Theor Chem Account (2008) 120:307–312
DOI 10.1007/s00214-007-0300-z

REGULAR ARTICLE

Assessment of Two Surface Monte Carlo (TSMC) method to find stationary points of $(H_2O)_{15}$ and $(H_2O)_{20}$ clusters

Pradipta Bandyopadhyay

Received: 15 November 2006 / Accepted: 15 February 2007 / Published online: 1 May 2007

Abstract The Two Surface Monte Carlo (TSMC) technique reduces computational cost by using a computationally cheap biasing potential, which guides the molecular system to explore the potential energy surface of interest. It was shown earlier that the Effective Fragment Potential (EFP) can be a good choice for this biasing potential (Bandyopadhyay, J Chem Phys 122:091102, 2005) when the potential energy surface of interest is quantum mechanical. This may help in expanding the applicability of TSMC, since finding a good biasing potential is a major challenge. In the present work, the viability of TSMC method in finding stationary points of large molecular system is investigated using EFP as the biasing potential and RHF theory as the potential of interest. TSMC is applied to find the stationary points of water clusters of size 15 and 20. A semi-automated method starting from random geometries, without using any chemical intuition, found several stationary points. The simulated annealing method was used to refine the structures obtained from TSMC. Among the several low-energy structures obtained for 15 water cluster, one minimum, about 1 kcal/mol higher than the global minimum, was found. However, for 20 water cluster, no structure very close to the global minimum was obtained. Several strategies, learned from the experience of the present work, are discussed for improving the TSMC method, including the acceptance between the two energy surfaces.

Keywords Monte Carlo · Effective fragment potential · Stationary points · Water clusters · Global minimum

---

Contribution to the Mark S. Gordon 65th Birthday Festschrift Issue.

P. Bandyopadhyay (✉)
Department of Biotechnology, Indian Institute of Technology,
Guwahati 781039, Assam, India
e-mail: pradipta@iitg.ernet.in; praban07@gmail.com

1 Introduction

Computational investigation of floppy molecular systems, such as water clusters and flexible proteins, is challenging due to the presence of a large number of stationary points, many of which may be close in energy and accessible during experiments at room temperature. In quantum chemistry, stationary points are typically determined by employing local minimization methods from reasonable starting structures constructed by chemical intuition. However, because of diversity of different conformations of large fluxional systems, finding a reasonable starting structure becomes increasingly difficult with increase in size and flexibility of the molecule. Molecular simulation techniques such as Molecular Dynamics (MD) and Monte Carlo (MC) can also be used to find stationary points. In this method, simulations are done to explore different structures of molecules and then local minimization methods are used to minimize selected low-energy structures from the simulation trajectory. One major problem with the simulation techniques is the quality of the potential function used. Ideally, one would like to use high quality energy function, including the quantum mechanical (QM) or quantum mechanical/molecular mechanical (QM/MM) ones. However, the use of QM or QM/MM energy functions in molecular simulation is prohibitively expensive for investigations in a routine manner. The use of approximate energy functions can make the simulation feasible; however, the quality of structures obtained from such simulations could be poor and may be far from the stationary points on the QM energy surface.

The Two Surface Monte Carlo (TSMC) technique is a promising method to circumvent this problem. In this method, a cheap potential function is used as a biasing potential to propagate the more expensive potential function of interest. Jumps between the cheap and the expensive potential functions ensure that the structures sampled are with the expensive potential. It can be shown that this method of using a cheap potential to propagate an expensive potential is fully rigorous in the framework of the Monte Carlo method. Since exploration of the structures is done mostly by the cheap energy function, drastic reduction of computational cost can be achieved. For instance, the expensive potential can be a QM potential and the cheap one can be a classical potential. Warshel et al. [1] used this method in several investigations. Later, Iftimie et al. [2,3] used this method for different problems including proton transfer reaction with QM as the expensive and MM as the cheap potential. Gelb [4] used the same method with model potentials. Bandyopadhyay used this method in conjunction with the local optimization method to find the stationary points of an amino acid plus 12 water system in a semi-automated manner [5] by using the Effective Fragment Potential (EFP) [6,7] as the cheap potential and QM/EFP as the expensive potential. The success of this method depends on finding a suitable cheap potential, because if the two potential functions are not close to each other, jumps between the two surfaces will be rarely accepted, which in turn will make TSMC an inefficient method. In most of the previous works, construction of the cheap potential energy function turned out to be a major issue. The use of EFP in Ref. [5] suggested that TSMC method can be used efficiently if the cheap potential is accurate and can be generated on the fly as is true for EFP.

However, before the TSMC method could be used in a routine manner, it is necessary to assess its performance in a diverse set of problems. This may give avenues where it could be improved. In this endeavor, stationary points of water clusters of size 15 and 20 are investigated in the present work. This system is taken because both global minimum and several other structures are known for these clusters [8], yet finding those is non-trivial.

Global minima and other low energy structures of water clusters of different sizes were investigated by different authors using simple to complex potential functions with different optimization algorithms. Day et al. [9] explored the global minima of water clusters of size 3–5 using quantum chemical methods. Belchior et al. [10] used Genetic Algorithm and Density Functional Theory (DFT) to determine the global minima of $(H_2O)_n$ for $2 \leq n \leq 13$. Bulusu et al. [11] examined the global minima structures for $(H_2O)_{11}$ and $(H_2O)_{13}$ using the searching/screening approach. Goddard et al. [12] used DFT for structure and energy calculations of water clusters up to size 19. The global minima of water clusters from size 6 to 20 were investigated by Day et al. [13] using EFP . The TIP4P model was used by Tsai and Jordan [14] to find the global minima of water clusters of size 8–20. The TIP5P model was used to get the global minima of water clusters of size up to 21 by Wales et al. [15]. A comparative study of three potentials, namely SPC/E, TIP3P and TIP4P, was made for water cluster structure prediction up to size 25 by Karbrede and Hentschke [16]. In another work, Karbrede [17] examined how vibrational modes can be used in search for the global minima of water clusters of size between 25 and 30. Kazimirski and Buch [18] investigated low energy structures of water clusters of size 20, 22, 123 and 293.

The purpose of the present work is to assess the quality of sampling using TSMC and compare the structures obtained from the present work with the other published works. To focus solely on the sampling quality of TSMC, no chemical intuition was used to build the starting geometry of the clusters or used during the simulation.

The paper is arranged in the following way. Section 2 describes the TSMC method and EFP. Section 3 describes the details of simulation. This paper ends with a result and discussion section.

2 Method

2.1 Monte Carlo technique

In the Markov chain Monte Carlo method, for canonical ensemble, the states are generated with the limiting distribution $\exp(-\beta E_i)$ where $\beta = 1/kT$, and k, T and E_i are the Boltzmann constant, absolute temperature and energy of the state i, respectively. To get the correct limiting distribution, it is sufficient to enforce the condition of detailed balance as given by

$$P(x)T(x \to y) = P(y)T(y \to x) \tag{1}$$

where $P(x)$ is the probability of realizing the state x and $T(x \to y)$ is the transition probability of going from x to y. The Metropolis acceptance/rejection criterion in a canonical Monte Carlo simulation can be derived from (1) as

$$acc(x \to y) = \min\left(1, \frac{p(y)}{p(x)}\right) = \min\left(1, \exp(-\Delta E/kT\right) \tag{2}$$

where $\Delta E = E_y - E_x$.

In the case of two energy functions, the acceptance/rejection criterion between the two energy functions becomes [2].

$$acc(x \to y) = \min\left(1, \exp(-\Delta\Delta E/kT\right) \tag{3}$$

with $\Delta\Delta E = (E_{\text{EXP}}(y) - E_{\text{CHEAP}}(y)) - (E_{\text{EXP}}(x) - E_{\text{CHEAP}}(x))$.

2.2 Effective fragment potential method

The EFP method has been developed to represent interactions between the solute (treated quantum mechanically) and the solvent (treated by EFP) and also between the solvent molecules (EFP–EFP interactions) by considering fundamental intermolecular interactions. In the case of QM–EFP interaction, one electron terms are added to the *ab initio* Hamiltonian. In the original EFP model (EFP1), there are three types of terms. The first two terms are electrostatics and induction. The third term takes care of exchange–repulsion and charge transfer interactions. For details of the EFP methodology and application readers are referred to the Refs. [6,7].

3 Details of the Monte Carlo simulation

The initial structures of 15 and 20 water clusters were built manually by putting the water molecules close to each other. No chemical intuition was used other then not allowing the water molecules to overlap. Water molecules were represented by the RHF level of theory and EFP was used as the biasing potential (cheap potential). Short simulations with different basis sets showed that the 6-31+G* and DH(d,p) basis sets were giving maximum acceptance between the EFP and RHF energy. Subsequently, the 6-31+G* basis set was used for the simulation. No symmetry was used during the simulation. Internal geometry of the water molecules was fixed as in the EFP model. Several different simulation protocols were used. At first simulations were done at 300 K by putting the water molecules in boxes of size 12 × 12 × 12, and $8\times8\times8\,\text{Å}^3$ for 20 and 15 waters, respectively. Only one water molecule, chosen randomly, was moved in each Monte Carlo step. Both rotational and translation moves were used. The maximum displacements of rotational and translational moves were varied from 0.2 to 1.0 rad and Å, respectively, depending on the rate of acceptance. Different box sizes were also used to check the effect of box size on the rate of acceptance. Another set of simulation was performed at temperature 2,000 K. Altogether, 200,000 steps of simulation were done for 15 water cluster and 130,000 steps were done for 20 water cluster. The jump between the two energy surfaces was considered for every 400 steps. So, the number of RHF calculations done was 500 and 325 for 15 and 20 water clusters respectively. From the Monte Carlo structures obtained for the RHF energy function, local minimization technique was applied to the low energy structures to get the stationary points. Refinement of the structures obtained from the TSMC method was done with simulated annealing with EFP. All the calculations were done with a modified local version of GAMESS [19].

4 Results and discussion

4.1 Acceptance in Monte Carlo simulation

There are two types of acceptance in the performed TSMC simulation. First is the acceptance between two different geometries having EFP as the energy function (EFP–EFP). The second one is the acceptance between the two different energy functions (EFP–RHF) with the same geometry. It was found that the EFP–EFP acceptance was around 80% or more. However, the EFP–RHF acceptance was initially found to be poor. At 300 K, the acceptance was less than 10% for both 15 and 20 water clusters. Moreover, very often the structures were confined to some particular geometry from which no EFP–RHF acceptance was obtained for several thousand steps.

Several measures were taken to increase the acceptance. First, the simulation was run at 2,000 K. Second, when there was no acceptance for a long time, the two surface simulation was turned off and a regular MC simulation was performed with EFP before resuming TSMC. Also, it was found that the maximum step lengths for translational and rotational steps had important effects on the acceptance between the two surfaces. All these factors made the acceptance more than 30% for both 15 and 20 water clusters. It is interesting to note that although the EFP water model is known to be very close to the RHF theory, getting reasonable acceptance is non-trivial. One reason can be that while these two potentials are very similar near the equilibrium geometries, for geometries far from equilibrium, there might be some differences. It is to be noted that changing temperature and other parameters during the simulation breaks the Markov Chain of the Monte Carlo simulation, so no thermodynamic properties can be calculated from the obtained structures. However, this can still be used for getting stationary points on the RHF energy surface.

4.2 Stationary points

Several stationary points of 5–10 kcal/mol higher in energy than the global minimum of 15 water cluster were obtained from local optimization. For 20 water cluster, the stationary points are of even higher energy than the global minimum. The global minima for these two clusters are taken from Ref. [8]. Most of the structures obtained are of the cage type, not like the highly symmetric global minima for these two clusters (see Fig. 1b and 2b). To refine the structures, simulated annealing was done starting from several structures obtained by TSMC. Since TSMC is not yet implemented with simulated annealing, EFP was used in simulated annealing as implemented in GAMESS. Monte Carlo with minimization method of Li and Scherega [20] was used in the simulated annealing runs. In this method, Monte Carlo simulation and local minimization are combined. This effectively reduces the search space as a collection of local minima. Two different starting temperatures, 2,000 and 500 K, were used. Several of the runs could find structures close to the global minimum for 15 water cluster at the EFP level. One structure was 1.2 kcal/mol higher than the global minimum at the

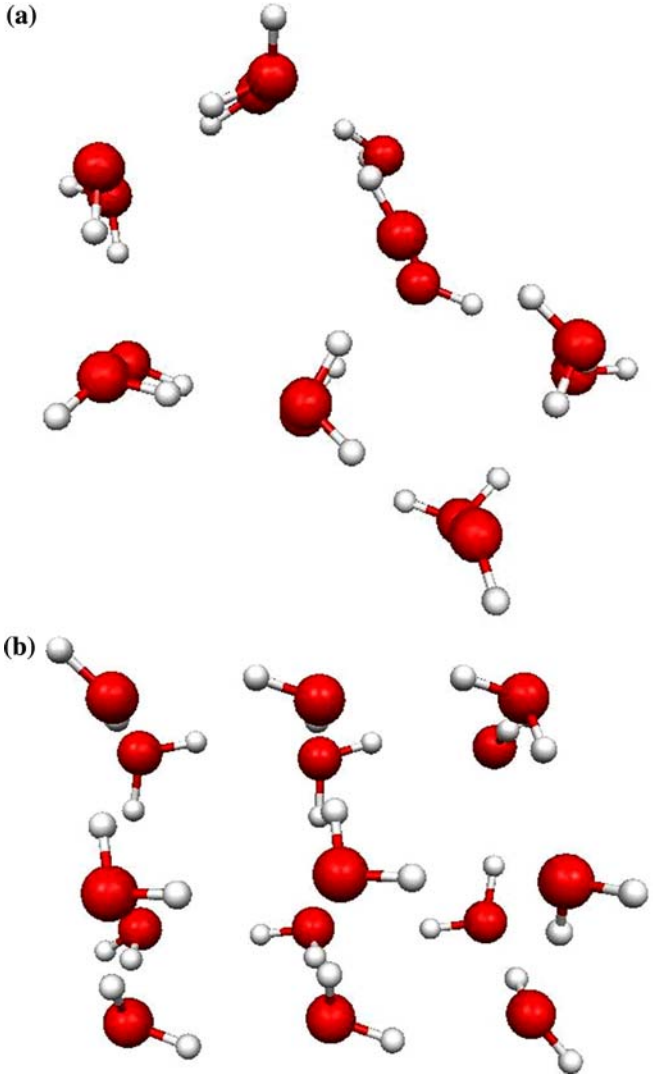

Fig. 1 **a** Lowest energy structure obtained for 15 water cluster in this work. **b** Global minimum of 15 water cluster as reported in Ref. [8]

RHF/6-31G** level (this basis set was used for comparison with the structure reported in Ref. [8]). Frequency calculations confirmed that it is a minimum. This structure is shown in Fig. 1a. By comparing with the global minimum (shown in Fig. 1b), it can be seen that the new structure can be thought to be two rings fused together; while, the structure in Fig. 1a is like three almost symmetric layers. For 20 water cluster, the minimum energy structure obtained was 10.4 kcal/mol higher than the global minimum at the RHF/6-31G** level. This structure is also confirmed to be a minimum by frequency calculation. This structure is shown in Fig. 2a along with the global minimum in Fig. 2b, which is much more symmetric. The ten lowest energy structures obtained from RHF/6-31G** optimization are shown in Table 1 for both 15 and 20 water clusters. It is to be noted that several other low energy structures were obtained at the EFP level (like several structures close to the structure shown in Fig. 1a); however, not all structures were optimized at the RHF level. TIP5P structures obtained by Wales [15] were also optimized at the RHF/6-31G** level of theory and it was found that the 15 and 20 water clusters were 3.1 and 3.6 kcal/mol higher in energy, respectively, than the global minima of Ref. [8].

Apart from the deficiency of the sampling technique, one reason for not finding any structure close to the global minimum of 20 water cluster might be that no symmetry was used in our simulation. It might be difficult to get a highly symmetric structure by an automated procedure, as used in the current study.

Though the success of finding global minimum in this work can be termed as modest, it is to be noted that the method described in this work is semi-automated. Moreover, no chemical intuition was used in simulation, and the simulation runs are by no means exhaustive. Preliminary investigations only with one TSMC simulation run for each cluster and subsequent simulated annealing method detected structure very close to the global minimum of 15 water cluster, though it did not find any structure close to the global minimum of 20 water cluster.

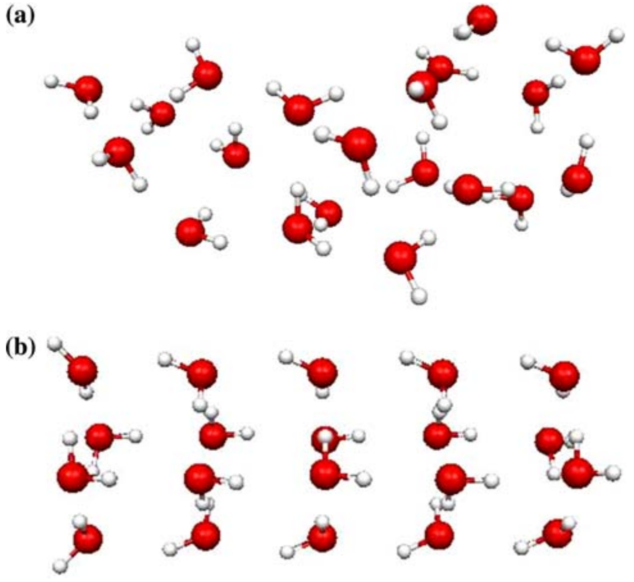

Fig. 2 **a** Lowest energy structure obtained for 20 water cluster in this work. **b** Global minimum of 20 water cluster as reported in Ref. [8]

Table 1 Relative energies (from the global minimum structures of Ref. [8]) of the optimized structures at the RHF/6-31G** level for 15 and 20 water clusters

Serial number	Relative energies/kcal/mol	
	$(H_2O)_{15}$	$(H_2O)_{20}$
1	1.2	10.4
2	8.3	15.4
3	8.7	17.5
4	9.0	21.9
5	12.0	22.2
6	12.0	22.3
7	12.5	23.6
8	12.7	24.8
9	13.2	26.5
10	13.7	28.0

One pertinent issue is the advantage of using TSMC as opposed to doing simulation of EFP waters alone and then optimizing those with RHF theory. Since, EFP water potential is very close to the QM, EFP only simulation may give structures similar to the QM ones for water clusters. However, it is interesting to note that at structures far from equilibrium, there are some differences between EFP and QM. Moreover, in a more general sense, if both solute and waters are present, the difference between EFP and QM is likely to increase, and TSMC method can play an important role in getting structures for these systems. The current work uses only waters, since this is one of the best case scenarios as far as the closeness of two potential functions are concerned.

4.3 Final conclusion

This work describes the viability of using the TSMC method in getting stationary points of large molecular systems, using an automated procedure. Previously studied 15 and 20 water clusters were taken as our test system. Effective Fragment Potential, which is a very accurate potential for water molecules, was taken as the biasing potential. Preliminary examination of long MC simulation reveals that local optimizations of the structures obtained from the simulation did not give any structure very close to the global minimum for either 15 or 20 water clusters. Examination of the acceptance between the two energy surfaces shows that even with a good potential like EFP, at 300 K the acceptance rate is poor. Only after performing the simulation at 2,000 K and tuning other simulation parameters, an acceptance rate over 30% was obtained. However, the naïve implementation of TSMC did not find structures close to the global minimum. Simulated annealing, starting from the several structures obtained from TSMC, found structure about 1 kcal/mol higher in energy than the global minimum of 15 cluster, although no structure close to the global minimum of 20 water cluster was obtained. A more extensive investigation of the same system by TSMC and simulated annealing is in progress, which will be reported in a future publication. However, the current manuscript showed some important issues regarding the use of TSMC in geometry optimization of large molecules, which will be used in subsequent works.

Acknowledgment The author is thankful to Prof. T. Head-Gordon for introducing him to the TSMC method. He is grateful to Prof. N. Sathyamurthy and Mr. C. N. Ramachandran for their helpful discussions. This work is dedicated to Prof. Mark S. Gordon on the occasion of his 65th birthday.

References

1. Muller RP, Warshel A (1995) J Phys Chem 99:17516
2. Iftimie R, Salahub D, Wei D, Schofield J (2000) J Chem Phys 113:4852
3. Iftimie R, Schofield J (2001) J Chem Phys 114:6763
4. Gelb LD (2003) J Chem Phys 118:7747
5. Bandyopadhyay P (2005) J Chem Phys 122:091102
6. Gordon MS, Freitag M, Bandyopadhyay P, Jensen JH, Kairys V, Stevens WJ (2001) J Phys Chem A 105:293
7. Bandyopadhyay P, Gordon MS (2000) J Chem Phys 113:1104
8. Maheshwary S, Patel N, Sathyamurthy N, Kulkarni AD, Gadre SR (2001) J Phys Chem A 105:10525
9. Day MB, Kirschner KN, Shields GC (2005) J Phys Chem A 109:6773
10. Silva ESA, Duarte HA, Belchior JC (2006) Chem Phys 323:553
11. Bulusu S, Yoo S, Apra E, Xantheas S, Zeng XC (2006) J Phys Chem A 110:11781
12. Su JT, Xu X, Goddard WA III (2004) J Phys Chem A 108:10518
13. Day PN, Pachter R, Gordon MS, Merill GN (2000) J Chem Phys 112:2063

14. Tsai CJ, Jordan KD (1993) J Phys Chem 97:5208
15. James T, Wales DJ, Hernandez-Rojas J (2005) Chem Phys Lett 415:302
16. Kabrede H, Hentschke R (2003) J Phys Chem B 107:3914
17. Kabrede H (2006) Chem Phys Lett 430:336
18. Kazimirski JK, Buch V (2003) J Phys Chem A 107:9762
19. Schmidt MW, Baldridge KK, Boatz JA, Elbert ST, Gordon MS, Jensen JH, Koseki S, Matsunaga N, Nguyen KA, Su SJ, Windus TL (1993) J Comp Chem 14:1347
20. Li Z, Scheraga HA (1987) PNAS 84:6611