

Development of a perturbative approach for Monte Carlo simulations using a hybrid ab initio QM/MM method

Thanh N. Truong^{*}, Eugene V. Stefanovich¹

Department of Chemistry, University of Utah, Salt Lake City, UT 84112, USA

Received 6 February 1996; in final form 26 March 1996

Abstract

We present the development and assessment on the accuracy of a perturbative approach for Monte Carlo simulations using a hybrid ab initio quantum mechanical/molecular mechanics (QM/MM) potential. The central idea of this approach is that for most solvent moves, the changes in the solute wavefunction are small, thus one can use the perturbation theory to approximate the energy change without having to perform full quantum calculation at each Monte Carlo step. Consequently, the computational demand can be reduced by several orders of magnitude while maintaining a reasonable level of accuracy. This opens new possibilities for using more accurate levels of theory to describe both solute and solvent molecules in Monte Carlo simulations.

1. Introduction

In the last few years, the hybrid quantum mechanical/molecular mechanics (QM/MM) approach [1–19] has been emerging as a powerful tool for studying processes in condensed media. In this approach, the system is divided into two regions. The primary region requires to be treated quantum mechanically. This region can be either the solute in studying solvation, the active site of the protein in studying enzyme mechanism, or the adsorbate plus the local region of the adsorption site in studying adsorption and chemisorption processes. Interactions in the remaining region are approximated by molecular mechanics force fields. This approach rivals ab initio

dielectric continuum methods [20–22] in overcoming the limitation of the molecular mechanics method for studying reactions in solution which involve bond-breaking and bond-forming processes. In addition, it includes the solute electronic polarization that is ignored in classical force fields. A rather complete collection of references on development and applications of the QM/MM approach was given in a recent excellent review by Gao [6].

In the QM/MM approach, the total effective Hamiltonian is partitioned as

$$H_{\text{eff}} = H^{\text{QM}} + H^{\text{QM/MM}} + H^{\text{MM}}, \quad (1)$$

where H^{QM} represents interactions within the quantum region, H^{MM} describes classical interactions within the MM region, and $H^{\text{QM/MM}}$ denotes interactions between the QM and MM regions. Due to the computational demand of solving the electronic Schrödinger equation, most applications of the QM/MM methods to date were limited to the use of

^{*} Corresponding author.

¹ On leave from the Institute of Chemical Physics, University of Latvia, Riga LV1586, Latvia.

semiempirical methods, such as the generalized valence bond [23] or semiempirical molecular orbital method [6]. The use of ab initio Hartree–Fock/MM (HF/MM) coupled potential in molecular dynamics (MD) simulations was first introduced by Singh and Kollman [12]. Merz and co-workers recently introduced the DFT/MM coupled potential [13,14] and provided a more detailed assessment on the accuracy of the HF/MM potential [19]. It is important to point out that in MD simulations, full self-consistent field (SCF) calculations of both energy and derivatives are required at every time step. Thus, it is computationally expensive to reach adequate convergence of the calculated dynamic and thermodynamic properties. This limits the range of applicability of QM/MM methods to small quantum regions or to low levels of theory and small basis sets.

Unlike MD simulations where all atoms move at each time step, in Monte Carlo (MC) simulations only one molecule moves at a time. For solvent move that does not change the solute wavefunction significantly, one can calculate the corresponding change in the quantum energy contribution using the wavefunction from the previously accepted step without having to perform full SCF calculation again. This will greatly improve the performance of MC simulations. The present idea is referred to as the perturbative MC approach. In this study, we present the theoretical basis for the perturbative MC simulations and examine their accuracy. While this manuscript was in preparation, two reports appeared in print employing this approach, however, no theoretical basis and systematic assessment on its accuracy was given. One report is by Gao [24] using the AM1/MM potential, and the other by Tuñón et al. [17] using the DFT/MM coupled potential.

2. Perturbative approach for MC simulations using QM/MM potentials

In the QM/MM method, the $H^{\text{QM/MM}}$ Hamiltonian is given by

$$H^{\text{QM/MM}} = \sum_i \sum_{\alpha} -\frac{q_{\alpha}}{R_{i\alpha}} + \sum_A \sum_{\alpha} \frac{q_{\alpha} Z_A}{R_{A\alpha}} + \sum_{\alpha} \sum_A \left(\frac{A_{\alpha A}}{(R_{\alpha A})^{12}} - \frac{B_{\alpha A}}{(R_{\alpha A})^6} \right), \quad (2)$$

where the indices i and A run over the solute electrons and nuclei, respectively, while α runs over all solvent atoms. The first two terms represent electrostatic interactions between solvent atoms and solute electrons and nuclei, respectively. The last term describes solvent–solute van der Waals interactions, denoted as E_{s-s}^{vdW} below. $A_{\alpha A}$ and $B_{\alpha A}$ are van der Waals parameters. The total energy is then expressed as

$$E_{\text{tot}} = \left\langle \Psi \left| H^{\text{QM}} + \sum_i \sum_{\alpha} -\frac{q_{\alpha}}{R_{i\alpha}} \right| \Psi \right\rangle + \sum_A \sum_{\alpha} \frac{q_{\alpha} Z_A}{R_{A\alpha}} + E_{s-s}^{\text{vdW}} + E^{\text{MM}}, \quad (3)$$

where E^{MM} is the interaction energy within the MM region and Ψ is the solute wavefunction.

In MC simulation using the QM/MM Hamiltonian given by Eq. (1), when the solvent molecule m is moved, the effective Hamiltonian in the first term of Eq. (3) is changed by

$$\Delta H = \sum_i \sum_{\alpha \in m} -q_{\alpha} \left(\frac{1}{R'_{i\alpha}} - \frac{1}{R_{i\alpha}} \right). \quad (4)$$

If the perturbation given by Eq. (4) is small, we can approximate the corresponding energy change within the first order perturbation theory by

$$\left\langle \Psi \left| \sum_i \sum_{\alpha \in m} -q_{\alpha} \left(\frac{1}{R'_{i\alpha}} - \frac{1}{R_{i\alpha}} \right) \right| \Psi \right\rangle,$$

where Ψ is the solute wavefunction from the previously accepted full SCF MC step. The change in the total energy for such a perturbative MC step can be expressed as

$$\Delta E_{\text{tot}} = \sum_{\mu\nu} P_{\mu\nu} \left\langle \mu \left| \sum_{\alpha \in m} -q_{\alpha} \left(\frac{1}{R'_{i\alpha}} - \frac{1}{R_{i\alpha}} \right) \right| \nu \right\rangle + \sum_A \sum_{\alpha} q_{\alpha} Z_A \left(\frac{1}{R'_{A\alpha}} - \frac{1}{R_{A\alpha}} \right) + \Delta E_{s-s}^{\text{vdW}} + \Delta E^{\text{MM}}, \quad (5)$$

where $P_{\mu\nu}$ is the density matrix element between the basis functions $|\mu\rangle$ and $|\nu\rangle$, $\Delta E_{s-s}^{\text{vdW}}$ and ΔE^{MM} are the changes in the solvent–solute van der Waals and solvent–solvent interactions, respectively. Eq. (5) indicates that the main computational cost in calculating ΔE_{tot} is the evaluation of a small number of

one-electron integrals. Consequently, one can expect significant speed up comparing to MC simulations when SCF calculation is done at each MC step. In practice, one needs to perform full SCF calculations to update the solute wavefunction after a certain number of perturbative moves. For convenience in the discussion below, we denote the number of perturbative moves between two full SCF calculations as NMOVE.

The accuracy of this perturbative approach critically depends on the assumption that the perturbation given in Eq. (4) is small relative to the unperturbed Hamiltonian. To estimate the magnitude of errors, we performed MC simulations for the enthalpies of binding at 25°C of several ion–H₂O complexes: Li⁺–H₂O, Na⁺–H₂O, Na⁺–(H₂O)₂, F[–]–H₂O, Cl[–]–H₂O and Cl[–]–(H₂O)₂. These strongly bound complexes present the most severe tests of the perturbative approach since more distant water molecules in solution would yield smaller perturbations. Furthermore, using these ion–water complexes rather than full solvation studies allows us (1) to separate methodological errors resulting from approximations made in the theory (Eq. (5)) from model errors associated with the periodic boundary condition; (2) to estimate the lower bound for the parameter NMOVE for simulations in solutions and (3) to compare with reliable experimental data. The computational costs of both cluster and liquid-phase simulations should be similar since the major computational demand is in the quantum calculations. Calculations of binding enthalpies were done using our local interface between the G92/DFT [25] and BOSS [26] programs. The TIP3P model [27] was used for

water molecules. OPLS parameters [28] were used for the solvent–solute van der Waals interaction. An equilibration period consisted of 10⁵ steps in which ions were treated as classical point charges followed by 10⁴ quantum equilibration steps in which ions were treated quantum mechanically at the HF/6-31G^{**} level. For each equilibrated complex, several 10⁴-step averaging runs were performed with different values of NMOVE. We found that 10⁴ averaging steps yield a standard deviation of about 0.2 kcal/mol in the binding enthalpies which is sufficient for our purpose.

3. Results

In this study, we focus on the accuracy of calculated enthalpies of binding of the ion–water complexes and the convergence of the calculated results with respect to the parameter NMOVE. The calculated enthalpies of binding are listed in Table 1. At each full SCF step, we calculated the error in the differential solvent–solute electron electrostatic energy by comparing ΔE_{tot} (Eq. (5)) calculated with the density matrix from the current full SCF calculation and that from the previous full SCF step. Root mean square (RMS) errors of this energy contribution are also listed in Table 1 and plotted versus NMOVE in Fig. 1.

The HF/6-31G^{**}/OPLS calculated enthalpies of binding of anion–water complexes are in good agreement with experimental data [29,30], though generally smaller by about 1 kcal/mol. The calculated enthalpies of binding of Na⁺–(H₂O)_{1,2} com-

Table 1
Enthalpies of binding (kcal/mol) for ion–(H₂O)_n complexes

NMOVE ^a	Li ⁺ –H ₂ O	Na ⁺ –H ₂ O	Na ⁺ –(H ₂ O) ₂	F [–] –H ₂ O	Cl [–] –H ₂ O	Cl [–] –(H ₂ O) ₂
1	–31.95 ± 0.03			–22.12 ± 0.04		
10	–31.93 (0.004) ^b	–26.79 (0.011)	–52.67 (0.010)	–22.73 (0.125)	–13.40 (0.052)	–26.46 (0.056)
100	–31.93 (0.006)	–26.45 (0.019)	–52.65 (0.021)	–22.67 (0.156)	–13.42 (0.076)	–26.48 (0.127)
500	–31.92 (0.014)	–26.39 (0.054)	–52.56 (0.045)	–22.62 (0.200)	–13.33 (0.140)	–26.44 (0.189)
1000	–31.89 (0.020)	–26.42 (0.055)	–52.61 (0.077)	–22.41 (0.352)	–13.34 (0.139)	–26.30 (0.170)
2000	–31.90 (0.025)	–26.42 (0.106)	–52.51 (0.041)	–22.38 (0.391)	–13.27 (0.111)	–25.92 (0.231)
Exp. ^c	–34 (est.)	–24	–44	–24	–14.7	–27.7

^aNMOVE is the number of MC steps between two successive full SCF calculations.

^bValues in parentheses are the RMS errors (kcal/mol) of the solvent–solute electron electrostatic interactions.

^cRefs. [29,30].

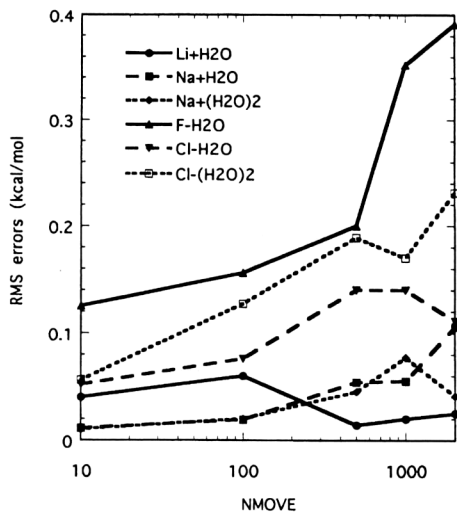


Fig. 1. Plot of the RMS errors (kcal/mol) of the solvent–solute electron electrostatic interaction as functions of NMOVE, the number of MC steps between two successive full SCF calculations.

plexes are overestimated by more than 2 kcal/mol while the predicted value for $\text{Li}^+ - \text{H}_2\text{O}$ is 2 kcal/mol too small. The differences between our calculated results and experiments can be due to several sources: (1) the deficiency in the description of the QM wavefunction, i.e. such as neglecting electron correlation effects and the use of small basis set; and (2) inconsistency between OPLS van der Waals parameters and the QM/MM level of theory. The later point was raised recently by Stanton et al. [19]. This supports the need for reoptimization of the van der Waals parameters for QM/MM methods.

The most significant result of this study, however, is a technical one. The calculated enthalpies of binding were found remarkably stable with respect to NMOVE even when NMOVE is as large as 2000 in all cases considered here (see Table 1). The RMS errors of the solvent–solute electron electrostatic interaction are larger for anion–water complexes compared to the cation–water complexes as shown in Fig. 1. This is expected due to the larger electronic polarizability for anions. This difference is also reflected in the average induced dipole moments of solutes which are less than 0.02 D for cations and about 0.1 D for anions. The errors generally increase as NMOVE increases. The largest error is about 0.4 kcal/mol for the $\text{F}^- - \text{H}_2\text{O}$ complex.

Since the computational cost is roughly proportional to the number of full SCF calculations and inversely proportional to NMOVE, the computational cost for NMOVE = 1000 is three orders of magnitude less compared to the NMOVE = 1 case. Note that in liquid-phase simulations, NMOVE can be much larger since the majority of MC steps will involve displacements of solvent molecules which are farther from the solute. It is possible to introduce an error weighting function to optimize the variable NMOVE as the simulation proceeds. We will address this issue in a future study.

4. Conclusions

We have shown that the perturbative approach for MC simulations of solvation processes using ab initio QM/MM potential offers significant computational advantage. The computational cost can be reduced by several orders of magnitude without sacrificing accuracy. This opens new possibilities for high-level ab initio electronic structure theory treatment for solutes and for more accurate treatment of solvent molecules such as using the pseudopotential [31] or frozen density [32,33] methods in MC simulations of reactions and spectroscopic properties in solution. All these possibilities are being investigated in our laboratory.

Acknowledgements

This work was supported in part by the University of Utah and the National Science Foundation via a Young Investigator Award to TNT.

References

- [1] J. Aqvist and A. Warshel, *Chem. Rev.* 93 (1993) 2523.
- [2] D. Bakowies and W. Thiel, *J. Comput. Chem.* 17 (1996) 87.
- [3] C.S. Carmer, B. Weiner and M. Frenklach, *J. Chem. Phys.* 99 (1993) 1356.
- [4] J.J. Field, P.A. Bash and M. Karplus, *J. Comput. Chem.* 11 (1990) 700.
- [5] M.J. Field, in: *Computer simulation of biomolecular systems: theoretical and experimental applications*, eds. W.F. van Gunsteren, P.K. Weiner and A.J. Wilkinson (ESCOM, Leiden, 1993) p. 82.

- [6] J. Gao, in: *Reviews in computational chemistry*, eds. K.B. Lipkowitz and D.B. Boyd (VCH Publishers, New York, 1996) p. 119.
- [7] J. Gao, *J. Phys. Chem.* 96 (1992) 537.
- [8] J.L. Gao and X.F. Xia, *Science* 258 (1992) 631.
- [9] V. Luzhkov and A. Warshel, *J. Comput. Chem.* 13 (1992) 199.
- [10] F. Maseras and K. Morokuma, *J. Comput. Chem.* 16 (1995) 1170.
- [11] J.H. McCreery, R.E. Christoffersen and G.G. Hall, *J. Am. Chem. Soc.* 98 (1976) 7198.
- [12] U.C. Singh and P.A. Kollman, *J. Comput. Chem.* 7 (1986) 718.
- [13] R.V. Stanton, D.S. Hartsough and K.M. Merz, Jr., *J. Phys. Chem.* 97 (1993) 11868.
- [14] R.V. Stanton, D.S. Hartsough and K.M. Merz, Jr., *J. Comput. Chem.* 16 (1995) 113.
- [15] O. Tapia and G. Johannin, *J. Chem. Phys.* 75 (1981) 3624.
- [16] M.A. Thompson, *J. Phys. Chem.* 99 (1995) 4794.
- [17] I. Tuñón, T.C. Martins-Costa, C. Millot, M.F. Ruiz-López and Rivail, *J. Comput. Chem.* 17 (1996) 19.
- [18] A. Warshel and M. Levitt, *J. Mol. Biol.* 103 (1976) 227.
- [19] R.V. Stanton, L.R. Little and K.M. Merz Jr., *J. Phys. Chem.* 99 (1995) 17344.
- [20] C.J. Cramer and D.G. Truhlar, in: *Solvent effects and chemical reactivity*, eds. O. Tapia and J. Bertrán (Kluwer, Dordrecht, 1995) in press.
- [21] T.N. Truong and E.V. Stefanovich, *Chem. Phys. Letters* 240 (1995) 253.
- [22] J. Tomasi and M. Persico, *Chem. Rev.* 94 (1994) 2027.
- [23] A. Warshel and R.M. Weiss, *J. Am. Chem. Soc.* 102 (1980) 6218.
- [24] J. Gao, *J. Am. Chem. Soc.* 117 (1995) 8600.
- [25] M.J. Frisch, G.W. Trucks, H.B. Schlegel, P.M.W. Gill, B.G. Johnson, M.W. Wong, J.B. Foresman, M.A. Robb, M. Head-Gordon, E.S. Replogle, R. Gomperts, J.L. Andres, K. Raghavachari, J.S. Binkley, C. Gonzalez, R.L. Martin, D.J. Fox, D.J. DeFrees, J. Baker, J.J.P. Stewart and J.A. Pople, *GAUSSIAN 92/DFT*, Version G.3 (Gaussian, Inc., Pittsburgh, PA, 1993).
- [26] W.L. Jorgensen, *BOSS*, Version 3.6 (Yale University, New Haven, CT, 1995).
- [27] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey and M.L. Klein, *J. Chem. Phys.* 79 (1983) 926.
- [28] W.L. Jorgensen and J. Tirado-Rives, *J. Am. Chem. Soc.* 110 (1988) 1657.
- [29] A.W. Castleman, Jr. and R.G. Keese, *Chem. Rev.* 86 (1986) 589.
- [30] P. Kebarle, *Ann. Rev. Phys. Chem.* 28 (1977) 445.
- [31] N. Vaidehi, T.A. Wesolowski and A. Warshel, *J. Chem. Phys.* 97 (1992) 4264.
- [32] T.A. Wesolowski and A. Warshel, *J. Phys. Chem.* 97 (1993) 8050.
- [33] E.V. Stefanovich and T.N. Truong, *J. Chem. Phys.* 104 (1995) 2946.