JCTC Journal of Chemical Theory and Computation

Article

Cite This: J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

pubs.acs.org/JCTC

Rapid Convergence of Energy and Free Energy Profiles with Quantum Mechanical Size in Quantum Mechanical−Molecular Mechanical Simulations of Proton Transfer in DNA

Susanta Das,[†] Kwangho Nam,[‡,§] and Dan Thomas Major*[,†]

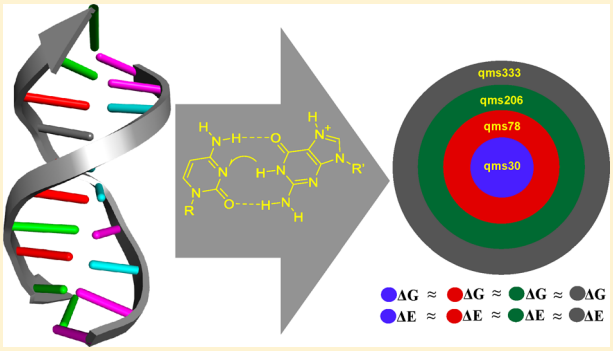
[†]Department of Chemistry, Bar-Ilan University, Ramat-Gan 5290002, Israel

[‡]Department of Chemistry, Umeå University, 901 87 Umeå, Sweden

[§]Department of Chemistry and Biochemistry, University of Texas at Arlington, Arlington, Texas 76019-0065, United States

Supporting Information

ABSTRACT: In recent years, a number of quantum mechanical−molecular mechanical (QM/MM) enzyme studies have investigated the dependence of reaction energetics on the size of the QM region using energy and free energy calculations. In this study, we revisit the question of QM region size dependence in QM/MM simulations within the context of energy and free energy calculations using a proton transfer in a DNA base pair as a test case. In the simulations, the QM region was treated with a dispersion-corrected AM1/d-PhoT Hamiltonian, which was developed to accurately describe phosphoryl and proton transfer reactions, in conjunction with an electrostatic embedding scheme using the particle-mesh Ewald summation method. With this rigorous QM/MM potential, we performed rather extensive QM/MM sampling, and found that the free energy reaction profiles converge rapidly with respect to the QM region size within ca. ±1 kcal/mol. This finding suggests that the strategy of QM/MM simulations with reasonably sized and selected QM regions, which has been employed for over four decades, is a valid approach for modeling complex biomolecular systems. We point to possible causes for the sensitivity of the energy and free energy calculations to the size of the QM region, and potential implications.



INTRODUCTION

In the past four decades, hybrid quantum mechanics−molecular mechanics (QM/MM)[1,2] has become the method of choice for the study of condensed phase biomolecular reactions.[3−9] Using this approach, large systems can be addressed effectively at a subnanoscale level by treating only a carefully selected limited region of interest using electronic structure theory, whereas the remaining part of the system is represented by a computationally inexpensive classical force field.[3−9] This efficiency, in conjunction with the increased accuracy of force fields, allows routine studies of large biomolecular systems with explicit solvent molecules and enhanced sampling techniques to extract reaction and activation free energies.

Despite the conceptual simplicity of the QM/MM approach, its reliability is highly sensitive to a proper computational protocol. In particular, the choice of an accurate QM method and sufficiently large QM region that includes all moieties essential to described chemistry are crucial.[10,11] The latter point of the size of the QM region has received significant recent attention, due to emerging massively parallel supercomputers with PFLOPS capabilities and linear-scaling algorithms.[12,13] Using specialized hardware and appropriate software, one can treat large QM regions including thousands of atoms surrounded by additional MM charges. However, different groups have reached diverging conclusions (see the discussion below),[14−16] and the question of the required minimal size of QM regions is yet to be settled. To resolve this question, additional rigorous and systematic analysis of the size sensitivity of QM/MM calculations is necessary.

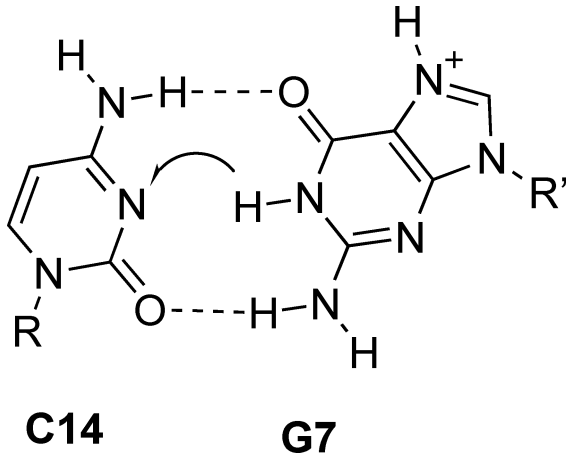
Recently, several studies have contended with the issue of QM system size dependence in cluster[17−19] and QM/MM calculations.[15,17−25] The dependence of electronic and energetic properties in QM cluster calculations on the QM size is readily understood, as the explicit nature of the medium surrounding the cluster is ignored. However, the claimed QM region size dependence in QM/MM calculations is more elusive, as in principle the environmental effects are accounted for via the QM/MM interaction terms. Therefore, if a reasonable QM region is chosen (e.g., accounting for the major electronic structure changes), one might expect that electronic and energetic properties would be rather insensitive to the QM region size. This was indeed found by Sumowski and Ochsenfeld, who investigated the convergence of QM/MM isomerization energies of peptide systems and concluded that the hybrid QM/MM results converge much faster with the size

Received: September 14, 2017
Published: February 15, 2018

DOI: 10.1021/acs.jctc.7b00964
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

of the QM system than the pure cluster approaches.[20] However, these authors also concluded that a rather large QM region is needed to obtain converged energetics within the QM/MM framework. Ryde et al. further drew similar conclusions in a study of proton transfer in [Ni, Fe] hydrogenase.[22] This was also the conclusion of Liao and Thiel in their study of tungsten dependent acetylene hydratase.[23]

Scheme 1. Single Proton Transfer within a Protonated Cytosine−Guanine Base Pair[a]



[a]Proton transfer occurs from guanine (residue 7) to cytosine (residue 14).

The sensitivity of the QM definition in QM/MM free energy simulations of enzymatic reactions was described by Fuxreiter and co-workers in their study of a proton transfer in lysozyme.[21] Although some dependence of the free energy profiles on the QM region size was observed, the effects were rather small (2−3 kcal/mol) considering the highly polar environment in the lysozyme active site. In contrast, a more recent study of *O*-methyltransferase by Kulik et al. suggested that QM regions of up 500−600 atoms were required to obtain converged activation energies.[15] This result was challenged by Jindal and Warshel, who studied the activation free energy for the same system and found that less than 50 atoms are required to reach convergence.[16] Similarly, one of the current authors found that even highly charged active site moieties can be treated adequately as MM atoms in QM/MM free energy simulations of two terpene synthases.[26]

In this work, we investigated the proton transfer reaction within a DNA base pair (Scheme 1), which is a part of the larger double helix embedded in explicit solvent and ions. This reaction has been extensively studied by theory over the years employing a range of methods.[27−39] A recent study by Roßbach and Ochsenfeld addressed the effect of QM embedding schemes, QM method, and size of QM region on the computed proton transfer energetics in a DNA double helix.[24] In particular, these authors performed a systematic study of the energetics of proton transfer with increasing QM regions, and observed that, to get converged results using electrostatic embedding, a total of 1150 QM atoms were required. This result, which was based on MM energy minimizations of stationary points along the reaction coordinate for a single DNA configuration, is somewhat discouraging as it has been established for many biomolecular systems that extensive sampling is required to obtain converged reaction free energy profiles. Indeed, if reaction energy profiles converge slowly with the size of the QM region, it raises the question of why the QM/MM approach has been successful for so long using relatively small QM regions chosen based on an often-intuitive understanding of the underlying chemistry.

To address this question, we investigated the convergence of the free energy profile with QM region size for the DNA proton transfer studied by Roßbach and Ochsenfeld. We employed the semiempirical AM1/d-PhoT method of Nam et al., which was specially developed for phosphorus-containing biological compounds and for accurate proton affinities of biologically relevant molecules.[40] The AM1/d-PhoT Hamiltonian gives good agreement with density functional theory (DFT) results in the gas phase and in solution. This method not only has been applied in QM/MM simulations of transphosphorylation reaction where proton transfer plays a major role,[41,42] but also has been extended to include rigorous long-range electrostatics with the particle-mesh Ewald (PME) summation method,[43−45] allowing a size-consistent electrostatic environment for QM/MM simulations with different QM sizes. Hence, the AM1/d-PhoT Hamiltonian in conjunction with its PME extension provides a computationally economical, yet reliable method to study the effect of QM size on free energy profiles.

COMPUTATIONAL MODELING

System Setup and MM Molecular Dynamics. The initial X-ray structure (1ZEW),[46] which contains 10 base pairs of B-DNA, was used to construct the initial configurations for the present work. The system was soaked in a well-equilibrated box of three point charge TIP3P water of dimensions ca. 50 × 50 × 50 Å^3,[47] and neutralized with 18 randomly placed Na^+ ions. All atoms in the DNA−solvent model were first relaxed in a stepwise fashion using energy minimization with the adopted basis Newton−Raphson (ABNR) method[48] at the MM level of theory. The energy minimized system was then heated systematically from 48 to 298 K for 25 ps and equilibrated at 298 K over the course of 10 ns at the MM level of theory to obtain proper solvation and distribution of ions. These MM molecular dynamics (MD) simulations were performed with the CHARMM software package[48] applying the CHARMM 27 force field.[49,50] Periodic boundary conditions (PBC) were applied throughout the MD simulation studies,[51] and long-range electrostatic interactions were treated using the particle-mesh Ewald (PME) summation technique (50 × 50 × 50 FFT grid, κ = 0.340 Å^{-1}).[43] A 12.0 Å group-based cutoff was applied for real-space electrostatic and van der Waals interactions with a switching function between 10 and 12 Å. A larger cutoff was also tested (vide infra). The simulations of the solvated DNA were performed using the constant particle−pressure−temperature (*NPT*) ensemble at 298 K and 1 atm. The target pressure and temperature were controlled by the extended system pressure[52] and the Hoover thermostat algorithms (P_γ = 25 ps^{-1}, P_{mass} = 500 amu, T_{mass} = 1000 kcal·mol^{-1}·ps^2).[53] The leapfrog Verlet integration scheme[54] was applied to propagate the equations of motion. The SHAKE algorithm[55] was used to constrain all MM bonds involving hydrogen atoms, allowing a time step of 1 fs.

Hybrid AM1/d-PhoT QM/MM Potential Energy Surface. The proton transfer reactions in DNA between guanine 7 (G7) and cytosine 14 (C14) were treated with a hybrid QM/MM potential energy surface:[1]

DOI: 10.1021/acs.jctc.7b00964
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

$$\hat{H} = \hat{H}_{QM} + \hat{H}_{MM} + \hat{H}_{QM/MM} \quad (1)$$

To study the convergence of the QM size, we prepared four systems, viz., qms30, qms78, qms206, and qms333, with increasing number of QM atoms. The QM/MM partitioning scheme is shown in Figure 1. The QM region for all four setups

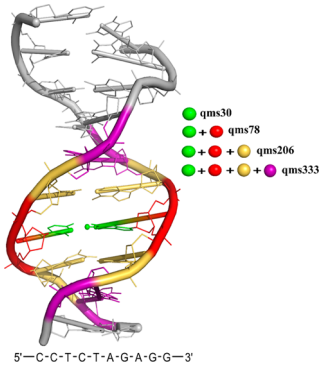


Figure 1. QM partitioning scheme. The QM regions for different QM/MM setups are shown in color. Proton transfer occurs in the green region, and the transferring proton is shown as a sphere. Protons of the DNA double strand are not shown for clarity. Atoms treated at the MM level are shown in gray. Solvent molecules and ions are not shown for clarity.

was treated by the AM1/d-PhoT Hamiltonian. To account for dispersion, we implemented the D3 empirical dispersion correction method of Grimme[56] in the MNDO97 module[57] of CHARMM. This implementation was based on the dispersion corrected tight-binding DFT in CHARMM.[58] The smallest setup, qms30, consisted of only 30 QM atoms of the DNA double strand, i.e., the nucleobases of G7 and C14. For the remaining three systems, the QM region was systematically increased, yielding setups for qms78, qms206, and qms333 consisting of 78, 206, and 333 QM atoms, respectively. The generalized hybrid orbital (GHO) method was employed for the boundary C atoms of DNA.[59] For example, for the qms30 setup, the C1′ atoms of residue 7 (G7) and residue 14 (C14) were treated as GHO atoms. In all QM/MM simulations, the QM/MM−PME method of Nam et. al,[44,45] was employed. The qms30 system was initiated from the preequilibrated MM system (vide supra), equilibrated at the QM/MM level of theory for 300 ps prior to commencing free energy simulations. Subsequent simulations with larger QM regions used the qms30 system as starting point. Additional energy calculations were performed using the link-atom approach as well, to compare more directly with ref 24. In this latter case, the sizes of the QM regions were qms$X$ (X = 28, 64, 192, 319), due to slightly different QM/MM region cuts.

Potential of Mean Force Simulations. The classical mechanical potential of mean force (CM-PMF) associated with the DNA proton transfer reaction was obtained employing adaptive umbrella sampling (US) MD simulations,[60] combined with the weighted histogram analysis method (WHAM).[61,62] The reaction coordinate was defined as the antisymmetric stretch coordinate between the donor nitrogen−hydrogen bond (N1−H) of G7 and acceptor nitrogen−hydrogen bond (N3−H) of C14:

$$\zeta_{asym} = R(\text{N1}-\text{H}) - R(\text{N3}-\text{H}) \quad (2)$$

This type of reaction coordinate has been widely used in the literature for QM/MM free energy simulations and especially for $H^+/H^-/H^\bullet$ transfer reactions.[63] All PMF simulations employed the QM(AM1/d-PhoT)/MM or QM(AM1/d-PhoT-D3)/MM Hamiltonians (Figures 2−4 and Figures S1−S7), in which D3 refers to the Hamiltonian with D3 dispersion corrections.

In all four setups, a series of 11 evenly spaced, discrete regions ("windows") separated by 0.25 Å were used to span the antisymmetric stretch reaction coordinate. Each window was subject to an appropriate harmonic restraint (5−10 kcal/mol·Å^2), which keeps ζ_{asym} in the desired region. In addition, the antisymmetric coordinate was subject to an umbrella potential, which is approximately the negative of the PMF obtained in the preceding series of simulations. The 11 US windows for qms30 were sampled in 5−10 series each, wherein each series was equilibrated for 2 ps, followed by 50−300 ps of production simulations, resulting in a total of 0.6 ns production MD for each window. Each US simulation started from the previous series in the same window, and the 2 ps equilibration data were not included in the PMF statistics. Similarly, the aggregated simulation times for the other three setups (i.e., qms78, qms206, and qms333) were 0.6 ns, 0.2 ns, and 25 ps, respectively, per US window. Hence, the total simulation time for each of PMF was 6.6, 6.6, 2.2, and 0.3 ns for systems qms30, qms78, qms206, and qms333, respectively. During the production simulations, we collected the probability densities of configurations along the reaction coordinate, sorted into bins of width 0.01 Å. The unbiased probability distribution, $\rho(\zeta)$, for each system was then computed by combining the different windows using WHAM to get the combined classical mechanical potential of mean force, W^{cm}:

$$W^{cm}(\zeta) = -\beta^{-1} \ln \rho(\zeta) + C \quad (3)$$

Sampling was continued until the PMF was invariant to within approximately ±1.0 kcal mol^{-1}, which is the expected statistical error in such simulations (see analysis in Figures S1−S4). A detailed error analysis was performed using the bootstrapping approach (Figure S5).[64] We tested both 12.0 and 16.0 Å group-based cutoffs for real-space electrostatic and van der Waals interactions, and the results were nearly identical (Figure S6).

Energy Calculations. We calculated the single-point energy of 10 different configurations in the framework of QM/MM density functional theory (DFT) using the 6-31+G(d)[65] basis set and the M06-2X functional, i.e., QM-(M06-2X)/MM,[66] as well as single-point QM(AM1/d-PhoT)/MM calculations. The 10 configurations were chosen from 500 ps qms30 QM(AM1/d-PhoT)/MM MD simulations of the reactant state. In these calculations, the systems were first geometry optimized using the smallest QM region (qms30) setup with QM(AM1/d-PhoT)/MM, followed by the above-mentioned single-point calculations for all QM region sizes (results presented in Figures 5 and 6 and Figures S8−S11). In these geometry optimizations, the MM region was fixed, while there were no restraints imposed on QM region atoms other than the reaction coordinate restraint (using the RESD module of CHARMM). RESD was employed to generate reactant, transition, and product states from configurations from the MM MD simulations. The positions of these stationary points along the reaction coordinate were taken as the stationary points obtained from the qms30 PMF simulation (Figure S7). We also performed single-point calculations for the reactant and

DOI: 10.1021/acs.jctc.7b00964
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

product states using configurations from the MM MD simulations that were geometry optimized only at the MM level, followed by single-point energy calculations (without reoptimization at the used QM/MM level), as explained above (results presented in Figure 7).

In cases where geometry optimizations are implicated, the ABNR method implemented in CHARMM was employed, with a convergence criterion of 0.005 kcal/mol·Å.

All simulations described in this work used the CHARMM software, and the semiempirical QM calculations employed the MNDO97 program interfaced with CHARMM,[48,57,67] while DFT calculations used Q-Chem interfaced with CHARMM.[68]

RESULTS

In the present study, the convergence of QM(AM1/d-PhoT-D3)/MM free energy profiles with the size of QM region was investigated using the proton transfer reaction within a protonated C−G base pair (Scheme 1) as a model system. The QM partitioning protocol is shown in Figure 1, in which each different QM system is denoted as qms*X*, *X* being the number of QM atoms. The corresponding PMF profiles for the DNA proton transfer at 25 °C are shown in Figure 2 and

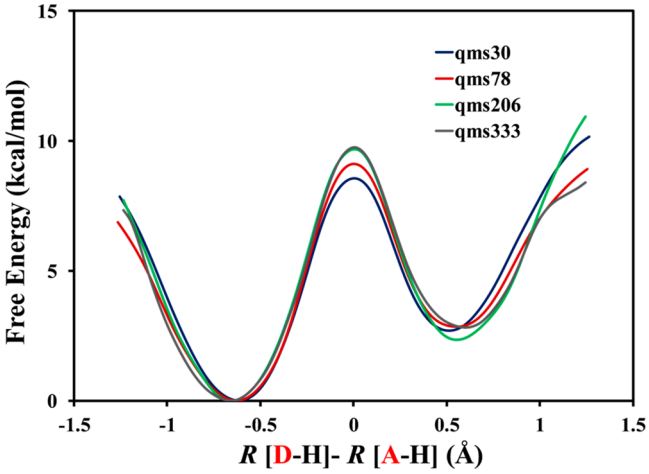


Figure 2. Free energy profiles for the proton transfer in a C−G base pair in DNA. The simulations used the QM(AM1/d-PhoT-D3)/MM Hamiltonian with GHO boundary atoms.

summarized in Table 1. In qms30, we considered only the nucleobases of residues 7 (G7) and 14 (C14) as the QM region, resulting in a mere 30 QM atoms, with a QM charge of +1. The activation free energy ($\Delta G^{\ddagger}$) for the proton transfer was 8.54 ± 0.03 kcal/mol, and the reaction free energy (ΔG_r) was 2.70 ± 0.04 kcal/mol. The second QM partitioning scheme (qms78) contained 78 QM atoms, consisting of the entire G7

Table 1. Different Setups Based on the Number of QM Atoms and Corresponding Free Energies of Activation and Reaction for a Proton Transfer in a C−G Base Pair in DNA[a]

setup	QM atoms	QM charge	$\Delta G^{\ddagger}$ (kcal/mol)	ΔG_r (kcal/mol)
qms30	30	+1	8.54 ± 0.03	2.70 ± 0.04
qms78	78	−3	9.11 ± 0.01	2.86 ± 0.01
qms206	206	−7	9.69 ± 0.05	2.35 ± 0.09
qms333	333	−11	9.8 ± 0.1	2.8 ± 0.2

[a]The simulations used the QM(AM1/d-PhoT-D3)/MM Hamiltonian with GHO boundary atoms.

and C14 and several atoms of residues 6, 8, 13, and 15, which neighbor the C−G base pair in the DNA sequence. This setup included the phosphate groups of G7 and C14 and the total QM charge was −3. The $\Delta G^{\ddagger}$ and ΔG_r values for this system were 9.11 ± 0.01 kcal/mol and 2.86 ± 0.01 kcal/mol, respectively. When comparing qms30 and qms78, the difference between the two systems is 0.57 kcal/mol for $\Delta G^{\ddagger}$ and 0.16 kcal/mol for ΔG_r, both being within the statistical sampling errors (see Figures S1−S5).

Next, we added four more residues (A6, A8, T13, and T15) to include additional base pairs above and below the qms78 setup, resulting in a total of 206 atoms in the QM region, and the total charge of the QM region increased to −7. The free energy barrier for this system was 9.69 ± 0.05 kcal/mol, which is slightly higher than for the smaller setups. The ΔG_r for qms206 was 2.35 ± 0.09 kcal/mol, which is marginally different from those for qms30 and qms78. In qms333, we further increased the QM region by 127 atoms including the A−T (T5 and A16) and C−G (G9 and C12) base pairs of DNA. The C3′ atoms of C4 and C11 and C5′ atoms of G10 and G17 were defined as GHO atoms, and the total charge of the QM region was −11. The free energy barrier for qms333 was 9.8 ± 0.1 kcal/mol, which is slightly higher than the value for qms206, whereas ΔG_r for qms333 was found to be 2.8 ± 0.2 kcal/mol. Overall, we consider the PMF profiles and the QM-size dependence of the reaction barrier and free energy to be well converged (<1 kcal/mol).

To probe the effect of QM dispersion on the computed results, the QM(AM1/d-PhoT)/MM free energy simulations were performed using a protocol similar to the one above, but without the dispersion corrections within the QM region. The $\Delta G^{\ddagger}$ values for the proton transfer reaction were 9.3 ± 0.2, 9.3 ± 0.5, 10.8 ± 0.1, and 11.6 ± 0.2 kcal/mol for the setups qms30, qms78, qms206, and qms333, respectively (Table S1, Figure S7). The ΔG_r values for the proton transfer reaction for setups qms30, qms78, qms206, and qms333 were 4.6 ± 0.2, 4.0 ± 0.5, 4.6 ± 0.1, and 6.0 ± 0.2 kcal/mol, respectively. In comparison to the nice convergence with dispersion corrections (Figure 1), the slight divergence for the larger systems can be ascribed to the lack of dispersion between the base pairs, which caused some perturbation of the structure. Likely the treatment of dispersion interactions is better at the MM level than at the QM level in this case, and hence increasing the QM region did yield slightly divergent results in comparison with the small QM regions without dispersion.

Next, we attempted to understand how choosing an asymmetric QM region influences the free energy profiles. To this end, we chose two scenarios, with QM regions consisting of (a) the nucleobase of C14 and the entire nucleotide of G7 plus the A8 phosphate group (qms54a), and (b) the nucleobase of G7 and the entire nucleotide of C14 plus the T15 phosphate (qms54b). In both cases the total number of QM atoms was 54, and the charge of the QM region was −1, and we employed QM(AM1/d-PhoT-D3)/MM. Seemingly, using such asymmetric QM regions, asymmetry is introduced into the PMF profiles (Figure 3). The barriers obtained are 10.4 and 7.8 kcal/mol for cases a and b, respectively. The reaction free energy was found to be 5.8 and 1.0 kcal/mol, respectively, for cases a and b. The change in the reaction free energy can be attributed to insufficient stabilization of the protonated base in the reactant state for the qms54b case and in the product state for qms54a, and the change in the barrier height follows Hammond's

DOI: 10.1021/acs.jctc.7b00964
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

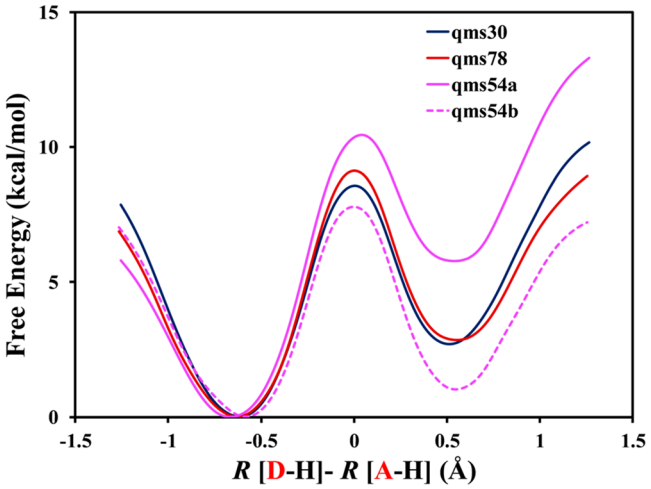


Figure 3. Free energy profiles for the proton transfer in a C−G base pair in DNA. Solid pink line indicates only nucleobase of cytosine and full guanine monomer (i.e., qms54a). Dashed pink line indicates full cytosine monomer and only nucleobase of guanine (i.e., qms54b). The simulations used the QM(AM1/d-PhoT-D3)/MM Hamiltonian with GHO boundary atoms.

postulate. Overall, these results show the importance of including chemical intuition when selecting the QM region.

Until now, we only included DNA in the QM region and treated all waters at the MM level of theory. To examine how the water molecules near the reaction center influence the proton transfer energetics, 20 water molecules were included in the QM region based on the qms78 system using the FIRES protocol in CHARMM.[69,70] The results in Figure 4 suggest a

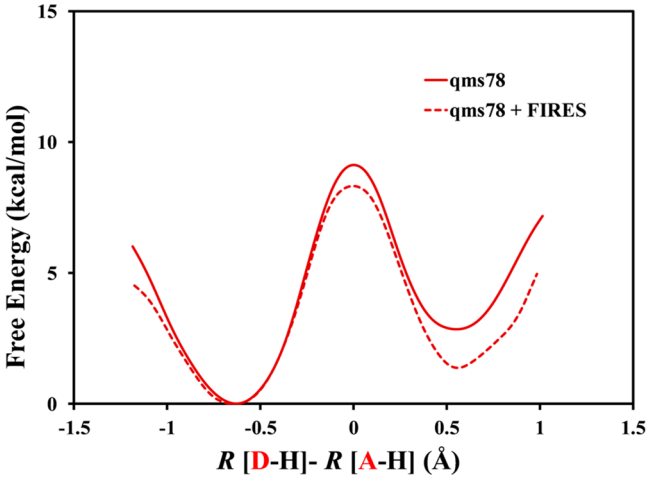


Figure 4. Potential of mean force profiles of qms78 and qms78 with 20 nearby water molecules in the QM region (FIRES approach) for the proton transfer in a C−G base pair in a DNA double strand. The simulations used the QM(AM1/d-PhoT-D3)/MM Hamiltonian.

rather small effect of adding a significant number of water molecules. The barriers were 8.3 kcal/mol using FIRES and 9.1 kcal/mol without QM water molecules. The reaction free energy was 1.4 kcal/mol using FIRES, while it was 2.9 kcal/mol with qms78. We note that, while using the FIRES protocol, it was necessary to apply mild harmonic restraints (1 kcal/mol·Å^2) on the backbone of the DNA chains, as the anisotropic distribution of water molecules around the G7−C14 base pair, in conjunction with the FIRES forces, resulted in structural distortions.

To further understand the QM size dependence and convergence behavior, we investigated energetics (activation barrier ($\Delta E^{\ddagger}$) and reaction energy (ΔE_r)) for the DNA proton transfer considering different configurations of the system. Based on the reactant state of the qms30 simulations, we selected 10 different geometries.[71] Each structure was fully optimized at the same QM(AM1/d-PhoT)/MM level (no dispersion) used in the qms30 MD simulations. Then, applying the general distance restraints (RESD)[72] functionality in CHARMM to the transferring proton, we optimized the transition state (TS) and the product state (PS), allowing flexibility only for the QM region, and calculated the $\Delta E^{\ddagger}$ and ΔE_r for different QM sizes (i.e., qms30, qms78, qms206, and qms333). The results are shown in Figure 5. For every configuration in these calculations, $\Delta E^{\ddagger}$ and ΔE_r converged rapidly, supporting the small effect of the QM region size. For better understanding of these data, a statistical analysis was also performed. Figure S8 presents the minimum, maximum, and average values, as well as standard deviations of $\Delta E^{\ddagger}$ and ΔE_r for all four QM setups (qms30, qms78, qms206, and qms333). Indeed, the system energetics are well converged even with 30 atoms in the QM region (typically <2 kcal/mol).

Then, to compare whether the same trend is observed using DFT-based QM/MM, we repeated the same minimization protocol for the small QM region using H-link atoms instead of GHO atoms (qms28) and performed single-point energy calculations at each determined qms28 minimum for all qms$X$ systems (X = 64, 192, 319). In this case, we employed the AM1/d-PhoT Hamiltonian, as well as the DFT M06-2X/6-31+G* method for the QM region (Figure 6 and Figure S9), while the Ewald summation method was not used. Also in these cases, the reaction energetics are relatively well converged with the tested QM region size.

Finally, we performed QM/MM single-point calculations to get the reaction energy on the structures obtained from the initial pure MM simulations that were minimized at the MM level of theory (Figure 7). Interestingly, with this strategy, which was adopted in ref 24, we observed a greater size dependence on the QM region, as seen by Roßbach and Ochsenfeld. Indeed, with an increase of the size of the QM system, we observed convergence toward a given value for each of the 10 configurations investigated. This behavior is seen for both DFT and semiempirical QM methods employed here (AM1/d-PhoT and M06-2X). The most extreme behavior is observed for the small qms30 system, which is, however, considerably smaller than the smallest system employed by Roßbach and Ochsenfeld. These results therefore serve as a cautious warning against protocols that use MM structures for QM/MM single-point calculations, especially when one is attempting to obtain quantitative energetic information.

DISCUSSION

Various studies have suggested that results of QM/MM calculations depend on the size of the QM region.[15,24] In the current study of a proton transfer in DNA, we found that the free energy profiles are rather insensitive to the size of the QM region, when the QM region is selected based on reasonable chemical principles underlying the reaction studied. Indeed, employing the well-balanced QM(AM1/d-PhoT-D3)/MM model and chemically intuitive QM regions, we obtained very well converged (differences <1 kcal/mol) free energy profiles

DOI: 10.1021/acs.jctc.7b00964
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

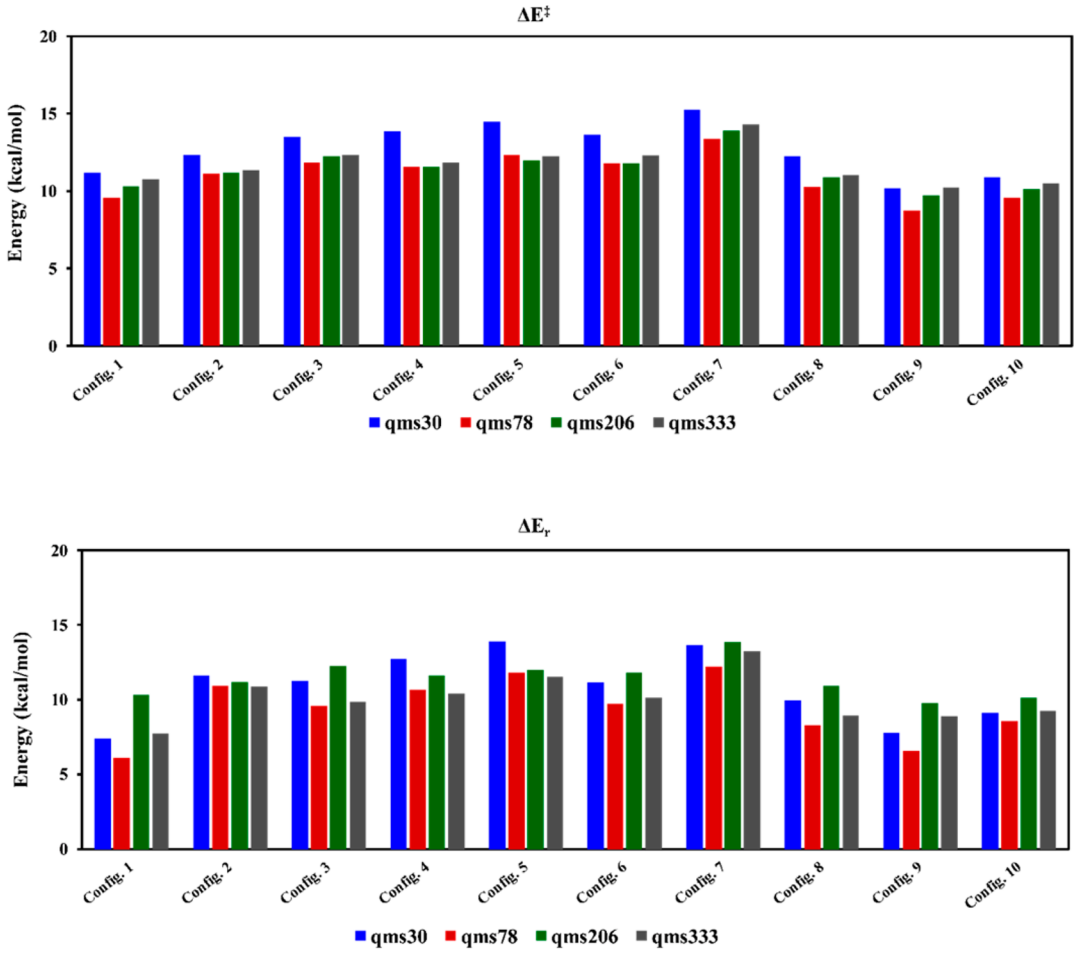


Figure 5. Activation energies ($\Delta E^{\ddagger}$) and reaction energies (ΔE_r) for the proton transfer in a C−G base pair in DNA for 10 configurations. The calculations used the QM(AM1/d-PhoT)/MM Hamiltonian with GHO boundary atoms, and the structures were minimized using qms30.

with QM regions ranging from 30 to 333 atoms (Figure 2). Further, we found that when comparing reaction energetics for individual configurations obtained with different QM region sizes, the results were rather similar. On the other hand, the reaction energetics varied significantly between different configurations (Figures 5 and 6 and Figure S9); the difference between configurations was often larger than the difference between different QM regions in a single configuration (Figure S8). We also conclude that the sensitivity of the reaction energetics to the size of QM region is not due to the use of QM/MM boundary atom methods, not due to the long-range electrostatic treatment, and not due to the electronic structure method employed.

On the other hand, we observed greater variability in the determined PMF profiles using the QM(AM1/d-PhoT-D3)/MM method with less intuitive QM region selections, such as asymmetric portions of each DNA strand. Similarly, when dispersion correction was removed from the QM region, we likely introduced an imbalance in the dispersion interaction energies in the QM and MM regions, and the resulting free energy profiles were slightly divergent. Moreover, we note that QM region dependence could occur when performing single-point-energy calculations on geometries that were optimized at different levels of theory, which potentially leads to unphysically high energy geometries (such as different bond lengths and intermolecular distances).[24]

In another related work of Kulik et al., different QM regions with an increasing number of QM atoms were investigated based on the reaction catalyzed by *O*-methyltransferase.[15] In the reaction studied by Kulik et al., careful treatment of Mg^{2+} ion coordination in the active site is critical to obtaining reliable energetics. Indeed, in all QM/MM calculations, the hexacoordinated Mg^{2+} ion was included. However, in the small QM regions, the Mg^{2+} coordination sphere was not properly saturated with QM ligands. One might speculate that it might make chemical sense to first saturate the Mg^{2+} with QM ligands, and this was indeed observed by Jindal and Warshel.[16] Based on the present study, we expect that such a strategy of applying chemical intuition to QM region selection might show better energetics convergence with QM region size.

One could imagine that lack of charge polarization in standard force fields and ignoring charge transfer could lead to discrepancies when using different sized QM regions. However, numerous simulation studies have shown that MM force fields successfully reproduce many experimental observables, including ligand binding affinities,[73−75] protein−protein interactions,[76,77] protein conformational change,[77] and solvation free energies.[74,78] Recent studies have also shown that the effect of using polarizable force fields in QM/MM calculations are small.[79−81] These studies suggest that intermolecular interactions are well captured by the MM potential energy functions. Therefore, if a proper, chemically meaningful QM region is chosen with well-calibrated intermolecular QM−MM

DOI: 10.1021/acs.jctc.7b00964
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

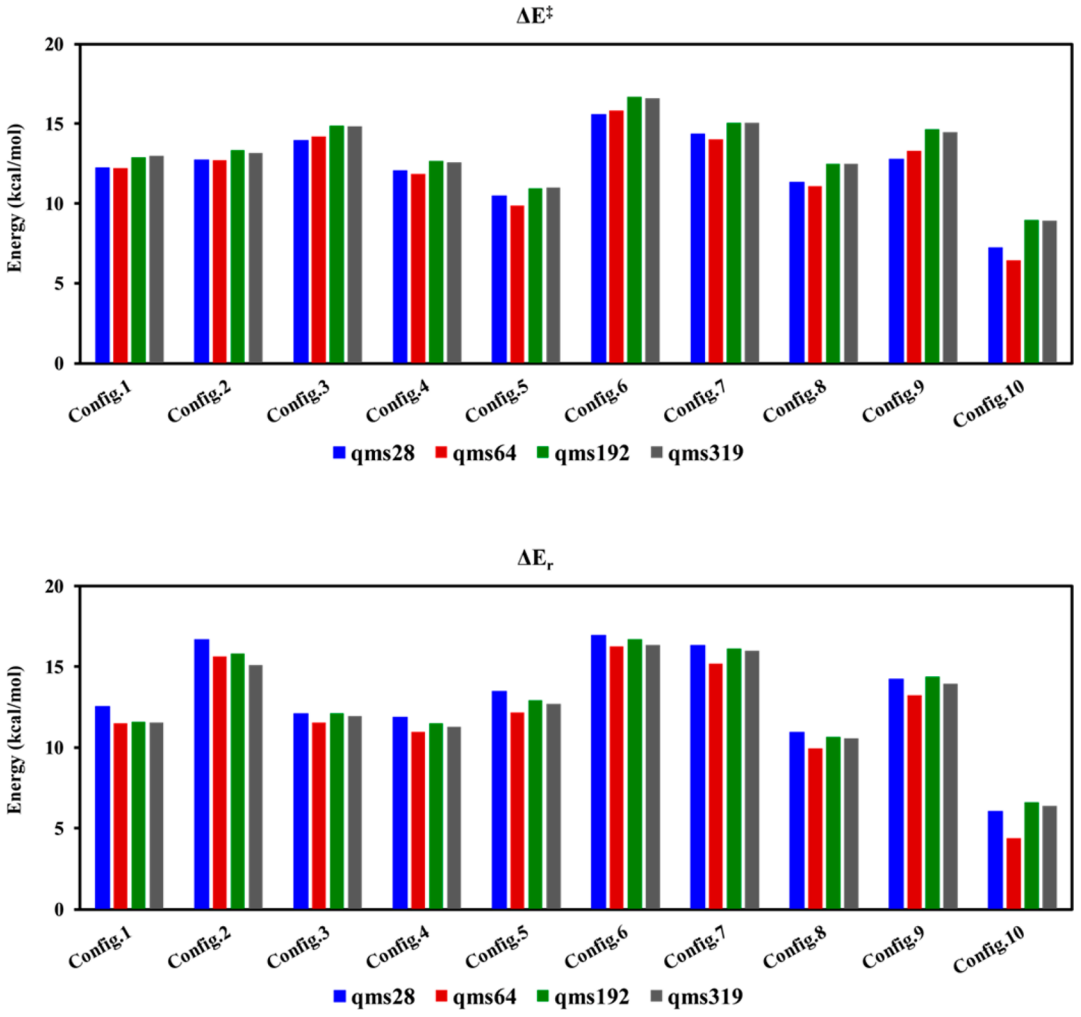


Figure 6. Activation energies ($\Delta E^{\ddagger}$) and reaction energies (ΔE_r) for the proton transfer in a C−G base pair in DNA for 10 configurations. The calculations used the QM(M06-2X)/MM Hamiltonian with link atoms. The structures were minimized using QM(AM1/d-PhoT)/MM with link atoms using qms30.

interactions, one might expect to observe a relatively small QM size dependence on the determined biocatalytic barriers and reaction free energies. Indeed, this was realized early on by Gao, who studied the effects of fine-tuning QM/MM interactions,[82,83] and this approach has been adopted extensively since.[84−89]

Based on the earlier work of Jindal and Warshel[16] and the current results, it seems that one needs to carefully check that perceived differences due to the size of the QM region indeed necessitates increasing the extent of the QM region. Here we show that QM region size effects could be due to imbalanced QM−MM treatments, counterintuitive QM region selections, multiple conformations, and/or particular solvent configurations.[8,20,24,90,91] Additionally, the level of theory employed for the QM region can often have a greater influence than the size of the QM region, and therefore it might be preferable to use a smaller QM region with a higher level of theory.[92] Hence, before increasing the size of the QM region, with the accompanying computational overhead, it seems prudent to check other possible cheaper routes for accurate QM/MM simulations. Moreover, based on experience,[9,63,93] including the present study, we recommend performing extensive configurational sampling rather than relying on a few selected biomolecular configurations.

CONCLUSIONS

In the current work, we studied the effect of the size of the QM region in QM/MM free energy simulations of a proton transfer between two nucleobases in DNA. Four different QM regions were considered, with the number of QM atoms ranging between 30 and 333. Different QM/MM methods and protocols were employed to test different hypotheses that potentially underlie the QM region size dependencies reported. The results showed that the changes in the activation and reaction free energies between the tested systems were relatively small (ca. ±1 kcal/mol), and often within the statistical errors of the simulations. Reaction and activation energies also showed a similar convergence behavior, regardless of the type of QM region boundary employed or the electronic structure method adopted. However, when chemically counterintuitive QM regions, dispersion-free QM methods, or MM-based structures were used, we did observe sensitivity to the size of the QM region. These findings imply that choosing an optimal QM/MM theoretical method, chemically intuitive QM region selection, and a proper simulation protocol might be more important than QM region size when studying biomolecular reactions. Based on the current results, previous results by Jindal and Warshel,[16] and also one of the current authors,[26] we, therefore, conclude that both QM/MM free

DOI: 10.1021/acs.jctc.7b00964
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

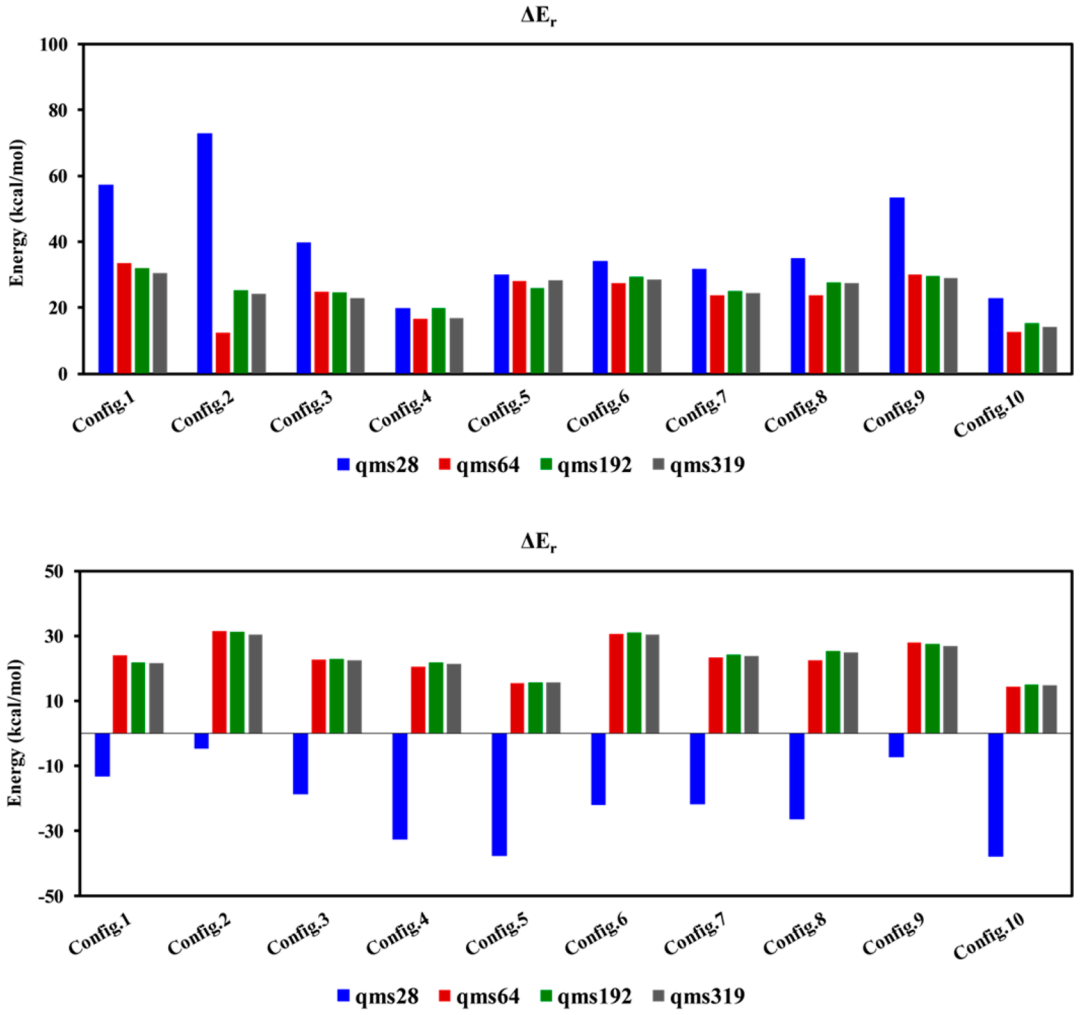


Figure 7. Reaction energies (ΔE_r) for the proton transfer in a C−G base pair in DNA for 10 configurations. The calculations used the QM(AM1/d-PhoT)/MM (top) QM(M06-2X)/MM (bottom) Hamiltonians with link atoms. The structures were minimized at the MM level of theory.

energy simulations and reaction energetics converge rather rapidly with QM region size when using proper modeling protocols.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.7b00964.

Convergence of free energy profiles, dependence on cutoff, statistical analysis of energy barriers and reaction energies, pictures of selected configurations, (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: majort@biu.ac.il.

ORCID

Kwangho Nam: 0000-0003-0723-7839

Dan Thomas Major: 0000-0002-9231-0676

Funding

This work has been supported by the Israel Science Foundation (Grant 2146/15 to D.T.M.), the Swedish Research Council (VR 2015-04114 to K.N.), and the University of Texas at Arlington (K.N.).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Prof. Q. Cui for providing us with the code for the FIRES implementation in CHARMM.

■ REFERENCES

(1) Warshel, A.; Levitt, M. Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.* **1976**, *103*, 227−49.

(2) Warshel, A. Multiscale modeling of biological functions: from enzymes to molecular machines (Nobel Lecture). *Angew. Chem., Int. Ed.* **2014**, *53*, 10020−10051.

(3) Warshel, A. *Computer Modeling of Chemical Reactions in Enzymes and Solutions*; John Wiley & Sons: New York, 1991.

(4) Gao, J. Methods and Applications of Combined Quantum Mechanical and Molecular Mechanical Potentials. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH Publishers: New York, 1996; Vol. 7, pp 119−185. DOI: 10.1002/9780470125847.ch3.

(5) Friesner, R. A.; Guallar, V. Ab Initio Quantum Chemical and Mixed Quantum Mechanics/Molecular Mechanics (QM/MM) Methods for Studying Enzymatic Catalysis. *Annu. Rev. Phys. Chem.* **2005**, *56*, 389−427.

DOI: 10.1021/acs.jctc.7b00964
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

(6) Hu, H.; Yang, W. Free energies of chemical reactions in solution and in enzymes with ab initio quantum mechanics/molecular mechanics methods. *Annu. Rev. Phys. Chem.* **2008**, *59*, 573−601.
(7) Senn, H. M.; Thiel, W. QM/MM methods for biomolecular systems. *Angew. Chem., Int. Ed.* **2009**, *48*, 1198−1229.
(8) van der Kamp, M. W.; Mulholland, A. J. Combined quantum mechanics/molecular mechanics (QM/MM) methods in computational enzymology. *Biochemistry* **2013**, *52*, 2708−2728.
(9) Dixit, M.; Das, S.; Mhashal, A. R.; Eitan, R.; Major, D. T. Practical Aspects of Multiscale Classical and Quantum Simulations of Enzyme Reactions. *Methods Enzymol.* **2016**, *577*, 251−286.
(10) Ochsenfeld, C.; Kussmann, J.; Koziol, F. Ab initio NMR spectra for molecular systems with a thousand and more atoms: a linear-scaling method. *Angew. Chem., Int. Ed.* **2004**, *43*, 4485−4494.
(11) Hartman, J. D.; Neubauer, T. J.; Caulkins, B. G.; Mueller, L. J.; Beran, G. J. Converging nuclear magnetic shielding calculations with respect to basis and system size in protein systems. *J. Biomol. NMR* **2015**, *62*, 327−340.
(12) Goedecker, S.; Scuseria, G. E. Linear scaling electronic structure methods in chemistry and physics. *Comput. Sci. Eng.* **2003**, *5*, 14−21.
(13) Fattebert, J. L.; Gygi, F. Linear scaling first-principles molecular dynamics with controlled accuracy. *Comput. Phys. Commun.* **2004**, *162*, 24−36.
(14) Zhang, J.; Kulik, H. J.; Martinez, T. J.; Klinman, J. P. Mediation of donor-acceptor distance in an enzymatic methyl transfer reaction. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 7954−7963.
(15) Kulik, H. J.; Zhang, J.; Klinman, J. P.; Martinez, T. J. How Large Should the QM Region Be in QM/MM Calculations? The Case of Catechol O-Methyltransferase. *J. Phys. Chem. B* **2016**, *120*, 11381−11394.
(16) Jindal, G.; Warshel, A. Exploring the Dependence of QM/MM Calculations of Enzyme Catalysis on the Size of the QM Region. *J. Phys. Chem. B* **2016**, *120*, 9913−9921.
(17) Velichkova, P.; Himo, F. Methyl transfer in glycine N-methyltransferase. A theoretical study. *J. Phys. Chem. B* **2005**, *109*, 8216−8225.
(18) Hopmann, K. H.; Himo, F. Quantum Chemical Modeling of the Dehalogenation Reaction of Haloalcohol Dehalogenase. *J. Chem. Theory Comput.* **2008**, *4*, 1129−1137.
(19) Liao, R. Z.; Yu, J. G.; Himo, F. Quantum Chemical Modeling of Enzymatic Reactions: The Case of Decarboxylation. *J. Chem. Theory Comput.* **2011**, *7*, 1494−1501.
(20) Sumowski, C. V.; Ochsenfeld, C. A convergence study of QM/MM isomerization energies with the selected size of the QM region for peptidic systems. *J. Phys. Chem. A* **2009**, *113*, 11734−11741.
(21) Solt, I.; Kulhanek, P.; Simon, I.; Winfield, S.; Payne, M. C.; Csanyi, G.; Fuxreiter, M. Evaluating boundary dependent errors in QM/MM simulations. *J. Phys. Chem. B* **2009**, *113*, 5728−5735.
(22) Hu, L.; Soderhjelm, P.; Ryde, U. On the Convergence of QM/MM Energies. *J. Chem. Theory Comput.* **2011**, *7*, 761−777.
(23) Liao, R. Z.; Thiel, W. Comparison of QM-Only and QM/MM Models for the Mechanism of Tungsten-Dependent Acetylene Hydratase. *J. Chem. Theory Comput.* **2012**, *8*, 3793−3803.
(24) Roßbach, S.; Ochsenfeld, C. Influence of Coupling and Embedding Schemes on QM Size Convergence in QM/MM Approaches for the Example of a Proton Transfer in DNA. *J. Chem. Theory Comput.* **2017**, *13*, 1102−1107.
(25) Ryde, U. How many conformations need to be sampled to obtain converged QM/MM energies? The curse of exponential averaging. *J. Chem. Theory Comput.* **2017**, *13*, 5745−5752.
(26) Major, D. T. Electrostatic Control of Chemistry in Terpene Cyclases. *ACS Catal.* **2017**, *7*, 5461−5465.
(27) Clementi, E.; Mehl, J.; von Niessen, W. Study of the Electronic Structure of Molecules. XII. Hydrogen Bridges in the Guanine−Cytosine Pair and in the Dimeric Form of Formic Acid. *J. Chem. Phys.* **1971**, *54*, 508−520.
(28) Scheiner, S.; Kern, C. W. Theoretical-Study of Proton Transfers between Base-Pairs of DNA. *Chem. Phys. Lett.* **1978**, *57*, 331−333.
(29) Lipiński, J.; Gorzkowska, E. Double proton transfer and charge-transfer transitions in biological hydrogen-bonded systems guanine-cytosine (G-C) and G-C-Mg2+ systems. *Chem. Phys. Lett.* **1983**, *94*, 479−482.
(30) Florian, J.; Leszczynski, J. Spontaneous DNA mutations induced by proton transfer in the guanine cytosine base pairs: An energetic perspective. *J. Am. Chem. Soc.* **1996**, *118*, 3010−3017.
(31) Colominas, C.; Luque, F. J.; Orozco, M. Tautomerism and protonation of guanine and cytosine. Implications in the formation of hydrogen-bonded complexes. *J. Am. Chem. Soc.* **1996**, *118*, 6811−6821.
(32) Zhanpeisov, N. U.; Leszczynski, J. Specific Solvation Effects on the Structures and Properties of Neutral and One-Electron Oxidized Formamidine−Formamide Complexes. A Theoretical ab Initio Study. *J. Phys. Chem. A* **1999**, *103*, 8317−8327.
(33) Chandra, A. K.; Nguyen, M. T.; Uchimaru, T.; Zeegers-Huyskens, T. Protonation and deprotonation enthalpies of guanine and adenine and implications for the structure and energy of their complexes with water: Comparison with uracil, thymine, and cytosine. *J. Phys. Chem. A* **1999**, *103*, 8853−8860.
(34) Zoete, V.; Meuwly, M. Double proton transfer in the isolated and DNA-embedded guanine-cytosine base pair. *J. Chem. Phys.* **2004**, *121*, 4377−4388.
(35) Chen, H. Y.; Kao, C. L.; Hsu, S. C. Proton transfer in guanine-cytosine radical anion embedded in B-form DNA. *J. Am. Chem. Soc.* **2009**, *131*, 15930−15938.
(36) Matsui, T.; Sato, T.; Shigeta, Y.; Hirao, K. Sequence-dependent proton-transfer reaction in stacked GC pair II: The origin of stabilities of proton-transfer products. *Chem. Phys. Lett.* **2009**, *478*, 238−242.
(37) Villani, G. Theoretical investigation of the coupling between hydrogen atoms transfer and stacking interaction in guanine-cytosine dimers. *Phys. Chem. Chem. Phys.* **2013**, *15*, 19242−19252.
(38) Villani, G. Theoretical investigation of hydrogen transfer mechanism in the guanine-cytosine base pair. *Chem. Phys.* **2006**, *324*, 438−446.
(39) Xiao, S.; Wang, L.; Liu, Y.; Lin, X.; Liang, H. Theoretical investigation of the proton transfer mechanism in guanine-cytosine and adenine-thymine base pairs. *J. Chem. Phys.* **2012**, *137*, 195101−195110.
(40) Nam, K.; Cui, Q.; Gao, J.; York, D. M. Specific Reaction Parametrization of the AM1/d Hamiltonian for Phosphoryl Transfer Reactions: H, O, and P Atoms. *J. Chem. Theory Comput.* **2007**, *3*, 486−504.
(41) Nam, K.; Gao, J.; York, D. M. Quantum mechanical/molecular mechanical simulation study of the mechanism of hairpin ribozyme catalysis. *J. Am. Chem. Soc.* **2008**, *130*, 4680−91.
(42) Ojeda-May, P.; Li, Y.; Ovchinnikov, V.; Nam, K. Role of Protein Dynamics in Allosteric Control of the Catalytic Phosphoryl Transfer of Insulin Receptor Kinase. *J. Am. Chem. Soc.* **2015**, *137*, 12454−7.
(43) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089−10092.
(44) Nam, K.; Gao, J.; York, D. M. An Efficient Linear-Scaling Ewald Method for Long-Range Electrostatic Interactions in Combined QM/MM Calculations. *J. Chem. Theory Comput.* **2005**, *1*, 2−13.
(45) Nam, K. Acceleration of Ab Initio QM/MM Calculations under Periodic Boundary Conditions by Multiscale and Multiple Time Step Approaches. *J. Chem. Theory Comput.* **2014**, *10*, 4175−83.
(46) Hays, F. A.; Teegarden, A.; Jones, Z. J.; Harms, M.; Raup, D.; Watson, J.; Cavaliere, E.; Ho, P. S. How sequence defines structure: a crystallographic map of DNA structure and conformation. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 7157−7162.
(47) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926−935.
(48) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM - a Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comput. Chem.* **1983**, *4*, 187−217.

DOI: 10.1021/acs.jctc.7b00964
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

(49) Foloppe, N.; MacKerell, A. D., Jr. All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. *J. Comput. Chem.* **2000**, *21*, 86−104.
(50) MacKerell, A. D.; Banavali, N. K. All-atom empirical force field for nucleic acids: II. Application to molecular dynamics simulations of DNA and RNA in solution. *J. Comput. Chem.* **2000**, *21*, 105−120.
(51) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Oxford University Press: Oxford, U.K., 1989.
(52) Andersen, H. C. Molecular dynamics simulations at constant pressure and/or temperature. *J. Chem. Phys.* **1980**, *72*, 2384−2393.
(53) Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A: At., Mol., Opt. Phys.* **1985**, *31*, 1695−1697.
(54) Hockney, R. W. The Potential Calculation and Some Applications. *Methods Comput. Phys.* **1970**, *9*, 135−211.
(55) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977**, *23*, 327−341.
(56) Brandenburg, J. G.; Hochheim, M.; Bredow, T.; Grimme, S. Low-Cost Quantum Chemical Methods for Noncovalent Interactions. *J. Phys. Chem. Lett.* **2014**, *5*, 4275−4284.
(57) Ojeda-May, P.; Nam, K. Acceleration of Semiempirical QM/MM Methods through Message Passage Interface (MPI), Hybrid MPI/Open Multiprocessing, and Self-Consistent Field Accelerator Implementations. *J. Chem. Theory Comput.* **2017**, *13*, 3525−3536.
(58) Christensen, A. S.; Elstner, M.; Cui, Q. Improving intermolecular interactions in DFTB3 using extended polarization from chemical-potential equalization. *J. Chem. Phys.* **2015**, *143*, 084123−11.
(59) Gao, J. L.; Amara, P.; Alhambra, C.; Field, M. J. A generalized hybrid orbital (GHO) method for the treatment of boundary atoms in combined QM/MM calculations. *J. Phys. Chem. A* **1998**, *102*, 4714−4721.
(60) Torrie, G. M.; Valleau, J. P. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.* **1977**, *23*, 187−199.
(61) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* **1992**, *13*, 1011−1021.
(62) Doron, D.; Kohen, A.; Major, D. T. Collective Reaction Coordinate for Hybrid Quantum and Molecular Mechanics Simulations: A Case Study of the Hydride Transfer in Dihydrofolate Reductase. *J. Chem. Theory Comput.* **2012**, *8*, 2484−96.
(63) Gao, J.; Ma, S.; Major, D. T.; Nam, K.; Pu, J.; Truhlar, D. G. Mechanisms and free energies of enzymatic reactions. *Chem. Rev.* **2006**, *106*, 3188−3209.
(64) Hub, J. S.; de Groot, B. L.; van der Spoel, D. g_wham-A Free Weighted Histogram Analysis Implementation Including Robust Error and Autocorrelation Estimates. *J. Chem. Theory Comput.* **2010**, *6*, 3713−3720.
(65) Hehre, W. J.; Ditchfield, R.; Pople, J. A. Self-Consistent Molecular-Orbital Methods 0.12. Further Extensions of Gaussian-Type Basis Sets for Use in Molecular-Orbital Studies of Organic-Molecules. *J. Chem. Phys.* **1972**, *56*, 2257−2261.
(66) Zhao, Y.; Truhlar, D. G. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Acc.* **2008**, *120*, 215−241.
(67) Brooks, B. R.; Brooks, C. L., 3rd; Mackerell, A. D., Jr.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: the biomolecular simulation program. *J. Comput. Chem.* **2009**, *30*, 1545−614.
(68) Woodcock, H. L., III; Hodoscek, M.; Gilbert, A. T. B.; Gill, P. M. W.; Schaefer, H. F., III; Brooks, B. R. Interfacing Q-Chem and CHARMM to Perform QM/MM Reaction Path Calculations. *J. Comput. Chem.* **2007**, *28*, 1485−1502.
(69) Rowley, C. N.; Roux, B. The Solvation Structure of Na+ and K+ in Liquid Water Determined from High Level ab Initio Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2012**, *8*, 3526−3535.
(70) Jin, H.; Goyal, P.; Das, A. K.; Gaus, M.; Meuwly, M.; Cui, Q. Copper Oxidation/Reduction in Water and Protein: Studies with DFTB3/MM and VALBOND Molecular Dynamics Simulations. *J. Phys. Chem. B* **2016**, *120*, 1894−1910.
(71) Sumner, S.; Soderhjelm, P.; Ryde, U. Effect of Geometry Optimizations on QM-Cluster and QM/MM Studies of Reaction Energies in Proteins. *J. Chem. Theory Comput.* **2013**, *9*, 4205−4214.
(72) Eurenius, K. P.; Chatfield, D. C.; Brooks, B. R.; Hodoscek, M. Enzyme mechanisms with hybrid quantum and molecular mechanical potentials. I. Theoretical considerations. *Int. J. Quantum Chem.* **1996**, *60*, 1189−1200.
(73) Mackerell, A. D., Jr. Empirical force fields for biological macromolecules: overview and issues. *J. Comput. Chem.* **2004**, *25*, 1584−604.
(74) Shivakumar, D.; Williams, J.; Wu, Y.; Damm, W.; Shelley, J.; Sherman, W. Prediction of Absolute Solvation Free Energies using Molecular Dynamics Free Energy Perturbation and the OPLS Force Field. *J. Chem. Theory Comput.* **2010**, *6*, 1509−19.
(75) Genheden, S.; Ryde, U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin. Drug Discovery* **2015**, *10*, 449−61.
(76) Gohlke, H.; Case, D. A. Converging free energy estimates: MM-PB(GB)SA studies on the protein-protein complex Ras-Raf. *J. Comput. Chem.* **2004**, *25*, 238−50.
(77) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E.; Mittal, J.; Feig, M.; Mackerell, A. D., Jr. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone phi, psi and side-chain chi(1) and chi(2) dihedral angles. *J. Chem. Theory Comput.* **2012**, *8*, 3257−3273.
(78) Baker, C. M.; Lopes, P. E.; Zhu, X.; Roux, B.; Mackerell, A. D., Jr. Accurate Calculation of Hydration Free Energies using Pair-Specific Lennard-Jones Parameters in the CHARMM Drude Polarizable Force Field. *J. Chem. Theory Comput.* **2010**, *6*, 1181−1198.
(79) Boulanger, E.; Thiel, W. Solvent Boundary Potentials for Hybrid QM/MM Computations Using Classical Drude Oscillators: A Fully Polarizable Model. *J. Chem. Theory Comput.* **2012**, *8*, 4527−38.
(80) Boulanger, E.; Thiel, W. Toward QM/MM Simulation of Enzymatic Reactions with the Drude Oscillator Polarizable Force Field. *J. Chem. Theory Comput.* **2014**, *10*, 1795−809.
(81) Ganguly, A.; Boulanger, E.; Thiel, W. Importance of MM Polarization in QM/MM Studies of Enzymatic Reactions: Assessment of the QM/MM Drude Oscillator Model. *J. Chem. Theory Comput.* **2017**, *13*, 2954−2961.
(82) Gao, J. L.; Xia, X. F. A Priori Evaluation of Aqueous Polarization Effects through Monte-Carlo Qm-Mm Simulations. *Science* **1992**, *258*, 631−635.
(83) Gao, J. Computation of Intermolecular Interactions with a Combined Quantum Mechanical and Classical Approach. In *Modeling the Hydrogen Bond*; Smith, D. A., Ed.; ACS Symposium Series 569; American Chemical Society: Washington, DC, 1994; pp 8−21. DOI: 10.1021/bk-1994-0569.ch002.
(84) Riccardi, D.; Li, G.; Cui, Q. Importance of van der Waals Interactions in QM/MM Simulations. *J. Phys. Chem. B* **2004**, *108*, 6467−6478.
(85) Major, D. T.; York, D. M.; Gao, J. Solvent polarization and kinetic isotope effects in nitroethane deprotonation and implications to the nitroalkane oxidase reaction. *J. Am. Chem. Soc.* **2005**, *127*, 16374−5.
(86) Major, D. T.; Gao, J. A combined quantum mechanical and molecular mechanical study of the reaction mechanism and alpha-amino acidity in alanine racemase. *J. Am. Chem. Soc.* **2006**, *128*, 16345−57.

DOI: 10.1021/acs.jctc.7b00964
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

(87) Major, D. T.; Heroux, A.; Orville, A. M.; Valley, M. P.; Fitzpatrick, P. F.; Gao, J. Differential quantum tunneling contributions in nitroalkane oxidase catalyzed and the uncatalyzed proton transfer reaction. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 20734−9.
(88) Doron, D.; Major, D. T.; Kohen, A.; Thiel, W.; Wu, X. Hybrid Quantum and Classical Simulations of the Dihydrofolate Reductase Catalyzed Hydride Transfer Reaction on an Accurate Semi-Empirical Potential Energy Surface. *J. Chem. Theory Comput.* **2011**, *7*, 3420−37.
(89) Vardi-Kilshtain, A.; Major, D. T.; Kohen, A.; Engel, H.; Doron, D. Hybrid Quantum and Classical Simulations of the Formate Dehydrogenase Catalyzed Hydride Transfer Reaction on an Accurate Semiempirical Potential Energy Surface. *J. Chem. Theory Comput.* **2012**, *8*, 4786−96.
(90) Hu, L.; Eliasson, J.; Heimdal, J.; Ryde, U. Do quantum mechanical energies calculated for small models of protein-active sites converge? *J. Phys. Chem. A* **2009**, *113*, 11793−800.
(91) Lonsdale, R.; Harvey, J. N.; Mulholland, A. J. A practical guide to modelling enzyme-catalysed reactions. *Chem. Soc. Rev.* **2012**, *41*, 3025−3038.
(92) van der Kamp, M. W.; Zurek, J.; Manby, F. R.; Harvey, J. N.; Mulholland, A. J. Testing High-Level QM/MM Methods for Modeling Enzyme Reactions: Acetyl-CoA Deprotonation in Citrate Synthase. *J. Phys. Chem. B* **2010**, *114*, 11303−11314.
(93) Klahn, M.; Braun-Sand, S.; Rosta, E.; Warshel, A. On Possible Pitfalls in ab Initio Quantum Mechanics/Molecular Mechanics Minimization Approaches for Studies of Enzymatic Reactions. *J. Phys. Chem. B* **2005**, *109*, 15645−15650.

DOI: 10.1021/acs.jctc.7b00964
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX