Asian Journal of Chemistry Vol. 21, No. 6 (2009), 4831-4842

Effects of Molecular Numbers and Time Steps on the Estimation of Diffusion Coefficients in Molecular Dynamic Simulation

QINGZHI LIU
College of Science, Qingdao Nongye University, 266109 Qingdao, P.R. China
E-mail: liuqz2001@163.com

Using molecular dynamics simulation, the diffusions of glycine, methane, ammonia and carbon dioxide in water were simulated in the canonical NVT ensemble. This study shows that the molecular number of the sample and the time step used are two important factors for the simulation. The diffusion coefficients increase with increasing molecular numbers and gradually approach to the experimental value. When two samples are big enough that their difference is not obvious, the sample with smaller molecules should be applied. Time step also plays a significant role in the estimation of diffusion coefficients. The estimated diffusion coefficients increase with time step and reach maximum before they start to decline. This paper suggests several time steps between 0.1 fs to 1.5 fs be taken in the simulation of a special system and take 70 % of the time step corresponding to maximal diffusion coefficient value as the reasonable time step.

Key Words: Time step, Molecular dynamic simulation, Diffusion coefficient, Sample.

INTRODUCTION

In the past few years, the growth of computational power and the development of theoretical models have increased the possibility of describing various physical processes by means of computer simulations. It is well known that the diffusion coefficients play an important role in the description of chemical processes in solution.

Much work has been done to get this parameter. For example, the diffusion coefficients of amino acids in aqueous solution were estimated using holographic interferometric technique[1,2] and the diffusion coefficients of L-proline, L-threonine and L-arginine in aqueous solution were estimated using metallic diaphragm cell method[3,4]. Similarly, Wise and Houghton[5] measured the diffusion coefficients of neon, krypton, xenon, carbon monoxide and nitric oxide in water at 10-60 ºC by microscopically following the collapse of small stationary bubbles against a horizontal wall. Verhallen[6] has measured the diffusion coefficients of helium, hydrogen, oxygen and nitrogen in water at temperatures between 10 and 60 ºC by taking the permeability of a stagnant liquid layer in the quasi-steady state method.

However, above methods are time consuming and expensive and still lack experimental values. If computer simulations could present ideal structure of system and reasonable motion parameter of molecule, we can avoid expensive experiments, especially for systems at critical situation which experiment can't implement. Now, this field is mature enough that it can provide reliable predictions in many instances and it shows an unprecedented ability to extrapolate prediction from small to complex molecules, from low to high pressures and from low temperatures to high temperatures. As such, molecular simulation is now considered in the chemical industry as capable of filling the gap between experimental data and engineering models in various circumstances: unknown chemicals, extreme conditions of temperature or pressure, toxic compounds, among others[7]. Many researches have been directed toward studying small molecule water system in the literature. An infinite dilute aqueous solution of formaldehyde has been carried out by molecular dynamics calculation[8,9]. Mean square displacement is calculated with different potentials and diffusion coefficient is evaluated from the slope of the mean square displacement according to Einstein's relationship. A diffusion of 0.6 M potassium nitrate aqueous solution has been calculated by molecular dynamics and Monte Carlo calculation[10]. Molecular dynamics is used to calculate the pure diffusivities and MC is used to calculate the effective diffusivities. The diffusion coefficients also follow the Einstein's relationship.

In the molecular dynamics (MD) method, the atomic positions and the velocities of atoms are evaluated after time steps conveniently small through the numerical integration of Newton's equation of motion. The sequence of events recorded at atomistic resolution results in a dynamic representation of the overall transport phenomena predicting the whole transport process. Molecular dynamics can in fact allow the estimation of the fraction of small molecules that can cross the potential energy barriers between adjacent locations of the water matrix.

The current restriction on fully atomistic modeling is due to the size of the system studied and to the time scale of the process. Previous studies have taken periodic boundary condition to make up the lack of fully atomistic modeling. Long time step is taken to save CPU time in order to broaden time scale of the process. Periodic boundary condition enables a simulation to be performed using a relatively small number of particles, in such a way, that the particles experience forces as if they were in bulk fluid. However, the limited number of particles can not represent the real system. This may bring error to simulation result. Time step is an important parameter in simulation process, too small and the trajectory will cover only a limited proportion of the phase space; too large and instabilities may arise in the integration algorithm due to high energy overlaps between atoms. Till now, there are no hard and fast rules for calculating the most appropriate time step to use in a molecular dynamics simulation. In this study, different samples and time steps will be taken for several systems (glycine, oxygen, ammonia, carbon dioxide and methane in water) and the effects of samples and time steps will be studied.

The main goal of our molecular dynamics simulations is essentially to make research of the effects of samples and time steps on diffusion coefficients and to give deep insight in the mechanism of simulating method.

EXPERIMENTAL

The simulation was performed by means of MS/Discover software. For the potential energy, the COMPASS[11] force-field is applied to all the simulations. The calculation is performed on Pentium4 3.0E. Non-bonded interactions are described as the sum of Lennard-Jones and coulomb interactions. Water in the COMPASS force-field is modeled as a flexible molecule with a single Lennard-Jones site centered on the oxygen atom and partial charges of -0.82 e on the oxygen and +0.41 e on each hydrogen.

A residue based cut-off of 10 Å is used for all interactions. Molecular dynamics runs are performed at temperature of 298.15 K for a closed system at constant volume. The system is thermostated by Andersen method. Periodic velocity rescaling is used to approximate a constant (NVT) trajectory. The initial configuration of this system is composed of an orthorhombic cell with periodic-boundary in three directions. The initial direction of water is random. The number of configurations is 10. The initial position of gas is taken random. The initial velocity of each molecular in cell is set by Maxwell distributing method, then the velocity was reconfirmed to ensure the total quantum of system to be zero. Total simulation time is at least 40 ps and frame output every 250 steps. The time step is 0.1 femtosecond, 0.5 femtosecond, 0.7 femtosecond, 1.0 femtosecond and 1.5 femtosecond, respectively when considering the effect of time step.

During the simulation processes, the calculation of non-bonded parameter may take a lot of CPU time. In normal situation, atom-based method is chosen for vdW interaction and Ewald method for static interaction. Considering that present system is mainly composed of small molecules, we chose group-based method for both vdW and static interaction. It has been proved that this method can save much CPU time without interfering with the precision of simulation result.

Calculation of diffusion coefficient: The MSD (mean square displacement) curves are usually obtained by analyzing the particle positions recorded in molecular dynamics simulation. There are two methods for analyzing the MSD curves: the Einstein model and the differentiation-interval variation method.

According to Einstein model for calculating diffusion coefficient:

$$D_\alpha = \frac{1}{6N_\alpha} \lim \frac{d}{dt} \sum_{i=1}^{N\alpha} < [r_i(t) - r_i(0)]^2 >$$

Data averaging is performed over a large time scale. However, for MSD curves, there are always irrational molecule movements as shown in Fig. 1. The Einstein model treats these data as normal data. Consequently,they are involved in the calculation of the final results, leading to errors in the calculated diffusion coefficients.

As a new method, the differentiation-interval variation analysis[12] can overcome the defects of the Einstein model. In this method, the MSD curves are obtained through mean squared displacement analysis; then the average diffusion coefficient $\overline{D}$ can be obtained by analyzing the curves based on Differentiation-Interval variation analysis.

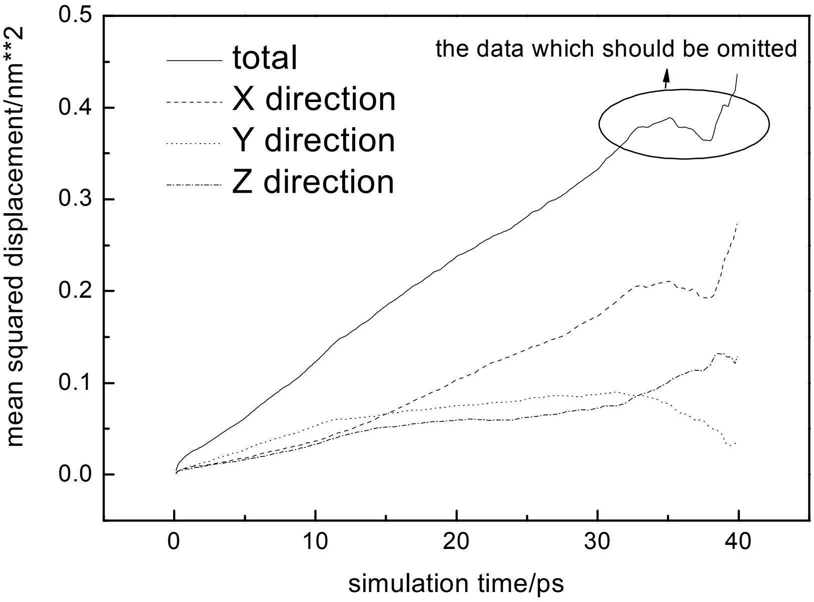

Fig. 1. Mean square displacement curves of aqueous oxygen solution

In the differentiation-interval variation analysis method, an MSD curve is first divided into n segments. The diffusion coefficient of the jth segment is as follows:

$$D_j = \frac{1}{2}\left[\frac{y_{j+1} - y_j}{x_{j+1} - x_j} + \frac{y_j - y_{j-1}}{x_j - x_{j-1}}\right] \tag{1}$$

The diffusion coefficient of the entire system is the average of the above diffusion coefficients.

$$\overline{D} \equiv \frac{1}{n}\sum_{j=1}^{n} D_j \tag{2}$$

$$\sigma^2 \equiv \frac{1}{n-1}\sum_{j=1`}^{n}(D_j - \overline{D})^2 \tag{3}$$

Define function Z as

$$Z \equiv \frac{\sigma^2 t_{max}^2}{b(b-a)} \tag{4}$$

The restrictive conditions are assumed to be as follows: The values of $[(b\text{-}a)/t_{max}]$ and b/t_{max} are maxima and that of Z is minimum. Then the value of diffusion coefficient $\overline{D}$ can be estimated by changing the values of a and b, which are time points on the simulation line. They also represent the reasonable initial and end times on the MSD curve.

RESULTS AND DISCUSSION

Calculation and analysis of the diffusion coefficient of glycine: Three glycine aqueous solution systems with different concentrations (0.5, 0.7 and 0.9 mol/L) were simulated. The position of every glycine molecule was recorded after simulation. The MSD curve is shown in Fig. 2. It is evident that the slope/diffusion coefficient decreases with the increase of the concentration of glycine in water from 5 to 25 ps time interval. This is in accordance with experiment. This also illustrates the fact that simulating method can provide a reasonable value for certain systems. Relationships between different systems can be described through the MSD curve.

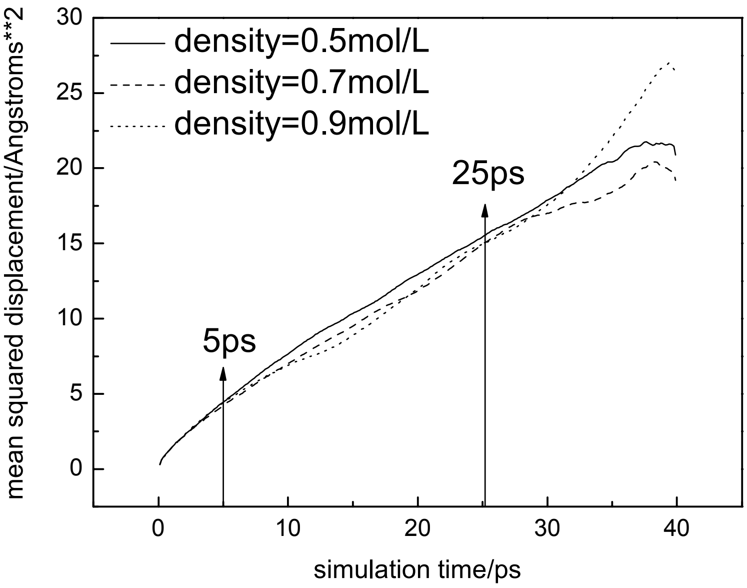

Fig. 2. Comparison of mean square displacement curves of aqueous glycine solution with different concentration

In order to get the accurate diffusion coefficient value, MSD data need to be analyzed by differentiation-interval variation analysis. The curves produced from this method are shown in Fig. 3a. The value of Z is at the minimum and that of $[(b-a)/t_{max}]$ is at the maximum when a and b are 0.375 and 34.125 ps, respectively, t_{max} is 40 ps. $\overline{D}$ is the final diffusion coefficient. Diffusion coefficients of different concentrations can also be derived using the same method (Figs. 3b and c). Table-1 shows the contrast of simulation values and experimental values of glycine in water. The last column lists the simulation values of the three different concentrations and the second column lists the corresponding experimental values.The simulated values are only a little smaller than the experimental values, with an error of less than ± 7 %.

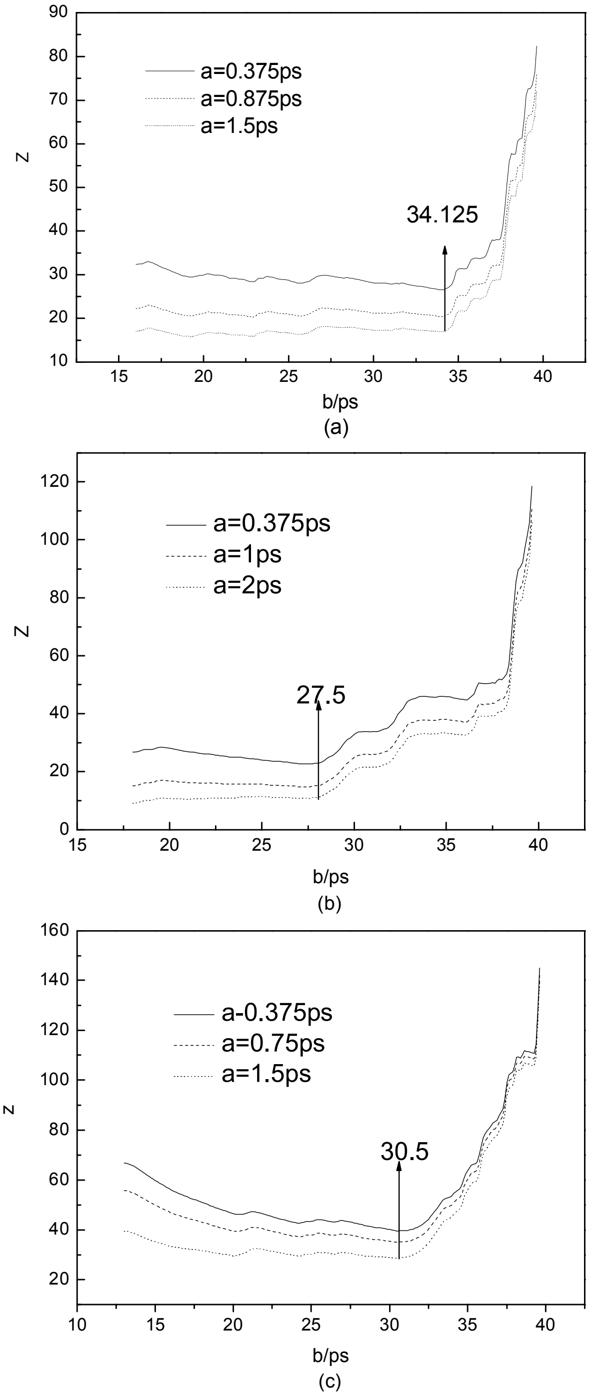

Fig. 3. Differentiation-interval variation analysis curves of aqueous glycine solution at different concentrations. Panel (a), (b) and (c) show the curves at the concentration of 0.5, 0.7 and 0.9 mol/L, respectively

TABLE-1
SIMULATION VALUES AND EXPERIMENTAL VALUE OF GLYCINE IN WATER

Concentration (mol/L)	Experimental value (10^{-6} cm^2 s^{-1}) [Ref. 3] (298.15 K)	Simulation value (10^{-6} cm^2 s$^{-1)}$ (298.15 K)
0.5	9.73	9.62
0.7	9.57	9.54
0.9	9.43	9.23

Effect of sampled molecular numbers: Periodic boundary conditions enable the simulation to be performed using a relatively small number of particles. However, the limited number of particles might not represent the real system. In order to examine whether the sampled molecular numbers can affect the simulation result in the periodic boundary condition and to know the extent of this effect, two glycine systems with different concentrations (0.7 mol/L, 0.9 mol/L) were taken for simulation. Different molecular numbers were used for the two systems. 3, 5, 7, 10 and 12 molecules were used for the 0.7 mol/L glycine and 4, 7, 9, 13 and 15 molecules were used for the 0.9 mol/L glycine. In order to ensure constant density of the system, water molecules were added proportionably. The results show that for both of the systems, the values of diffusion coefficients increase with the number of molecules and approach to a constant value gradually (Fig. 4). Take the 0.7 mol/L glycine system as an example, the diffusion coefficients of the system with 10 molecules is nearly equal to that with 12 molecules (9.43 and 9.54 for the sampling of 10 and 12 molecules, respectively).

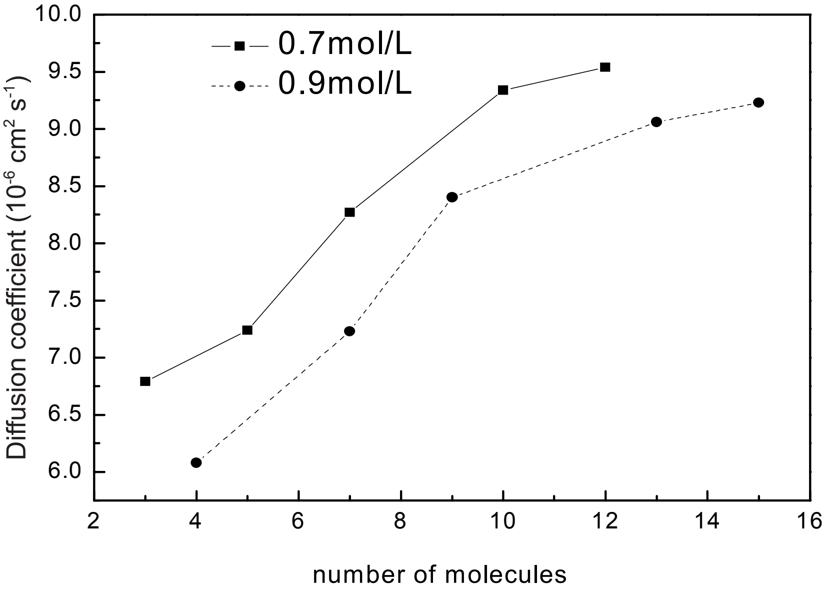

Fig. 4. Effect of molecule number sample on diffusion coefficient (glycine system)

Effect of time step: The selection of time step is vital for the accurate estimation of diffusion coefficients. For the molecular dynamics simulation process, a time step that's too small will cause the trajectory to cover only a limited proportion of the phase space. On the other hand, a time step that's too large will lead to instabilities in the integration algorithm due to high energy overlaps between atoms. However, time step has not been regarded as an important parameter for simulation before. For the systems that are mainly composed of small molecules, small time steps such as 0.5 fs have been widely used to avoid high speed circle and ensure the stability of function. In this study, the role of time steps on the estimation of the diffusion coefficients are tested. An aqueous methane system is used as an example.

First the time step of 0.5 femtosecond is used. The resulting MSD curve is shown in Fig. 5 and the corresponding differentiation-interval variation curve is shown in Fig. 6. The estimated value of $\overline{D}$ is 1.17, which is not very ideal compared with the experimental value of 1.49.

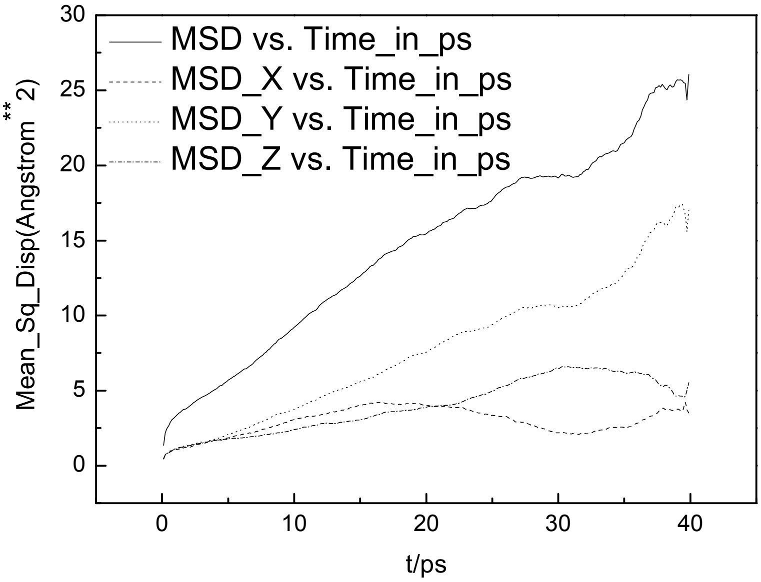

Fig. 5. MSD curve of methane in water

When the time step of 0.5 fs is used, there may be two reasons for the error. One is that 0.5 femtosecond time step is too long for this system, small time step should be taken the other is that 0.5 femtosecond is too short for this system. Simulating time is extended to make up for the error of small time step. The total simulating time is extended to 80 ps. Diffusion coefficient value is 1.23, there is no obvious difference between 40 ps and 80 ps total simulating time, which means that total time has no effect in final result over 40 ps. The total times in following simulation are all set 40 ps.

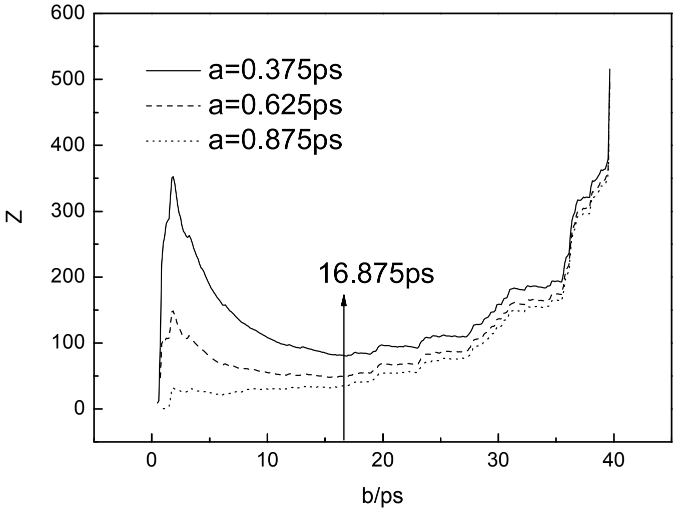

Fig. 6. Differentiation-interval variation curve of methane in water

Time steps of 0.1, 0.7 and 1.0 fs were then used for the system. All of their simulated values and corresponding experimental values are shown in Table-2. The results show that the time step has an obvious effect on diffusion coefficient of certain system. When 0.1 fs is used to make up for the error of large time step, the result is even worse than when the time step of 0.5 fs is used (Table-2). It seems that times steps of 0.7 and 1.5 fs, which produce results that are most close to the experimental values, are more appropriate than other time steps.

TABLE-2
SIMULATING VALUE OF DIFFUSION COEFFICIENT OF METHANE IN WATER WITH DIFFERENT TIME STEP

Time step (fs)	0.1	0.5	0.5	0.7	1.0	1.5
Simulating time (ps)	40	40	80	40	40	40
Simulating value of diffusion coeff. (10^{-9} m^2 s^{-1})	0.45	1.17	1.23	1.45	1.94	1.52
Exp. value of diffusion coeff. (10^{-9} m^2 s^{-1})	1.49 [**Ref. 13**]					

The estimated diffusion coefficients increase with time steps and maximize at about 1.0 fs, after which, they start to decline (Fig. 7). The diffusion coefficients when the time steps of 0.7 and 1.5 fs are used (neighboring on the maximum) are most close to the experimental value.

Simulation values of ammonia and carbon dioxide aqueous systems show a similar trend (Tables 3 and 4). For the ammonium system, 0.5 femtosecond and 1.0 femtosecond are reasonable time steps (Table-3). For the carbon dioxide system, 0.7 femtosecond and 1.5 femtosecond are reasonable time steps. One shared feature of all these systems is that the ideal time steps are all close to the time step that produces maximum value.

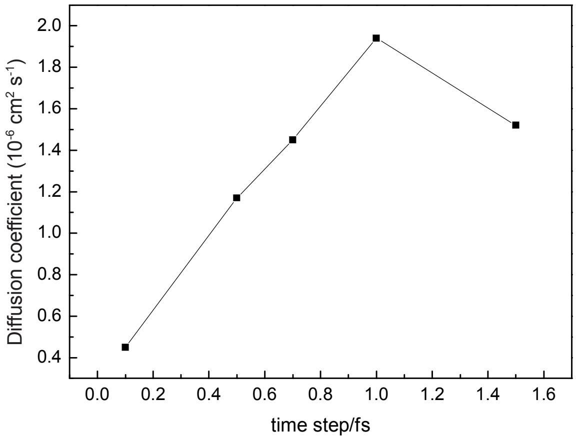

Fig. 7. Curve of diffusion coefficients with different time step

TABLE-3
SIMULATING VALUE OF DIFFUSION COEFFICIENT OF AMMONIA IN
WATER WITH DIFFERENT TIME STEP

Time step (fs)	0.1	0.1	0.1	0.5	0.7	1.0
Simulating time(ps)	40	80	120	40	40	40
Simu. value of diffusion coeff. (10^{-9} m^2 s^{-1})	0.55	0.52	0.62	1.37	2.24	1.41
Exp. value of diffusion coeff. (10^{-9} m^2 s^{-1})	1.64 [Ref. 13]					

TABLE-4
SIMULATING VALUE OF DIFFUSION COEFFICIENT OF CARBON
DIOXIDE IN WATER WITH DIFFERENT TIME STEP

Time step (fs)	0.1	0.5	0.7	1.0	1.5	1.0
Simulating time (ps)	40	40	40	40	40	40
Simu. value of diffusion coeff. (10^{-9} m^2 s^{-1})	0.54	1.07	1.36	3.07	1.89	1.41
Exp. value of diffusion coeff. (10^{-9} m^2 s^{-1})	1.92[13]					

Considering the effect of time step in the simulation, this paper advises several time steps between 0.1 fs to 1.5 fs be taken in the simulation of a special system and take 70 % of the time step corresponding to maximum diffusion coefficient value as the reasonable value.

Conclusion

The study was intended to find a good simulation method for deriving diffusion coefficient of certain kinds of molecules. The molecular numbers of the sample and the time steps are considered as important factors. Using molecular dynamics (MD) simulation, diffusion of glycine, oxygen, methane and ammonia in water are simulated in constant (NVT) ensemble trajectory. Results showed that differentiation-interval variation analysis method is effective for analyzing simulating data.

Molecular numbers of the sample has an evident effect on diffusion coefficients. The value of diffusion coefficient increases with the number of sampled molecules and approaches to a constant value gradually. In most cases, the sample with the most molecules produces values most close to the experimental value than other molecular numbers. Since more time will be consumed for big sample system than small sample systems, when two samples are big enough that their difference is not obvious the small sample should be applied.

Different time steps are chosen in the simulation process. Result shows that the value of time step has an evident effect on diffusion coefficient. The estimated diffusion coefficients increase with time step and maximize before they decline. The two time steps that are about 70 and 150 % of the time step corresponding to the maximal diffusion coefficients, produce diffusion coefficients that are most close to the experimental value. Considering the effect of time step in the simulation, this paper advises several time steps between 0.1 fs to 1.5 fs be taken in the simulation of a special system and take 70 % of the time step corresponding to maximum diffusion coefficient value as the reasonable value.

We demonstrate that in order to capture the sensitive dependence of the diffusivities on molecular number sample and time step, it is necessary to use system with larger molecular number for simulation and run simulations with multiple time steps to find the reasonable time step for a certain system. This study shows that 0.5 and 0.7 fs time step might be typical for small molecular system.

Nomenclature

a,b	The certain time among simulating process, b > a as usual, ps
$\overline{D}$, D_α, D_j	Mean diffusion coefficient value of differentiation-interval variation analysis method, diffusion coefficient value of Einstein method, diffusion coefficient of the jth time space of differentiation-interval variation analysis method, $Å^2.ps^{-1}$
E_{elec} E_{vdW}	Static interaction energy, vdW interaction energy, kcal mol^{-1}
ε_{ij}	Dielectric constant
n	The number of segment divided in differentiation-interval variation analysis method
N_α	Atom number in system
σ^2	The square difference of D_j and $\overline{D}$, m^{-2}
q_i, q_j	Interactive charge
$r_{i(t)}$, $r_{i(0)}$, r_{ij}	The displacement at t time, the initial displacement, the effective radii of vdW interaction, Å
t_{max}	Simulating time, ps
x_{j-1}, x_j, x_{j+1}	Value of different time on mean square displacement curve, ps
y_{i-1}, y_i, y_{i+1}	Corresponding value of mean square displacement with different time, $Å^2$
Z	User-defined function.

REFERENCES

1. Y.X. Wu, P.S. Ma, Y.Q. Liu and S.F. Li, *Fluid Phase Equilib.*, **186**, 27 (2001).
2. C.W. Zhao, J.D. Li, P.S. Ma and S.Q. Xia, *Chin. J. Chem. Eng.*, **13**, 285 (2005).
3. Y.G. Ma, C.Y. Zhu, P.S. Ma and K.T. Yu, *J. Chem. Eng. Data*, **50**, 1192 (2005).
4. C.W. Zhao and P.S. Ma, *Chin. J. Chem. Eng.*, **12**, 267 (2004).
5. D.L. Wise and G. Houghton, *Chem. Eng. Sci.*, **23**, 1211 (1968).
6. P.T.H.M. Verhallen, L.J.P. Oomen, A.J.J.M.V.D. Elsen, A.J. Kruger and J.M.H. Furtuin, *Chem. Eng. Sci.*, **39**, 1535 (1984).
7. P. Ungerer, C. Nieto-Draghi, B. Rousseau, G. Ahunbay and V. Lachet, *J. Mol. Liqui.*, **134**, 71 (2007).
8. S. Tolosa and J.A. Sansdn, *J. Chem. Phys.*, **213**, 203 (1996).
9. S. Tolosa and J.A. Sansdn, *J. Chem. Phys.*, **223**, 251 (1997).
10. Florian Fergg and Frerich J. Keil, *Comp. Chem. Eng.*, **22**, S611 (1998).
11. H. Sun, *J. Phys. Chem. B*, **139**, 7338 (1998).
12. M.B. Chen, R.X. Zhu and J.H. Zhou, Variation Principle in Simulation of Diffusion Coefficient, The Ninth National Quantum Chemistry Academic Meeting, Theory Group Report, Chinese Chemical Society, GuiLin, p. 1 (2005).
13. J. Zhou, X.H. Lu, Y.R. Wang and J. Shi, *J. Chem. Eng. Chinese. Univ.*, **14**, 1 (2000).

(*Received*: 15 November 2008; *Accepted*: 4 April 2009) AJC-7400