# Supporting Information: Biomolecular QM/MM Simulations: What are some of the “Burning Issues”?

Qiang Cui,[*,†] Tanmoy Pal,[‡] and Luke Xie[‡]

[†]*Departments of Chemistry, Physics and Biomedical Engineering, Boston University, 590 Commonwealth Avenue, Boston, MA 02215*

[‡]*Department of Chemistry, Boston University, 590 Commonwealth Avenue Boston, MA 02215*

E-mail: qiangcui@bu.edu,Tel:(+1)-617-353-6189

## 1 Computational Detail for the Two-body and Three-Body Energy Calculations

The total energy of N interacting water molecules can be expressed using the many-body expansion.[1]

$$E(1,\ldots,N) = \sum_{i}^{N} V^{1B}(i) + \sum_{i<j}^{N} V^{2B}(i,j) + \sum_{i<j<k}^{N} V^{3B}(i,j,k) + \ldots + V^{NB}(1,\ldots,N) \tag{1}$$

Two-body energies $V^{2B}$ are calculated by subtracting the total energies of monomers from the dimer energy.

$$V^{2B} = E(1,2) - \sum_{i=1}^{2} E(i) \tag{2}$$

Three-body energies are obtained by subtracting the pairwise dimer energies from the trimer energy and adding monomer energies to the result.

$$V^{3B} = E(1,2,3) - \sum_{i<j}^{3} E(i,j) + \sum_{i=1}^{3} E(i) \tag{3}$$

The database of water molecules were taken from Medders et al.[2] It includes geometries of 1349 water dimers and 500 water trimers, as well as their two-body and three-body interaction energies, respectively, calculated using BSSE corrected[3] CCSD(T)/aug-cc-pVTZ.[4,5] Single point energy calculations at DFTB2, DFTB3, DFTB3+D3, DFTB3+D3(3OBw)[6] and DFTB3+D3H5[7] levels were conducted using DFTB+[8] release 19.1. The CHARMM package[9] (Developmental Version 43a1) was used to calculate energies with DFTB3(CPEQ)[10] method, which includes extended polarization from chemical-potential equalization. We used xTB package (Version 6.1)[11,12] to calculate energies with GFN2-xTB.[12] PM6,[13] PM6+D3, PM6+D3H4,[14] and PM7[15] calculations were performed using the MOPAC2016[16] program.

# 2 Computational Detail for the Second Virial Coefficient Calculations

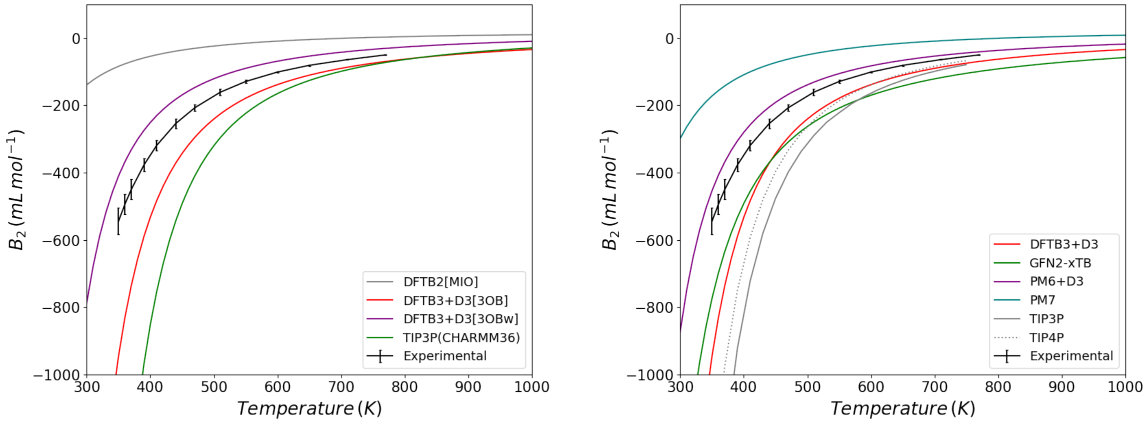
The second virial coefficient can be derived from the virial equation of state. The virial equation of state expresses $\frac{p}{k_B T}$ as a power series of density $\frac{N}{V}$.

$$\frac{p}{k_\mathrm{B}T} = \frac{N}{V}\left[1 + B_2\frac{N}{V} + B_3\left(\frac{N}{V}\right)^2 + \dots\right] \tag{4}$$

The virial coefficients $B_2, B_3, \dots$ of a fluid measure deviations from the ideal gas behavior. The second virial coefficient $B_2$ can be calculated by integrating the average Mayer function $\langle f_{12}\rangle_{\Omega_1,\Omega_2}$ over a distance grid of monomer centers of mass. The Mayer function only depends on pair-wise interaction energies, and the average is taken over angular orientations of molecules.

$$B_2(T) = -2\pi \int \mathrm{d}R_{12} R_{12}^2 \left\langle f_{12} \right\rangle_{\Omega_1,\Omega_2} \ , \ f_{12} = \mathrm{e}^{-\beta V^{2\mathrm{B}}(R_{12},\Omega_1,\Omega_2)} - 1 \tag{5}$$

In this study, the radial distance grid is made of at least 300 grid points at an interval of 0.05 Å, spanning the distance from 2.2 Å to 17.5 Å. At each point on the radial grid, $4 \times 10^4$ random angular orientations are performed independently for each monomer to achieve convergence of the angularly averaged Mayer function. Radial integrals are performed using the Simpson rule to calculate $B_2$ for temperatures ranging from 100-1100 K, at an interval of 100 K. Recommended experimental (labelled as Experimental) $B_2$ values are obtained from Landolt-Brnstein Numerical Data and Functional Relationships in Science and Technology,[17] where $B_2$ have been fit by a weighted least square fit of selected experimental values.



(a) (b)

Figure S1: $B_2$ for water: (a) Shows only experiment and different versions of DFTB; (b) Shows experiment, previously calculated[18] $B_2$ for TIP3P and TIP4P water models, DFTB3+D3, and other relevant semiempirical methods.

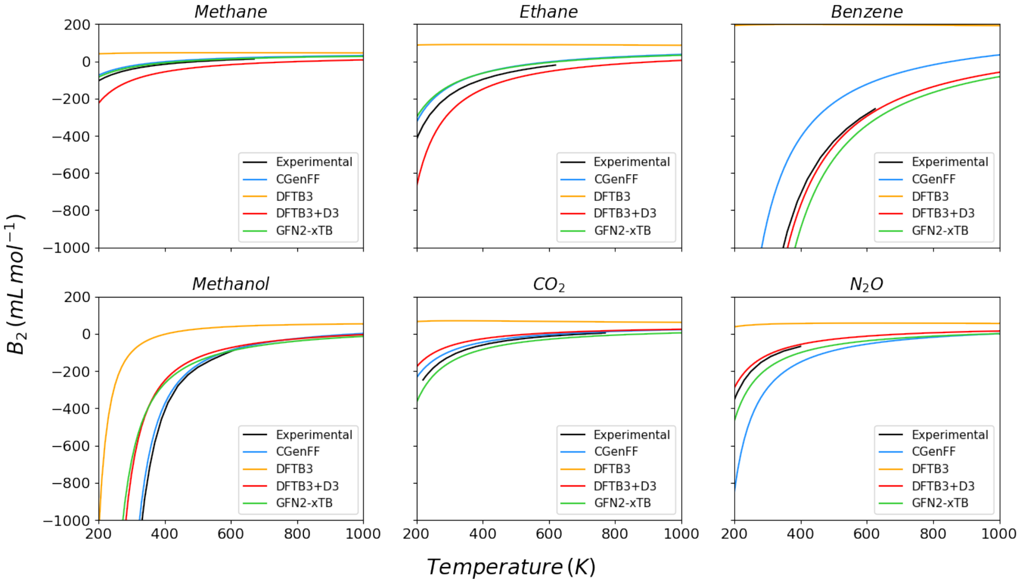


Figure S2: Calculated and experimental $B_2$ for polar and non-polar small molecules.

# 3 Free Energy Calculations for Methyl Phosphate Hydrolysis

## 3.1 Simulation Setup

All molecular dynamics (MD) simulations for methyl phosphate hydrolysis were performed using the GROMACS 5.0[19] package. The DFTB3 implemented[20] source code of GROMACS 5.0 was patched with PLUMED 2.4.5[21] to run restrained and biased MD simulations. PLUMED 2.4.5 was further modified to include biasing forces from deep neural networks. We performed QM(DFTB3/3OB)/MM MD simulations with 12 quantum atoms (methyl phosphate and one water molecule) at the center of the simulation box. The initial box dimensions were 2.43 nm on each side. The box was packed with 495 TIP3P[22] water molecules, and two sodium ions were included to maintain overall charge neutrality. A position restraint with force constant $2 \times 10^5\, kJ\, mol^{-1}\, nm^{-2}$ was applied in all three dimensions to keep the phosphorus atom fixed at the origin of the box. Additionally, an angular restraint of magnitude $2 \times 10^4\, kJ\, mol^{-1}\, rad^{-2}$ was imposed on $O_{attack} - P - O_{lg}$ to keep it linear. $P - O_{attack}$

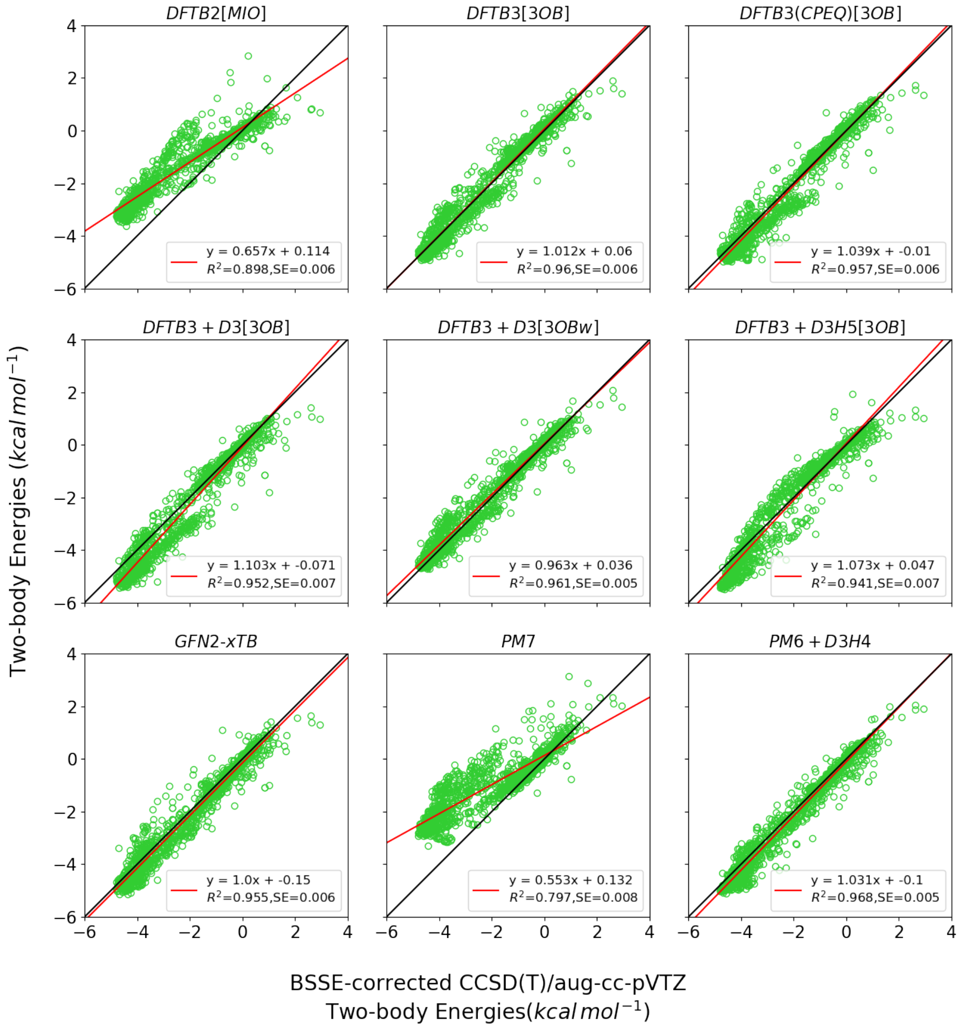


Figure S3: Correlation plots of two-body interaction energies for water calculated at different semi-empirical QM levels in comparison to BSSE corrected CCSD(T) calculations.

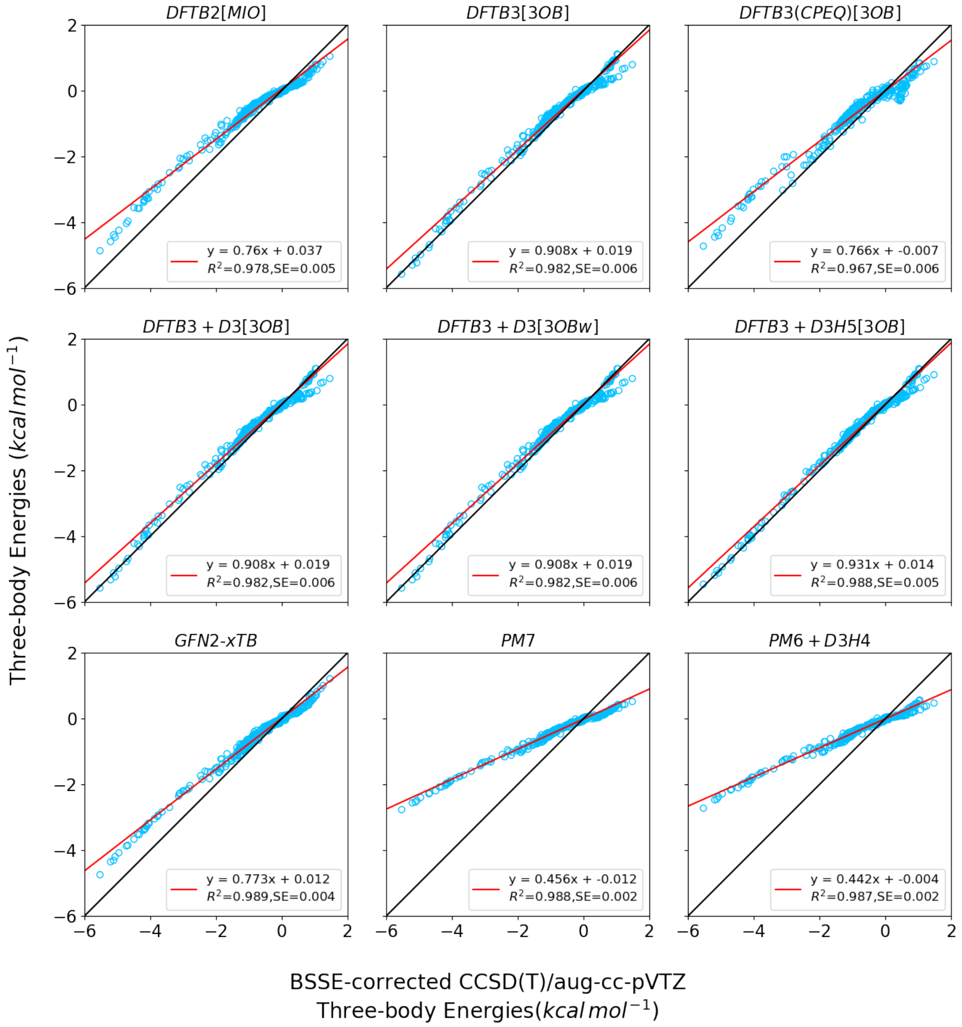


Figure S4: Correlation plots of three-body interaction energies for water at different semi-empirical QM levels in comparison to BSSE corrected CCSD(T) calculations.

and $P - O_{lg}$ distances were restrained with quadratic upper bounds at 3.0 Å and 3.5 Å. All other bonded oxygens were restrained with the same potential with upper bounds 2.0 Å from the P atom. Periodic boundary conditions were applied in all three dimensions. A short range Coulomb cutoff of 0.9 nm is used for electrostatic interactions in the real space. Long-range electrostatics were treated with the smooth Particle-Mesh-Ewald method.[23] Van der Waals cutoff radius was taken to be 0.9 nm. We used velocity-rescale thermostat[24] with a time constant of 0.1 ps to maintain the system temperature at 300 K. Pressure coupling at 1 atm was achieved with Parrinello-Rahman barostat[25] with a relaxation time 1.5 ps and compressibility $4.5 \times 10^{-5}\, bar^{-1}$. The time step used in the simulations was 0.5 fs.

## 3.2 Free Energy Calculations

In this work, we use the reinforced learning scheme[26] to compute free energy as a function of three collective variables (CVs). The three CVs chosen to study the hydrolysis reaction are: 1. the distance between the phosphorus atom and the attacking water oxygen ($O_{attack}$) 2. the distance between the phosphorus atom and leaving group oxygen ($O_{lg}$) 3. the asymmetric stretcth coordinate of proton transfer given by the distance difference between $O_{donor} - H$ and $O_{acceptor} - H$. In our case, $O_{donor}$ is the same atom as $O_{attack}$. The preprocessing transformation applied to these CVs is $\mathcal{P}(x) = x^3$. The initial guess of the free energy surface (FES) is constructed from a 50 ps unbiased MD simulation. Each reinforced dynamics step was initiated with a 50 ps biased MD. An ensemble of six deep neural networks (DNNs), each with three hidden layers of sizes (48, 24, 12) was used to estimate the error in free energy prediction. The uncertainty levels[26] used for the DNNs were $e_0 = 400\, kJ\, mol^{-1}\, nm^{-1}$ and $e_1 = 440\, kJ\, mol^{-1}\, nm^{-1}$. The DNNs were trained using Adam SGD algorithm in Tensorflow with a batch size of 20 for 12500 epochs. A maximum of 64 (initially) or 48 restrained MDs, each 25 ps long, were performed at CV values belonging to the regions of high uncertainty. The force constant for the restrained MDs was chosen to be $4 \times 10^5\, kJ\, mol^{-1}\, nm^{-2}$. A total of 42 biased MDs (in total 2.1 ns) and 1458 restrained MDs (in total ~36.5 ns) before convergence is achieved.

# References

(1) Hankins, D.; Moskowitz, J.; Stillinger, F. Water molecule interactions. *J. Chem. Phys.* **1970**, *53*, 4544–4554.

(2) Medders, G. R.; Babin, V.; Paesani, F. A critical assessment of two-body and three-body interactions in water. *J. Chem. Theory Comput.* **2013**, *9*, 1103–1114.

(3) Boys, S. F.; Bernardi, F. The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors. *Mol. Phys.* **1970**, *19*, 553–566.

(4) Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. A fifth-order perturbation comparison of electron correlation theories. *Chem. Phys. Lett.* **1989**, *157*, 479–483.

(5) Dunning Jr, T. H. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007–1023.

(6) Goyal, P.; Qian, H.-J.; Irle, S.; Lu, X.; Roston, D.; Mori, T.; Elstner, M.; Cui, Q. Molecular simulation of water and hydration effects in different environments: Challenges and developments for DFTB based models. *J. Phys. Chem. B* **2014**, *118*, 11007–11027.

(7) Rezac, J. Empirical self-consistent correction for the description of hydrogen bonds in DFTB3. *J. Chem. Theory Comput.* **2017**, *13*, 4804–4817.

(8) Aradi, B.; Hourahine, B.; Frauenheim, T. DFTB+, a sparse matrix-based implementation of the DFTB method. *J. Phys. Chem. A* **2007**, *111*, 5678–5684.

(9) Brooks, B. R.; III, C. L. B.; Mackerell, A. D.; 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*, 1545–1614.

(10) 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) Grimme, S.; Bannwarth, C.; Shushkov, P. A robust and accurate tight-binding quantum chemical method for structures, vibrational frequencies, and noncovalent interactions of large molecular systems parametrized for all spd-block elements (Z= 1–86). *J. Chem. Theory Comput.* **2017**, *13*, 1989–2009.

(12) Bannwarth, C.; Ehlert, S.; Grimme, S. GFN2-xTB—An accurate and broadly parametrized self-consistent tight-binding quantum chemical method with multipole electrostatics and density-dependent dispersion contributions. *J. Chem. Theory Comput.* **2019**, *15*, 1652–1671.

(13) Stewart, J. J. Optimization of parameters for semiempirical methods V: modification of NDDO approximations and application to 70 elements. *J. Mol. Model.* **2007**, *13*, 1173–1213.

(14) Řezáč, J.; Hobza, P. A halogen-bonding correction for the semiempirical PM6 method. *Chem. Phys. Lett.* **2011**, *506*, 286–289.

(15) Stewart, J. J. Optimization of parameters for semiempirical methods VI: more modifications to the NDDO approximations and re-optimization of parameters. *J. Mol. Model.* **2013**, *19*, 1–32.

(16) MOPAC2016, J. J. Stewart, Stewart Computational Chemistry, Colorado Springs, CO, USA, 2016.

(17) Dymond, J. D.; Marsh, K. N.; Wilhoit, R. C. In *Virial coefficients of pure gases and mixtures*; Fenkel, M., Marsh, K. N., Eds.; Springer Landord-Bornstein, 2003; Vol. 21.

(18) Benjamin, K. M.; Schultz, A. J.; Kofke, D. A. Gas-phase molecular clustering of TIP4P and SPC/E water models from higher-order virial coefficients. *Ind. Eng. Chem. Res.* **2006**, *45*, 5566–5573.

(19) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1*, 19–25.

(20) Kubar, T.; Welke, K.; Groenhof, G. New QM/MM implementation of the DFTB3 method in the gromacs package. *J. Comput. Chem.* **2015**, *36*, 1978–1989.

(21) Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; R. Broglia et al., PLUMED: A portable plugin for free-energy calculations with molecular dynamics. *Comput. Phys. Commun.* **2009**, *180*, 1961–1972.

(22) 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.

(23) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577–8593.

(24) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.

(25) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182–7190.

(26) Zhang, L.; Wang, H.; E, W. Reinforced dynamics for enhanced sampling in large atomic and molecular systems. *J. Chem. Phys.* **2018**, *148*, 124113.