

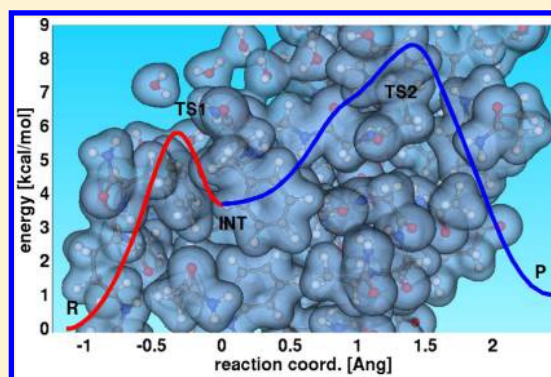
# Large-Scale First-Principles Molecular Dynamics Simulations with Electrostatic Embedding: Application to Acetylcholinesterase Catalysis

Jean-Luc Fattebert,<sup>\*,†</sup> Edmond Y. Lau,<sup>‡</sup> Brian J. Bennion,<sup>‡</sup> Patrick Huang,<sup>‡,§</sup> and Felice C. Lightstone<sup>‡</sup>

<sup>†</sup>Center for Applied Scientific Computing, Lawrence Livermore National Laboratory, Livermore, California 94550, United States

<sup>‡</sup>Physical and Life Sciences, Lawrence Livermore National Laboratory, Livermore, California 94550, United States

**ABSTRACT:** Enzymes are complicated solvated systems that typically require many atoms to simulate their function with any degree of accuracy. We have recently developed numerical techniques for large scale first-principles molecular dynamics simulations and applied them to the study of the enzymatic reaction catalyzed by acetylcholinesterase. We carried out density functional theory calculations for a quantum-mechanical (QM) subsystem consisting of 612 atoms with an  $O(N)$  complexity finite-difference approach. The QM subsystem is embedded inside an external potential field representing the electrostatic effect due to the environment. We obtained finite-temperature sampling by first-principles molecular dynamics for the acylation reaction of acetylcholine catalyzed by acetylcholinesterase. Our calculations show two energy barriers along the reaction coordinate for the enzyme-catalyzed acylation of acetylcholine. The second barrier (8.5 kcal/mol) is rate-limiting for the acylation reaction and in good agreement with experiment.



## INTRODUCTION

Molecular dynamics (MD) is a widely used method in computational biology. Most often classical potentials are used, and interactions between atoms are simply described by elementary rules based on distances and angles between atoms. However, very important chemical phenomena, such as bond breaking and bond formation, require accurate quantum-mechanical (QM) models and cannot be fully described by classical potentials.

To avoid the high computational cost of representing the whole system of interest at the QM level, one can restrict QM modeling to a subset of atoms around the chemical reaction of interest and treat the remaining system with a more approximate model, using either a continuum or classical particles. Combined quantum mechanics/molecular mechanics (QM/MM) methods are very popular for studying enzymatic reactions.<sup>1,2</sup> QM/MM methods come in many flavors, depending on the QM approach, the MM force field, their coupling, and the termination of the bonds cut at the QM–MM boundary. The size of the QM region obviously matters, as charge transfer and polarization effects in the vicinity of the region of interest can have a large influence on the phenomenon of interest. Thus, fairly large QM regions (150–1000 atoms) are required in order to achieve reliable accuracy.<sup>3,4</sup> Liao and Thiel<sup>5</sup> found that a QM region containing 408 atoms was required in order to get converged energies for the reaction catalyzed by acetylene hydratase. Hu et al.<sup>6</sup> recently carried out calculations with QM regions including as

many as 930 atoms to study the reaction energies for hydrogenase.

Even with a restricted subset of atoms at the quantum level, QM/MM molecular dynamics is limited to a few picoseconds, which is insufficient for accurate sampling of the available phase space for complex biological and chemical molecular systems. Often classical MD is performed to obtain sampling over a longer time scale, and QM calculations are carried out on a limited number of snapshots.<sup>6,7</sup> However, considering large QM regions introduces new problems, such as the question of whether geometry optimization is to be used for these structures and the introduction of a large number of local conformational minima.

In this paper, we are particularly interested in evaluating the free energy barrier between two given states in a chemical reaction. If one has a good guess of the chemical reaction coordinate, thermodynamic integration can be used to compute the free energy along that path.<sup>8</sup> Often a simple reaction coordinate can be chosen as a scalar parameter  $\lambda$  to parametrize that path. The free energy difference can then be calculated by integrating over energy changes along the path:

$$F(\lambda_2) - F(\lambda_1) = \int_{\lambda_1}^{\lambda_2} \left\langle \frac{\partial U(\lambda')}{\partial \lambda'} \right\rangle_{\lambda'} d\lambda' \quad (1)$$

Received: June 26, 2015

Published: October 22, 2015

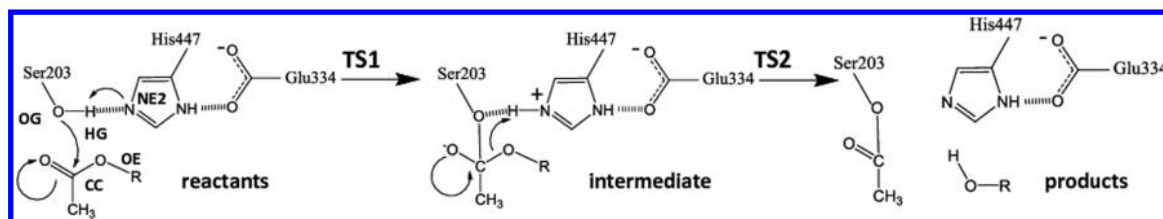


Figure 1. Acylation reaction of acetylcholine catalyzed by acetylcholinesterase.

where  $U$  denotes a parametrized potential energy and  $\langle \dots \rangle_{\lambda'}$  denotes a statistical average over the equilibrium ensemble associated with the parameter value  $\lambda'$ .

In practice, we perform this calculation by sampling a series of equilibrium configurations for a finite set of values of  $\lambda'$  using the multiconfiguration thermodynamic integration approach proposed by Straatsma and McCammon.<sup>9</sup> For each value of  $\lambda'$ , first-principles Born–Oppenheimer molecular dynamics in the  $NVT$  ensemble is used to obtain an ensemble average of the derivative of the energy of the system with respect to  $\lambda$ . Finally, we obtain the free energy by numerically integrating the derivative in eq 1. Other authors have used similar thermodynamic integration techniques to compute free energy profiles in biomolecular reactions using ab initio MD.<sup>10</sup>

In our numerical approach, we obtain our ensemble average by carrying out first-principles molecular dynamics (FPMD) for a subset of atoms comprising the biological system of interest (the QM region) embedded in a frozen electrostatic environment. We use an electrostatic embedding model similar to the one commonly used for QM/MM, but instead of including charges from MM, we precalculate a static external potential that we use throughout our molecular dynamics simulation, as proposed by Fattbert et al.<sup>11</sup> We use this electrostatic embedding model in combination with an  $O(N)$  complexity FPMD approach<sup>12,13</sup> to calculate energy barriers in a chemical reaction with a large QM region involving hundreds of atoms. We describe the reaction path by reaction coordinates corresponding to distances between atoms and distance differences between two pairs of atoms. We use constraints on the reaction coordinates to sample the reaction path, and we integrate the average forces required to enforce these constraints to obtain energy barriers.<sup>14</sup> Fox et al.<sup>7</sup> also used  $O(N)$  density functional theory (DFT) methods for large-scale first-principles calculations. Contrary to our approach, however, those authors used a classical potential for MD sampling and performed QM calculations only on a finite number of snapshot configurations.

Embedding in an external electrostatic potential is particularly important when a QM region contains a net charge. While linear combination of atomic orbitals (LCAO) methods restrict the electronic wave functions within the vicinity of the QM atoms by an artifact of the basis set, more general approaches, such as the finite-difference method (as used in this study) or the plane waves approach, do not confine electrons. While this absence of artificial confinement is a desirable feature from a general numerical point of view, this can lead to various problems if the net charge of the QM system is incorrect.

Using the methods described above, we chose to study the enzymatic reaction catalyzed by acetylcholinesterase (AChE). AChE plays a vital role in the central nervous system by hydrolyzing the neurotransmitter acetylcholine (ACh), resulting in termination of impulse signaling.<sup>15</sup> The catalytic rate of hydrolysis of ACh by AChE is one of the highest in nature and

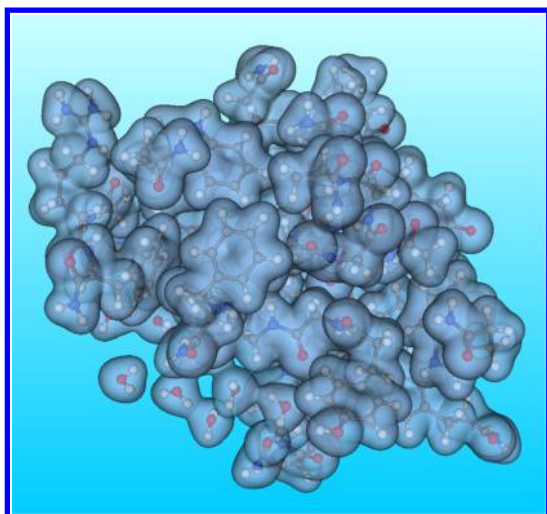
close to diffusion-limited (turnover number =  $7.4 \times 10^5 \text{ min}^{-1}$ ).<sup>16</sup> The active site is deeply buried within the enzyme, and a combination of hydrogen bonds and hydrophobic regions is used to position ACh in close proximity to the catalytic triad of Ser203–His447–Glu334 (human numbering) for facile reactivity.<sup>17</sup> AChE is among the most widely investigated enzymes by both experimental and computational methods. Of important interest is the binding and inhibition of AChE by organophosphates.<sup>18,19</sup>

The reaction of ACh catalyzed by this enzyme occurs in two stages, acylation and deacylation.<sup>20</sup> Transition state and intermediate structures are very difficult to study experimentally but are amenable to computational studies. In this contribution, we use large-scale QM/MD simulations to study the structures and energetics of the acylation reaction of ACh catalyzed by AChE (see Figure 1) and compare our results with those of previous studies.

## MODELING AND NUMERICAL METHODS

**Molecular System Preparation.** We performed all of the classical molecular dynamics calculations using the program NAMD<sup>21</sup> with the CHARMM 27 force field.<sup>22</sup> Acetylcholine was modeled into the active site of the apo human acetylcholinesterase structure (PDB ID 1B41)<sup>17</sup> using the program Autodock (version 3.05).<sup>23</sup> We solvated the protein in an  $81.8 \text{ Å} \times 70.1 \text{ Å} \times 87.8 \text{ Å}$  TIP3P<sup>24</sup> water box, whose size was sufficient to have at least  $12 \text{ Å}$  of water between the protein and interface. We added counterions ( $\text{Na}^+$  and  $\text{Cl}^-$ ) to the solution to physiological concentrations to neutralize the system. The final system contained a total of 46 879 atoms. We optimized the geometry of the system by energy minimization using 100 steps of the conjugate-gradient algorithm. We performed molecular dynamics starting from the energy-minimized system. The electrostatics was treated by particle mesh Ewald summation with a grid spacing of  $1 \text{ Å}$ .<sup>25</sup> Nonbonded interactions were cut off at  $12 \text{ Å}$ , and a  $2 \text{ fs}$  time step was used for the simulation. The system was heated to  $300 \text{ K}$  and equilibrated for  $1.2 \text{ ns}$ . We used the final structure from this MD simulation as the starting point for the QM calculations.

To perform our first-principles calculations, we divided the atomistic system into two parts: a primary (QM) system (encompassing the active site and several of its neighboring residues) and a secondary subsystem (the remaining atoms) that was not included in the QM calculation. All of the residues within  $9 \text{ Å}$  of the hydroxyl oxygen in Ser203 were considered to be part of the active site of the enzyme and were included in the QM region (612 atoms; Figure 2). For the secondary subsystem, we generated a list of atomic positions and associated partial charges (using CHARMM parameters). On the basis of that list of point charges, we generated a charge distribution given by a linear combination of Gaussian charge distributions associated with the individual atomic point

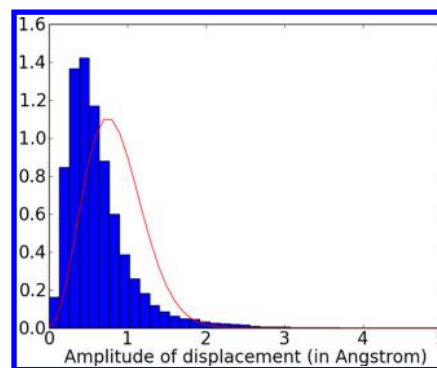


**Figure 2.** QM region of the AChE active-site in the reactant state. The isosurface of the calculated electronic density is overlaid on the 612 atomic centers.

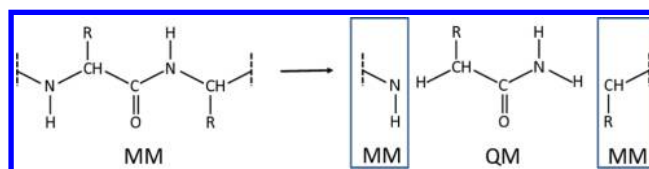
charges. The total charge of this distribution is equal to the sum of the point charges and is the opposite of the total charge of the QM system, resulting in an overall neutral system. We evaluated the sum of all of these Gaussian charges at each node of a uniform mesh covering the physical domain. We used a mesh spacing of  $\sim 0.12$  Å, resulting in a global uniform mesh of size  $672 \times 576 \times 704$ . We then calculated the Coulomb potential associated with that charge field by solving the Poisson equation on this mesh by finite differences using a multigrid solver. We used the resulting Coulomb potential as an external potential for the QM calculations, as proposed by Fattebert et al.<sup>11</sup>

By substituting a Gaussian charge distribution for each point charge associated with an MM atom, this model takes into account the average fluctuations of the electrostatic field generated by the MM charges. This model assumes that the charges are simply oscillating around the positions given by the MM snapshot chosen for our initial conditions and is a good approximation of what can be observed in practice for the system of interest on the time scale of a few picoseconds used for FPMD. Figure 3 illustrates this fact by comparing the Gaussian distribution used for the charges with the atomistic fluctuations observed over four snapshots of a 4 ps MM simulation of the molecular system of interest in this paper.

To prepare the QM system, we needed to properly terminate the bonds cut in the process of partitioning the atomistic simulation into primary and secondary subsystems. We followed the protocol described in our previous paper.<sup>11</sup> Briefly, partitioning of a QM residue occurs at NH–C $\alpha$  bonds. The bond between the NH and C $\alpha$  within the residue was cleaved, and a hydrogen atom was used to cap the C $\alpha$  to form a methylene group. We also cleaved the bond between the NH and C $\alpha$  in the trailing residue and capped it with a hydrogen at the NH to form an amide group (see Figure 4). Proper termination of the protein chain minimized disruption of the local electronic structure. Water molecules were either fully included in the QM region or not included at all. The resulting QM region was composed of 612 atoms. Within our pseudopotential approximation and neglecting spin, we ended up with 848 electronic wave functions to calculate.



**Figure 3.** Plot of the normalized distribution of atomic displacements for four AChE snapshots relative to an initial time point. The five coordinates were taken in 1 ps intervals. The red line shows the displacement distribution one could associate with the Gaussian charge distribution used to build the electrostatic field. It should be noted that water molecules were excluded from the statistics because they can diffuse and be interchanged without physical consequences.



**Figure 4.** Partitioning of a residue in our QM/MM scheme.

We carried out our quantum calculation on a fraction of the original domain used for the MM system. We used a QM domain with dimensions of  $36.57$  Å  $\times$   $30.82$  Å  $\times$   $31.69$  Å, which corresponds to a subset of the mesh points used for the Coulomb solver. The resulting mesh covering the QM computation domain was of size  $304 \times 264 \times 264$ . Since the QM mesh was a subset of the global mesh, the external electrostatic field was defined in the QM calculation by a simple injection of the global Coulomb potential.

Once this external potential was set, we optimized the geometry of the terminal atoms in this Coulomb field with all the other atoms fixed. Finally, we locked the positions of a shell of atoms at the boundary of the QM region and allowed only the atoms in the inner regions to move freely, including the active-site residues. This resulted in a total of 212 free QM atoms. Figure 2 shows an example of the electronic density isosurface calculated for the resulting QM problem.

**First-Principles Molecular Dynamics.** Our first-principles molecular dynamics approach is based on the  $O(N)$  DFT methodology proposed by Fattebert and Gygi.<sup>12,13</sup> In that approach, for  $N$  doubly occupied states, the DFT equations are formulated in a general form without orthonormality constraints as follows:

$$E_{\text{KS}}[\{\phi_i\}_{i=1}^N] = \sum_{i,j=1}^N (S^{-1})_{ij} \int_{\Omega} \phi_i(\mathbf{r}) (-\nabla^2) \phi_j(\mathbf{r}) d\mathbf{r} + \frac{1}{2} \int_{\Omega} \int_{\Omega} \frac{\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} d\mathbf{r}_1 d\mathbf{r}_2 + E_{\text{XC}}[\rho] + \sum_{i,j=1}^N 2(S^{-1})_{ij} \int_{\Omega} \phi_i(\mathbf{r}) V_{\text{ext}}(\mathbf{r}) \phi_j(\mathbf{r}) d\mathbf{r} \quad (2)$$

where the terms on the right side are the kinetic energy, the electrostatic energy, the exchange and correlation energy, and



the potential energy of the electrons in the potential  $V_{\text{ext}}$  modeling the atomic cores, respectively. The electron density is given by

$$\rho(\mathbf{r}) = 2 \sum_{i,j=1}^N (S^{-1})_{ij} \phi_i(\mathbf{r}) \phi_j(\mathbf{r}) \quad (3)$$

We used the pseudopotential approximation in its Kleinman–Bylander form<sup>26</sup> and the Perdew–Burke–Ernzerhof (PBE) exchange–correlation functional.<sup>27</sup> We discretized the energy functional (eq 2) by finite differences on a uniform mesh. We confined each electronic orbital  $\{\phi_i\}_{i=1}^N$  to a specific spherical region of radius  $R_c$ . These confinement regions limit the number of degrees of freedom associated with each orbital to a constant that does not increase with system size. This leads to  $O(N)$  degrees of freedom to represent  $N$  electronic wave functions and an  $O(N)$  computational complexity for the most computationally demanding operations. This truncation leads to an approximate solution with an error on the calculated atomic forces that decays exponentially fast as a function of the localization region radius.<sup>12</sup> We chose these regions to be large enough to reduce this error to an acceptable tolerance. For this study, we used  $R_c = 4.23 \text{ \AA}$ , which led to average cutoff errors of  $0.6 \text{ kcal mol}^{-1} \text{ \AA}^{-1}$  for the atomic forces according to our benchmarks tests. Confined electronic orbitals are typically localized on a chemical bond or where a lone pair would reside. Figure 5 illustrates our approach by displaying isosurfaces of a few of these localized orbitals as obtained in our QM system.

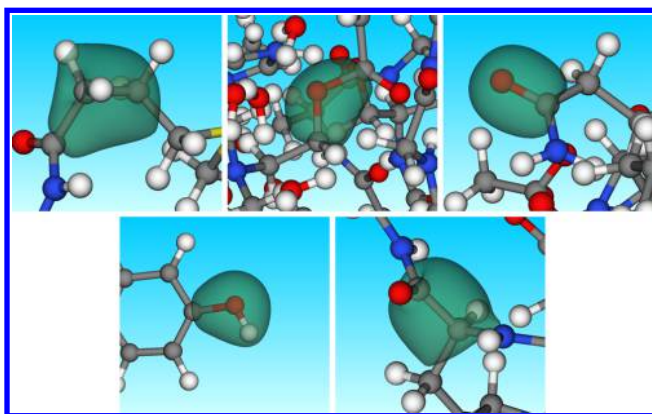


Figure 5. Isosurfaces of a few localized orbitals.

We used a Langevin thermostat to obtain a target temperature of 300 K. Deuterium masses were used for the hydrogen atoms and normal masses for all of the other elements. A time step of 0.36 fs was used. In our approach, the electronic orbital confinement regions evolve with the dynamics of the atoms. At every MD step, they are adapted on the basis of the center of charge of the latest computed orbitals.<sup>13</sup> We carefully monitored the amplitude of the displacement of each of these centers. When a charge center move is much larger than average, it typically means that the solution of the DFT equations has fallen from a local minimum associated with the orbital truncations<sup>13</sup> into another minimum of lower energy. This is not a cause for concern in itself since those minima lead to very similar atomic forces. However, such changes reduce convergence of the electronic structure iterative solver. To remedy this slowdown issue, we added an extra adaptation step and an extra iterative cycle to solve for the electronic wave

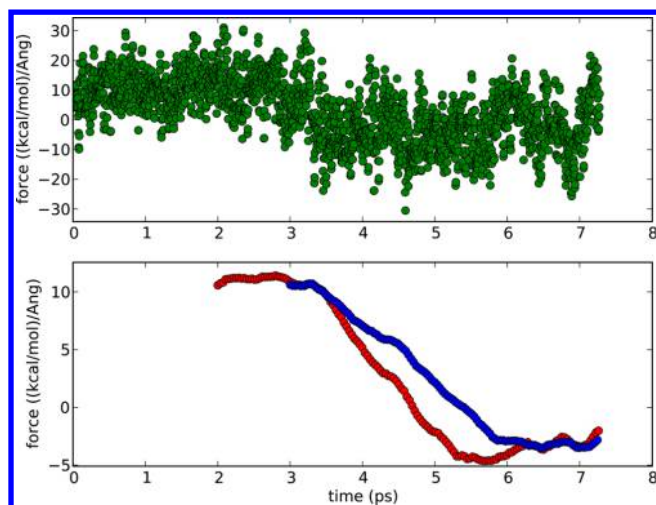
functions every time a move is larger than a set tolerance ( $0.026 \text{ \AA}$ ).

**Free Energy Barrier Computation.** As described in the Introduction, we used thermodynamic integration to compute the free energy along the reaction path.<sup>8,14</sup> Acylation of Ach by AChE goes through two energy barriers with a tetrahedral intermediate in between.<sup>28,29</sup> In the first part of the reaction, the Ser203 (human numbering) hydroxyl oxygen (OG) approaches the carbonyl carbon (CC) of acetylcholine and bonds with it. Simultaneously, the hydroxyl hydrogen (HG) is transferred to the neighboring nitrogen atom in the imidazole (NE2) of His447. To sample the first part of the acylation reaction (reactants to intermediate), we used the OG–CC distance as the reaction coordinate and ran several simulations at fixed (constraint) values along the reaction coordinate.

In the second part of the acylation reaction, the former hydroxyl hydrogen (HG) on the NE2 of His447 is transferred to the ester oxygen atom of the acetylcholine (OE), resulting in bond breaking with the carbonyl carbon (CC). The reaction path was obtained by constraining the difference in distance between HG and the atom to which it is bonded in the intermediate (NE2) and the distance between HG and the resulting hydroxyl oxygen to which it is bonded in the final choline product (OE). That is, the reaction coordinate is  $d(\text{HG} - \text{NE2}) - d(\text{HG} - \text{OE})$ . We sampled this path by choosing a discrete number of values for that reaction coordinate and running independent simulations for each of these values as a constraint.

We imposed constraints between atoms using the SHAKE algorithm.<sup>30</sup> For each constraint value  $\lambda$ , we ran a molecular dynamics simulation at 300 K for at least 4 ps. We obtained our initial conditions by geometry optimizations under the given constraints, producing the starting geometries of the reactants, products, and intermediate states. To calculate a free energy barrier by thermodynamic integration, the forces required to satisfy the constraints corresponding to  $\lambda$  must be determined. We calculated the average values of these forces over the last 3 ps of a simulation, considering the earlier time steps as the equilibration time. It should be noted that for the second reaction coordinate, we included a correction to the forces according to Sprik and co-workers.<sup>14,31</sup> We observed, however, that these corrections were negligible. Three picoseconds of sampling was typically not sufficient to converge statistical errors on an average of correlated data, but it did give some reasonable sampling of the configuration space for the QM system. In particular, this amount of time was sufficient to sample the vibrational modes of the QM system in the configuration restricted by the QM atoms frozen in place and the static embedding electrostatic field. The sampling also provided enough time for the system to forget about possible biased initial conditions. We used the running-average forces to estimate whether the system had reached a stable regime on the time scale considered. If we observed a significant drift in the running average, we extended the simulation for a few extra picoseconds until the system settled into a more stable regime. We stopped the simulations when the 3 ps running-average fluctuation was below a tolerance of  $1.2 \text{ kcal mol}^{-1} \text{ \AA}^{-1}$  over 1 ps. Figure 6 shows an example of such behavior that required over 6 ps to reach a nondrifting running-average force.

Typically, the QM/MD simulations were carried out using 363 CPUs of a Linux cluster (Intel Xeon EP X5660). With this resource, a single MD step took about 100 s (wall-clock time).

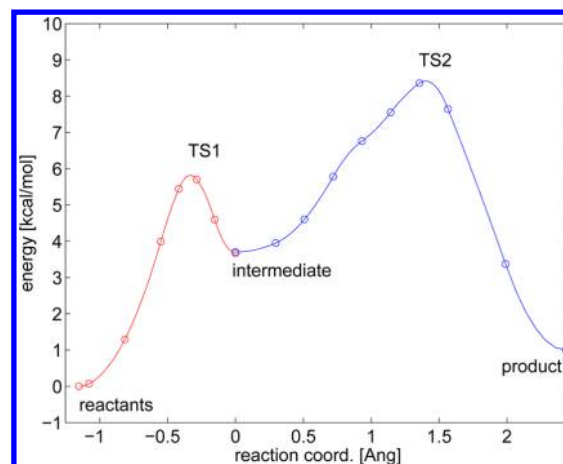


**Figure 6.** Illustration of (top) the instantaneous force acting to maintain a constraint and (bottom) the 2 ps (red) and 3 ps (blue) running averages of that force as functions of time for the MD run closest in energy to the second transition state (TS2). Only every other 10 points are shown for the instantaneous force and every other hundred points for the running averages. The data show a shift in the force due to a slight change in the geometry toward a more productlike structure for TS2.

## RESULTS AND DISCUSSION

The geometry of the reactant ACh in the active site of AChE is well-positioned for nucleophilic attack by the hydroxyl oxygen of Ser203. Two hydrogen bonds are formed between the backbone amides (NH) of Gly121 and Gly122 (the oxyanion hole) and the carbonyl oxygen (O) of ACh at distances of 2.93 and 2.79 Å, respectively. The average distance from OG of Ser203 to CC of ACh is 2.71 Å. The distance between OG and NE2 of His447 is 2.69 Å, which is reduced relative to the crystal structure distance of 2.90 Å. At the first transition state (TS1), the hydroxyl oxygen of Ser203 (OG) has lost its hydrogen (HG) to NE2 of His447. The HG–OG distance has increased from 1.01 Å in the reactant geometry to 1.47 Å in the TS1 structure. The loss of the hydrogen greatly enhances the nucleophilicity of the hydroxyl oxygen, and the separation between OG and CC of ACh is significantly reduced to 1.85 Å.

On the basis of the average forces measured on the constraints during the MD simulations as well as the average values of the reaction coordinates in the reactant, intermediate, and product states, an energy profile can be computed by a simple second-order polynomial interpolation between data points along the reaction coordinate (see Figure 7). The free energy barrier for the first transition state is about 6.0 kcal/mol. This is lower than the activation barriers of 10.5 kcal/mol calculated by Zhang et al.<sup>32</sup> and 12.4 kcal/mol by Zhou et al.<sup>33</sup> A charged tetrahedral intermediate structure is formed after passage through the first transition state. This intermediate is formed between ACh and Ser203 and is 2 kcal/mol lower in energy than TS1. Previous calculations have also found an intermediate structure that is 1–2 kcal/mol lower in energy than TS1.<sup>32,33</sup> The OG to CC distance in the intermediate is 1.56 Å. This charged intermediate is stabilized by the hydrogen bonds from the oxyanion hole, which have distances almost identical to those in the reactant geometry. Additionally, a third hydrogen bond is formed between the formerly carbonyl oxygen (O) of ACh and the backbone amide of Ala204 to provide further stabilization of the extra negative charge. The



**Figure 7.** Energy barrier along the reaction coordinate. The rate-determining step is predicted to be the second transition state with a barrier of 8.5 kcal/mol, which is comparable to the value of 11.8 kcal/mol derived from experiment. The calculations also predict a slightly endothermic acylation reaction.

distance between the amide NH of Ala204 and O of ACh contracts from 3.32 Å in the reactant structure to 2.82 Å in the intermediate structure. If unconstrained, the intermediate is unstable at  $T = 300$  K, and its structure reverts to the reactants. It is necessary to fix the OG–CC distance at a value of 1.57 Å to sample this geometry in the MD simulation.

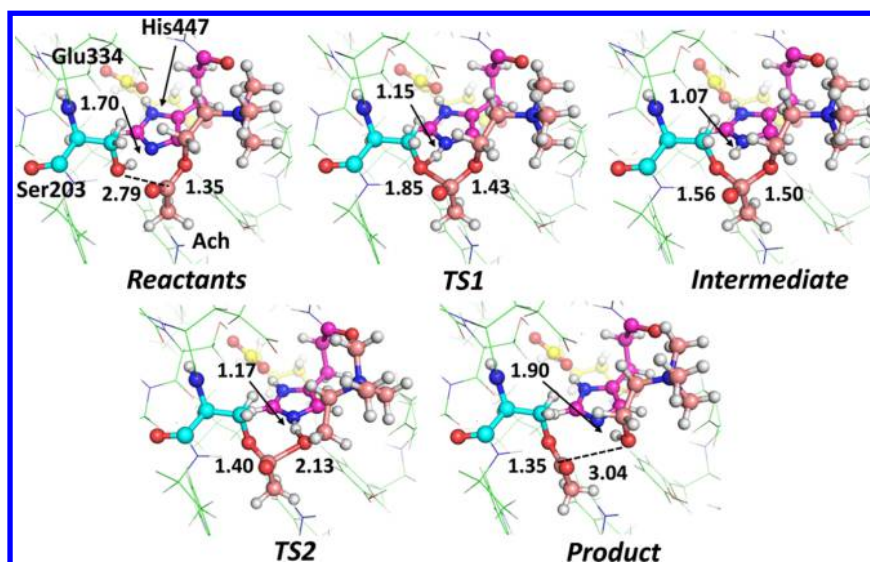
A second transition state (TS2) is formed when the imidazole of His447 rotates away from Ser203 and forms a hydrogen bond with the ester linkage of ACh (HG–OE distance = 1.41 Å). This change in position is commonly seen in crystal structures in which Ser203 has formed a covalent adduct.<sup>34,35</sup> This new interaction causes the CC–OE bond in ACh to elongate from 1.50 Å in the intermediate structure to 1.91 Å in the TS2 structure. The TS2 structure is 4.5 kcal/mol higher in energy than the intermediate. The acylation reaction is complete when HG is transferred from His447 to the ester oxygen of ACh, causing cleavage of the bond between CC and OE. These important distances are summarized in Table 1, and the active-site geometries are shown in Figure 8.

We calculated the free energy barrier for the enzymatically catalyzed acylation of acetylcholine to be 8.5 kcal/mol (see Figure 7), which is in reasonable agreement with the experimental value of 11.8 kcal/mol estimated from the application of simple transition state theory<sup>36</sup> to the acylation rate of  $1.7 \times 10^6 \text{ min}^{-1}$  published by Froede and Wilson.<sup>28</sup> The rate-limiting step for the acylation reaction on the basis of our calculations is the transition from the intermediate to the product. This is in qualitative agreement with QM/MM calculations by Zhou et al.,<sup>33</sup> although quantitatively their activation barrier is larger (13.2 kcal/mol). Additionally, Zhang and co-workers showed that the activation barrier for the acylation reaction can vary by as much as 5–6 kcal/mol for different starting structures depending on the size of the QM system and the level of theory.<sup>37</sup>

The observed difference in free energies can come from many sources. It has been shown experimentally that charges distant from the active site can influence the binding affinity of the enzyme.<sup>38</sup> We expect our electrostatic embedding scheme, even if static, to handle charge effects correctly. However, DFT calculations tend to underestimate transition state barrier heights.<sup>39,40</sup> Additionally, the PBE functional underestimates

**Table 1.** Interatomic Distances (in Å) and Angles (in deg) Describing Important Interactions for the Reactants (ES), First Transition State (TS1), Intermediate (INT), Second Transition State (TS2), and Product (EP)

	ES	rms	TS1	rms	INT	rms	TS2	rms	EP	rms
Distances (Å)										
OG–CC	2.79	0.19	1.85	0.0007	1.56	0.0002	1.40	0.04	1.35	0.02
OG–HG	1.02	0.03	1.44	0.16	1.59	0.09	2.44	0.14	2.97	0.24
HG–NE2	1.70	0.12	1.15	0.09	1.07	0.03	1.17	0.02	1.90	0.16
HG–OE	3.11	0.20	2.67	0.11	2.58	0.16	1.33	0.02	1.00	0.02
CC–OE	1.35	0.03	1.43	0.03	1.50	0.04	2.13	0.30	3.04	0.19
OE(E334)–NE2	2.66	0.07	2.62	0.06	2.60	0.06	2.82	0.09	2.88	0.11
CO–N(Gly121)	2.95	0.10	2.95	0.08	2.92	0.11	3.15	0.17	3.27	0.14
CO–N(Gly122)	2.78	0.10	2.77	0.07	2.78	0.08	2.84	0.10	2.88	0.09
CO–N(Ala204)	3.35	0.15	3.02	0.11	2.85	0.09	2.86	0.11	2.88	0.11
O(E202)–HE2(H447)	4.57	0.23	4.45	0.26	4.90	0.23	4.60	0.21	4.18	0.23
O(E202)–H(Wat)	1.76	0.14	1.82	0.16	1.78	0.11	1.78	0.12	1.80	0.15
Angles (deg)										
NE2–HG–OE	87		113		86		167		169	

**Figure 8.** Active-site geometries for the reactants, TS1, intermediate, TS2, and product. Atoms of the catalytic triad and ACh are shown in ball-and-stick representation, and the carbon atoms are colored as follows: Ser203 (cyan), His447 (magenta), Glu334 (yellow), and ACh (salmon). Selected active-site distances undergoing changes are shown, and the values are in angstroms.

reaction barriers relative to B3LYP.<sup>41</sup> In the calculation of the deacylation reaction, Zhou et al. used two different functionals (B3LYP and PBE0). The PBE0 functional<sup>42</sup> is a hybrid version of PBE. The details of the reaction did not change, but the relative energies differed. The B3LYP transition state and product energies were 2.6 and 4.8 kcal/mol higher in energy respectively, than the corresponding PBE0 energies. Although molecular dynamics was used to sample the phase space, it may not have completely removed the bias associated with the starting geometry. However, this is still better than 0 K simulations, where the optimized geometries strongly depend on the initial conditions and the forces required to maintain the constraints on a given reaction coordinate can differ significantly among the various local minima.

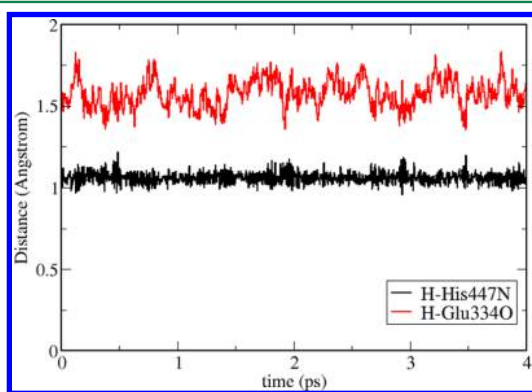
A comparison of our acylation reaction coordinate and active-site geometries with those of Zhou et al.<sup>33</sup> reveals similarities and several important differences. The geometries of the reactant, TS1, and intermediate states are in general agreement. The tetrahedral intermediate state is more stabilized in our calculations compared with those of Zhou et al. This increased stability is due to the more productlike TS2 geometry

(see Table 1). In unconstrained MD simulations, the tetrahedral intermediate readily transitions to the reactants, which also suggests that TS2 is more productlike. This shift in the TS2 geometry towards the product causes the overall energy of the acylation reaction in our calculations to be slightly endothermic. However, the reaction coordinate of Zhou et al. reveals a significantly exothermic acylation reaction (−6.0 kcal/mol). We were unable to find experimental results to either support or refute the final energetics. However, once the choline is displaced from the active site, only the forward reaction is possible. Our TS2 geometry as described by the distances shown in Table 1 is markedly different than that of Zhou. The cause of the difference between the two reports is not clear, although inspection of the chosen reaction coordinates suggests that different restraints may be one possible source.

The nature of the hydrogen-bonding interaction between the Glu334 carboxylate O atom and the His447 imidazole NE atom has been debated in the literature.<sup>43</sup> We observe the hydrogen atom to be tightly bonded to the His447 NE1 atom for the entire MD simulation (4 ps) of the tetrahedral acyl



intermediate (Figure 9). This observation is in contrast to the results published by Zhou et al.,<sup>33</sup> which show frequent sliding



**Figure 9.** Hydrogen-bonding interaction between the Glu334 carboxylate O atom and the His447 NE1 atom for the tetrahedral acyl intermediate. We show the H–N and H–O distances. On the basis of these data, there is no evidence that the hydrogen shifts between the N and O atoms of His447 and Glu334, respectively.

of the hydrogen between the Glu334 carboxylate O atom and the His447 imidazole NE atom during their 30 ps MD simulation. This difference in the results could be due to the functionals used for these calculations. The PBE functional is more accurate for the calculation of hydrogen-bonding strengths,<sup>44</sup> and the B3LYP functional is known to underestimate hydrogen-bonding energies.<sup>41</sup>

## CONCLUDING REMARKS

In this paper, we have presented a new numerical approach for calculating free energy barriers with large QM regions. We carried out simulations on a large QM region (612 atoms) embedded in a static electrostatic field to model the acetylcholine acylation reaction catalyzed by AChE. This application of our first-principles calculations shows general agreement with previous experimental and computational studies. The reaction barrier heights and rate-limiting step of the acylation reaction are correctly predicted within the uncertainties of experiment and theory.

We have compared our numerical results with the results of recent calculations by Zhou et al.<sup>33</sup> There are some differences between our results and theirs, in particular with respect to the height of the energy barrier and the geometry of TS2, which may be explained by the size of our QM region, which is an order of magnitude larger than in previous simulations.<sup>33,37</sup> In our model, even if quite a few atoms are fixed, the larger number of atoms treated at the QM level allows for greater QM interactions with atoms surrounding the catalytic active site, including electron screening and polarization.

While we do not have the capability to do it at this time, we think it would be very interesting to make a direct comparison between our approach and a traditional QM/MM approach with the same QM region and QM model. One significant difference between our model and a traditional QM/MM approach is the fact that our electrostatic embedding potential is static and represents an average effective potential, while MM atoms are continuously moving in a QM/MM approach. It is not clear that these fluctuations matter for transition state barrier calculations, particularly in the light of the conclusions of Schwartz and Schramm<sup>45</sup> that the barrier crossing time is on the order of 100 fs and the lifetime of a transition state is about

5 fs. Thus, on this time scale MM atoms do not move significantly.

It is also noteworthy that the speed at which these reactions happen once the system has reached a favorable configuration<sup>45</sup> justifies the use of short MD simulations to sample only the vibrational modes of the system in that configuration. Classical MD can be used on a larger system and a longer time scale to drive the system into a catalytically reactive configuration. Once we reach a favorable configuration for the reaction to occur, QM should be used to sample a restricted phase space for a few picoseconds in order to more accurately compute the energy barrier of interest at the DFT level.

Although the accuracy of our approach is ultimately limited by the accuracy of DFT in this context, we conclude that our approach is quite robust. This technique will be soon applicable to larger systems with faster computers and advanced scalable  $O(N)$  electronic structure methods.

## AUTHOR INFORMATION

### Corresponding Author

\*E-mail: [fattebert1@llnl.gov](mailto:fattebert1@llnl.gov).

### Present Address

<sup>§</sup>P.H.: Department of Chemistry & Biochemistry, California State University, East Bay, Hayward, CA 94542.

### Funding

We thank the Defense Threat Reduction Agency for funding (CBS.SCIC.01.10.LLNL.004). We also thank the Livermore Computing Grand Challenge for computer time. This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

We thank Eric R. Schwegler for helpful discussions.

## REFERENCES

- (1) Senn, H. M.; Thiel, W. *Angew. Chem., Int. Ed.* **2009**, *48*, 1198–1229.
- (2) Hu, H.; Yang, W. *J. Mol. Struct.: THEOCHEM* **2009**, *898*, 17–30.
- (3) Sumowski, C. V.; Ochsenfeld, C. *J. Phys. Chem. A* **2009**, *113*, 11734–11741.
- (4) Hu, L.; Eliasson, J.; Heimdal, J.; Ryde, U. *J. Phys. Chem. A* **2009**, *113*, 11793–11800.
- (5) Liao, R.-Z.; Thiel, W. *J. Comput. Chem.* **2013**, *34*, 2389–2397.
- (6) Hu, L.; Söderhjelm, P.; Ryde, U. *J. Chem. Theory Comput.* **2013**, *9*, 640–649.
- (7) Fox, S. J.; Pittock, C.; Tautermann, C. S.; Fox, T.; Christ, C.; Malcolm, N. O. J.; Essex, J. W.; Skylaris, C.-K. *J. Phys. Chem. B* **2013**, *117*, 9478–9485.
- (8) Carter, E.; Ciccotti, G.; Hynes, J. T.; Kapral, R. *Chem. Phys. Lett.* **1989**, *156*, 472–477.
- (9) Straatsma, T.; McCammon, J. *J. Chem. Phys.* **1991**, *95*, 1175–1188.
- (10) Kelly, E.; Seth, M.; Ziegler, T. *J. Phys. Chem. A* **2004**, *108*, 2167–2180.
- (11) Fattebert, J.-L.; Law, R.; Bennion, B.; Lau, E.; Schwegler, E.; Lightstone, F. *J. Chem. Theory Comput.* **2009**, *5*, 2257–2264.
- (12) Fattebert, J.-L.; Gygi, F. *Comput. Phys. Commun.* **2004**, *162*, 24–36.
- (13) Fattebert, J.-L.; Gygi, F. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2006**, *73*, 115124.
- (14) Sprick, M.; Ciccotti, G. *J. Chem. Phys.* **1998**, *109*, 7737–7744.

- (15) Dvir, H.; Silman, I.; Harel, M.; Rosenberry, T. L.; Sussman, J. L. *Chem.-Biol. Interact.* **2010**, *187*, 10–22.
- (16) Wilson, I. B.; Harrison, M. A. *J. Biol. Chem.* **1961**, *236*, 2292–2295.
- (17) Kryger, G.; Harel, M.; Giles, K.; Toker, L.; Velan, B.; Lazar, A.; Kronman, C.; Barak, D.; Ariel, N.; Shafferman, A.; Silman, I.; Sussman, J. L. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2000**, *56*, 1385–1394.
- (18) Bennion, B. J.; Lau, E. Y.; Fattebert, J.-L.; Huang, P.; Schwegler, E.; Corning, W.; Lightstone, F. C. *Military Medical Sci. Lett.* **2013**, *82*, 102–114.
- (19) Bennion, B. J.; Essiz, S. G.; Lau, E. Y.; Fattebert, J.-L.; Emigh, A.; Lightstone, F. C. *PLoS One* **2015**, *10*, e0121092.
- (20) Quinn, D. M. *Chem. Rev.* **1987**, *87*, 955–979.
- (21) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781–1802.
- (22) MacKerell, A. D.; et al. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
- (23) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. *J. Comput. Chem.* **1998**, *19*, 1639–1662.
- (24) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (25) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
- (26) Kleinman, L.; Bylander, D. *Phys. Rev. Lett.* **1982**, *48*, 1425–1428.
- (27) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.
- (28) Froede, H. C.; Wilson, I. B. *J. Biol. Chem.* **1984**, *259*, 11010–11013.
- (29) Silman, I.; Millard, C.; Ordentlich, A.; Greenblatt, H.; Harel, M.; Barak, D.; Shafferman, A.; Sussman, J. *Chem.-Biol. Interact.* **1999**, *119–120*, 43–52.
- (30) Ryckaert, J.; Ciccotti, G.; Berendsen, H. J. *Comput. Phys.* **1977**, *23*, 327–341.
- (31) Meijer, E. J.; Sprik, M. *J. Am. Chem. Soc.* **1998**, *120*, 6345–6355.
- (32) Zhang, Y.; Kua, J.; McCammon, J. A. *J. Am. Chem. Soc.* **2002**, *124*, 10572–10577.
- (33) Zhou, Y.; Wang, S.; Zhang, Y. *J. Phys. Chem. B* **2010**, *114*, 8817–8825.
- (34) Millard, C. B.; Koellner, G.; Ordentlich, A.; Shafferman, A.; Silman, I.; Sussman, J. L. *J. Am. Chem. Soc.* **1999**, *121*, 9883–9884.
- (35) Millard, C. B.; Kryger, G.; Ordentlich, A.; Greenblatt, H. M.; Harel, M.; Raves, M. L.; Segall, Y.; Barak, D.; Shafferman, A.; Silman, I.; Sussman, J. L. *Biochemistry* **1999**, *38*, 7032–7039.
- (36) Fuxreiter, M.; Warshel, A. *J. Am. Chem. Soc.* **1998**, *120*, 183–194.
- (37) Zhang, Y.; Kua, J.; McCammon, J. A. *J. Phys. Chem. B* **2003**, *107*, 4459–4463.
- (38) Tougu, V. *Curr. Med. Chem.: Cent. Nerv. Syst. Agents* **2001**, *1*, 155–170.
- (39) Zhao, Y.; Gonzalez-Garcia, N.; Truhlar, D. G. *J. Phys. Chem. A* **2005**, *109*, 2012–2018.
- (40) Cohen, A. J.; Mori-Sanchez, P.; Yang, W. *Chem. Rev.* **2012**, *112*, 289–320.
- (41) Zhao, Y.; Truhlar, D. G. *Acc. Chem. Res.* **2008**, *41*, 157–167.
- (42) Adamo, C.; Barone, V. *J. Chem. Phys.* **1999**, *110*, 6158–6170.
- (43) Bachovchin, W. W. *Magn. Reson. Chem.* **2001**, *39*, S199–S213.
- (44) Rao, L.; Ke, H.; Fu, G.; Xu, X.; Yan, Y. *J. Chem. Theory Comput.* **2009**, *5*, 86–96.
- (45) Schwartz, S. D.; Schramm, V. L. *Nat. Chem. Biol.* **2009**, *5*, 551.