Dual-topology/dual-coordinate free-energy simulation using QM/MM force field

Hao Hu and Weitao Yang

Citation: The Journal of Chemical Physics **123**, 041102 (2005); doi: 10.1063/1.1990113
View online: http://dx.doi.org/10.1063/1.1990113
View Table of Contents: http://scitation.aip.org/content/aip/journal/jcp/123/4?ver=pdfcov
Published by the AIP Publishing

Dual-topology/dual-coordinate free-energy simulation using QM/MM force field

Hao Hu and Weitao Yang[a)]
Department of Chemistry, Duke University, Durham, North Carolina 27708

(Received 9 March 2005; accepted 9 June 2005; published online 3 August 2005)

We have developed a dual-topology/dual-coordinate free-energy simulation method for use with a QM/MM force field. By combining two parallel processes into one alchemical process, we are able to compute the double free-energy difference ($\Delta\Delta F$) within a single simulation, which eliminates half of the expensive quantum-mechanical simulation in general. The method has been tested in computing the solvation free-energy differences of several molecular pairs and shows close agreement with experimental results. © *2005 American Institute of Physics*.
[DOI: 10.1063/1.1990113]

Free energy and related thermodynamic state functions are key properties for understanding many important physical, chemical, and biological processes. Free-energy (FE) simulation has become a most important component of molecular-dynamics (MD) and Monte Carlo (MC) simulation methods. With the use of highly efficient classical molecular-mechanics (MM) force fields, applications of the free-energy simulation method have made substantial achievement in various fields, such as drug and inhibitor design, and in the study of enzymatic reactions.[1–6]

With the successful development of hybrid quantum-mechanics/molecular-mechanics (QM/MM) force field[10] and the subsequent proposal of many methods for treatment of QM/MM boundary,[8,9,7,11,12] free-energy simulation using a QM/MM or QM force field becomes increasingly important and challenging in recent years. The free-energy change for a chemical process has been addressed in several ways.[13–15] When the QM calculations are at the level of semiempirical or tight-binding approximation and are thus not expensive, direct umbrella-sampling-type free-energy calculations are feasible and are routinely used for reactions in solutions and in enzymes.[14,16,17] When the QM method is at the first-principles level, it is a much more challenging issue to combine the QM calculations, which are expensive, with statistical-mechanics calculations of free energies which require extensive sampling. For reactions in solution, the QM-FE approach developed by Chandrasekhar *et al.*[13] can be viewed as a QM/MM approach to the free-energy calculations in which the reaction path is determined in vacuum without the environment and the free energies along the reaction path is determined by the free-energy perturbation (FEP) calculations with the QM systems fixed. Warshel and co-workers have developed several methods to attack this problem,[18–21] of which the most important approach is to use a simplified reference potential to extend the phase-space sampling of the QM system in the free-energy calculation. The significance of the use of a reference potential is that it is applicable to both the physical process and the nonphysical process such as those frequently employed in MM-based alchemical free-energy simulations. For reactions in enzymes, Zhang *et al.* built on the QM-FE approach and developed the QM/MM-FE method.[15] In the QM/MM-FE method, different from the QM-FE approach, the reaction path is determined by *ab initio* QM/MM methods in the realistic enzyme environment, with the polarization effects of the enzyme environment on the QM part considered, and hence there is no need to artificially map different structures along the reaction path into the enzyme environment described by MM. Similar to the QM-FE approach, the QM/MM-FE method uses FEP method to calculate the free-energy change along the reaction coordinate. The applications of these FE methods have shown great success in reactions in solutions and in enzymes.

When the details of the reaction process is not of importance, or simulating the chemical process bears too much difficulties, formalism is sought for a direct free-energy simulation between two states. For instance, to predict the pK_a of a titratable group, one would prefer direct calculation of the free-energy difference between protonated and deprotonated states, while the physical process of proton loading/unloading is irrelevant and can hardly be simulated in practice. Mindful that a direct QM sampling is expensive, except for fast semiempirical methods, one feasible way is to use a reference potential to mimic the QM energy surface and allow the extensive phase-space sampling that ensures proper convergence of free energy.[21–25] Even though this approach has been widely used, the question of how to directly apply the free-energy perturbation–like approach with a QM/MM force field remains to be addressed. The difficulty of direct-modeling the transformation between two QM molecular systems in FEP simulation has been reviewed by Li[26] *et al.*, who have pointed out the crucial difference between the use of a pure MM force field and a hybrid QM/MM force field in the free-energy simulation. That is, the QM/MM force field is nonadditive and thus suffers from numerical complications such as end-point problems. To solve the problem, they have developed the "dual-topology/single-coordinate" approach and applied it to several case studies.[26,27] On the other hand,

[a)]Electronic mail: weitao.yang@duke.edu

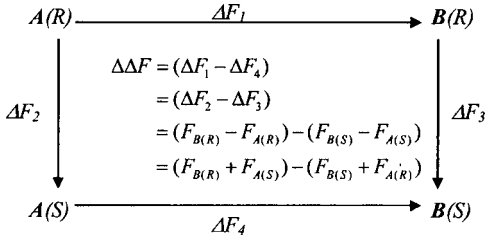

FIG. 1. Free-energy transformation cycle. *A* and *B* stand for two (QM) molecules, while subscripts *R* and *S* refer to two states.

to protect the emergence and annihilation of QM atoms, Yang *et al.* introduced a "chaperone" molecule at each simulation ending state.[28] This scheme may be used as a general method in the QM/MM free-energy simulation studies, even though MM descriptions for the mutating structural parts are needed and usually constructed artificially. Reported results have shown that the accuracy of the QM/MM free-energy simulation results is comparable,[26–29] if not better, with that of the classical MM-based simulations, despite the fact that they may still suffer from minor problems such as the absence of polarization of the MM environment and incompatible vdW parameters obtained from MM force fields.[30]

We develop here a new algorithm for the efficient simulation of the QM/MM free-energy differences. For a single process connecting two states, for example, the states in gas phase and in solution, computing the QM absolute free-energy difference (ΔF) remains problematic because the calculation is sensitively dependent on the basis set and the level of theory used. In many practical biochemical processes, however, the difference ($\Delta\Delta F$, Fig. 1) of the free-energy differences of two related processes is much more meaningful. Two simulations are usually performed to obtain two free-energy differences for each individual process. The computed value of the double difference ($\Delta\Delta F$) is usually more accurate because of error cancellation between two parallel processes.

Several observations motivate us to develop a new "dual-topology/dual-coordinate" algorithm. First, when quantum mechanics is introduced into the force field, the computation of the QM Hamiltonian generally becomes the bottleneck, except in the case of small molecules with a semiempirical QM method. Second, for the thermodynamic cycle shown in Fig. 1, processes ΔF_1 and ΔF_4 are simulated more frequently than are ΔF_2 and ΔF_3, even though both combinations, i.e., ΔF_1 and ΔF_4 or ΔF_2 and ΔF_3, should give exactly the same value of $\Delta\Delta F$. The two parallel processes ΔF_1 and ΔF_4 are preferentially selected for the MM force field simulation because the environment does not change during each process. In contrast, the processes ΔF_2 and ΔF_3 essentially simulate the exchange of the environments; the drastic differences between them make the simulation difficult to be carried out in practice. However, when the processes ΔF_1 or ΔF_4 are simulated with a QM/MM force field, *A* and *B* are always characterized by expensive quantum mechanics. The QM description of molecule *A* (and *B*) becomes replicated in the two processes, ΔF_1 and ΔF_4. Therefore, instead of performing two simulations in parallel, we propose here to carry out one single simulation to obtain $\Delta\Delta F$.

We start by expressing the Hamiltonian of molecule *A* in the QM/MM force field as[7]

$$H_A = H_{\mathrm{QM},A} + H_{\mathrm{QM/MM},A} + H_{\mathrm{MM}}, \tag{1}$$

in which the first term is the QM Hamiltonian of molecule *A*, the second term is the QM/MM interaction between molecule *A* and the MM environment, and the third term is the MM Hamiltonian of the environment. Recognizing that the first term $H_{\mathrm{QM},A}$ does not change between states, and that the change of the MM part (also the related QM/MM part) essentially determines the state of the system, we couple the transformation of the molecular Hamiltonian to the "alchemical" variable λ as

$$H_A(\lambda) = H_{\mathrm{QM},A} + [H_{\mathrm{QM/MM},A}(\lambda) + H_{\mathrm{MM}}(\lambda)], \tag{2}$$

with $H_A(0)=H_{A(R)}$ and $H_A(1)=H_{A(S)}$. The subscripts here represent the molecule (i.e., *A* or *B*) at certain states (i.e., *R* or *S*). For generality, the varying $H_{\mathrm{QM/MM},A}$ and H_{MM} are expressed here as functions of λ, instead of the common linear scaling by λ. One should also note that the term $H_{\mathrm{QM},A}$ is usually solved together with $H_{\mathrm{QM/MM},A}$, and thus implicitly depends on λ. We may now construct a system consisting of two noninteracting molecular systems with the Hamiltonian

$$\begin{aligned} H(\lambda) &= H_A(\lambda) + H_B(\lambda) \\ &= H_{\mathrm{QM},A} + [H_{\mathrm{QM/MM},A}(\lambda) + H_{\mathrm{MM}}(\lambda)] \\ &\quad + H_{\mathrm{QM},B} + [H'_{\mathrm{QM/MM},B}(\lambda) + H'_{\mathrm{MM}}(\lambda)], \end{aligned} \tag{3}$$

in which $H_A(\lambda)$ and $H_B(\lambda)$ are the Hamiltonians of molecule *A* and *B*, respectively. The two different MM states of the QM molecules are distinguished by the symbols *H* and *H'*. For the two end points of the simulation, we have

$$\begin{aligned} H(0) &= H_{A(R)} + H_{B(S)} \\ &= H_{\mathrm{QM},A} + [H_{\mathrm{QM/MM},A(R)} + H_{\mathrm{MM},A(R)}] \\ &\quad + H_{\mathrm{QM},B} + [H_{\mathrm{QM/MM},B(S)} + H_{\mathrm{MM},B(S)}], \end{aligned}$$

$$\begin{aligned} H(1) &= H_{A(S)} + H_{B(R)} \\ &= H_{\mathrm{QM},A} + [H_{\mathrm{QM/MM},A(S)} + H_{\mathrm{MM},A(S)}] \\ &\quad + H_{\mathrm{QM},B} + [H_{\mathrm{QM/MM},B(R)} + H_{\mathrm{MM},B(R)}]. \end{aligned} \tag{4}$$

That is, the simulation system starts from the molecule *A* in state *R* and the molecule *B* in state *S* ($\lambda=0$). The two molecules, *A* and *B*, gradually exchange their environments during the λ-coupled simulation process so that at the end point ($\lambda=1$) the molecule *A* is in state *S* and the molecule *B* is in state *R*. Therefore, a dual-topology/dual-coordinate description of the system is required for starting the free-energy simulation. However, the environments *R* and *S* are decoupled, as also are *A* and *B*. The motions of *A* and *B* are always governed by quantum mechanics without any need of a MM description. Thus only one simulation from $\lambda=0$ to $\lambda=1$ is required to obtain $\Delta\Delta F$. Since the derivatives $\delta H_A(\lambda)/\delta\lambda$ and $\delta H_B(\lambda)/\delta\lambda$ can also be computed during the simulation

TABLE I. Solvation free-energy differences (kcal/mol).

Transformation	$\Delta\Delta F^{a}$	$\Delta\Delta F^{b}$
MeOH→EtOH	−0.08±0.1	−0.05
Acem→NMA[c]	−0.13±0.2	−0.5[d]
$AcCH_3 \rightarrow$ Acem	−6.45±0.1	−5.9
$C_6H_6 \rightarrow C_6H_5OH$	−4.6±0.2	−5.62
MeOH→EtH	4.5±0.1	6.9

[a]Results from current simulation.
[b]Experimental results reported in literature (Ref. 39).
[c]Abbreviations: Acem=acetamide, NMA=N-methylacetamide.
[d]Data from Ref. 40.

process, it is possible to compute the individual free-energy differences.

Solvation free-energy differences between several molecular pairs have been simulated to test this method. All simulations were carried out with the molecular-dynamics simulation program SIGMA[31,32] combined with the self-consistent charge-density-functional tight-binding (SCCDFTB) method.[33,34] Each molecule was solvated using the TIP3P water model[35] in a cubic box of $40\times40\times40$ Å^3. The multiple-time-step algorithm was used with dual cutoffs of 8 and 12 Å, with time step of 1 and 6 fs, respectively. The vdW parameters were taken from the CHARMM22 force field.[36] The long-range electrostatic interactions were neglected here since all molecules are neutral. The slow-growth method was used for simulating the free-energy change, with each individual transformation lasting for 360 ps. For each simulation system, five cycles of forward and backward simulations were performed. The results were averaged and the mean deviation was used for the estimation of simulation uncertainties.[37,38]

The results are shown in Table I. Most systems show close agreement with experiments, even for the difficult case of MeOH→EtOH. The only exception with large deviation is the transformation between methanol and ethane. Although not certain at this time, several factors may contribute to this deviation, including the SCCDFTB method used in the current study and the vdW parameters. Considering the fact that the SCCDFTB method is not optimized for the purpose of free-energy simulation, nor are the QM/MM vdW interactions, the agreement between simulation and experimental results indicates the validity and effectiveness of the current approach.

Clearly, there are two main advantages of the current method: the computational expense is reduced by one half and no artificial chaperone molecule is required. For high-level QM methods, computing H_{QM} is usually much more time consuming than calculating $H_{\mathrm{QM,MM}}$ and H_{MM}. This assessment is based on the fact that in many QM/MM calculations most of the QM calculation time is spent on the self-consistent-field (SCF) iterations, while $H_{\mathrm{QM/MM}}$ is only calculated once per MD step. Therefore, by alchemically mixing the two environmental states, we save half of the effort for computing H_{QM}. With the increasing demands of achieving high simulation accuracy by using expensive QM methods, the superiority of this method will become more evident. Meanwhile, since the intramolecular interactions of each QM subsystem are preserved during the transformation process, we have no need for the ghost chaperone molecule to protect the QM atoms. Subsequently, there is no need for computing the thermodynamic contribution of chaperone molecules. Even when molecular-mechanical restrictions are necessary for retaining the global position and/or the orientation of the molecules, analytic contributions can be computed.[41] In most cases, they will essentially cancel out if the type and strength of the restrictions are carefully chosen.

We note that other simulation methods mentioned earlier may be invoked when the single free-energy difference (ΔF) is sought. Interestingly, in practice this will involve mostly reactions of one molecular species (i.e., no creation/annihilation of QM atoms). In such a case, the method proposed by Li *et al.* could be easy to use and implement.[26] In many practical applications of the free-energy simulation such as in drug and inhibitor design, however, relative free energy will be more useful. One immediate application of the method proposed here will be computing the binding free-energy difference of protein-ligand complexes.

Before extending the current method into practical applications, we realized there are two technical difficulties that need to be addressed. First, like many previous studies, we used a fast semiempirical QM method. When more rigorous methods such as Hartree-Fock or density-functional theory are used, the computational cost still makes the free-energy simulation prohibitive even with our time-saving approach proposed here. A more elegant and efficient method should be developed for that case. Second, the test calculations reported here are for solvation free energy, which has simplified environmental effects; extra care must be taken when the method is applied into enzymatic systems which have more complicated interactions in the active site.

In summary, we have developed a dual-topology/dual-coordinate free-energy simulation method for use with a QM/MM force field. The method facilitates the computation by eliminating half of the expensive QM calculations. Even though it has been tested with a semiempirical QM method in this article, the principle is independent of the form of the QM method and thus applicable to other high-level QM methods. This method can be used in many different applications of free energy and QM/MM simulations, such as refinement of QM/MM interaction forms and parameters (e.g., vdW parameters), and finally in the study of protein-ligand interactions.

ACKNOWLEDGMENTS

We thank Dr. Jan Hermans, Dr. Andrés Cisneros, Zhenyu Lu, and Jerry Parks for helpful discussions. The work is supported by the National Institute of Health.

[1] D. L. Beveridge and F. M. DiCapua, in *Computer Simulation of Biomolecular Systems*, edited by W. F. van Gunsteren and P. K. Weiner (ESCOM, Leiden, 1989), Vol. 1, p. 1.
[2] W. F. van Gunsteren, in *Computer Simulations of Biomolecular Systems*, edited by W. F. van Gunsteren and P. K. Weiner (ESCOM, Leiden, The Netherlands, 1989), p. 27.
[3] P. A. Kollman, Chem. Rev. (Washington, D.C.) **93**, 2395 (1993).
[4] T. Simonson, G. Archontis, and M. Karplus, Acc. Chem. Res. **35**, 430 (2002).

[5] W. L. Jorgensen, Science **303**, 1813 (2004).
[6] A. Shurki and A. Warshel, Adv. Protein Chem. **66**, 249 (2003).
[7] A. Warshel and M. Levitt, J. Mol. Biol. **103**, 227 (1976).
[8] U. C. Singh and P. A. Kollmann, J. Comput. Chem. **7**, 718 (1986).
[9] M. J. Field, P. A. Bash, and M. Karplus, J. Comput. Chem. **11**, 700 (1990).
[10] A. Warshel, Chem. Phys. Lett. **55**, 454 (1978).
[11] V. Théry, D. Rinaldi, J.-L. Rivail, B. Maigret, and G. G. Ferenczy, J. Comput. Chem. **15**, 269 (1994).
[12] Y. Zhang, T.-S. Lee, and W. Yang, J. Chem. Phys. **110**, 46 (1999).
[13] J. Chandrasekhar, S. F. Smith, and W. L. Jorgensen, J. Am. Chem. Soc. **107**, 154 (1985).
[14] J.-K. Hwang, G. King, S. Creighton, and A. Warshel, J. Am. Chem. Soc. **110**, 5297 (1988).
[15] Y. Zhang, H. Liu, and W. Yang, J. Chem. Phys. **112**, 3483 (2000).
[16] H. Liu and Y. Shi, J. Comput. Chem. **15**, 1311 (1994).
[17] Q. Cui, M. Elstner, and M. Karplus, J. Phys. Chem. B **106**, 2721 (2002).
[18] T. Wesolowski and A. Warshel, J. Phys. Chem. **98**, 5183 (1994).
[19] T. Wesolowski, R. P. Muller, and A. Warshel, J. Phys. Chem. **100**, 15444 (1996).
[20] R. P. Muller and A. Warshel, J. Phys. Chem. **99**, 17516 (1995).
[21] J. Bentzien, R. P. Muller, J. Florián, and A. Warshel, J. Phys. Chem. B **102**, 2293 (1998).
[22] V. Luzhkov and A. Warshel, J. Comput. Chem. **13**, 199 (1992).
[23] M. Strajbl, G. Hong, and A. Warshel, J. Phys. Chem. B **106**, 13333 (2002).
[24] R. H. Wood, E. M. Yezdimer, S. Sakane, J. A. Barriocanal, and D. J. Doren, J. Chem. Phys. **110**, 1329 (1999).
[25] S. Sakane, E. M. Yezdimer, W. Liu, J. A. Barriocanal, D. J. Doren, and R. H. Wood, J. Chem. Phys. **113**, 2583 (2000).
[26] G. Li, X. Zhang, and Q. Cui, J. Phys. Chem. B **107**, 8643 (2003).
[27] G. Li and Q. Cui, J. Phys. Chem. B **107**, 14521 (2003).
[28] W. Yang, R. Bitetti-Putzer, and M. Karplus, J. Chem. Phys. **120**, 9450 (2004).
[29] M. R. Reddy, U. C. Singh, and M. D. Erion, J. Am. Chem. Soc. **126**, 6224 (2004).
[30] D. Riccardi, G. Li, and Q. Cui, J. Phys. Chem. B **108**, 6467 (2004).
[31] G. Mann, R. H. Yun, L. Nyland, J. Prins, J. Board, and J. Hermans, in *Computational Methods for Macromolecules: Challenges and Applications*, Proceedings of the 3rd International Workshop on Algorithms for Macromolecular Modelling, edited by T. Schlick and H. H. Gan, (Springer, New York, 2002), p. 129.
[32] H. Hu, M. Elstner, and J. Hermans, Proteins: Struct., Funct., Genet. **50**, 451 (2003).
[33] M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, T. Frauenheim, S. Suhai, and G. Seifert, Phys. Rev. B **58**, 7260 (1998).
[34] M. Elstner, T. Frauenheim, E. Kaxiras, G. Seifert, and S. Suhai, Phys. Status Solidi B **217**, 357 (2000).
[35] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, J. Chem. Phys. **79**, 926 (1983).
[36] A. D. MacKerell, Jr., D. Bashford, M. Bellott, R. L. Dunbrack, Jr., J. D. Evanseek, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, III, B. Roux, M. Schlenkrich, J. C. Smith, R. Stotc, J. Straub, M. Watanbe, J. Wiórkiewcz-Kuczera, D. Yin and M. Karplus, J. Phys. Chem. B **102**, 3586 (1998).
[37] J. Hermans, J. Phys. Chem. **95**, 9029 (1991).
[38] H. Hu, R. H. Yun, and J. Hermans, Mol. Simul. **28**, 67 (2002).
[39] S. Cabani, P. Gianni, V. Mollica, and L. Lepori, J. Solution Chem. **10**, 563 (1981).
[40] P. A. Kollman, Acc. Chem. Res. **29**, 461 (1996).
[41] J. Hermans and L. Wang, J. Am. Chem. Soc. **119**, 2707 (1997).