THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

Efficient Parallel Implementations of QM/MM-REMD (Quantum Mechanical/Molecular Mechanics-Replica-Exchange MD) and Umbrella Sampling: Isomerization of HO in Aqueous Solution

Dmitri G. Fedorov, Yuji Sugita, and Cheol Ho Choi

J. Phys. Chem. B, **Just Accepted Manuscript** • DOI: 10.1021/jp4029529 • Publication Date (Web): 11 Jun 2013

The Journal of Physical Chemistry B is published by the American Chemical Society.
1155 Sixteenth Street N.W., Washington, DC 20036

Efficient Parallel Implementations of QM/MM-REMD (Quantum Mechanical/Molecular Mechanics-Replica-exchange MD) and Umbrella Sampling: Isomerization of H_2O_2 in Aqueous Solution

Dmitri G. Fedorov[1], Yuji Sugita[2] and Cheol Ho Choi[3,]*

[1]NRI, National Institute of Advanced Industrial Science and Technology (AIST), Central 2, Umezono 1-1-1, Tsukuba, 305-8568, Japan; [2]RIKEN Advanced Science Institute, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan; [3]Department of Chemistry and Green-Nano Materials Research Center, College of Natural Sciences, Kyungpook National University, Taegu 702-701, South Korea

AUTHOR EMAIL ADDRESS: d.g.fedorov@aist.go.jp, sugita@riken.jp

RECEIVED DATE

TITLE RUNNING HEAD: Replica-exchange MD for QM/MM

[*] Corresponding author's E-mail: cchoi@knu.ac.kr, TEL: +82-53-950-5332

ABSTRACT. An efficient parallel implementation of QM/MM based replica-exchange molecular dynamics (REMD) as well as umbrella samplings techniques was proposed by adopting generalized distributed data interface (GDDI). Parallelization speed-up of 40.5 on 48 cores was achieved, making our QM/MM-MD engine a robust tool for studying complex chemical dynamics in solution. They were comparatively used to study the torsional isomerization of hydrogen peroxide in aqueous solution. All of results by QM/MM-REMD and QM/MM umbrella sampling techniques yielded nearly identical potentials of mean force (PMFs) regardless of the particular QM theories for solute, showing that the overall dynamics are mainly determined by solvation. Although the entropic penalty of solvent rearrangements exists in *cisoid* conformers, it was found that both strong intermolecular hydrogen bonding and dipole-dipole interactions preferentially stabilize them in solution, reducing the torsional free energy barrier at 0° by about 3 kcal/mol as compared to gas phase.

KEYWORDS: QM/MM MD, replica-exchange, umbrella sampling, molecular dynamics

I. Introductions

Classical molecular dynamics (MD) based on AMBER[1], CHARMM[2] , OPLS[3], and GROMACS[4] force fields has been very successful and become an indispensible tool in chemistry. However, these techniques are not directly applicable to chemical reactions, where electronic structure calculations are necessary to describe the formation and cleavage of chemical bonds. Although semiempirical[5] or *ab initio* Molecular Dynamics (AIMD) simulations such as Car-Parrinello MD (CPMD)[6], fragment molecular orbital (FMO) MD[7], Quickstep[8] and GPU-accelerated AIMD[9] provide the most direct tool for studying these events, AIMD simulations beyond ~100ps for systems with realistic size are still not practical.

To overcome this challenge, recently Choi et al.[10] have demonstrated that the hybrid Quantum Mechanical/Effective Fragment Potential (QM/EFP) approach[11] is a very efficient and accurate way of doing quantum mechanical molecular dynamics when it is properly combined with well-developed classical MD techniques. Slipchenko and co-workers[12,13] have also shown the advantages of EFP for quantum mechanical MD studies. EFP is an advanced polarizable force field with a charge transfer model, used similarly to other force fields in molecular mechanics (MM), so that QM/EFP is a particular case of a more general QM/MM.[14-17] QM/MM-MD at various QM levels and using both EFP and TIP5P[18] parameters has been also successfully applied for NaCl association/dissociation dynamics,[19] hydrophobic and hydrophilic association of methanol dimer[20], and anharmonic vibrational properties[21] in aqueous solutions. Especially in the study of anharmonic vibrational properties, TIP5P water model was also successfully adopted. One of the important terms of hybrid method is the coupling interactions between QM and MM. The effective Gaussian repulsion potentials similar to those of EFP1-HF model was used for the coupling interaction between QM and TIP5P. The detailed descriptions of our approach can be found in the previous papers. In short, the unique combination of the hybrid QM/EFP (or QM/TIP5P) and the traditional sampling techniques can be both practical and highly accurate route for

solvation dynamics. This approach comprises of QM-QM, QM-MM as well as MM-MM interaction terms, which are ingredients for the correct dynamic descriptions of solute in solvents. Current implementation allows us to perform beyond 1 ns QM/MM-MD simulations with 300 ~ 400 water molecules (1000 atoms) rather easily on a small-scale cluster, which are sufficient for a statistically appropriate samplings of pico second dynamics in solution.

Although QM/MM-MD is much faster than conventional AIMD, an enhanced sampling and efficient parallelization are desirable for the study of difficult chemical problems, requiring longer simulations and larger size. As an extension of our QM/MM-MD implementation, the replica-exchange molecular dynamics (REMD)[22] and umbrella sampling (US)[23] are parallelized in this work with the two level generalized distributed data interface (GDDI).[24] In the demonstration of the efficiency of our approach, we study the isomerization dynamics of H_2O_2 in aqueous solution.

Hydrogen peroxide plays an important role in various chemical processes such as reactions in atmosphere[25] and in biological environments[26]. Compared with its isoelectronic ethane, hydrogen peroxide exhibits a double well torsional energy curve where skew conformers are favored over *trans* and *cis* conformers. Clearly, the enhanced stabilizing n -> σ* negative hyperconjugation effect and destabilizing repulsion among lone pairs, complicates the conformational analysis. Furthermore, the torsional energy curves of hydrogen peroxide are dependent not only on hyperconjugative stabilization and steric repulsion but also on structural and electronic relaxations. The gas phase *ab initio* study showed that H_2O_2 has C_2 symmetry with a torsional angle of 112.5°, the OOH angle of 100°, and O-O bond distance of 145.4 pm.[27] Its full 6-variable PES has been computed at a high level of theory.[28] The CCSD(T)/CBS level leads to a energy barrier for the "*cis*" barrier of 7.13 kcal/mol[29] in agreement with the experimental value of 7.33 kcal/mol.[30] Song *et al.* [31] have found that that both the hyperconjugative and steric interactions lower the energy of skew conformers and eventually form low barriers from skew to *trans* conformers and high barriers from skew to *cis* conformers in H_2O_2. Gas phase H_2O_2-$(H_2O)_n$ complexes have been studied for $n \leq 3$ (Ref. 32) and $n \leq 6$ (Ref. 33).

Martins-Costa and Ruiz-Lopez[34] have performed a mixed quantum/classical molecular dynamics (QM/MM) simulation of H_2O_2 in water. They have noted that the H-O-O-H torsional angle distribution peaks around ±110° and essentially never attains the "*cisoid*" (near 0°) configuration; and the propensity for H_2O_2 to be 4-coordinated. They have shown that interconversion between energy minima always proceeds through a *transoid* transition state. However, proper samplings near *cisoid* configuration have not been obtained, since the interconversion via the *cisoid* conformer would have a much slower dynamics. In order to obtain a statistically meaningful torsional propensity and its related solvation dynamics of H_2O_2, enhanced sampling techniques are necessary.

The replica-exchange Monte Carlo (MC) method has been originally developed by Swendsen et al.[35] as a parallel tempering method. Independently, Geyer et al.[36] and Hukushima et al.[37] have developed a similar algorithm also in the Monte Carlo scheme, as multiple Markov chain MC method and replica-exchange method. Afterward, Sugita and Okamoto[22] introduced the velocity rescaling scheme for the MD version of the replica-exchange method and this is usually known as replica-exchange molecular dynamics or REMD. In this method, N copies of the original system at different temperatures are simulated simultaneously and independently in MD. Then, configurations at different temperatures are exchanged on the basis of the Metropolis criterion. The idea of this method is to make configurations at high temperatures available to the simulations at low temperatures and vice versa, yielding a very robust ensemble in which both low and high energy configurations are sampled. The configuration space sampled at the whole temperature range is made available to the replica with the target temperature (such as 300 K in this study), whose properties are then averaged over the whole trajectory. In this way, thermodynamical properties such as the specific heat, which is in general not accurately computed in the canonical ensemble, can be computed with high precision. The transition probability $w(X \to X')$ between the generalized-ensemble states X and X' can be written by the Metropolis criterion as,

$$w(X \to X') \equiv w\left(x_m^{[i]} \middle| x_n^{[j]}\right) = \begin{cases} 1, & \text{for } \Delta \le 0 \\ \exp(-\Delta), & \text{for } \Delta > 0 \end{cases}, \qquad (1)$$

where *i and j* represent replicas, and *m and n* label temperature, $\Delta = (\beta_n - \beta_m)(E_i - E_j)$ is the factor determining the exchange of replicas *i* and *j*, with the inverse temperatures β_m and β_n , respectively, where $\beta = 1/k_B T$. T is temperature and k_B is the Boltzmann constant. E_i is the potential energy of replica *i*, which depends on the coordinates x_i and momenta of all atoms.

Because each replica is independent until the temperature exchange, the parallelization of REMD itself appears rather straightforward. However, as compared to the classical MD computer programs such as GROMACS[4] which take the uniformity advantage of the system description (force fields are used for all atoms) and deliver high scalability on massively parallel computers, the parallelization of QM methods in general is a much more challenging task. More complications arise for QM/MM, because the parallelization strategies of QM and MM are very different, and the parallelization of QM/MM is made more difficult by a common scenario of using various add-on interfaces executing separate QM and MM programs. A modern, reasonably scalable QM/MM engine is thus apparently needed, which is developed in this work.

II. Parallelization and Computational Details

A. Parallelization of QM/MM-REMD and umbrella sampling

In order to achieve a high parallel scalability of QM/MM-REMD, we employed the generalized distributed data interface (GDDI)[24] in GAMESS,[38] which has been used to parallelize FMO[39,40] and divide-and-conquer (DC)[41] methods. In GDDI, all computer cores are divided into groups, and each group is assigned a computational task (the intergroup level of parallelization). Consequently, computations in each group are also parallelized (the intragroup level), so that GDDI is known as a hierarchical parallelization scheme, at present limited to two levels. Such parallelization is coarse-grained and the heuristic static load balancing (HSLB) [42] can be used to improve parallel efficiency.

In REMD, we assign one replica with a given temperature to one GDDI group. Because all replica

trajectories have an essentially identical computational cost, it generally suffices to use a simple static load balancing; however, in more complicated computer configurations when mixed hardware is used, for instance, in grid computing, one would have to either switch to HSLB or to dynamic load balancing. In our implementation, each GDDI group performs a single replica calculation for the whole duration of MD. Within the group, the parallelization is done for each geometry, by assigning various work loads such as integrals to different CPU cores. If the QM region is small, one cannot use many CPU cores efficiently for intragroup parallelization, and thus the two level GDDI scheme is a very efficient solution to utilize massively parallel computers.

The REMD-GDDI scheme is shown in Figure 1. The tests of parallel efficiency were performed on the Soroban cluster, consisting of 4 dual hexa-core 2.67 GHz Xeon nodes (48 CPU cores total) equipped with 24 GB RAM per node and connected by Gigabit Ethernet. Each group master writes out its own log and trajectory files, whose temperature varies along the trajectory. Next, we use GDDI to parallelize umbrella sampling (US) in MD, by performing one US window on one GDDI group. Within the group, the original parallelization of MD in GAMESS was used. This enables sampling all US windows in a single GDDI run, for example, one can scan a range of dihedral angles independently constrained with umbrella sampling, using multiple trajectories in one calculation. The parallelization of US is shown in Figure 1. Because MD trajectories running with different sets of constraints are completely independent and never exchange any data or have any synchronization points, GDDI implementation was straightforward and we do not describe it in more detail.

Finally, we summarize the factors which limit the parallel efficiency of our method. Firth of all, QM calculations never take exactly the same time, if the structure is different. This is because a different number of iterations may be required to converge the electronic state. In addition, there is integral screening and other techniques, whose efficiency may differ depending on the geometry. Therefore, the time required to do different replicas is not uniform, and, because there is a synchronization of all groups at some points to exchange data, some efficiency loss takes place. The other limiting factor is

that typically one does not use too many replicas. This means that we cannot use any arbitrarily large number of CPU cores without having to resort to very large GDDI groups, which would decrease the performance because the QM parallelization efficiency for the solute is not perfect. Thus, our present method is most useful for a small or medium parallel computer (PC cluster) with a few dozens or at most hundreds of CPU cores.

B. Details of MD simulations

A spherical system with a radius of 11 Å containing one H_2O_2 molecule (treated with QM) surrounded by 290 water molecules was prepared for both QM/MM-REMD and umbrella sampling (US) simulations. In order to prevent evaporation of water during long time simulations, we applied a harmonic restraint potential with the force constant of 2 kcal mol^{-1} $Å^{-2}$ for the boundary solvent molecules, following the implementation in CHARMM.[2] The center of mass of the solute was also constrained to remain in the center of the sphere with the force constant of 1 kcal mol^{-1} $Å^{-2}$. The QM molecule was described by HF/3-21G, HF/6-31G(d), MP2/6-31G(d) and MP2/aug-cc-pVDZ levels of theory in the umbrella sampling simulations. In the case of REMD, only HF/3-21G level of theory was used for QM molecule. The solvent waters were treated with TIP5P[18] for both REMD and US. The initial structure was equilibrated in MD for 10 *ps* at 300K. Subsequent, 1 ~ 2 *ns* production runs of REMD were performed with the temperatures between 300 ~ 700 K (for 32 replicas) and 300 ~ 1400 K (for 64 replicas).[43] The REMD temperatures were exchanged for every 100 fs.

The umbrella samplings along the torsional angle of H_2O_2 were performed with the harmonic constraint potential of

$$V(\tau) = \frac{1}{2} k(\tau - \tau_0)^2 . \qquad (2)$$

Simulations on seven windows of τ_0 =0, 30, 60, 90, 120, 150 and 180° were performed with the

force constant (k) of 10 kcal mol^{-1} Å^{-2}. The potentials of mean force (PMF) from both REMD and umbrella samplings were obtained using the weighted histogram analysis method (WHAM).[44] Initially, NVT runs of QM/MM-MD simulations over 10ps were performed on each of the seven windows at 300 K to equilibrate the systems. Then, 100 ps QM/MM-MD production runs for the NVT ensemble at 300K were continued from the final structures after the equilibration. We implemented umbrella sampling and replica-exchange molecular dynamics (or REMD) in GAMESS-US, extending the existing single trajectory MD code developed for EFP. [45,46] These improvements are now available in the distributed version of GAMESS.

III. Results and Discussions

A. The dynamics of H_2O_2 in aqueous solution

As described in the previous section, umbrella samplings along the torsional isomerization of hydrogen peroxide were performed with the hybrid QM/MM-MD. The QM hydrogen peroxide was treated at the HF/3-21G, HF/6-31G*, MP2/6-31G* and MP2/aug-cc-pVDZ levels of theory, while the TIP5P parameters were used for all water molecules. The obtained potentials of mean force (PMFs) along the torsional angle are presented in Figure 2a. The PMFs of three REMD runs and four US runs are nearly identical with each other. In the same figure, we also plotted the potential energy surface of MP2/aug-cc-pVDZ along with the same torsional angle. This suggests that the torsional barrier in solution is mostly determined by the detailed interplay between solute and solvent interactions rather than the solute dynamics alone.

As seen in the static gas phase potential energy curve obtained with a constrained MP2/aug-cc-pVDZ, geometry optimization, the major and minor torsional barriers appear at 0 and 180° in the free energy surfaces. However, in solution the major barrier height is reduced to about 4 kcal/mol and the torsion angle at the energy minimum is also reduced to 90°. Our results contradict the earlier QM/MM-

MD study by Martins-Costa and Ruiz-Lopez[34], in which the H-O-O-H torsional angle distribution peaks around ±110°. According to our MD results, the barrier ~ 4 kcal/mol at "*cisoid*" (near 0°) transition is too large so that any interconversion through the "*cisoid*" cannot occur within their 50 ps time scale. Consequently, our results demonstrate the importance of enhanced samplings for the correct statistical analysis.

In order to further analyze the solvation effects, the radial distribution functions (RDF) $g(r)$ between the hydrogen peroxide oxygen and water oxygen were computed and the results are presented in Figure 2b. The $g(r)$ at 0, 90 and 180° umbrella sampling windows were computed using the MP2/6-31G*//TIP5P results. These RDFs show the average distribution of water molecules around the solute at the corresponding torsional angles of hydrogen peroxide. The peaks at 2.9 Å increase in the order of 90 (minimum), 180 (*transoid*) and 0° (*cisoid*) torsional angle, indicating that more ordered first hydration shell appears in the main torsional barrier. As a result, the *cisoid* transition state is the least favorable entropically. The hydrogen bonding between the solute and solvent molecules is one of the main solvent effects.

To investigate the binding in detail, *ab initio* calculations with MP2/aug-cc-pVTZ level of theory were performed on the H_2O_2—H_2O complex and the results are shown in Figure 3. The torsion angle of the *cis* conformer (3b) was fixed to 0 during geometry optimizations. While the minimum structure of Figure 3a has two typical intermolecular hydrogen bonds, the *cis* conformer has an uncommon dual intermolecular hydrogen bonding. The corresponding intermolecular bindings are also seen in MD trajectories as major species. The complex formation stabilization energies for the structures shown in Figures 3a and 3b are 7.5 and 9.2 kcal/mol, respectively, showing the preferential enthalpic stabilization of the *cis* conformer by additional water. Although this simple H_2O_2—H_2O complex cannot represent the dynamic situations in solution, it can certainly show that the direct intermolecular binding strength between the solute and solvent depends on the solute's torsional angle.

In addition to the hydrogen bonding, dipole-dipole interactions between the solute and solvent are

also important solvation effects, which are substantial at long distances. The gas phase and solution dipole moments of hydrogen peroxides as a function of solute's torsional angle are comparatively presented in Figure 4. The gas phase dipole moments were obtained by constraint optimizations at the MP2/aug-cc-pVDZ level, while the solution dipole moments are the averaged dipole moments of each umbrella sampling windows. The dipole moment enhancements by solvation are clearly seen. The maximum of the dipole moment is at 0°, indicating that the *cisoid* conformers in solution is more stabilized by the strong dipole-dipole interactions. In short, although the solvent entropy is reduced, both the strong intermolecular hydrogen bonding and dipole-dipole interactions preferentially stabilize the *cisoid* hydrogen peroxide conformers in solution, reducing the torsional barrier at 0° relative to the gas phase value.

In addition to the torsional angle, the O—O bond distance and H—O—O angle of hydrogen peroxide are also important molecular parameters in solution. Two-dimensional PMFs were obtained on the basis of one-dimensional umbrella sampling windows and the results are presented in Figure 5. Figure 5a shows the 2D PMF between the O—O bond distance and torsion, which shows a minimum O—O bond distance at 1.4 Å, being 0.5 Å shorter than that in gas phase. On the other hand, the minimum H—O—O angle appears at 104 °, which is 4 ° larger than the gas phase experimental value (Figure 5b). These show how solvation affects the molecular geometries of the solute.

B. The performance of QM/MM-REMD

Compared to umbrella sampling with which one can selectively study particular regions of the free energy surface including highly unfavorable reaction barriers or transition states, REMD relies on the high temperature replicas for overcoming reaction barriers. The two methods have their own merits. The umbrella sampling is much more efficient, if the reaction coordinates can be uniquely defined such as the torsional angle of hydrogen peroxide in this study. However, it is not always the case: for example,

when one has to find out various conformers in a complex system, in which a unique reaction coordinate cannot be easily defined. For such a situation, REMD is more effective.

The PMFs of production REMD simulations on hydrogen peroxide dynamics are shown in Figure 2. The umbrella sampling and REMD results agree well with each other. The midpoint of the main barrier at dihedral angle τ=0 is around 4 kcal/mol, differing less than 1 kcal/mol between all MD levels including US results. Both doubling the number of replicas and doubling the MD duration has a small effect on the results, indicating that 1*ns* simulation with 32 replicas is sufficient. It is generally seen that the free energy surface near τ=0 by umbrella samplings is much more smooth, which is due to the unique torsional reaction coordinate during the simulations. On the other hand, despite the fact that REMD does not assume any particular reaction coordinate, it produces remarkably accurate PMFs.

The REMD parallelization results are shown in Figure 6. In order to show the importance of the two-level parallelization, we performed a REMD calculation on 48 CPU cores, divided into 1, 2, 4, 8, 12, 24 and 48 GDDI groups of equal size. Each group did a separate trajectory, but all used the same temperature of 300 K. The length of each trajectory was set in such a way that the total duration of all trajectories was 480 fs. For instance, for 1 GDDI group there was a single 480 fs trajectory, and for 48 groups there were 48 trajectories of 10 fs. The former case corresponds to a one-layer parallelization, and it has the worst performance: the total wall-clock timing is 109.4 sec (Figure 6a). The best performance is achieved for 48 groups: the timing is 14.1 sec, and this is an acceleration by the factor of 7.8 achieved by using coarse grained parallelization (relatively to the plain parallelization on the same number of CPU cores).

Next, the speed-up data showing the effect of the number of CPU cores per group are shown in Figure 6b for a 10 fs trajectory at 300 K, extracted from the MD results displayed in Figure 6a. It can be seen that the calculation is essentially saturated at 24 cores per group. Increasing the group size further leads to no noticeable improvement in the speed-up. This is related to the fact that the underlying QM system is very small, just 4 atoms.

The total speed-up is shown in Figure 6c, obtained for the cumulative REMD trajectory of 480 fs on 1-48 CPU cores, divided 1 CPU core per GDDI group. Although the value of 40.5 on 48 CPU cores may seem modest, the reason for this is that we used multiple core CPUs (dual hexa-core CPUs), so that running just 1 process on 1 node very efficiently uses CPU cache and memory bus, whereas running 12 compute processes per node incurs some overhead. Another factor reducing the scaling is the disk I/O. MD simulations write on disk quite substantial amounts of data. On our hardware platform we had only 2 HDDs per node, so that in the case of 12 cores running MD per node, 12 compute processes wrote to disk nearly simultaneously.

IV. Conclusions

An efficient parallel QM/MM-REMD engine by two-level hierarchical approach (GDDI).[24] is now available within GAMESS program, which extends the applicability of QM/MM MD to the more challenging chemical dynamics in condensed phase. The integrated and parallelized QM/MM-REMD package within GAMESS developed in this work is unique and should be useful for many future applications. The free energy surfaces along the HOOH torsional angle of 1 ns simulation by QM/MM-REMD with 32 replicas are nearly identical to those of 100 ps simulations of QM/MM MD with 7 umbrella sampling windows, proving that these two independent sampling methods are essentially converged in sampling the configurational space.

According to our converged free energy surfaces, the static gas phase rotation barrier of H_2O_2 isomerization at the MP2/aug-cc-pVDZ level is reduced from about 7.7 kcal/mol in gas phase (gas phase experimental value:[30] 7.33 kcal/mol) to about 4.0 kcal/mol in solution, which were attributed to the factors leading to the preferential stabilization of the *cisoid* isomers: stronger hydrogen bonding as seen by the peaks of the radial distribution functions and the dipole-dipole interactions between the solute and solvent, as seen by the magnitude of the dipole moment.

The performance of replica-exchange MD and umbrella sampling techniques for the hybrid QM/MM has been enhanced by coarse grained parallelization, so that individual replicas in REMD or windows in umbrella sampling can be run on independent groups of CPU cores. We have achieved the parallelization speed-up of 40.5 on 48 cores, making our QM/MM-MD engine a robust tool for studying chemical dynamics in solution on medium-sized parallel computers such as PC clusters.

The developed REMD engine is general so that it can be extended easily to FMO-MD[7] or DC-MD.[41] FMO in particular is already highly parallelized with the two-level hierarchical approach (GDDI).[24] Adding another dimension (multiple replicas or windows) and developing a three-level hierarchical parallelization (all nodes divided into companies, companies divided into groups, and groups into nodes), one can take the advantage of peta and even exascale parallel computers.

ACKNOWLEDGMENTS. This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No. 2007-0056341 and No. 2012-0004821). DGF thanks the Next Generation SuperComputing Project, Nanoscience Program and Strategic Programs for Innovative Research (MEXT, Japan) for the financial support.

REFERENCES

(1) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber Biomolecular Simulation Programs. *J Comput Chem* **2005**, *26*, 1668–1688.

(2) Mackerell, A. D., Jr; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J. A.; Guo, H.; Ha, S. A. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J Phys Chem B* **1998**, *102*, 3586–3616.

(3) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.

(4) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. GROMACS: Fast, Flexible, and Free. *J Comput Chem* **2005**, *26*, 1701–1718.

(5) Xie, W.; Orozco, M.; Truhlar, D. G.; Gao, J. X-Pol Potential: An Electronic Structure-Based Force Field for Molecular Dynamics Simulation of a Solvated Protein in Water. *J. Chem. Theory Comput.* **2009**, *5*, 459–467.

(6) Car, R.; Parrinello, M. Unified Approach for Molecular Dynamics and Density-Functional Theory. *Phys. Rev. Lett.* **1985**, *55*, 2471–2474.

(7) Komeiji, Y.; Nakano, T.; Fukuzawa, K.; Ueno, Y.; Inadomi, Y.; Nemoto, T.; Uebayasi, M.; Fedorov, D. G.; Kitaura, K. Fragment Molecular Orbital Method: Application to Molecular Dynamics Simulation, "Ab Initio FMO-MD." *Chemical Physics Letters* **2003**, *372*, 342–347.

(8) Vandevondele, J.; Krack, M.; Mohamed, F.; Parrinello, M.; Chassaing, T.; Hutter, J. Quickstep: Fast and Accurate Density Functional Calculations Using a Mixed Gaussian and Plane Waves Approach. *Computer*

Physics Communications **2005**, *167*, 103–128.
(9) Ufimtsev, I. S.; Martinez, T. J. Quantum Chemistry On Graphical Processing Units. 3. Analytical Energy Gradients, Geometry Optimization, and First Principles Molecular Dynamics. *J. Chem. Theory Comput.* **2009**, *5*, 2619–2628.
(10) Choi, C. H.; Re, S.; Feig, M.; Sugita, Y. Quantum Mechanical/Effective Fragment Potential Molecular Dynamics (QM/EFP-MD) Study on Intra-Molecular Proton Transfer of Glycine in Water. *Chemical Physics Letters* **2012**, *539-540*, 218–221.
(11) Gordon, M. S.; Fedorov, D. G.; Pruitt, S. R.; Slipchenko, L. V. Fragmentation Methods: A Route to Accurate Calculations on Large Systems. *Chem. Rev.* **2012**, *112*, 632–672.
(12) Hands, M. D.; Slipchenko, L. V. Intermolecular Interactions in Complex Liquids: Effective Fragment Potential Investigation of Water– Tert-Butanol Mixtures. *J Phys Chem B* **2012**, *116*, 2775–2786.
(13) Kosenkov, D.; Slipchenko, L. V. Solvent Effects on the Electronic Transitions of P-Nitroaniline: A QM/EFP Study. *J Phys Chem A* **2011**, *115*, 392–401.
(14) 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.
(15) Gao, J. A Molecular-Orbital Derived Polarization Potential For Liquid Water. *J Chem Phys* **1998**, *109*, 2346–2354.
(16) Lin, H.; Truhlar, D. G. QM/MM: What Have We Learned, Where Are We, And Where Do We Go from Here? *Theoretical Chemistry Accounts: Theory, Computation, And Modeling (Theoretica Chimica Acta)* **2007**, *117*, 185–199.
(17) 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.
(18) Mahoney, M. W.; Jorgensen, W. L. A Five-Site Model for Liquid Water and the Reproduction of the Density Anomaly By Rigid, Nonpolarizable Potential Functions. *J Chem Phys* **2000**, *112*, 8910–8922.
(19) Ghosh, M. K.; Re, S.; Feig, M.; Sugita, Y.; Choi, C. H. Interionic Hydration Structures Of NaCl in Aqueous Solution: A Combined Study of Quantum Mechanical Cluster Calculations and QM/EFP-MD Simulations. *J Phys Chem B* **2013**, *117*, 289–295.
(20) Ghosh, M. K.; Uddin, N.; Choi, C. H. Hydrophobic and Hydrophilic Associations of a Methanol Pair in Aqueous Solution. *J Phys Chem B* **2012**, *116*, 14254–14260.
(21) Ghosh, M. K.; Lee, J.; Choi, C. H.; Cho, M. Direct Simulations of Anharmonic Infrared Spectra Using Quantum Mechanical/Effective Fragment Potential Molecular Dynamics (Qm/Efp-Md): Methanol in Water. *J Phys Chem A* **2012**, *116*, 8965–8971.
(22) Sugita, Y.; Okamoto, Y. Replica-Exchange Molecular Dynamics Method for Protein Folding. *Chemical Physics Letters* **1999**, *314*, 141–151.
(23) Torrie, G. M.; Valleau, J. P. Nonphysical Sampling Distributions in Monte Carlo Free-Energy Estimation: Umbrella Sampling. *Journal Of Computational Physics* **1977**, *23*, 187–199.
(24) Fedorov, D. G.; Olson, R. M.; Kitaura, K.; Gordon, M. S.; Koseki, S. A New Hierarchical Parallelization Scheme: Generalized Distributed Data Interface (GDDI), and an Application to the Fragment Molecular Orbital Method (FMO). *J Comput Chem* **2004**, *25*, 872–880.
(25) Vione, D.; Maurino, V.; Minero, C.; Pelizzetti, E. The Atmospheric Chemistry of Hydrogen Peroxide: A Review. *Ann Chim* **2003**, *93*, 477–488.
(26) Stone, J. R.; Yang, S. Hydrogen Peroxide: A Signaling Messenger. *Antioxidants & Redox Signaling* **2006**, *8*, 243–270.
(27) Koput, J. An Ab Initio Study On the Equilibrium Structure and Torsional Potential Energy Function of Hydrogen Peroxide. *Chemical Physics Letters* **1995**, *236*, 516–520.
(28) Koput, J.; Carter, S.; Handy, N. C. Potential Energy Surface and Vibrational-Rotational Energy Levels of Hydrogen Peroxide. *J Phys Chem A* **1998**, *102*, 6325–6330.
(29) Halpern, A. M.; Glendening, E. D. An Intrinsic Reaction Coordinate Calculation of the Torsion-Internal Rotation Potential of Hydrogen Peroxide and Its Isotopomers. *J Chem Phys* **2004**, *121*, 273–279.
(30) Flaud, J.-M.; Camy-Peyret, C.; Johns, J.; Carli, B. The Far Infrared Spectrum Of H_2O_2. First Observation of the Staggering of the Levels and Determination of the cis Barrier. *J Chem Phys* **1989**, *91*, 1504–1510.
(31) Song, L.; Liu, M.; Wu, W.; Zhang, Q.; Mo, Y. Origins of Rotational Barriers in Hydrogen Peroxide and Hydrazine. *J. Chem. Theory Comput.* **2005**, *1*, 394–402.
(32) Ju, X. H.; Xiao, J. J.; Xiao, H. M. Theoretical Study on Intermolecular Interactions and Thermodynamic Properties of Water–Hydrogen Peroxide Clusters. *J Mol Struc-Theochem* **2003**, *626*, 231–238.
(33) Kulkarni, A. D.; Pathak, R. K.; Bartolotti, L. J. Structures, Energetics, and Vibrational Spectra Of $H_2O_2\cdots$(H

$_2O)_n$, N= 1−6 Clusters: Ab Initio Quantum Chemical Investigations. *J Phys Chem A* **2005**, *109*, 4583–4590.
(34) Martins-Costa, M. T. C.; Ruiz-López, M. F. Molecular Dynamics Of Hydrogen Peroxide in Liquid Water Using A Combined Quantum/Classical Force Field. *Chem Phys* **2007**, *332*, 341–347.
(35) Swendsen, R. H.; Wang, J.-S. Replica Monte Carlo Simulation Of Spin-Glasses. *Phys. Rev. Lett.* **1986**, *57*, 2607–2609.
(36) Geyer, C. J. (1991) In *Computing Science And Statistics*, Proceedings Of the 23rd Symposium on the Interface, American Statistical Association, New York; P. 156.
(37) Hukushima, K.; Nemoto, K. Exchange Monte Carlo Method and Application to Spin Glass Simulations. *J. Phys. Soc. Jpn.* **1996**, *65*, 1604–1608.
(38) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. General Atomic and Molecular Electronic Structure System. *J Comput Chem* **1993**, *14*, 1347–1363.
(39) Kitaura, K.; Ikeo, E.; Asada, T.; Nakano, T.; Uebayasi, M. Fragment Molecular Orbital Method: an Approximate Computational Method for Large Molecules. *Chemical Physics Letters* **1999**, *313*, 701–706.
(40) Fedorov, D. G.; Nagata, T.; Kitaura, K. Exploring Chemistry with the Fragment Molecular Orbital Method. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7562–7577.
(41) Katouda, M.; Kobayashi, M.; Nakai, H.; Nagase, S. Two-Level Hierarchical Parallelization of Second-Order Møller-Plesset Perturbation Calculations in Divide-And-Conquer Method. *J Comput Chem* **2011**, *32*, 2756–2764.
(42) Alexeev, Y.; Mahajan, A.; Leyffer, S.; Fletcher, G. D.; Fedorov, D. *Proc. Supercomputing 2012*, Ieee Computer Society, Salt Lake City, 2012.
(43) Van Der Spoel, D.; Seibert, M. Protein Folding Kinetics and Thermodynamics from Atomistic Simulations. *Phys. Rev. Lett.* **2006**, *96*, 238102.
(44) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. Multidimensional Free-Energy Calculations Using The Weighted Histogram Analysis Method. *J Comput Chem* **1995**, *16*, 1339–1350.
(45) Netzloff, H. M.; Gordon, M. S. The Effective Fragment Potential: Small Clusters and Radial Distribution Functions. *J Chem Phys* **2004**, *121*, 2711–2714.
(46) Netzloff, H. M.; Gordon, M. S. Fast Fragments: The Development of a Parallel Effective Fragment Potential Method. *J Comput Chem* **2004**, *25*, 1926–1935.

Figure Captions

Figure 1: Parallelization scheme for REMD and MD/US. Each replica in REMD and window in MD/US is run by a separate group of CPU cores. In REMD, the temperature T of each replica is adjusted every K time steps (we used K=100), represented by the crossed rectangle, at which point all groups exchange current temperatures T_I^i and energies of each replica I, and a new set T_I^{i+1} is determined. In MD/US each group of CPUs runs a separate window independently without any communications to other groups. CPU cores are shown as black filled circles enclosed in rectangles representing GDDI groups.

Figure 2: (a) Potentials of Mean Force (PMF) as obtained by umbrella sampling techniques as well as from 1 ns, 2ns replica exchange (RHF/3-21G) with 32 replicas, and 1ns simulation with 64 replicas. Static potential energy surface as obtained with a constrained MP2/aug-cc-pVDZ geometry optimization is also shown. (b) The radial distribution functions, $g(r)$ as a function of the $O(H_2O_2)$-$O(H_2O)$ distance. They were obtained at 0, 90 and 180° umbrella sampling windows at the MP2/6-31G*//TIP5P level.

Figure 3: Optimized structures of H_2O_2—H_2O as obtained with MP2/aug-cc-pVTZ level of theory. (a) The global minimum structure. (b) The complex of *cis* conformer with C_{2v} symmetry.

Figure 4: The gas phase and solution dipole moments of hydrogen peroxides as a function of solute's torsional angle. The gas phase dipole moments were obtained by constraint optimizations at the MP2/aug-cc-pVDZ level, while the solution dipole moments are the averaged dipole moments of each umbrella sampling windows of the MP2/6-31G*//TIP5P results.

Figure 5: (a) 2D PMF as a function of the O—O bond distance and the torsional angle, and (b) 2D PMF as a function of the H—O—O angle and the torsional angle. In order to generate 2D PMF, the all seven windows were used with the zero umbrella sampling force constant for the additional dimension.

Figure 6. Parallel REMD efficiency for QM/MM at the RHF/3-21G/TIP5P level. a) Dependence of the wall-clock timing on the number of GDDI groups, for 48 CPU cores and 480 fs (combining all trajectories). b) Dependence of the computational speed-up on the group size (the number of CPU cores), for a single trajectory of 10 fs. c) Total speed-up for the total trajectory of 480 fs (each replica is run on a single CPU core).

Figure 1

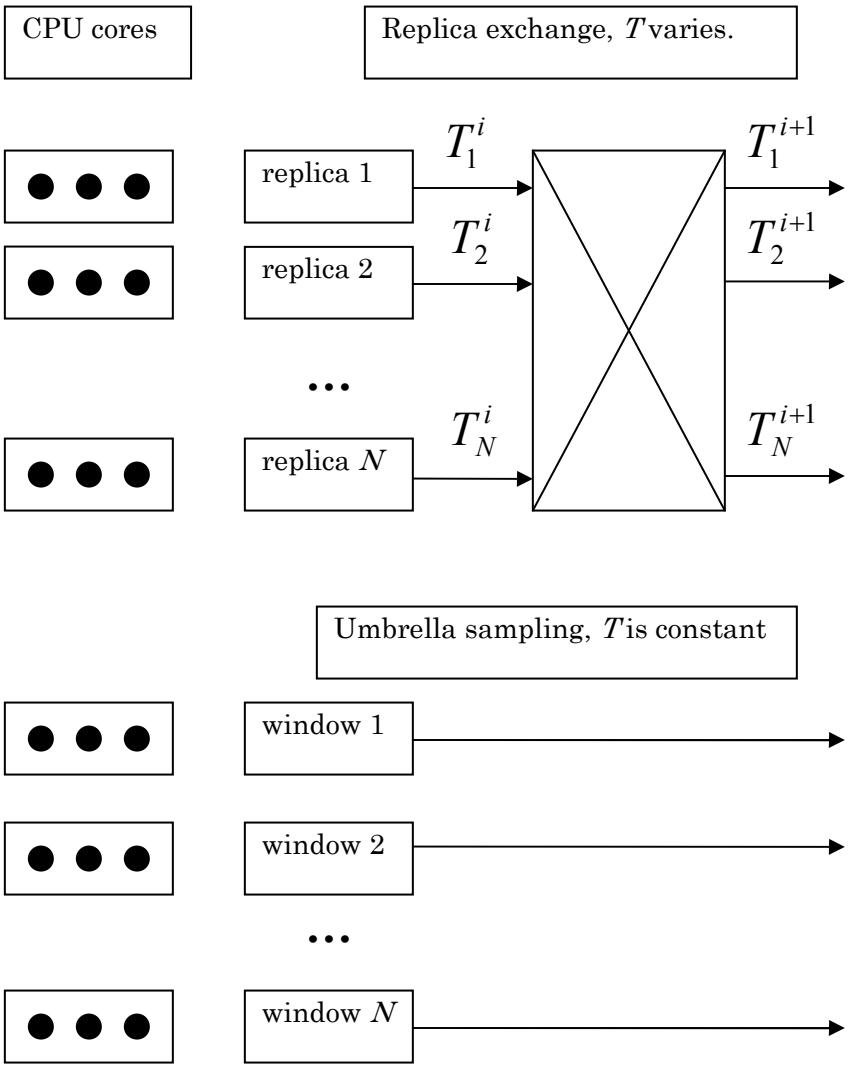
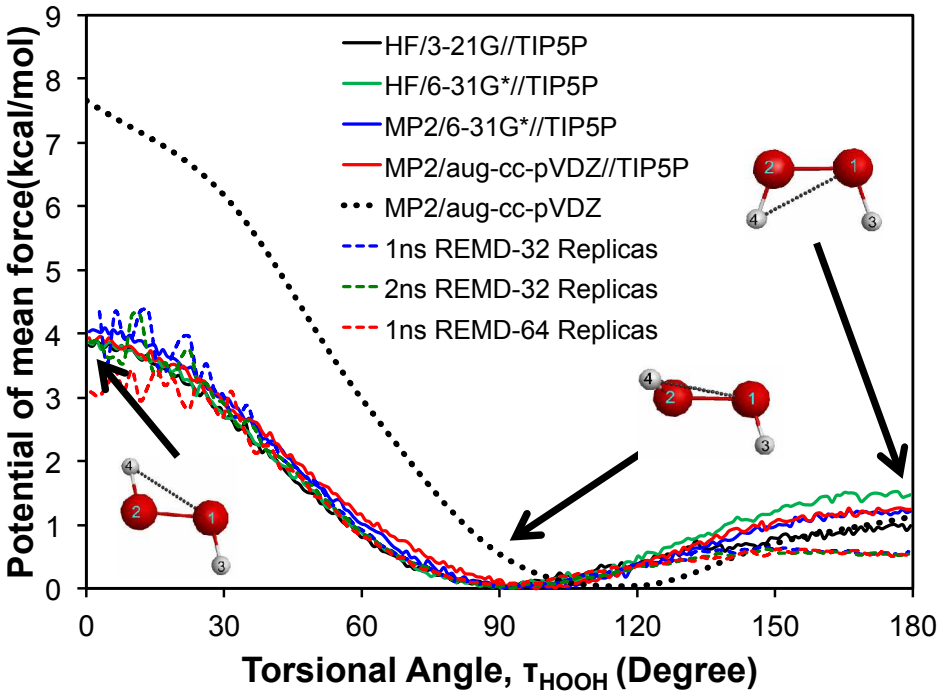

CPU cores
Replica exchange, T varies.
replica 1
replica 2
replica N
T_1^i
T_2^i
T_N^i
T_1^{i+1}
T_2^{i+1}
T_N^{i+1}
Umbrella sampling, T is constant
window 1
window 2
window N

Figure 2

a


HF/3-21G//TIP5P
HF/6-31G*//TIP5P
MP2/6-31G*//TIP5P
MP2/aug-cc-pVDZ//TIP5P
MP2/aug-cc-pVDZ
1ns REMD-32 Replicas
2ns REMD-32 Replicas
1ns REMD-64 Replicas
Potential of mean force(kcal/mol)
Torsional Angle, τ_{HOOH} (Degree)
0
1
2
3
4
5
6
7
8
9
0
30
60
90
120
150
180

b

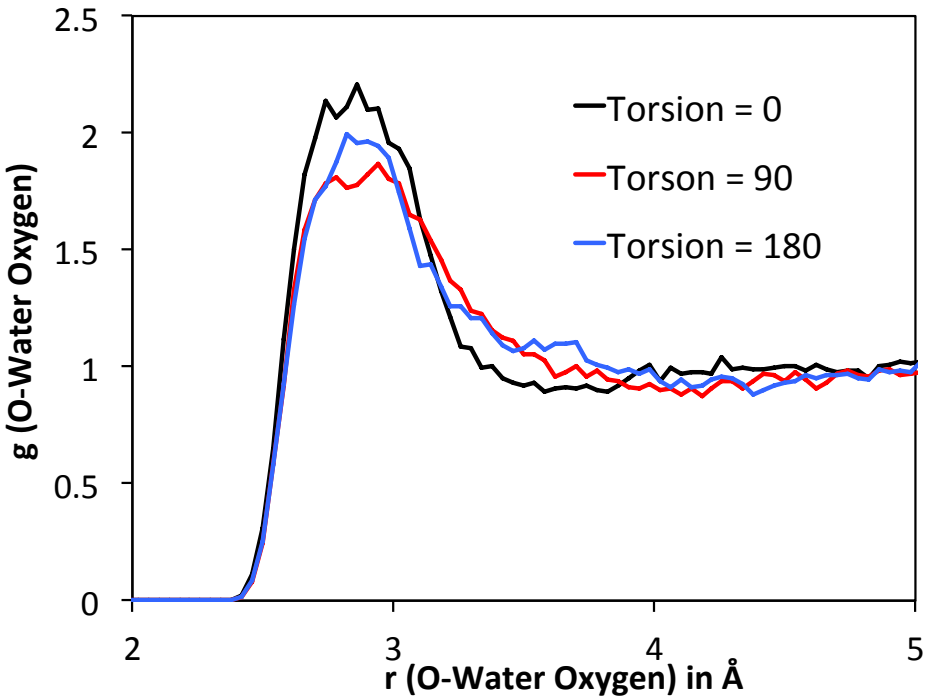

Torsion = 0
Torson = 90
Torsion = 180
g (O-Water Oxygen)
r (O-Water Oxygen) in Å
0
0.5
1
1.5
2
2.5
2
3
4
5

Figure 3

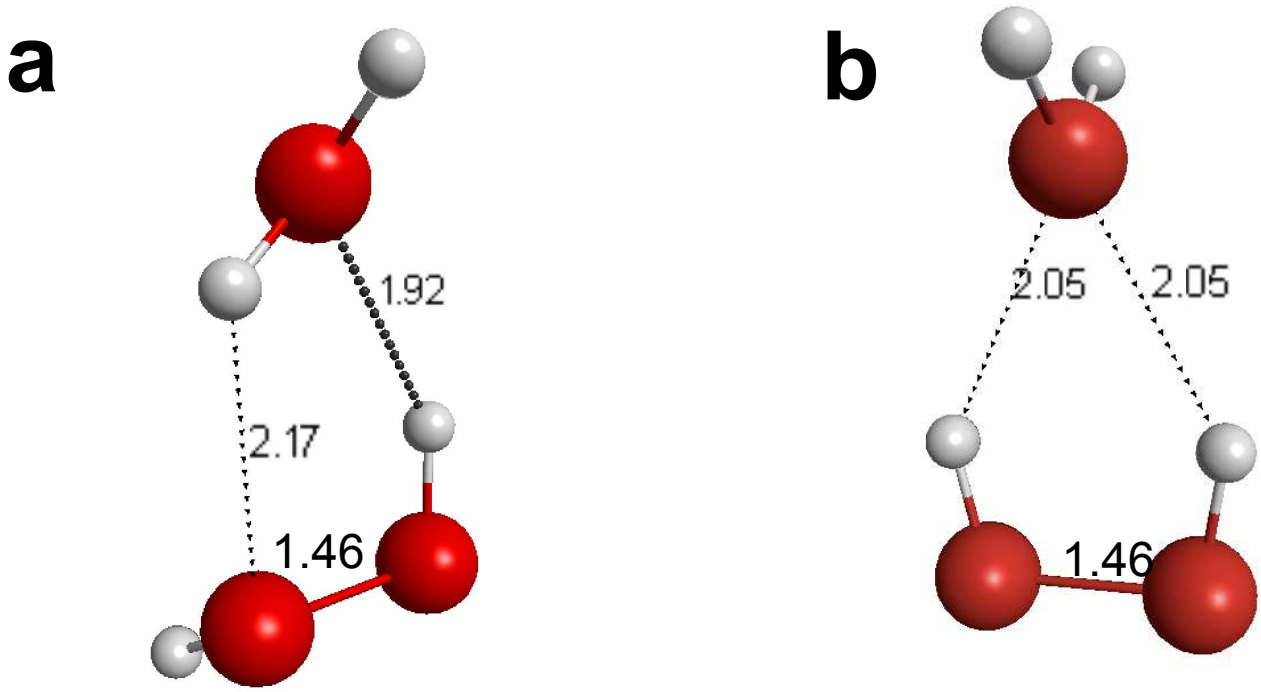


Figure 4

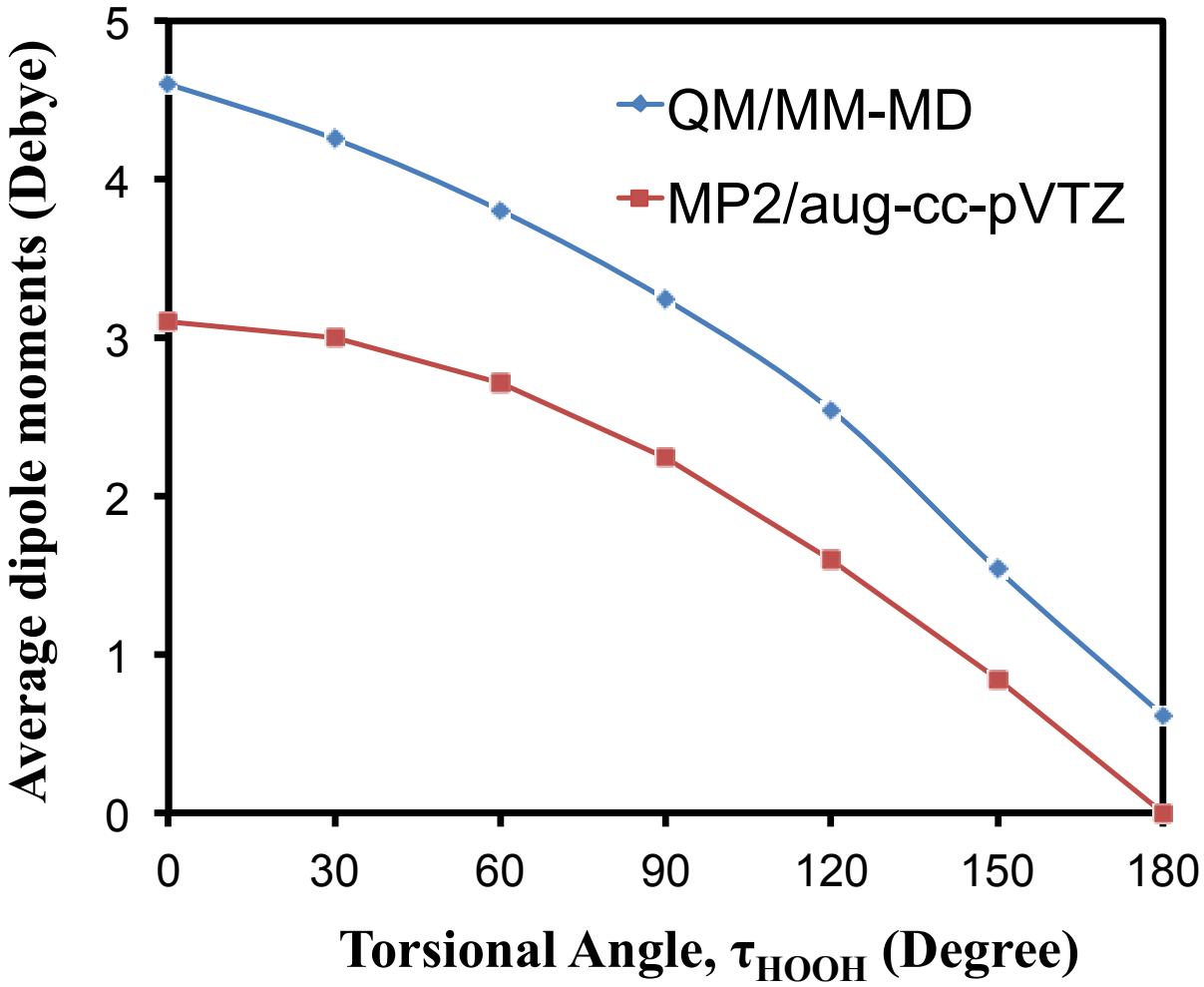

Average dipole moments (Debye)
0
1
2
3
4
5
0
30
60
90
120
150
180
Torsional Angle, τ_{HOOH} (Degree)
QM/MM-MD
MP2/aug-cc-pVTZ

Figure 5

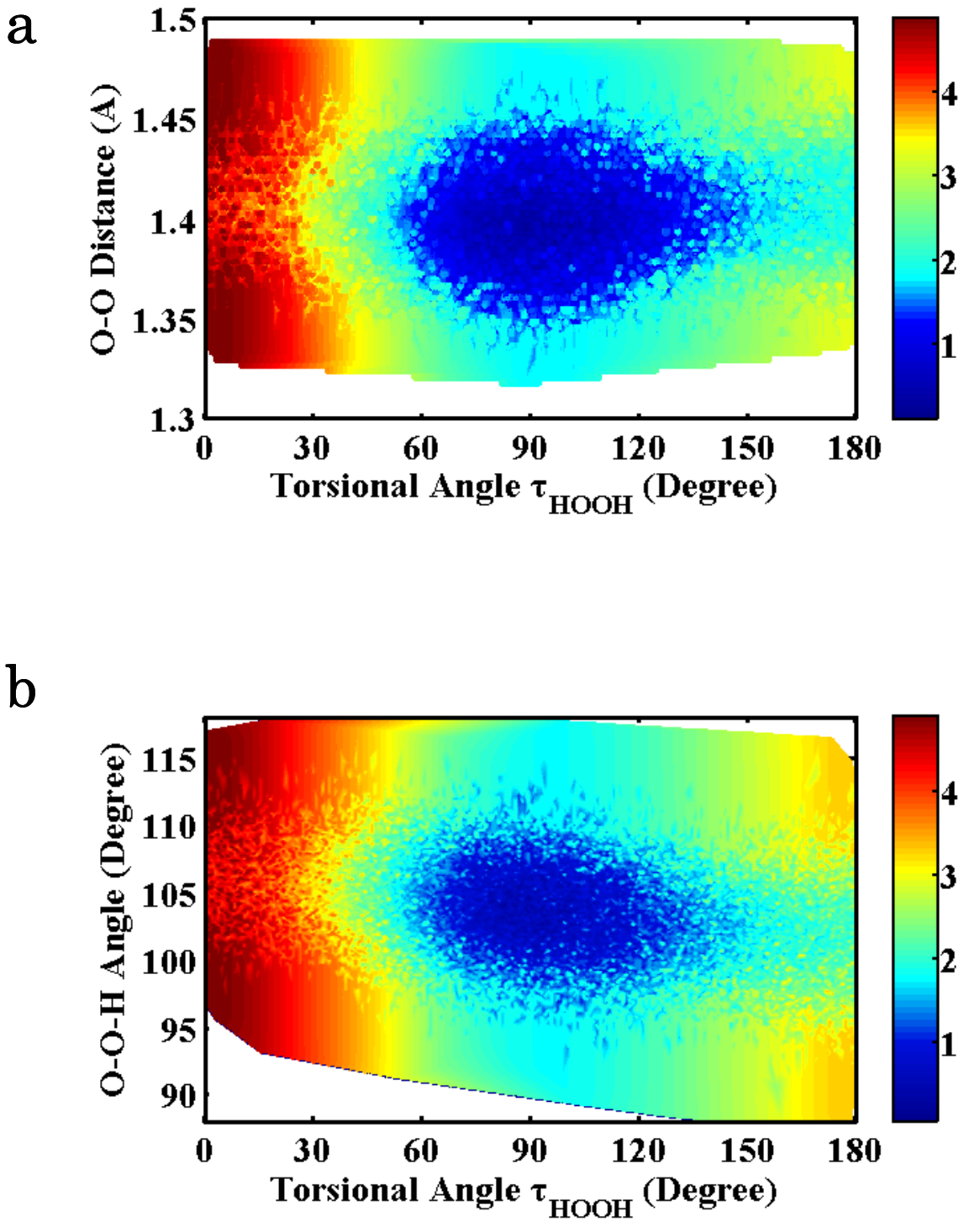
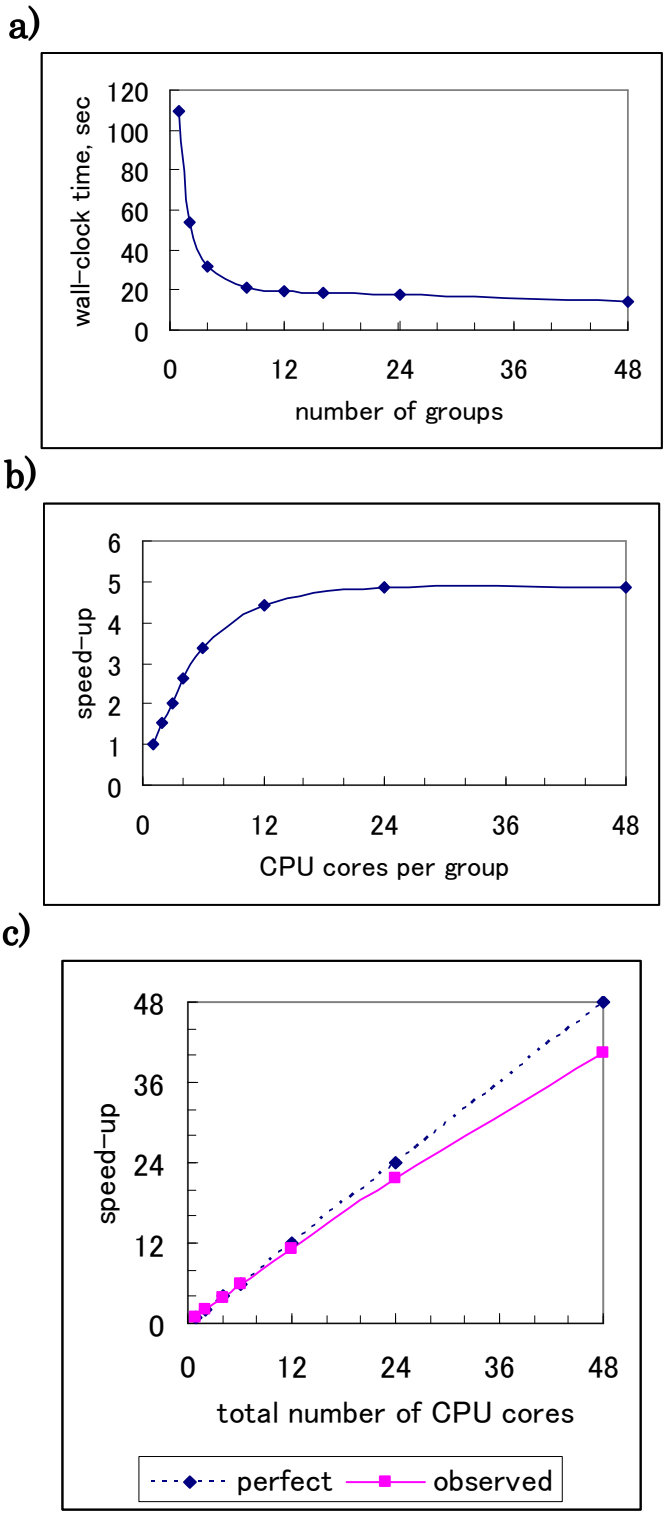
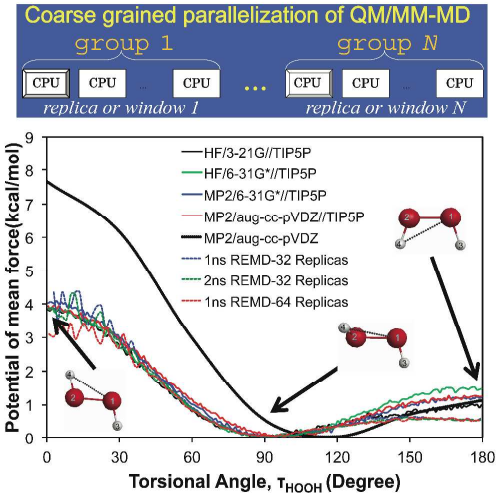

a
O-O Distance (Å)
1.5
1.45
1.4
1.35
1.3
0
30
60
90
120
150
180
Torsional Angle τ HOOH (Degree)
4
3
2
1
b
O-O-H Angle (Degree)
115
110
105
100
95
90
0
30
60
90
120
150
180
Torsional Angle τ HOOH (Degree)
4
3
2
1

Figure 6


a)
wall-clock time, sec
120
100
80
60
40
20
0
0
12
24
36
48
number of groups
b)
speed-up
6
5
4
3
2
1
0
0
12
24
36
48
CPU cores per group
c)
speed-up
48
36
24
12
0
0
12
24
36
48
total number of CPU cores
perfect
observed

TOC graphic


Coarse grained parallelization of QM/MM-MD
group 1
group N
CPU
CPU
CPU
CPU
CPU
CPU
replica or window 1
replica or window N
HF/3-21G//TIP5P
HF/6-31G*//TIP5P
MP2/6-31G*//TIP5P
MP2/aug-cc-pVDZ//TIP5P
MP2/aug-cc-pVDZ
1ns REMD-32 Replicas
2ns REMD-32 Replicas
1ns REMD-64 Replicas
Potential of mean force(kcal/mol)
Torsional Angle, τHOOH (Degree)