand, intermolecular distances between the groups involved in the catalytic reaction do not undergo large changes. As shown in Figure 10, the atoms that form the glycosidic bond in the reactant state, i.e., O(40) and C(41) (the numbers in parentheses are the atom indices defined in Figure S1, Supporting Information), keep their mutual distance of 3.18 Å in the product state upon the free energy optimization (see also Figure S6, Supporting Information). Moreover, the distance between the glycosidic oxygen atom, O(40), and the O(12) atom of Glu200, which is the proton donor for the dissociation of the glycosidic bond, becomes even shorter than that optimized on the potential energy surface, 2.75 Å \rightarrow 2.73 Å. Their relatively short non-bonding distances compared with those in a nonreactive condition (e.g., 3.59 Å for O...C estimated with the LJ parameters of Amber force field) manifest a strong electronic interaction in the reaction core region. Note that the product state of the reaction step considered in the present study corresponds to an intermediate of the overall enzymatic reaction as mentioned above. Thus, the reaction core region in the product state is still reactive, and the partial bonding character for the seemingly nonbonding interactions among those reaction core atoms remains. The present method is therefore capable of describing such complex electronic interactions quantum mechanically in a thermally fluctuating environment.

ES Potential of the Catalytic Site. Figure 11 shows the ES potential produced by the MM region acting on the QM catalytic site, which plays a role in the enzymatic catalysis. The extensive relaxations of the protein observed in the free energy optimization alter the ES potential largely. Figure 11a displays differences in the mean field ES potentials acting on the QM atoms between the initial and final MM distributions of the sequential sampling, i.e., $\Delta_{ss-opt} V_A = V_A(\text{final step}) - V_A(\text{initial step})$. The initial MM distribution was obtained for the geometry and the charges of the QM molecules determined by the QM/MM potential energy geometry optimization without reweighting and thus corresponds to that used in an approximate scheme for estimation of the QM/MM reaction free energy.^{47,48} Large negative peaks of the ES potential differences around O(18) of Asp264 and C(50) of glucose (−1) are due to the dissociation of a hydrogen bond of Asp264 with a hydroxy group of the neighboring glucose (−2) (see Figure 9). As the absolute values of the ES potential in the QM region are positive because of a negative net charge (−2) of the QM region, the negative differences of the ES potentials indicate decreases of the positive ES potentials. One can also discern large changes of the ES potentials on glucose(+1), which originate from the large movements of loops, L1 and L2 (see Figure 9).

Figure 11b shows differences in the mean field ES potentials between the reactant and the product at the free energetically optimized states, i.e., $\Delta_{\text{reaction}} V_A = V_A(\text{product}) - V_A(\text{reactant})$, which identify important ES interactions for the catalysis. A large decrease and increase of ES potentials on Asp174 and Glu200, respectively, represent ES reorganizations of the protein for changes of protonated states of those groups upon the reaction (see Figure 1). Several large peaks resulting from ES reorganization are also found in glucose(−1). Furthermore, increases of the ES potential are extended over the regions of Asp264 and glucose(+1). In order to examine the origins of the ES potential differences, we evaluated differences in $\Delta_{\text{reaction}} V_A$ between the

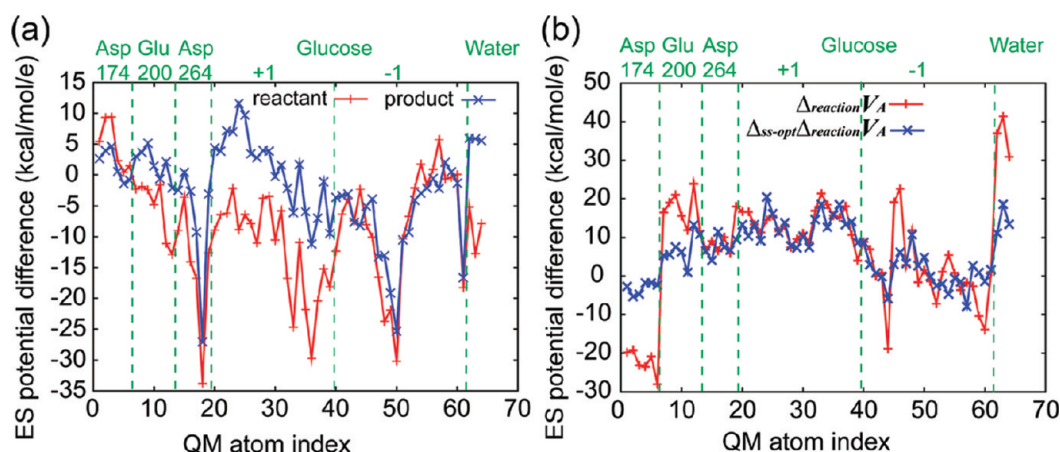


Figure 11. (a) Differences in the mean field ES potentials acting on the QM atoms between the initial and final MM distributions of the sequential sampling, $\Delta_{ss-opt}V_A = V_A(\text{final step}) - V_A(\text{initial step})$. (b) Differences in the mean field ES potentials between the reactant and the product. Red and blue lines indicate differences in the mean field ES potentials between the reactant and the product at the free energetically optimized states, $\Delta_{\text{reaction}}V_A = V_A(\text{product}) - V_A(\text{reactant})$, and differences in $\Delta_{\text{reaction}}V_A$ between the initial and final MM distributions of the sequential sampling, $\Delta_{ss-opt}\Delta_{\text{reaction}}V_A = \Delta_{\text{reaction}}V_A(\text{final step}) - \Delta_{\text{reaction}}V_A(\text{initial step})$, respectively.

initial and final MM distributions of the sequential sampling, i.e., $\Delta_{ss-opt}\Delta_{\text{reaction}}V_A = \Delta_{\text{reaction}}V_A(\text{final step}) - \Delta_{\text{reaction}}V_A(\text{initial step})$, which express ES reorganization due to the conformational changes resulting from the sequential sampling. Unlike other regions, $\Delta_{\text{reaction}}V_A$ of Asp264 and glucose(+1) are close to $\Delta_{ss-opt}\Delta_{\text{reaction}}V_A$, indicating that the ES reorganization comes from the large conformational changes of the L1 and L2 loops. On the other hand, the large $\Delta_{\text{reaction}}V_A$'s of Asp174 are mainly attributed to ES reorganization in a linear response regime because of small $\Delta_{ss-opt}\Delta_{\text{reaction}}V_A$'s which indicate that the ES reorganization is already present in the initial MM distribution of the sequential sampling. It is noteworthy that large positive $\Delta_{ss-opt}\Delta_{\text{reaction}}V_A$'s are found for the reaction core atoms, O(12) (13.2 kcal/mol/e) and H(13) (10.8 kcal/mol/e) of Glu200, and the glycosidic oxygen, O(40) (8.7 kcal/mol/e), which contribute to catalysis of the reaction by stabilization of the developing negative partial charges on those atoms upon the reaction. As the L2 loop is in close proximity to those reaction core atoms, the large movement of L2 accompanied by the reaction is responsible for the generation of the positive ES potential differences.

Comparison with Other QM/MM Free Energy Methods.

The present QM/MM-RWFE-SCF method is based on a combination of theories developed by Yamamoto³⁰ and Yang and co-workers.²⁵ As demonstrated above, the highly improved computational efficiency furnished by the combination allows one to search a QM/MM optimized structure on an unprecedentedly extensive free energy surface. In order to clarify the significant feature of the present method, we compare the present method with other QM/MM methods for the examination of the reaction free energy profile.

Aguilar and co-workers proposed a MM mean field QM/MM method called ASEP/MD.^{28,29} In this method, the MM ES mean field is represented by point charges on grid points around the QM region that are fitted so as to reproduce the mean field acting on the QM region. This contraction of the MM ES mean field reduces the computational cost for time-consuming direct statistical averaging of the ES one electron integral given by eq 11. Theoretically, the description of the QM/MM ES mean field interaction used in the ASEP/MD method based on a form of

eq 11 is more precise than that with the RESP operator introduced in the present study if the mean field is accurately reproduced by the contraction. However, the method suffers from a drawback of inconsistency of the MM thermal distribution. In the case that the QM-MM ES interaction is described in a one electron integral form of eq 11, one needs to compute the time-consuming one electron integrals at each step of a MD trajectory calculation for evaluation of the MM distribution in order to keep the consistency of the MM distribution. The computation of one electron integrals increases drastically the computational time of the MD simulation and thus limits severely the sampling time. In the ASEP/MD method, therefore, the QM-MM ES interaction in the MD simulation is approximately evaluated in a classical Coulombic form with RESP charges of the QM atoms. However, the approximate description of the QM-MM ES interaction in the MD simulation introduces inconsistency with that in the QM/MM geometry optimization in the mean field of MM distribution and in fact led to a slow convergence behavior of the geometry optimization.²⁹ Such a poor convergence might make calculation of the Hessian matrix required for transition state determination and vibrational modes difficult. Furthermore, since the reweighting scheme cannot be applied to the contracted ES mean field, the method requires very frequent updates of the MM distribution during the geometry optimization, which is not suitable for protein systems with slow relaxation (see also below).

Yamamoto developed a mean field QM/MM theory, QM/MM-FE with mean-field embedding,³⁰ on which the mean field ES interaction term of the Fock or KS equation, eq 14, employed in the present study is derived. Although the sequential sampling scheme is also utilized in the method, the MM distribution is not changed during a QM/MM free energy geometry optimization cycle. The iteration scheme simplifies the SCF cycle of the electronic function; the update of the ES mean field by eqs 16 and 17 at each SCF step is not necessary. However, as mentioned above, the unchanged MM distribution is no longer optimal once the electronic wave function (i.e., the charges) and the geometry of the QM region are updated in the QM/MM free energy geometry optimization cycle. Hence, the variational condition for the free energy functional is not satisfied at most steps of the

optimization cycles. The violation of the variational condition causes an arrested minimum search on the free energy surface, leading to slow convergence of the geometry optimization. Figure 6 depicts RMSDs of the first four optimization steps without the reweighting scheme. The RMSD step sizes that result from the optimization searches are considerably small compared with those from the present method. Furthermore, the Hessian matrix calculation by a finite differential method is more difficult. In principle, the MM conformational samples need to be obtained for each of the QM coordinates with small displacements, whereas the reweighting scheme allows one to employ the same samples throughout the calculation as described above.

Yang and co-workers developed the QM/MM-MFEP method,^{24,25} where the reweighting scheme for the QM/MM free energy optimization is introduced. Unlike the other methods described above and the present one, an electronic wave function variational to a free energy functional is not directly solved. Instead, an approximate Hamiltonian function is defined with a reference electronic wave function and its charge response kernel (CRK)^{49,50} that describes linear QM charge response to the MM ES field. Free energy geometry optimization with the sequential sampling is performed on the free energy surface of the approximate Hamiltonian. From a theoretical point of view, the method is advantageous over the mean field approximation since the treatment takes into account the linear response fluctuation of the QM charges in the free energy function, although the effect of the fluctuation of the QM charges is expected to be very minor for the present protein system, as described above. However, computationally, the method includes shortcomings. One is determination of the reference electronic wave function. The change in the ES interaction due to polarization of the QM wave function is represented only by the linear changes of the QM charges from values obtained for the reference electronic wave functions. Thus, the reference electronic wave function needs to be close to the optimal one in order to describe accurately the QM polarization by the linear response approximation. However, no reasonable way to determine such a good reference electronic wave function has been proposed. It is suggested that the present mean field method would provide a good reference wave function for the linear treatment of the QM polarization. Another drawback is that the method is not variational with respect to the electronic wave function. Hence, evaluation of the free energy gradient requires the calculation of coupled perturbed equations, which reduces computational efficiency. Furthermore, MD trajectory calculations for MM conformational samplings were very limited in the studies reported in refs 24 and 25 (approximately 100 ps for each sequential sampling). As revealed above, however, the reweighting scheme is very sensitive to the MM distribution and leads easily to the serious ill behavior of the reweighted distribution for an insufficiently sampled MM distribution. Careful examination of the reweighted MM distribution as carried out in the present study is suggested when the reweighting scheme is used.

Finally, QM/MM-MD simulations with empirical QM methods such as EVB,¹⁹ MCMM,^{20,21} and DFTB^{17,18} are mentioned. Those empirical methods are furnished with computational efficiency enough to carry out MD simulations for a relatively long time with accuracy attained through parametrizations of empirical Hamiltonians so as to reproduce energies and forces of high-level *ab initio* QM or QM/MM calculations. The direct MD simulations with QM/MM Hamiltonians allow one to sample

thermal distributions of the QM coordinates which cannot be obtained by the QM/MM geometry optimization methods described above. However, in general, accurate parametrizations of the Hamiltonian become difficult for complex reactions where the QM systems are strongly correlated with large conformational changes of the MM surroundings. It is therefore suggested that the present method is suitably employed to obtain the reference energies and geometries for the parametrization because the method provides more accurate descriptions of electronic wave function, geometry, and normal modes at the special states of reaction than conventionally used ones, i.e., gas phase QM methods or QM/MM ones based on potential energy surfaces.

CONCLUSION

We developed a QM/MM free energy optimization method by combining a mean field QM/MM theory³⁰ with a reweighting technique for the MM ensemble averaging.²⁵ This QM/MM-RWFE-SCF method features applicability for enzymatic reactions that involve extensive, heterogeneous, and slow thermal relaxation of the protein. Its high efficiency of computational scheme and precise description for long-range ES interaction using the Ewald summation technique enable one to explore an enzymatic reaction on an extensive free energy surface of the protein conformation. We demonstrated free energy geometry optimizations of the reactive substrate following global and non-linear protein conformational changes of the α -amylase protein on a time scale reaching the submicrosecond level. The free energy geometry optimizations revealed that a loop adjacent to the catalytic site forms in significantly different conformations in the reactant and the product, respectively, and produces a catalytic ES field for the enzymatic reaction. The method now opens the way for theoretical examination of a proposal on enzymatic catalysis by slow protein dynamics, which was under debate recently.^{51,52}

ASSOCIATED CONTENT

S Supporting Information. Index for atoms in the QM site, differences between Ewald ES potentials and Ewald ES potentials without interaction between the QM regions in real and reciprocal space, differences between Ewald ES forces and Ewald ES forces without interaction between the QM regions in real and reciprocal space, convergence of Ewald ES potentials with respect to the number of k vectors of reciprocal space, correlation between QM charges evaluated using the QM/MM-RWFESCF method and those using the average without the mean field approximation for 2000 conformations in the last MM distribution of the sequential sampling, and changes of distances between atoms in the QM sites during the sequential samplings. This material is available free of charge via the Internet at <http://pubs.acs.org>.

AUTHOR INFORMATION

Corresponding Author

*Tel.: +81-75-753-4006. Fax: +81-75-753-4000. E-mail: hayashig@kuchem.kyoto-u.ac.jp.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENT

The study was supported by research fellowships for young scientist from the Japan Society for the Promotion of Science (JSPS) to T.K., by a Grant-in-Aid for Scientific Research on Priority Areas (18074004) and that on Innovative Areas (23107717) from the Ministry of Education, Culture, Sports, Science, and Technology, Japan; by a Grant-in-Aid for Scientific Research from JSPS (23700580); by Research and Development of the Next-Generation Integrated Simulation of Living Matter; and the Global COE program "International Center for Integrated Research and Advanced Education in Materials Science". The molecular images were created with VMD.⁵³

REFERENCES

- (1) Fersht, A. *Enzyme Structure and Mechanism*, 2nd ed; W. H. Freeman and Company: New York, 1985.
- (2) Suckling, C. *Enzyme Chemistry: Impact and application*, 2nd ed; Chapman and Hall: London, U.K., 1990.
- (3) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227–229.
- (4) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *6*, 700–733.
- (5) Gao, J. *Acc. Chem. Res.* **1996**, *29*, 298–305.
- (6) Svensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K. *J. Phys. Chem.* **1996**, *100*, 19357–19363.
- (7) Monard, G.; Merz, K. M., Jr. *Acc. Chem. Res.* **1999**, *32*, 904–911.
- (8) Hess, B. *Phys. Rev. E* **2000**, *62*, 8438–8448.
- (9) Hess, B. *Phys. Rev. E* **2002**, *65*, 031910.
- (10) Klepeis, J. L.; Lindorff-Larsen, K.; Dror, R. O.; Shaw, D. E. *Curr. Opin. Struct. Biol.* **2009**, *19*, 120–127.
- (11) Nishihara, Y.; Kato, S.; Hayashi, S. *Biophys. J.* **2010**, *98*, 1649–1657.
- (12) Jimenez, R.; Fleming, G. R.; Kumar, P. V.; Maroncelli, M. *Nature* **1994**, *369*, 471–473.
- (13) Devi-Kesavan, L. S.; Gao, J. *J. Am. Chem. Soc.* **2003**, *125*, 1532–1540.
- (14) Ridder, L.; Rietjens, I. M. C. M.; Vervoort, J.; Mulholland, A. J. *J. Am. Chem. Soc.* **2002**, *124*, 9926–9936.
- (15) Ruiz-Pernía, J. J.; Silla, E.; Tuñón, I. *J. Phys. Chem. B* **2006**, *110*, 20686–20692.
- (16) Lameira, J.; Alves, C. N.; Moliner, V.; Martí, S.; Castillo, R.; Tuñón, I. *J. Phys. Chem. B* **2010**, *114*, 7029–7036.
- (17) Han, W.; Elstner, M.; Jalkanen, K. J.; Frauenheim, T.; Suhai, S. *Int. J. Quantum Chem.* **2000**, *78*, 459–479.
- (18) Cui, Q.; Elstner, M.; Frauenheim, T.; Kaxiras, E.; Karplus, M. *J. Phys. Chem. B* **2001**, *105*, 569–585.
- (19) Warshel, A.; Weiss, R. M. *J. Am. Chem. Soc.* **1980**, *102*, 6218–6226.
- (20) Kim, Y.; Corchado, J. C.; Villà, J.; Xing, J.; Truhlar, D. G. *J. Chem. Phys.* **2000**, *112*, 2718–2735.
- (21) Higashi, M.; Truhlar, D. G. *J. Chem. Theory Comput.* **2008**, *4*, 790–803.
- (22) Okuyama-Yoshida, N.; Nagaoka, M.; Yamabe, T. *Int. J. Quantum Chem.* **1998**, *70*, 95–103.
- (23) Okuyama-Yoshida, N.; Kataoka, K.; Nagaoka, M.; Yamabe, T. *J. Chem. Phys.* **2000**, *113*, 3519–3524.
- (24) Hu, H.; Lu, Z.; Yang, W. *J. Chem. Theory Comput.* **2007**, *3*, 390–406.
- (25) Hu, H.; Lu, Z.; Parks, J. M.; Burger, S. K.; Yang, W. *J. Chem. Phys.* **2008**, *128*, 034105.
- (26) Higashi, M.; Hayashi, S.; Kato, S. *Chem. Phys. Lett.* **2007**, *437*, 293–297.
- (27) Higashi, M.; Hayashi, S.; Kato, S. *J. Chem. Phys.* **2007**, *126*, 144503.
- (28) Sánchez, M. L.; Aguilar, M. A.; Olivares del Valle, F. *J. Comput. Chem.* **1997**, *18*, 313–322.
- (29) Galván, I. F.; Sánchez, M. L.; Martín, M. E.; Olivares del Valle, F. J.; Aguilar, M. A. *J. Chem. Phys.* **2003**, *118*, 255–263.
- (30) Yamamoto, T. *J. Chem. Phys.* **2008**, *129*, 244104.
- (31) Siddiqui, K. S.; Cavicchioli, R. *Annu. Rev. Biochem.* **2006**, *75*, 403–433.
- (32) D'Amico, S.; Gerday, C.; Feller, G. *J. Biol. Chem.* **2001**, *276*, 25791–25796.
- (33) D'Amico, S.; Gerday, C.; Feller, G. *J. Biol. Chem.* **2002**, *277*, 46110–46115.
- (34) D'Amico, S.; Marx, J.-C.; Gerday, C.; Feller, G. *J. Biol. Chem.* **2003**, *278*, 7891–7896.
- (35) Kosugi, T.; Hayashi, S. *Chem. Phys. Lett.* **2011**, *501*, 517–522.
- (36) Qian, M.; Nahoum, V.; Bonicel, J.; Bischoff, H.; Henrissat, B.; Payan, T. *Biochemistry* **2001**, *40*, 7700–7709.
- (37) Hayashi, S.; Ohmine, I. *J. Phys. Chem. B* **2000**, *104*, 10678–10691.
- (38) Essman, V.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577–8593.
- (39) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Pearlman, D. A.; Crowley, M.; Walker, R. C.; Zhang, W.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Wong, K. F.; Paesani, F.; Wu, X.; Brozell, S.; Tsui, V.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Mathews, D. H.; Schafmeister, C.; Ross, W. S.; Kollman, P. A. *AMBER 9*; University of California: San Francisco, CA, 2006.
- (40) Aghajari, N.; Roth, M.; Haser, R. *Biochemistry* **2002**, *41*, 4273–4280.
- (41) Kirschner, K. N.; Woods, R. J. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 10541–10545.
- (42) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (43) Schmidt, M. W.; Baldrige, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem.* **1993**, *14*, 1347–1363.
- (44) Kaukonen, M.; Söderhjelm, P.; Heimdal, J.; Ryde, U. *J. Chem. Theory Comput.* **2008**, *4*, 985–1001.
- (45) Sánchez, M. L.; Martín, M. E.; Galván, I. F.; Olivares del Valle, F. J.; Aguilar, M. A. *J. Phys. Chem. B* **2002**, *106*, 4813–4817.
- (46) Galván, I. F.; Martín, M. E.; Aguilar, M. A.; Ruiz-López, M. F. *J. Chem. Phys.* **2006**, *124*, 214504.
- (47) Zhang, Y.; Liu, H.; Yang, W. *J. Chem. Phys.* **2000**, *112*, 3483–3492.
- (48) Ishida, T.; Kato, S. *J. Am. Chem. Soc.* **2003**, *125*, 12035–12048.
- (49) Morita, A.; Kato, S. *J. Am. Chem. Soc.* **1997**, *119*, 4021–4032.
- (50) Lu, Z.; Yang, W. *J. Chem. Phys.* **2004**, *121*, 89–100.
- (51) Nagel, Z. D.; Klinman, J. P. *Nat. Chem. Biol.* **2009**, *5*, 543–550.
- (52) Kamerlin, S. C. L.; Warshel, A. *Proteins* **2010**, *78*, 1339–1375.
- (53) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graph.* **1996**, *14*, 33–38.