Accelerated Computation of Free Energy Profile at *Ab Initio* Quantum Mechanical/Molecular Mechanics Accuracy via a Semiempirical Reference Potential. 4. Adaptive QM/MM

Jia-Ning Wang,[†] Wei Liu,[†,⊥] Pengfei Li,[*,†,#] Yan Mo,[*,†,‡,¶] Wenxin Hu,[§] Jun Zheng,[§] Xiaoliang Pan,[‖] Yihan Shao,[‖] and Ye Mei[*,†,‡,¶]

†*State Key Laboratory of Precision Spectroscopy, School of Physics and Electronic Science, East China Normal University, Shanghai 200062, China*

‡*NYU-ECNU Center for Computational Chemistry at NYU Shanghai, Shanghai 200062, China*

¶*Collaborative Innovation Center of Extreme Optics, Shanxi University, Taiyuan, Shanxi 030006, China*

§*The Computer Center, School of Data Science & Engineering, East China Normal University, Shanghai 200062, China*

‖*Department of Chemistry and Biochemistry, University of Oklahoma, Norman Oklahoma 73019, United States*

⊥*Current address: Center for Integrative Imaging, Hefei National Laboratory for Physical Sciences at the Microscale, and School of Life Sciences, University of Science and Technology of China (USTC), Hefei, Anhui 230026, China*

#*Current address: Silicon Therapeutics (Suzhou) Co., Ltd., Suzhou, Jiangsu 215000, China*

E-mail: alan.pengfeili@gmail.com; ymo@phy.ecnu.edu.cn; samuel.y.mei@gmail.com

Abstract

Although Quantum Mechanical/Molecular Mechanics (QM/MM) methods are now routinely applied to the studies of chemical reactions in condensed phases and enzymatic reactions, they may confront technical difficulties when the reactive region is varying over time. For instance, when the solvent molecules are directly participating in the reaction, the exchange of water molecules between the QM and MM regions may occur on a time scale that is comparable to that of the reaction. Several adaptive QM/MM schemes have been proposed to cope with this situation. However, these methods either significantly increase the computational cost or introduce artificial restraints to the system. In this work, we developed a novel adaptive QM/MM scheme and applied it to a study of a nucleophilic addition reaction. In this scheme, the configuration sampling was performed with a small QM region (without solvent molecules), and the thermodynamic properties under other potential energy functions with larger QM regions (with a different number of solvent molecules and/or different level of QM theory) are computed via extrapolation using the reference-potential method. Our simulation results show that this adaptive QM/MM scheme is numerically stable, at least for the case studied in this work. Furthermore, this method also offers an inexpensive way to examine the convergence of the QM/MM calculation with respect to the size of the QM region.

Introduction

Hybrid Quantum Mechanical/Molecular Mechanical (QM/MM) methods are nowadays well accepted for the simulations of chemical reactions in condensed phases and enzymatic reactions.[1–8] However, applications of these methods are always hindered by their steep computational expense and complexity in domain partitioning. First, in order to determine the reaction mechanism, a long molecular dynamics simulation at *ab initio* levels is needed, from which statistical properties can be extracted reliably. With a sub-fs time step for propagation, 10^6 to 10^9 steps of energy and force evaluations are required to reach a ns to μs time scale. Second, the size of the QM region matters. Defining the QM region is often based on chemical intuition and is a compromise between accuracy and efficiency. A small QM region may lead to systematically biased results.[9–13] Last but not the least, the partitioning of the whole system into the QM and the MM regions is nontrivial, especially when solvent molecules are strongly involved in the reactions. By including some of the solvent molecules near the solute molecule into the QM region can capture the quantum mechanical interaction between the solute and the solvent molecules. However, it brings another technical difficulty in maintaining dynamic continuity when the exchange of solvent molecules between the QM and the MM regions takes place, especially when an abrupt on-the-fly repartitioning scheme of the QM and MM regions is adopted. In order to solve this difficulty, various schemes of adaptive QM/MM methods have been proposed,[14–17] which can be categorized broadly into restrained QM/MM schemes[18–20] and adaptive QM/MM schemes.[21–28] In the former class of schemes, solvent exchange between the QM and the MM regions is prevented by applying a restraining potential. However, the evolution of the system under study is no longer under a realistic Hamiltonian due to the introduction of an artificial restraint. In the adaptive QM/MM scheme, an effective QM/MM potential is adopted by a weighted average of the potentials from multiple means of partitioning of the system with varying combinations of solute and solvent molecules. This may significantly increase the computational expense.

Fortunately, if we are only interested in thermodynamic properties, for instance, the free energy profile, instead of real dynamics, these properties can be calculated indirectly via the reference-potential approach,[29–32] of which the idea has been applied to many studies.[33–57] Specifically, the QM/MM partitioning with a fixed number of solvent molecules in the QM region, although the specific water molecules may vary, has a constant functional form for the Hamiltonian, say H_1. The exchange of solvent molecules between the QM and MM regions is just a permutation of the state before the exchange takes place. Another partitioning scheme, with no solvent molecules in the QM region, adopts a Hamiltonian H_0, which serves as the reference potential. The ensemble average of an operator X under H_1 can be computed from the ensemble of H_0 via reweighting[58]

$$\langle X\rangle_1 = \frac{\int X e^{-\beta H_1}\,\mathrm{d}\mathbf{R}}{\int e^{-\beta H_1}\,\mathrm{d}\mathbf{R}} = \frac{\int X e^{\beta(H_0-H_1)} e^{-\beta H_0}\,\mathrm{d}\mathbf{R}}{\int e^{\beta(H_0-H_1)} e^{-\beta H_0}\,\mathrm{d}\mathbf{R}} = \frac{\left\langle X e^{\beta(H_0-H_1)}\right\rangle_0}{\left\langle e^{\beta(H_0-H_1)}\right\rangle_0},$$

where $\langle\cdot\rangle$ denotes the ensemble average or expectation, and the subscript 0 or 1 indicates the Hamiltonian under which the ensemble is calculated. Here, $\frac{e^{\beta(H_0-H_1)}}{\left\langle e^{\beta(H_0-H_1)}\right\rangle_0}$ can be considered as the weight under H_1 for the configurations sampled with H_0. For a generalized ensemble, the equation can be slightly more complicated, but the idea is the same. Recently, Jia et al proposed a reference potential method for the free energy calculations at an expensive level of theory using a unique Boltzmann ensemble.[41] Li et al extended this method to mixed ensembles from, but not limited to, umbrella sampling (US)[58] simulations.[47] In these methods, a long simulation using a less expensive Hamiltonian is performed to explore the phase space, and from this simulation, a free energy profile corresponding to this Hamiltonian can be estimated using well-established postprocessing methods such as Multistate Bennett Acceptance Ratio (MBAR)[59,60] and the Weighted Histogram Analysis Method (WHAM).[61–63] Next, a correction in the free energy from this inexpensive Hamiltonian to the Hamiltonian of interest is calculated using thermodynamic perturbation (TP).[64] In this way, expensive direct simulations at the high level Hamiltonian can be avoided. When calculating the correction for a mixed ensemble, weight factors from the MBAR analysis should be used.[47] Therefore, the TP should be carried out with nonuniform weights for the samples. Thermodynamic expectations of any structural properties can be computed in a similar way.[47,53]

In this work, we put forth a new method for the free energy calculations with an adaptive QM domain for the study of the intramolecular nucleophilic addition reaction of $\mathrm{Me_2N{-}(CH_2)_3{-}CH{=}O}$ (NCO) molecule utilizing the idea of energy reweighting in the reference-potential methods. It has been shown in a previous study that explicit solvation matters for the thermodynamic property calculations along the reaction.[65] The umbrella sampling simulations are carried out only at the semi-empirical level, specifically PM6[66]/MM level, without a single water molecule in the QM region. After the simulation, the trajectories are postprocessed for semi-empirical QM/MM or *ab initio* QM/MM energy calculations with a certain number of solvent molecules included in the QM region, from which the free energy profiles at these levels are obtained. This manuscript is organized in the following way. In the next section, the theory behind this method and the simulation setup are explained, and the results are presented subsequently with discussion. Finally, we end with a conclusion for this study.

Method

Multistate Thermodynamic Perturbation (MsTP) Method

The MsTP method, previously known as MBAR+wTP, was proposed by Li et al.[47] recently. Derivation of the MsTP method has been fully presented in ref. 47. In this method, enhanced sampling methods such as umbrella sampling simulations are conducted under a reference (and usually inexpensive) Hamiltonian, for instance, semi-empirical (SE) QM/MM. Thermodynamics properties under this reference Hamiltonian can be obtained using MBAR analysis, and then corrected to the target Hamiltonian using the weighted thermodynamic perturbation. These steps can be integrated into the MBAR formulation as explained in the following.

With trajectories from K simulations using different potential energy functions U_k as is typically done in Umbrella Sampling, thermodynamic properties, which depend only on coordinates, under another potential energy function U_t can be computed via

$$\langle \mathbf{A} \rangle_t = \frac{\sum\limits_{n=1}^{N} w_t(\mathbf{r}_n)\mathbf{A}(\mathbf{r}_n)}{\sum\limits_{n=1}^{N} w_t(\mathbf{r}_n)}, \tag{1}$$

in which $N = \sum\limits_{k} N_k$ and N_k is the number of configurations extracted from the kth simulation, and

$$w_t(\mathbf{r}_n) = \frac{\exp\left[-\beta U_t(\mathbf{r}_n)\right]}{\sum\limits_{k=1}^{K} N_k \exp\left[\beta f_k - \beta U_k(\mathbf{r}_n)\right]} \tag{2}$$

is the unnormalized weight of configuration $\mathbf{r}_n$ under U_k. Here, f_k is known as the free energy corresponding to U_k and can be obtained by iteratively solving the MBAR equations

$$f_i = -\beta^{-1} \ln \sum_{n=1}^{N} \frac{\exp\left[-\beta U_i(\mathbf{r}_n)\right]}{\sum\limits_{k=1}^{K} N_k \exp\left[\beta f_k - \beta U_k(\mathbf{r}_n)\right]}, \quad \forall i = 1, \ldots, K. \tag{3}$$

In US, the potential energy functions used for configuration sampling are

$$U_k(\mathbf{r}_n) = U_0(\mathbf{r}_n) + W_k(\mathbf{r}_n) \tag{4}$$

where $U_0(\mathbf{r})$ and $W_k(\mathbf{r})$ are the unbiased potential energy function and the biasing potential for the kth simulation, respectively. Equation 2 can be rewritten as

$$w_t(\mathbf{r}_n) = \frac{\exp\left[-\beta \Delta U_t(\mathbf{r}_n)\right]}{\sum\limits_{k=1}^{K} N_k \exp\left[\beta f_k - \beta W_k(\mathbf{r}_n)\right]}, \tag{5}$$

with $\Delta U_t(\mathbf{r}) = U_t(\mathbf{r}) - U_0(\mathbf{r})$. Further defining the free energy f_t corresponding to the potential energy function $U_t(\mathbf{r})$

$$f_t = -\beta^{-1} \ln \sum_{n=1}^{N} w_t(\mathbf{r}_n), \tag{6}$$

we have the normalized weight for configuration $\mathbf{r}_n$ under the potential energy function $U_t(\mathbf{r})$

$$\tilde{w}_t(\mathbf{r}_n) = \frac{\exp\left[\beta f_t - \beta \Delta U_t(\mathbf{r}_n)\right]}{\sum\limits_{k=1}^{K} N_k \exp\left[\beta f_k - \beta W_k(\mathbf{r}_n)\right]}, \tag{7}$$

and Eq. 1 can be simplified as

$$\langle \mathbf{A} \rangle_t = \sum_{n=1}^{N} \tilde{w}_t(\mathbf{r}_n)\mathbf{A}(\mathbf{r}_n). \tag{8}$$

It can be easily identified that for a single unbiased simulation with $K = 1$ and $W = 0$, Eq. 7 can be rewritten as the normal TP equation. Therefore, the idea behind Eq. 7 can be thought as multistate thermodynamics perturbation (MsTP). MsTP has been applied to the calculations of free energy profiles for chemical reactions in condensed phase rendered in both one dimensional[47] and two dimensional[53] reaction coordinates. The computational expense decreases by two orders of magnitude comparing with direct QM/MM calculations while maintaining a high accuracy.

Specifically for this adaptive QM/MM calculation, potential energy function $U_0(\mathbf{r})$ corresponds to the partitioning with a solvent-free QM region described by a semi-empirical Hamiltonian PM6. $W_k(\boldsymbol{\xi}(\mathbf{r}))$ is the restraining potential on a predefined collective variable (CV) $\boldsymbol{\xi}(\mathbf{r})$ that may enhance the phase space sampling in a certain region. $U'_M(\mathbf{r})$ is the potential energy function for the partitioning with M solvent molecules in the QM region described by either a semi-empirical QM or an *ab initio* QM level of theory. The prime sign here is to emphasize that the QM level of theory can be either the same as or different from the reference Hamiltonian. If $\mathbf{A}$ is an indication function δ of some chosen CV $\boldsymbol{\xi}(\mathbf{r})$

$$\delta(\boldsymbol{\xi}_m - \boldsymbol{\xi}(\mathbf{r})) = \begin{cases} 1, & \text{if } -\Delta\boldsymbol{\xi}/2 < \boldsymbol{\xi}_m - \boldsymbol{\xi}(\mathbf{r}) < \Delta\boldsymbol{\xi}/2 \\ 0, & \text{otherwise} \end{cases}, \tag{9}$$

we have the PMF for $U'_M(\mathrm{r})$ as

$$F_M(\boldsymbol{\xi}_m) = -\beta^{-1} \ln \sum_{n=1}^{N} \omega_M(\mathbf{r}_n)\delta(\boldsymbol{\xi}_m - \boldsymbol{\xi}(\mathbf{r}_n)) \tag{10}$$

defined up to an additive constant, and

$$w_M(\mathbf{r}_n) = \frac{\exp\left[-\beta\Delta U'_M(\mathbf{r}_n)\right]}{\sum\limits_{k=1}^{K} N_k \exp\left[\beta f_k - \beta W_k(\mathbf{r}_n)\right]} = \frac{\exp\left[-\beta(U'_M(\mathbf{r}_n) - U_0(\mathbf{r}_n))\right]}{\sum\limits_{k=1}^{K} N_k \exp\left[\beta f_k - \beta W_k(\mathbf{r}_n)\right]}. \tag{11}$$

Similarly, the PMF for $U_0(\mathbf{r})$ is

$$F_0(\boldsymbol{\xi}_m) = -\beta^{-1} \ln \sum_{n=1}^{N} \omega_0(\mathbf{r}_n)\delta(\boldsymbol{\xi}_m - \boldsymbol{\xi}(\mathbf{r}_n)), \tag{12}$$

in which

$$\omega_0(\mathbf{r}_n) = \frac{1}{\sum\limits_{k=1}^{K} N_k \exp\left[\beta f_k - \beta W_k(\mathbf{r}_n)\right]}. \tag{13}$$

Random noise in the potential of mean force from finite sampling was eliminated by a Gaussian smoothing on the density-of-states of $\Delta U'_M(\mathbf{r})$.[67] Gaussian processes regression (GPR) method[68] is used to eliminate the statistical noise in the free energy profile from the MsTP calculation.

Model Setup

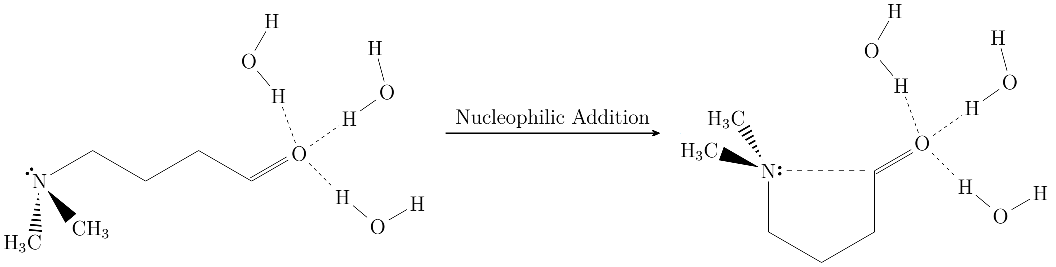

Figure 1: The nucleophilic addition reaction of $Me_2N-(CH_2)_3-CH{=}O$ (NCO) molecule studied in this work.

$Me_2N-(CH_2)_3-CH{=}O$ (NCO) was solvated in a TIP3P water[69] sphere with a radius of 20 Å centering at the NCO molecule, which contains 1020 water molecules. The whole system was optimized by 2000 steps of steepest descent algorithm and 3000 steps of conjugate gradient method. The optimized structure was heated up to 300 K in 1 ns and then further relaxed for 10 ns. Periodic boundary condition was not applied, and the water sphere was restrained by a soft half-harmonic potential with a force constant of 10 $\text{kcal} \cdot \text{mol}^{-1} \cdot \text{Å}^{-2}$ to avoid evaporation. The integration time step was set to 2 fs. The nonbonded interaction was fully counted without any truncation. The van der Waals (vdW) parameters for the NCO molecule were taken from the general AMBER force field (GAFF)[70] for the ring-opening structure, and the AM1bcc charges from the reactant configuration were assigned to the NCO molecule. The temperature was regulated to 300 K using the Langevin dynamics with a collision frequency of 1 ps^{-1}.[71]

Umbrella Sampling

The phase space exploration was assisted by umbrella sampling.[58] The distance between the nitrogen and carbonyl carbon atom of the NCO molecule was chosen as the CV $\boldsymbol{\xi}(\mathbf{r})$, which ranges from 1.50 to 5.00 Å with an increment of 0.05 Å. Overlap matrix proposed by Klimovich et al.[72] was used to monitor the degree of overlap between adjacent simulation windows. Extra windows were added when the overlap between neighboring windows are insufficient, resulting in 84 windows in total. The setup of the restraint potential in each window simulation can be found in the supporting information (SI). The central region contains only the NCO molecule, and PM6 was used for its interaction potential. For each US window, the whole system was optimized by 1000 steepest descent steps and 1000 conjugate gradient steps. The relaxed system was heated up to 300 K in 100 ps, followed by a 1-ns production simulation. The temperature was maintained at 300 K using the Langevin dynamics with a collision frequency of 1 ps^{-1}. The integration time step was set to 1 fs. The configurations were saved every 1 ps for subsequent free energy analysis. The free energy profile at this level was computed using the MBAR analysis method. After that, single point energies under PM6/MM and ωB97X-D[73]/6-31+G(d,p)/MM levels were obtained for the MsTP calculations. For the single point energy calculations, the QM region was augmented with $M = 0$, 2, 3, or 4 water molecules that are closest to the oxygen atom in the carbonyl group for each snapshot. The propagation of the molecular dynamics simulations and single point energy calculations were carried out using the AMBER 18 package suite.[74] Interfacing with Gaussian 16 package[75] was utilized when ωB97X-D energy calculations were requested.

Results and Discussion

Potential of Mean Force at the PM6 Levels

The free energy profiles at the PM6 level are shown in Fig. 2. Since in the US simulations the QM region had no water molecules included, the uncertainty of the free energy profile is very small. The free energy profile shows a shallow well at $d_{CN} = 4.54\,\text{Å}$ for the reactant but a deep well at $d_{CN} = 1.66\,\text{Å}$ for the product. The barrier for the forward reaction is 3.4 kcal/mol, and the reaction free energy is 4.2 kcal/mol. By including two water molecules into the QM region, the free energy profile shows only small difference from that with a solvent-free QM region. The locations of the reactant and the product are nearly unchanged, with the new locations being $d_{CN} = 4.52\,\text{Å}$ and $d_{CN} = 1.66\,\text{Å}$ for the reactant and the product, respectively. The barrier for the forward reaction becomes 3.5 kcal/mol, and the reaction free energy is 4.3 kcal/mol. The PM6 free energy profile with two water molecules in the QM region has already reached convergence by comparing it with those enveloping more water molecules into the QM region. The results indicate that under the PM6 level of theory, the solvent molecules play a role as an electrostatic perturbator to the NCO molecule that only weakly tunes the reaction. As shown in Fig. S1, the magnitude of charge transferred from the solute molecule to the solvent is smaller than $0.1e$, even when the C-N bond has formed.

Potential of Mean Force at the DFT Levels

Extrapolation to the DFT/MM level using the US trajectories from the PM6/MM simulations is also possible, and the free energies profiles are shown in Fig. 3. Similar to the results

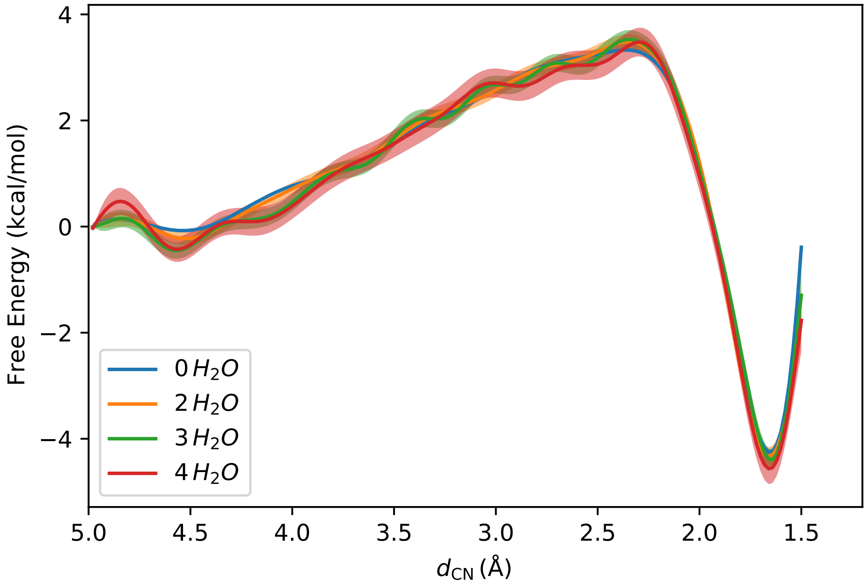

Figure 2: Free energy profiles at the PM6 levels with different numbers of water molecules in the QM region. The shaded areas are the 95% confidence intervals.

at the PM6/MM level, the locations of the reactant, the products and the transition states are nearly independent of the number of water molecules in the QM region. Without water molecules in the QM region, the free energy barrier for the forward reaction and the reaction free energy are 2.0 kcal/mol and -1.6 kcal/mol, respectively. Both of them are much smaller than those under the PM6/MM level of theory in their absolute values. When two water molecules are added to the QM region, the free energy profile shows large deviations from the one at the same level of theory but with no solvent molecules in the QM region, especially at the product side. When solvent molecules in the MM region are represented as background charges that polarize the electronic structure of the QM region, charge transfer is not allowed between the QM solute molecule and the MM solvent molecules. When two nearest solvent molecules are included in the QM region, the water molecules can accommodate the extra electrons around the oxygen atom in the carbonyl group, especially after the formation of the C−N bond. Therefore, the product is stabilized by 1.6 kcal/mol, and the reaction free energy becomes -3.3 kcal/mol. The free energy barrier for the forward reaction increases to 2.4 kcal/mol. Adding more water molecules into the QM region does not significantly change the profile. With three water molecules in the QM region, the reactant and the product are located at $d_{CN} = 4.52\,\text{Å}$ and $1.64\,\text{Å}$, and the free energy barrier and the reaction free energy are 2.3 kcal/mol and -3.3 kcal/mol. With four water molecules in the QM region, the reactant and the product are located at $d_{CN} = 4.50\,\text{Å}$ and $1.66\,\text{Å}$, and the free energy barrier and the reaction free energy are 2.1 kcal/mol and -3.1 kcal/mol. Considering the uncertainties in the free energy profiles, these numbers are statistically identical. Therefore, when two water molecules are included in the QM region, the free energies have converged.

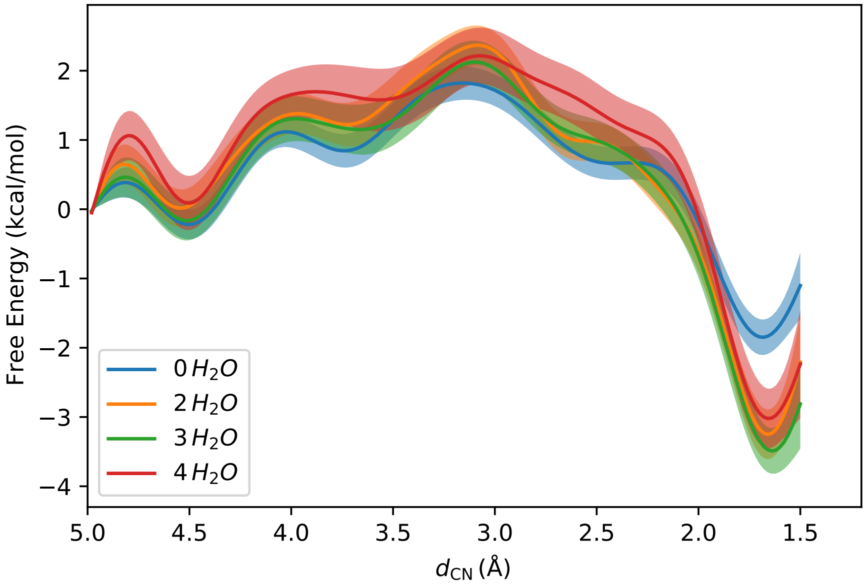

Figure 3: Free energy profiles at the DFT levels with different numbers of water molecules in the QM region. The shaded areas are the 95% confidence intervals.

The variations of the CM5 charges[76] of the polar atoms in the NCO molecule during the reaction are shown in Fig. 4, with the atomic charges of the bonded hydrogen atoms merged into those of the heavy atoms. It shows that when the nitrogen atom approaches the carbon atom in the carbonyl group, the lone pair electrons of the nitrogen atom become shared electrons between the nitrogen atom and the carbon atom, and push the shared electrons in the carbonyl group to the oxygen side. Some portion of the electrons drifts away from oxygen atom in the carbonyl group to the water molecules hydrogen-bonded with the carbonyl group. As a results, the CM5 charge of the nitrogen atom goes up (becomes less negative) by about $0.187e$, that of the oxygen atom in the carbonyl group goes down by about $0.162e$, and the four water molecules near the carbonyl oxygen atom accept about 0.228 electrons altogether. Since the carbonyl carbon atom accepts electrons from the nitrogen atom but donates electrons to the oxygen atom, its CM5 charge decreases by only $0.074e$. It agrees with the previous observation of the $\mathrm{N}^{+}|\mathrm{C}{-}\mathrm{O}^{-}$ pattern.[77] Since the bond order between the carbon and the oxygen atom in the carbonyl group decreases, the bond length increases during the reaction, as can be seen in Fig. 5. In the reactant region, the CO bond distance remains around 1.22 Å. When the distance between the nitrogen atom and the carbon atom is smaller than 3.0 Å (on the product side), the CO bond distance goes up quickly.

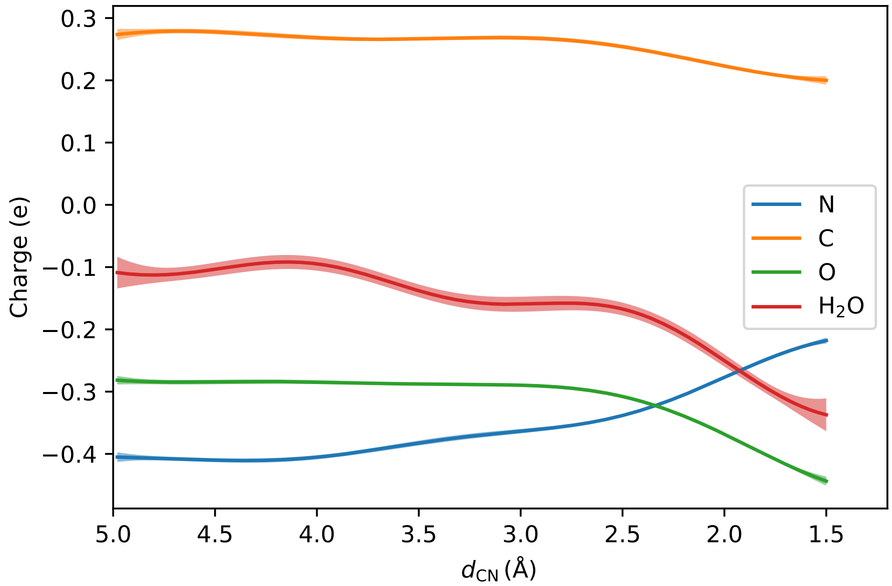

Figure 4: Alternations of atomic CM5 charges along the reaction at the DFT/MM level with 4 water molecules in the QM region. The shaded areas are the 95% confidence intervals.

If we define a buffer region between the QM and the MM regions, this method can be straightforwardly combined with the permuted adaptive partitioning (PAP) scheme[22] by

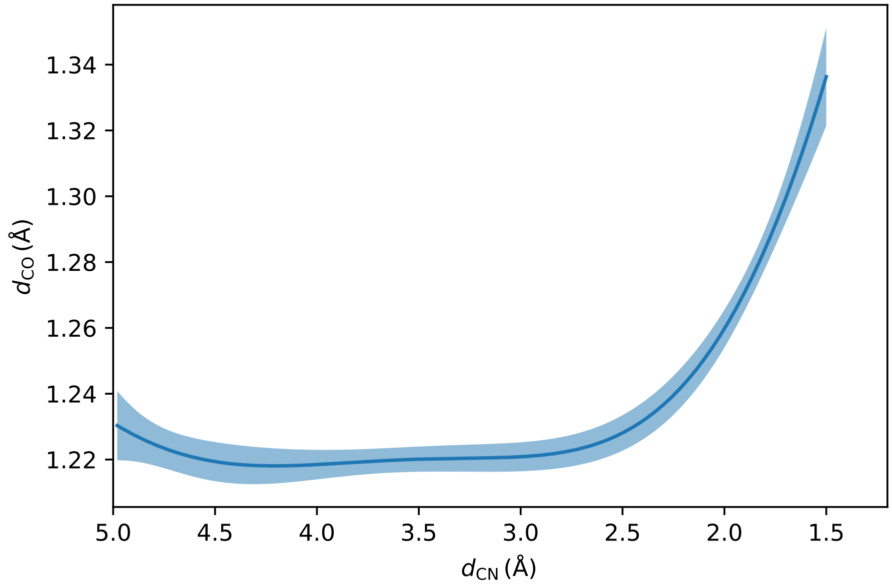

Figure 5: Variations of the C=O bond length along the reaction at the DFT/MM level with 4 water molecules in the QM region. The shaded area is the 95% confidence interval.

setting the PAP Hamiltonian at a certain QM level of theory as the target Hamiltonian. In spite of its rigor, PAP partition scheme is computationally very demanding and scales poorly with the size of the buffer region. Therefore, only one permutation is considered in this work, in which the nearest M solvent molecules are bracketed into the QM region. The consequence is the loss of the detailed balance and the continuity of the trajectory because of the introduction of human intervention in picking the water molecules to be included in the QM region. Other partition schemes can be adopted. However, as shown in Fig. S2, the permutation we have used (bracketing the nearest M solvent into the QM region) in this work has the largest impact on the free energy change from the reactant to the product, since it can best relax the extra electron density around the oxygen atom in the NCO molecule once the C−N bond has formed. While the other scheme (with the 2nd and 3rd nearest water molecules in the QM region) underestimates the reaction free energy, due that the magnitude of the electron transfer from the NCO molecule to the solvent is underestimated, especially at the product side.

Conclusion

In this work, we proposed a novel method for adaptive QM/MM simulations of chemical reaction in a homogeneous environment, which is based on the reference-potential method and can be easily implemented. With this method, extrapolations to a different level of theory or/and to a different size of QM region are made possible. The uncertainty increases with the "magnitude" of the extrapolation. Increasing the number of water molecules being bracketed into the QM region increases the standard deviation of the potential of mean force. Fortunately, this numerical difficulty can be easily solved by extending the length of the simulation at the low level of theory. This method also offers a convenient way to check the convergence of the QM/MM calculations with respect to the size of the QM region even in heterogeneous but invariant embedding environment. Semiempirical methods such as PM6 should be used with care, due to the difficulty in handling charge transfer effect with minimum basis sets.

Acknowledgement

Y. Mei wants to express his gratitude to Dr. Wei Yang at Florida State University and Dr. Hai Lin at University of Colorado Denver for some insightful discussion. Y. Mei is supported by the National Natural Science Foundation of China (Grant No. 21773066 and 22073030). Y. Mo is supported by the National Natural Science Foundation of China (Grant No. 21973030). W.H. is supported by the Fundamental Research Funds for the Central Universities. Y.S. is supported by the National Institutes of Health (Grant No. R01GM135392). CPU time was provided by the Supercomputer Center of East China Normal University (ECNU Public Platform for Innovation No. 001).

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*, 227–249.

(2) Field, M. J.; Bash, P. A.; Karplus, M. A Combined Quantum Mechanical and Molecular Mechanical Potential for Molecular Dynamics Simulations. *J. Comput. Chem.* **1990**, *11*, 700–733.

(3) Monard, G.; Merz, K. M. Combined Quantum Mechanical/Molecular Mechanical Methodologies Applied to Biomolecular Systems. *Acc. Chem. Res.* **1999**, *32*, 904–911.

(4) 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.

(5) Lin, H.; Truhlar, D. G. QM/MM: What Have We Learned, Where Are We, and Where Do We Go from Here? *Theor. Chem. Acc.* **2006**, *117*, 185.

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

(9) Solt, I.; Kulhánek, P.; Simon, I.; Winfield, S.; Payne, M. C.; Csányi, G.; Fuxreiter, M.

Evaluating Boundary Dependent Errors in QM/MM Simulations. *J. Phys. Chem. B* **2009**, *113*, 5728–5735.

(10) 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.

(11) Liao, R.-Z.; Thiel, W. Convergence in the QM-only and QM/MM Modeling of Enzymatic Reactions: A Case Study for Acetylene Hydratase. *J. Comput. Chem.* **2013**, *34*, 2389–2397.

(12) Kulik, H. J.; Zhang, J.; Klinman, J. P.; Martínez, 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.

(13) Mehmood, R.; Kulik, H. J. Both Configuration and QM Region Size Matter: Zinc Stability in QM/MM Models of DNA Methyltransferase. *J. Chem. Theory Comput.* **2020**, *16*, 3121–3134.

(14) Bulo, R. E.; Michel, C.; Fleurat-Lessard, P.; Sautet, P. Multiscale Modeling of Chemistry in Water: Are We There Yet? *J. Chem. Theory Comput.* **2013**, *9*, 5567–5577.

(15) Pezeshki, S.; Lin, H. In *Quantum Modeling of Complex Molecular Systems*; Rivail, J.-L., Ruiz-Lopez, M., Assfeld, X., Eds.; Springer International Publishing: Cham, 2015; pp 93–113.

(16) Zheng, M.; Waller, M. P. Adaptive Quantum Mechanics/Molecular Mechanics Methods. *WIREs Comput. Mol. Sci.* **2016**, *6*, 369–385.

(17) Duster, A. W.; Wang, C.-H.; Garza, C. M.; Miller, D. E.; Lin, H. Adaptive Quantum/Molecular Mechanics: What Have We Learned, Where Are We, and Where Do We Go from Here? *WIREs Comput. Mol. Sci.* **2017**, *7*, e1310.

(18) 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.

(19) Shiga, M.; Masia, M. Boundary Based on Exchange Symmetry Theory for Multilevel Simulations. I. Basic Theory. *J. Chem. Phys.* **2013**, *139*, 044120.

(20) Takahashi, H.; Kambe, H.; Morita, A. A Simple and Effective Solution to the Constrained QM/MM Simulations. *J. Chem. Phys.* **2018**, *148*, 134119.

(21) Kerdcharoen, T.; Morokuma, K. ONIOM-XS: An Extension of the ONIOM Method for Molecular Simulation in Condensed Phase. *Chem. Phys. Lett.* **2002**, *355*, 257–262.

(22) Heyden, A.; Lin, H.; Truhlar, D. G. Adaptive Partitioning in Combined Quantum Mechanical and Molecular Mechanical Calculations of Potential Energy Functions for Multiscale Simulations. *J. Phys. Chem. B* **2007**, *111*, 2231–2241.

(23) Bulo, R. E.; Ensing, B.; Sikkema, J.; Visscher, L. Toward a Practical Method for Adaptive QM/MM Simulations. *J. Chem. Theory Comput.* **2009**, *5*, 2212–2221.

(24) Bernstein, N.; Várnai, C.; Solt, I.; Winfield, S. A.; Payne, M. C.; Simon, I.; Fuxreiter, M.; Csányi, G. QM/MM Simulation of Liquid Water with an Adaptive Quantum Region. *Phys. Chem. Chem. Phys.* **2012**, *14*, 646–656.

(25) Takenaka, N.; Kitamura, Y.; Koyano, Y.; Nagaoka, M. The Number-adaptive Multiscale QM/MM Molecular Dynamics Simulation: Application to Liquid Water. *Chem. Phys. Lett.* **2012**, *524*, 56–61.

(26) Waller, M. P.; Kumbhar, S.; Yang, J. A Density-Based Adaptive Quantum Mechanical/Molecular Mechanical Method. *ChemPhysChem* **2014**, *15*, 3218–3225.

(27) Watanabe, H. C.; Kubař, T.; Elstner, M. Size-Consistent Multipartitioning QM/MM:

A Stable and Efficient Adaptive QM/MM Method. *J. Chem. Theory Comput.* **2014**, *10*, 4242–4252.

(28) Field, M. J. An Algorithm for Adaptive QC/MM Simulations. *J. Chem. Theory Comput.* **2017**, *13*, 2342–2351.

(29) Gao, J. Absolute Free Energy of Solvation from Monte Carlo Simulations Using Combined Quantum and Molecular Mechanical Potentials. *J. Phys. Chem.* **1992**, *96*, 537–540.

(30) Gao, J.; Xia, X. A Priori Evaluation of Aqueous Polarization Effects through Monte Carlo QM-MM Simulations. *Science* **1992**, *258*, 631–635.

(31) Muller, R. P.; Warshel, A. Ab Initio Calculations of Free Energy Barriers for Chemical Reactions in Solution. *J. Phys. Chem.* **1995**, *99*, 17516–17524.

(32) Bentzien, J.; Muller, R. P.; Florián, J.; Warshel, A. Hybrid *ab initio* Quantum Mechanics/Molecular Mechanics Calculations of Free Energy Surfaces for Enzymatic Reactions: The Nucleophilic Attack in Subtilisin. *J. Phys. Chem. B* **1998**, *102*, 2293–2301.

(33) Rod, T. H.; Ryde, U. Quantum Mechanical Free Energy Barrier for an Enzymatic Reaction. *Phys. Rev. Lett.* **2005**, *94*, 138302.

(34) Beierlein, F. R.; Michel, J.; Essex, J. W. A Simple QM/MM Approach for Capturing Polarization Effects in Protein-Ligand Binding Free Energy Calculations. *J. Phys. Chem. B* **2011**, *115*, 4911–4926.

(35) König, G.; Boresch, S. Non-Boltzmann Sampling and Bennett's Acceptance Ratio Method: How to Profit from Bending the Rules. *J. Comput. Chem.* **2011**, *32*, 1082–1090.

(36) Heimdal, J.; Ryde, U. Convergence of QM/MM Free-Energy Perturbations Based on

Molecular-Mechanics or Semiempirical Simulations. *Phys. Chem. Chem. Phys.* **2012**, *14*, 12592–12604.

(37) Polyak, I.; Benighaus, T.; Boulanger, E.; Thiel, W. Quantum Mechanics/Molecular Mechanics Dual Hamiltonian Free Energy Perturbation. *J. Chem. Phys.* **2013**, *139*, 064105.

(38) König, G.; Hudson, P. S.; Boresch, S.; Woodcock, H. L. Multiscale Free Energy Simulations: An Efficient Method for Connecting Classical MD Simulations to QM or QM/MM Free Energies Using Non-Boltzmann Bennett Reweighting Schemes. *J. Chem. Theory Comput.* **2014**, *10*, 1406–1419.

(39) Hudson, P. S.; Woodcock, H. L.; Boresch, S. Use of Nonequilibrium Work Methods to Compute Free Energy Differences Between Molecular Mechanical and Quantum Mechanical Representations of Molecular Systems. *J. Phys. Chem. Lett.* **2015**, *6*, 4850–4856.

(40) Hudson, P. S.; White, J. K.; Kearns, F. L.; Hodoscek, M.; Boresch, S.; Woodcock, H. L. Efficiently Computing Pathway Free Energies: New Approaches Based on Chain-of-Replica and Non-Boltzmann Bennett Reweighting Schemes. *Biochim. Biophys. Acta, Gen. Subj.* **2015**, *1850*, 944–953.

(41) Jia, X.; Wang, M.; Shao, Y.; Koenig, G.; Brooks, B. R.; Zhang, J. Z. H.; Mei, Y. Calculations of Solvation Free Energy through Energy Reweighting from Molecular Mechanics to Quantum Mechanics. *J. Chem. Theory Comput.* **2016**, *12*, 499–511.

(42) Shen, L.; Wu, J.; Yang, W. Multiscale Quantum Mechanics/Molecular Mechanics Simulations with Neural Networks. *J. Chem. Theory Comput.* **2016**, *12*, 4934–4946.

(43) Dybeck, E. C.; König, G.; Brooks, B. R.; Shirts, M. R. Comparison of Methods To Reweight from Classical Molecular Simulations to QM/MM Potentials. *J. Chem. Theory Comput.* **2016**, *12*, 1466–1480.

(44) Kearns, F. L.; Hudson, P. S.; Woodcock, H. L.; Boresch, S. Computing Converged Free Energy Differences between Levels of Theory via Nonequilibrium Work Methods: Challenges and Opportunities. *J. Comput. Chem.* **2017**, *38*, 1376–1388.

(45) Wang, M.; Li, P.; Jia, X.; Liu, W.; Shao, Y.; Hu, W.; Zheng, J.; Brooks, B. R.; Mei, Y. Efficient Strategy for the Calculation of Solvation Free Energies in Water and Chloroform at the Quantum Mechanical/Molecular Mechanical Level. *J. Chem. Inf. Model.* **2017**, *57*, 2476–2489.

(46) Li, P.; Liu, F.; Jia, X.; Shao, Y.; Hu, W.; Zheng, J.; Mei, Y. Efficient Computation of Free Energy Surfaces of Diels-Alder Reactions in Explicit Solvent at Ab Initio QM/MM Level. *Molecules* **2018**, *23*, 2487.

(47) Li, P.; Jia, X.; Pan, X.; Shao, Y.; Mei, Y. Accelerated Computation of Free Energy Profile at ab Initio Quantum Mechanical/Molecular Mechanics Accuracy via a Semi-Empirical Reference Potential. I. Weighted Thermodynamics Perturbation. *J. Chem. Theory Comput.* **2018**, *14*, 5583–5596.

(48) Wang, M.; Mei, Y.; Ryde, U. Predicting Relative Binding Affinity Using Nonequilibrium QM/MM Simulations. *J. Chem. Theory Comput.* **2018**, *14*, 6613–6622.

(49) Hudson, P. S.; Boresch, S.; Rogers, D. M.; Woodcock, H. L. Accelerating QM/MM Free Energy Computations via Intramolecular Force Matching. *J. Chem. Theory Comput.* **2018**, *14*, 6327–6335.

(50) König, G.; Brooks, B. R.; Thiel, W.; York, D. M. On the Convergence of Multi-Scale Free Energy Simulations. *Mol. Simulat.* **2018**, *44*, 1062–1081.

(51) Wang, M.; Mei, Y.; Ryde, U. Host-guest Relative Binding Affinities at Density-Functional Theory Level from Semiempirical Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2019**, *15*, 2659–2671.

(52) Pan, X.; Li, P.; Ho, J.; Pu, J.; Mei, Y.; Shao, Y. Accelerated Computation of Free Energy Profile at ab initio Quantum Mechanical/Molecular Mechanical Accuracy via a Semi-Empirical Reference Potential. II. Recalibrating Semi-Empirical Parameters with Force Matching. *Phys. Chem. Chem. Phys.* **2019**, *21*, 20595–20605.

(53) Li, P.; Liu, F.; Shao, Y.; Mei, Y. Computational Insights into Endo/Exo Selectivity of Diels–Alder Reaction in Explicit Solvent at ab Initio Quantum Mechanical/Molecular Mechanical Level. *J. Phys. Chem. B* **2019**, *123*, 5131–5138.

(54) Hudson, P. S.; Woodcock, H. L.; Boresch, S. Use of Interaction Energies in QM/MM Free Energy Simulations. *J. Chem. Theory Comput.* **2019**, *15*, 4632–4645.

(55) Giese, T. J.; York, D. M. Development of a Robust Indirect Approach for MM → QM Free Energy Calculations That Combines Force-Matched Reference Potential and Bennett's Acceptance Ratio Methods. *J. Chem. Theory Comput.* **2019**, *15*, 5543–5562.

(56) Piccini, G.; Parrinello, M. Accurate Quantum Chemical Free Energies at Affordable Cost. *J. Phys. Chem. Lett.* **2019**, *10*, 3727–3731.

(57) Chung, S.; Choi, S. M.; Lee, W.; Cho, K. H.; Rhee, Y. M. Free Energy Level Correction by Monte Carlo Resampling with Weighted Histogram Analysis Method. *Chin. J. Chem. Phys.* **2020**, *33*, 183–195.

(58) Torrie, G. M.; Valleau, J. P. Nonphysical Sampling Distributions in Monte Carlo Free-energy Estimation: Umbrella Sampling. *J. Comput. Phys.* **1977**, *23*, 187–199.

(59) Shirt, M. R.; Chodera, J. D. Statistically Optimal Analysis of Samples from Multiple Equilibrium States. *J. Chem. Phys.* **2008**, *129*, 124105.

(60) Shirts, M. R. Reweighting from the Mixture Distribution as a Better Way to Describe the Multistate Bennett Acceptance Ratio. *arXiv.org* **2017**, 1704.00891, https://arxiv.org/abs/1704.00891.

(61) Ferrenberg, A. M.; Swendsen, R. H. Optimized Monte Carlo Data Analysis. *Phys. Rev. Lett.* **1989**, *63*, 1195–1198.

(62) Souaille, M.; Roux, B. Extension to the Weighted Histogram Analysis Method: Combining Umbrella Sampling with Free Energy Calculations. *Comput. Phys. Commun.* **2001**, *135*, 40–57.

(63) Gallicchio, E.; Andrec, M.; Felts, A. K.; Levy, R. M. Temperature Weighted Histogram Analysis Method, Replica Exchange, and Transition Paths. *J. Phys. Chem. B* **2005**, *109*, 6722–6731.

(64) Zwanzig, R. W. High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. *J. Chem. Phys.* **1954**, *22*, 1420.

(65) Boereboom, J. M.; Fleurat-Lessard, P.; Bulo, R. E. Explicit Solvation Matters: Performance of QM/MM Solvation Models in Nucleophilic Addition. *J. Chem. Theory Comput.* **2018**, *14*, 1841–1852.

(66) Stewart, J. J. P. Optimization of Parameters for Semi-Empirical Methods V: Modification of NDDO Approximations and Application to 70 Elements. *J. Mol. Model* **2007**, *13*, 1173–1213.

(67) Hu, W.; Li, P.; Wang, J.-N.; Xue, Y.; Mo, Y.; Zheng, J.; Pan, X.; Shao, Y.; Mei, Y. Accelerated Computation of Free Energy Profile at Ab Initio Quantum Mechanical/Molecular Mechanics Accuracy via a Semiempirical Reference Potential. 3. Gaussian Smoothing on Density-of-States. *J. Chem. Theory Comput.* **2020**, doi: 10.1021/acs.jctc.0c00794.

(68) Rasmussen, C.; Williams, C. *Gaussian Processes for Machine Learning*; MIT Press, 2006; pp 7–32.

(69) Jorgensen, W. L.; Chandresekhar, 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.

(70) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.

(71) Langevin, P. Sur la Théorie du Mouvement Brownien. *C. R. Acad. Sci. (Paris)* **1908**, *146*, 530–533.

(72) Klimovich, P. V.; Shirts, M. R.; Mobley, D. L. Guidelines for the Analysis of Free Energy Calculations. *J. Comput. Aid. Mol. Des.* **2015**, *29*, 397–411.

(73) Chai, J.-D.; Head-Gordon, M. Long-range Corrected Hybrid Density Functionals with Damped Atom-atom Dispersion Corrections. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615–6620.

(74) Case, D. A.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; Cheatham, T. E., III; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E.; Ghoreishi, D.; Gilson, M. K.; Gohlke, H.; Goetz, A. W.; Greene, D.; Harris, R.; Homeyer, N.; Izadi, S.; Kovalenko, A.; Kurtzman, T.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Mermelstein, D. J.; Merz, K. M.; Miao, Y.; Monard, G.; Nguyen, C.; Nguyen, H.; Omelyan, I.; Onufriev, A.; Pan, F.; Qi, R.; Roe, D. R.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C. L.; Smith, J.; Salomon-Ferrer, R.; Swails, J.; Walker, R. C.; Wang, J.; Wei, H.; Wolf, R. M.; Wu, X.; Xiao, L.; York, D. M.; Kollman, P. A. AMBER 18, University of California, San Francisco. 2018.

(75) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-

Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. Gaussian 16 Revision B.01. 2016; Gaussian Inc. Wallingford CT.

(76) Marenich, A. V.; Jerome, S. V.; Cramer, C. J.; Truhlar, D. G. Charge Model 5: An Extension of Hirshfeld Population Analysis for the Accurate Description of Molecular Interactions in Gaseous and Condensed Phases. *J. Chem. Theory Comput.* **2012**, *8*, 527–541.

(77) Pilmé, J.; Berthoumieux, H.; Robert, V.; Fleurat-Lessard, P. Unusual Bond Formation in Aspartic Protease Inhibitors: A Theoretical Study. *Chem. - Eur. J.* **2007**, *13*, 5388–5393.

Graphical TOC Entry