[Article]

doi: 10.3866/PKU.WHXB201803161

www.whxb.pku.edu.cn

On the Simulation of Complex Reactions Using Replica Exchange Molecular Dynamics (REMD)

XIN Liang^{1,3}, SUN Huai^{1,2,3,*}¹ School of Chemistry and Chemical Engineering, Shanghai Jiao Tong University, Shanghai 200240, P. R. China.² Materials Genome Initiative Center, Shanghai Jiao Tong University, Shanghai 200240, P. R. China.³ State Key Laboratory of Inorganic Synthesis and Preparative Chemistry, Jilin University, Changchun 130012, P. R. China.

Abstract: A complex reaction, such as combustion, polymerization, and zeolite synthesis, involves a large number of elementary reactions and chemical species. Given a set of elementary reactions, the apparent reaction rates, population of chemical species, and energy distribution as functions of time can be derived using deterministic or stochastic kinetic models. However, for many complex reactions, the corresponding elementary reactions are unknown. Molecular dynamics (MD) simulation, which is based on forces calculated by using either quantum mechanical methods or pre-parameterized reactive force fields, offers a possibility to probe the reaction mechanism from the first principles. Unfortunately, most reactions take place on timescales far above that of molecular simulation, which is considered to

be a well-known rare event problem. The molecules may undergo numerous collisions and follow many pathways to find a favorable route to react. Often, the simulation trajectory can be trapped in a local minimum separated from others by high free-energy barriers; thus, crossing these barriers requires prohibitively long simulation times. Due to this timescale limitation, simulations are often conducted on very small systems or at unrealistically high temperatures, which might hinder their validity. In order to model complex reactions under conditions comparable with those of the experiments, enhanced sampling techniques are required. The replica exchange molecular dynamics (REMD) is one of the most popular enhance sampling techniques. By running multiple replicas of a simulation system using one or several controlling variables and exchanging the replicas according to the Metropolis acceptance rule, the phase space can be explored more efficiently. However, most published work on the REMD method focuses on the conformational changes of biological molecules or simple reactions that can be described by a reaction coordinate. The optimized parameters of such simulations may not be suitable for simulations of complex reactions, in which the energy changes are much more dramatic than those associated with conformational changes and the hundreds elementary reactions through numerous pathways are unknown prior to the simulations. Therefore, it is necessary to investigate how to use the REMD method efficiently for the simulation of complex reactions. In this work, we examined the REMD method using temperature (T-REMD) and Hamiltonian (H-REMD) as the controlling variable respectively. In order to quantitatively validate the simulation results against direct simulations and analytic solutions, we performed the study based on a simple replacement reaction ($A + BC = AB + C$) with variable energy barrier heights and reaction energies described using the ReaxFF functional forms. The aim was to optimize the simulation parameters including number, sequence, and swap frequency of the replicas. The T-REMD method was found to be efficient for modeling exothermic reactions of modest reaction energy ($<3 \text{ kcal}\cdot\text{mol}^{-1}$) or activation energy (*ca.* $< 20 \text{ kcal}\cdot\text{mol}^{-1}$). The efficiency was severely impaired for reactions with high activation and reaction energies. The analysis of the simulation trajectory revealed that the problem was intrinsic and could not be solved by adjusting the simulation parameters since the phase space sampled using T-REMD was localized in the region favored by high (artificial for speed-up) temperatures, which is different from the region favored by low (experimental) temperatures. This issue was

Received: February 8, 2018; Revised: March 13, 2018; Accepted: March 13, 2018; Published online: March 16, 2018.

*Corresponding author. Email: huaisun@sjtu.edu.cn.

The project was supported by the National Natural Science Foundation of China (21073119, 21173146, 21473112) and National Basic Research Program of China (973) (2014CB239702).

国家自然科学基金(21073119, 21173146, 21473112)与国家重点基础研究发展规划项目(973) (2014CB239702)资助

© Editorial office of Acta Physico-Chimica Sinica

aggravated in the case of endothermic reactions. On the other hand, the H-REMD run on a series of potential surfaces having different activation energies was demonstrated to be remarkably robust. Since the energy barrier only reduces the reaction rates, while the phase space controlled by the reaction energy differences remains unchanged at a fixed temperature, excellent results were obtained with fewer replicas by using H-REMD. It is evident that H-REMD is a more suitable method for the simulation of complex reactions.

Key Words: Replica exchange; Molecular dynamics; Complex reaction; Temperature; Hamilton

关于副本交换分子动力学模拟复杂化学反应的研究

辛亮^{1,3}, 孙淮^{1,2,3,*}

¹上海交通大学化学化工学院, 上海 200240

²上海交通大学材料基因组联合研究中心, 上海 200240

³吉林大学无机合成与制备化学国家重点实验室, 长春 130012

摘要: 本文研究用温度副本交换分子动力学(T-REMD)和哈密顿副本交换分子动力学(H-REMD)方法模拟复杂化学反应的问题。使用具有不同活化能和反应能的简单置换反应模型, 我们检验了上述两种方法用来预测反应平衡产物的效率和应用范围。T-REMD方法对具有适度活化能(约 $< 20 \text{ kcal}\cdot\text{mol}^{-1}$)或者反应能量($< 3 \text{ kcal}\cdot\text{mol}^{-1}$)的放热反应是有效的。由于在相空间的不完整采样, 对于同时具有高活化能和反应能量的反应其模拟效率有严重障碍, 并且对于吸热反应问题更为显著。另一方面, H-REMD对一系列具有不同活化能的反应能的模型表现出色, 与T-REMD相比, H-REMD可以使用更少的副本获得优异的结果。

关键词: 副本交换; 分子动力学; 复杂反应; 温度; 哈密顿
中图分类号: O645

1 Introduction

Complex reaction, such as combustion, polymerization and zeolite synthesis, consists of a large number of elementary reactions and chemical species¹. The apparent reaction kinetics can be derived using deterministic or stochastic models based on a knowledge of the elementary reactions². However, for many complex reactions the elementary reactions are unknown. Molecular dynamics (MD) simulation using either forces calculated quantum mechanically³ or pre-parameterized reactive force field⁴ offers a possibility to reveal the reaction mechanism from the first principles. However, most reactions take place in time-scales far above that of molecular simulation. The molecules may undergo numerous collisions and follow many pathways to find a favorable route to react. Often the simulation trajectory is trapped in a local minimum separated from others by high free-energy barriers, crossing the barrier requires prohibitively long simulation time. Because of the time-scale limitation, the simulations have been conducted on very small systems or at unrealistically high temperatures, both have hindered the validity of the simulation. To model complex reactions at conditions comparable with that of experiment, enhanced sampling techniques are required.

The replica exchange molecular dynamics (REMD)⁵ is one of the well-accepted techniques of enhanced sampling. By running multiple replicas of the simulation system in one or several controlling variables such as temperatures⁶ and potential energy surfaces⁷, and exchanging the replicas by Metropolis acceptance

rule, the phase space can be explored more thoroughly. The controlling variables are not limited to temperature and potential energy, others such as coulombic energy⁸, van der Waals (VDW) energy⁹, pH value¹⁰ and surface tension¹¹ have been applied. Although the kinetic information is not directly accessible because of the replica exchanges, recent developments have shown that kinetic information can be derived from REMD simulations^{12,13}.

Most published applications of the REMD method are focusing on conformational changes of biological molecules¹⁴⁻¹⁶ or simple reactions which can be described by a simple reaction coordinate^{17,18}. In this work we explore the feasibility of using REMD methods to simulate complex reactions in which the pathways are numerous and unknown prior to the simulation. We compared the temperature replica exchange (T-REMD) and Hamiltonian replica exchange (H-REMD) molecular dynamics methods. The simulations were carried out using a model reaction described using the ReaxFF functional forms.

In the following section we present how the two methods were carried out and the simulation parameters were optimized. It is followed by the results on different potential models and discussions on the robustness and limitation of the two REMD methods. Finally, we summarize the main findings of this work in the last section.

2 Model and method

A simple substitution reaction, $\text{AB} + \text{C} \rightleftharpoons \text{A} + \text{BC}$, is used in

this study. The potential energy function is written using the ReaxFF formula⁴ expressed in four terms:

$$E_{\text{system}} = E_{\text{bond}} + E_{\text{val}} + E_{\text{pen}} + E_{\text{VDW}} \quad (1)$$

The bond (E_{bond}), angle (E_{val}) and coordination penalty (E_{pen}) terms are written as functions of the bond order (BO), which is a function of interatomic distance.

The VDW term (E_{VDW}) is represented by a tapered Morse potential. By restricting the maximum coordination number to one for each particle, ternary species such as AB_2 and A_2B are excluded from the model. In addition, only A–B and B–C bonds are allowed, other pairs of atoms (A–A, B–B, C–C and A–C) are described by repulsive forces only. The model is further simplified by setting the same mass and VDW parameters for all particles.

A contour plot representing the potential energy surface (PES) is shown in Fig. 1 in terms of bond lengths, R_{AB} and R_{BC} , while the angle of A–B–C is fixed at 180° , the most favorable attack angle. The PES is similar to the London-Eyring-Polanyi-Sato PES¹⁹ with one minimum energy path, a single saddle point, and two minima. By adjusting the force field parameters, different PESs with different reaction energies (ΔE_{r}) and activation energies (E_{a}) are constructed, as shown in Fig. 1. The reaction energies range from 0 to $-10 \text{ kcal}\cdot\text{mol}^{-1}$ ($1 \text{ kcal}\cdot\text{mol}^{-1} = 4.1868 \text{ kJ}\cdot\text{mol}^{-1}$) and activation energies range from 15 to $50 \text{ kcal}\cdot\text{mol}^{-1}$. These potential models are denoted by (ΔE_{r} , E_{a}) in this paper.

The theoretical equilibrium constant can be calculated analytically for these models. Since all particles have the same masses and bond lengths, the translational and rotational partition functions of products and reactants are the same, then the ideal equilibrium constant can be calculated as:

$$K_{\text{i}} = \exp(-\Delta E_{\text{r}} / k_{\text{B}} T) \exp\left(\frac{(\Theta_{\text{v,AB}} - \Theta_{\text{v,BC}}) / 2T}{1 - \exp(-\Theta_{\text{v,AB}} / T)} - \frac{1 - \exp(-\Theta_{\text{v,BC}} / T)}{1 - \exp(-\Theta_{\text{v,AB}} / T)}\right) \quad (2)$$

where $\Theta_{\text{v,AB}}$ and $\Theta_{\text{v,BC}}$ are vibrational temperatures which can be derived by fitting the bond stretch potential curves to harmonic functions. The non-ideal equilibrium constant can be calculated from the ideal equilibrium constant by pressure correction:

$$K_{\text{r}} = K_{\text{i}} \langle p_{\text{i}} / p_{\text{c}} \rangle \quad (3)$$

The ideal pressure is calculated using the ideal gas law, and the computed pressure p_{c} is an ensemble average of the simulated pressure.

The cubic simulation box with periodic boundary condition was used in the simulations. The box contained 198 particles which were initially set as 66 diatomic (AB) molecules and 66 monoatomic C molecules. The mass of particle was 20 amu and the box edge was 2.5 nm, which resulted a density of $421 \text{ kg}\cdot\text{m}^{-3}$ for the simulation models. The fixed number, volume and temperature (NVT) MD simulations were carried out by using the reax/c module of LAMMPS²⁰. The time step was 0.25 fs, and the simulation temperature was controlled by using the Nosé-Hoover thermostat²¹. The initial configuration was generated by randomly distributing the molecules in the simulation box and subsequently relaxed for 100 ps at room temperature (300 K), before the reactive and replica-exchange MD simulations started at elevated temperatures.

In the T-REMD simulations, attempts were made to exchange the temperatures of adjacent replicas (i, j) in a specified period. The Metropolis criteria was used to accept or reject the attempted exchanges using probability of $\min[1, \exp(-\Delta_{ij})]$, where Δ_{ij} is defined as

$$\Delta_{ij} = (E_i - E_j) \left(\frac{1}{k_{\text{B}} T_i} - \frac{1}{k_{\text{B}} T_j} \right) \quad (4)$$

The quantities E_i , T_i and E_j , T_j are the potential energies and temperatures of replicas i and j , respectively, and k_{B} is the Boltzmann constant.

The temperature sequence of T-REMD simulations were determined by the canonical heat capacities of the reactive systems at different temperatures. Fig. 2 shows the heat capacity as a function of temperature calculated for models with the same reaction energy of $-3 \text{ kcal}\cdot\text{mol}^{-1}$ and various activation energies ranging from 15 to $50 \text{ kcal}\cdot\text{mol}^{-1}$. Each data point was calculated using the energy fluctuation obtained in a 500 ps normal MD simulation performed at the specified temperature. Each curve exhibits a maximum value at so called critical temperature, which indicates the system undergoes massive number of bond

Fig. 1 (Left) Contour plot of the potential energy surface for model reaction $\text{AB} + \text{C} \rightleftharpoons \text{A} + \text{BC}$ in terms of bond lengths r_{AB} and r_{BC} , the collision angle A–B–C is fixed at 180° which corresponds to the minimum energy path. (Right) The reactive model for a simple reaction $\text{AB} + \text{C} \rightleftharpoons \text{A} + \text{BC}$ with different activation energies and reaction energies.

1 Å = 0.1 nm

Fig. 2 The canonical heat capacities calculated for models (−3, E_a) with different activation energies in 15, 20, 25, 30, 40 and 50 kcal·mol^{−1}.
1 kcal·mol^{−1} = 4.1868 kJ·mol^{−1}

dissociations, transferring from a molecule-rich phase to a radical-rich phase. The critical temperature is correlated with the activation energy as shown in Fig. 2. Consequently, we set the high-end temperature to be 1000 K above the critical temperature and the low-end (T_{low}) at 1000 K, for each of the potential models.

Given the temperature range fixed, the number of replicas and the temperature sequences were determined to get an even distribution of the exchange acceptance ratio (EAR). The EAR between adjacent replicas can be calculated as²²

$$P_{\text{acc}}(T_i, T_{i+1}) = \text{erfc}\left[\sqrt{C(T_i)} \frac{T_{i+1} - T_i}{T_{i+1} + T_i}\right] \quad (5)$$

where $\text{erfc}[]$ is the complementary error function and $C(T_i)$ is the heat capacity at temperature T_i . Starting from the lowest temperature T_0 and a targeted EAR ($P_{\text{acc}} = 0.5$), initial sequence of temperature was obtained by recursively using Eq. (5). The number of replicas varies, for example, 24 replicas were required for models with activation energy of 40 kcal·mol^{−1}. The temperature sequence was then adjusted using the calculated EAR:

$$P_{\text{acc}}(T_i, T_{i+1}) = \iint g(E_i, T_i) g(E_{i+1}, T_{i+1}) \min[1, \exp(\Delta_{ij})] dE_i dE_{i+1} \quad (6)$$

Where $g(E_i, T_i)$ and $g(E_{i+1}, T_{i+1})$ are Gaussian functions representing the potential energy profiles of adjacent replicas simulated at T_i and T_{i+1} :

$$g(E_i, T_i) = \frac{1}{\sqrt{2\pi}\sigma(T_i)} \exp\left[-\frac{(E_i - \bar{E}(T_i))^2}{\sigma^2}\right] \quad (7)$$

The width $\sigma(T_i)$ and mean $\bar{E}(T_i)$ are functions of temperature. The Gaussians at the sequence temperatures were obtained by fitting the potential distributions, the Gaussians at temperatures off the sequence were estimated by spline interpolation. The new temperature sequence was then calculated by iteratively solving equations $P_{\text{acc}}(T_{i-1}, T_i) = P_{\text{acc}}(T_i, T_{i+1})$, with boundary conditions of $T_0 = T_{\text{min}}$ and of $T_{\text{max}} = T_{\text{high}}$.

The H-REMD simulations were carried out at one fixed temperature (1000 K) but on a series of different potential functions indexed by i :

$$E_i(\mathbf{q}) = E(H_i, \mathbf{q}) \quad (8)$$

where H_i is the i -th potential function and $\mathbf{q}$ represents the coordinates. At selected time interval, exchanges between neighboring replicas (i, j) were attempted, followed by the Metropolis acceptance probability of $\min[1, \exp(\Delta_{ij})]$ where the Δ_{ij} is defined as

$$\Delta_{ij} = \frac{1}{T} [(E_i(\mathbf{q}_i) - E_j(\mathbf{q}_j)) - (E_j(\mathbf{q}_i) - E_i(\mathbf{q}_j))] \quad (9)$$

Since the activation energy is the most sensitive factor that affects chemical reaction rate, we designed the sequence of potential functions by varying the activation energies while keeping the reaction energy constant. The range of activation energies was determined based on the canonical heat capacity curves (Fig. 3). At 1000 K, the activation energy of low-end potential (H_{low}) was set to be 15 kcal·mol^{−1}, the high-end (H_{high}) was the original potential.

With the range of activation energies fixed, the number of replicas was determined based on the same principle as stated above, aiming for an even distribution of EARs. Assuming the potential energy profile of replica running on potential surface $E_i(\mathbf{q})$ is given by an Gaussian distribution $g(E_i, H_i)$, the EAR of neighboring replicas i and j can be calculated by

$$P_{\text{acc}}(H_i, H_j) = \iint g(E_i, H_i) g(E_j, H_j) \min[1, \exp(\Delta_{ij})] dE_i dE_j \quad (10)$$

Because all replicas were simulated at the same temperature (1000 K) and the potential functions were different only in activation energies, the Gaussians are assumed to have the same width with adjustable mean $\bar{E}(H_i)$:

$$g(E_i, H_i) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(E_i - \bar{E}(H_i))^2}{\sigma^2}\right] \quad (11)$$

The sequence of activation energies was then adjusted using Eq. (10). After every 100 attempted exchanges, the potential energy profiles of all replicas were calculated and the new sequence of mean $\bar{E}(H_i)$ (and the activation energies) were estimated by iteratively solving equation $P_{\text{acc}}(T_{i-1}, T_i) = P_{\text{acc}}(T_i, T_{i+1})$ with the boundary condition of $H_0 = H_{\text{min}}$ and $H_{\text{max}} = H_{\text{high}}$.

Fig. 3 Convergence time measured by the number of reactants for a T-REMD simulation at 1000 K using model (−3, 25).

The blue line represents instantaneous number, the green line is the 25-ps block averaged data, and the red line represents the equilibrium value. The first time when the block averaged data reaches the equilibrium line is defined as the convergence time. Color online.

The exchange attempt frequency (EAF) is another critical factor that influences the simulation efficiency. It has been demonstrated that EAF can be determined based on the autocorrelation time of potential energy²³ which means each replica running on different local time (1/EAF). On the other hand, studies demonstrated that the EAF should be set as high as possible^{24,25}. In this work, we determined EAF empirically by evaluating the convergence time (t_c) which is defined as the minimum time required to reach equilibrium.

Fig. 3 shows the number of reactants smoothed by 25-ps block average as a function of simulation time. The equilibrium number of reactants is calculated using the data of last 250 ps simulation trajectory. The first time when the block averaged number reaches the equilibrium value is defined as the convergence time. Table 1 lists different EAF (in 1 ps^{-1}), convergence time (t_c) and total number of attempted exchanges (N_{ex}) obtained using model $(-3, 25)$. The convergence time decreases until the EAF increases to 4 ps^{-1} , and the convergence time is no longer reduced as the EAF is greater than 4 ps^{-1} . The last column lists the number of exchanges required to reach the convergence time. Since $N_{\text{ex}} = t_c \text{EAF}$, the higher EAF, the more exchanges are required for the same convergence time. Although the exchange does not cost too much computational resource, it is unnecessary to do it excessively.

The optimal EAF (4 ps^{-1}) is independent of potential models used in the simulation. As the low-boundary of the EAF is determined by the longest autocorrelation time²³, we examined the maximum autocorrelation time, defined as integral of the autocorrelation function obtained at the critical temperature²⁶:

$$\tau_{\text{corr},x} = \int_{t=0}^{\infty} R_{xx}(t) dt \approx \left(\frac{1}{2} + \sum_{n=1}^N R_{xx}(n\Delta t) \right) \Delta t \quad (12)$$

where Δt is time step, N is the total number of steps, and R_{xx} is the auto-correlation coefficient. The maximum correlation time was calculated based on 50-ps simulation trajectories, the results calculated for different models using both normal MD and T-REMD simulations are given in Table 2. Due to the short period of simulation data, the uncertainties are rather large. Nevertheless, it is evident that the maximum correlation times in either normal MD or T-REMD simulations are independent of the underlying potential models.

The maximum correlation time is independent of the activation energy deserves a discussion. The reason is that the maximum autocorrelation time is measured at the critical

Table 2 Maximum autocorrelation times (ps) obtained from normal MD and T-REMD simulations for different potential models.

	Activation energy	Reaction energy				
		0	-1.5	-3	-5	-10
MD	15	3.42	2.97	2.46	3.43	2.21
	20	2.42	2.83	3.19	2.53	2.50
	25	3.95	2.47	2.75	3.31	2.60
	30	2.45	2.38	3.22	2.39	4.62
	40	3.64	3.29	3.28	2.56	4.99
	50	3.21	2.42	3.31	3.42	3.70
T-REMD	15	0.31	0.30	1.51	0.79	0.91
	20	0.65	0.39	0.51	0.51	0.69
	25	0.31	0.31	0.22	0.36	0.44
	30	0.19	0.23	0.30	0.45	0.33
	40	0.13	0.21	0.22	0.28	0.44
	50	0.13	0.12	0.10	0.10	0.22

Activation and reaction energies in $\text{kcal} \cdot \text{mol}^{-1}$. $1 \text{ kcal} \cdot \text{mol}^{-1} = 4.1868 \text{ kJ} \cdot \text{mol}^{-1}$.

temperature, which is determined by the activation energy. In other words, the impact of activation energy is represented by the critical temperature; and at the critical temperature the maximum autocorrelation function is roughly the same regardless of the activation energy.

Therefore, the maximum autocorrelation time is an intrinsic quantity that characterizes the dynamics at the point of phase transition. With replica exchanges, the correlation is broken, therefore, the maximum correlation time is reduced by one order of magnitude as shown in Table 2. The same reduction would be seen in H-REMD simulation because the physics at the phase transition point is essentially the same. In addition, all replicas are running at the same temperature in our H-REMD simulations, consequently only the replicas near the low-end of the potential sequence are close to the critical temperature, all other replicas are running at non-critical temperatures and their autocorrelation times are lower than the maximum autocorrelation time.

3 Results and discussion

We first tested T-REMD against normal MD using model $(-3, 25)$. The relatively low energy barrier in this model enables a direct comparison with the normal MD simulations. Table 3 lists the equilibrium constants (in logarithm) predicted using normal MD and REMD simulations, together with the theoretical equilibrium constants for comparison. The uncertainties in predicted equilibrium constants are calculated using 300-ps block-average²⁷. The predicted equilibrium constants are consistent between T-REMD and normal MD, and the results agree reasonably well with the theoretical data. The convergence times range from 10 ps at 4000 K to 1000 ps at 1000 K for normal MD simulations, and 250 ps (at all temperatures) for the T-REMD simulation. At 1000 K, the acceleration gained by using T-REMD is modest: 4 times faster than the normal MD simulation. At lower temperatures, the acceleration would be more significant.

Table 1 Exchange attempt frequency (EAF), convergence time (t_c) and number of exchange attempts (N_{ex}) of T-REMD simulations on model $(-3, 25)$.

EAF/ ps^{-1}	t_c/ns	N_{ex}
1	1.00	1,000
2	0.50	1,000
4	0.25	1,000
8	0.25	2,000
40	0.25	10,000

Table 3 Logarithm equilibrium constants ($\ln K_e$) predicted using MD and T-REMD simulations and theoretical calculations ($\ln K_r$) at different temperatures (in Kelvin) for model (−3, 25).

MD			REMD			Theo. $\ln K_r$
T/K	$\ln K_e$	t_e/ps	T/K	$\ln K_e$	t_e/ps	
1000	1.52(0.12)	1000	1000	1.45(0.17)	250	1.41
2000	0.69(0.09)	100	2042	0.74(0.13)	250	0.69
3000	0.40(0.08)	30	2954	0.40(0.08)	250	0.33
4000	0.37(0.07)	10	4000	0.33(0.03)	250	0.26

Uncertainties are given in parenthesis.

The free energy map on the A—B and B—C bond lengths can be calculated from the equilibrium density profile:

$$F(\xi) = -k_B T \ln P(\xi) + C \quad (13)$$

where $P(\xi)$ is the density distribution on reaction coordinate ξ , and C is a constant. Using the density profiles calculated from the last 250 ps trajectory of T-REMD and the last 2 ns trajectory of normal MD, we estimated the free energy surfaces for the model compounds (−3, 25) at 4000 K and the results are shown in Fig. 4. The two FESs resemble the main shape of the PES (Fig. 1), and are similar except the T-REMD map shows more features in high energy regions than that of normal MD. The difference essentially reflects the fact that the T-REMD simulation explores the phase space more extensively than the normal MD simulation. A close comparison of the energy barriers show that the free activation energies are quite different from the potential activation energies. Table 4 lists the free energy barriers calculated by T-REMD and normal MD for model (−3, 25). The results are consistent between the two FESs (T-REMD and normal MD) and the barriers depend on temperature. At temperature lower than 2000 K, the free energy barrier are lower

Table 4 Free energy barriers of model (−3, 25) at different temperatures, obtained by T-REMD and normal MD simulation.

T/K	T-REMD	MD
1000	17.1	16.8
1937	22.8	23.1
2991	26.6	26.5
4000	30.8	30.9

The energy scale is in kcal·mol^{−1}. 1 kcal·mol^{−1} = 4.1868 kJ·mol^{−1}.

Fig. 4 Free energy contour maps obtained from BFMD (left) and T-REMD (right) for M(−3, 25) at 4000 K.

The energy scale is in kcal·mol^{−1}.

than the potential energy barriers. Above 3000 K the free barriers are higher than the potential energy barriers. At approximately 2500 K the two energy barriers are close to each other (25 kcal·mol^{−1}). The phenomenon indicates that the entropy contribution in FES is complicated, which is different to the transition state and the pre-reaction state at different temperatures.

Unfortunately, the success of T-REMD simulations depends on the potential models. Fig. 5 shows the discrepancies between the predicted and theoretical equilibrium constants for different models at 1000, 2000 and 4000 K. At 2000 and 4000 K, the discrepancies are in the range of uncertainties, although the errors are larger for models of high activation energies. At 1000 K, the discrepancies are significantly large for models with high reaction energy and activation energy. However, for models with either low reaction energy or low activation energy, the errors are acceptable.

The replica mixing efficiency is not the source of problem. We calculated the mixing ratio²⁸ defined for each replica i_r as:

$$\delta_{i_r} = n_{i_r, \text{up}} / (n_{i_r, \text{up}} + n_{i_r, \text{down}}) \quad (14)$$

The quantities $n_{i_r, \text{up}}$ and $n_{i_r, \text{down}}$ are the numbers of replicas entering the temperature slot upward and downward. If the replicas diffuse freely in the temperature space, δ is a straight line connecting two ends from 1 (high) to 0 (low). A slope steeper (more negative) than the straight line indicates a bottleneck in the replica diffusion. Fig. 6 shows the calculated mixing ratios as the index of temperature sequence for two sets

Fig. 5 Discrepancies between predicted and theoretical equilibrium constants in natural logarithm for different potential models (ΔE_r , E_a) in kcal·mol^{−1}.

Fig. 6 Replica mixing ratio obtained in T-REMD simulations of different reactive models.

The left panel shows the good cases, models $(-3, 25)$ and $(0, 40)$, the right panel shows the bad cases, models $(-10, 25)$ and $(-5, 40)$. The diagonal line represents the ideal case.

of potential models that yield good and poor predictions. The mixing efficiencies are essentially the same for both models.

The large discrepancies found in the prediction demonstrates the problem in the ergodicity of T-REMD. To illustrate the problem, we examined the Gibbs free energies as function of reaction quotient Q and temperature T using Van't Hoff equation:

$$\Delta_r G = -RT \ln K + RT \ln Q \quad (15)$$

where K is the theoretical equilibrium constant. The results are plotted in Fig. 7. The purple dots represent the equilibrium constants (when $Q = K$). The small dots are data extracted from a 2-ns T-REMD simulation trajectory. The combination (Q, T) represents a point in the phase space. Four models, $(-3, 15)$, $(-3, 40)$, $(-5, 15)$ and $(-5, 40)$ are compared to examine how the reaction energy and activation energy affect the samplings in the phase space. The impact of reaction energy can be seen by comparing the charts vertically. Increase the reaction energy (from -3 to -5 kcal·mol $^{-1}$) the phase space expands in Q coordinate. Comparison of the chart horizontally shows the

impact of activation energy. Increasing the activation energy suppresses the sampling range in Q coordinate. Therefore, a combination of high reaction energy and high activation energy (e.g., $-5, 40$) makes the simulation unable to cover the required phase space. The samplings are limited to a subdomain in the phase space despite high overlap in potential energies, sufficient acceptance ratios and high replica mixing efficiency. Although this kind of problem has been noticed for simulation of conformational change in biological system¹⁶, the problem is more pronounced in the simulation of chemical reactions at high concentration.

The ergodicity problem can be avoided by using the H-REMD simulations. Using twelve (12) replicas running on sequence of potentials with activation energy $E_a(i)$ spanning from 15 to 40 kcal·mol $^{-1}$, we carried out H-REMD simulations for model $(-5, 40)$. The actual sequence of potential energies was adjusted to maintain a smooth distribution of acceptance ratios. The optimized activation energies in the sequence of potentials are

Fig. 7 Sampling in phase space represented in reaction quotient (Q) and temperature (in K), obtained in 2-ns T-REMD simulations for models $(-3, 15)$, $(-3, 40)$, $(-5, 15)$ and $(-5, 40)$, respectively.

The purple dots represent the equilibrium constants, the small dots in different colors are simulation data from cold (blue) to hot (red) temperatures. Color online.

Table 5 Optimized activation energies in models ($-5, E_a(i)$) used for the H-REMD simulations.

i	$E_a(i)$
1	15.0
2	16.1
3	17.8
4	19.9
5	21.6
6	23.7
7	26.0
8	28.5
9	31.2
10	34.0
11	36.9
12	40.0

list in Table 5. Using this sequence of potentials, the average EAR in the H-REMD simulations is 67%.

The potential energy distribution obtained in the H-REMD simulations shows significant overlaps between the neighboring replicas as given in Fig. 8a. The diffusion ratios are given in Fig. 8b which indicate the replica diffuses smoothly in the Hamiltonian space. The double peaks in each replica are due to the accepted exchanges in different potential energies for the same configuration. This explains why less replicas can be used

Table 6 Comparison of H-REMD and T-REMD predictions of $\ln K$ on different potential models.

Model	T-REMD	H-REMD	Theo.
(-3, 25)	1.45 ± 0.17	1.44 ± 0.15	1.48
(-5, 40)	1.77 ± 0.21	2.40 ± 0.30	2.49
(-10, 50)	1.68 ± 0.20	4.90 ± 0.50	5.01

in H-REMD than that in T-REMD. The performance of replica exchange simulation depends on the overlap in potential energies. The overlaps in T-REMD are mostly determined by thermal fluctuations, but in H-REMD the fluctuations can be controlled by the underlying potential energy functions. In addition, the performance of T-REMD heavily depends on the size of simulation mode as the thermal fluctuation decreases as the size of system increase ($\sqrt{N}$), however, the performance of H-REMD could be less sensitive to the simulation size.

The equilibrium constants predicted using H-REMD for 3 models, (-3, 25), (-5, 40) and (-10, 50), are listed in Table 6 for comparison with the H-REMD and theoretical predictions. For (-3, 25) both H-REMD and T-REMD yield the same results, and the results agree well with the theoretical value. For (-5, 40), the T-REMD result of 1.77 is significantly lower than the values of 2.40 and 2.49 predicted using H-REMD and theoretical equation. For (-10, 50), the T-REMD prediction fails completely, the H-REMD yields good result of 4.9, in good

Fig. 8 The energy overlaps of neighbored replicas of H-REMD (left) and the replica-diffusive ratio (right) obtained from H-REMD simulation of model (-5, 40), the diagonal line represents the ideal case. **Fig. 9** Number of reactant molecules (AB) versus simulation time obtained from T-REMD (left) and H-REMD (right) simulations with different initial states: all AB (exothermic reaction) and all BC (endothermic reaction) comparing with theoretical values at equilibration for potential models ($\pm 5, 40$). The data points are block-averaged in 25-ps bin.

comparison with 5.0 predicted theoretically.

The discussions above are about exothermic reactions. For endothermic reactions, another factor further reduces the efficiency of T-REMD method. In an endothermic reaction, the product is higher in potential energy than the reactant. Since the products are more likely to be sampled at higher temperature, the combination of increase of potential energy and increase of temperature worsens the acceptance ratio (see Eq. (4)) and makes the convergence more difficult to reach. For H-REMD, this is not a problem because the temperature is a constant. In Fig. 9, the population of products are compared with both T-REMD and H-REMD simulations from different initial conditions: A-B molecules only for forward (exothermic) reaction and B-C molecules only for backward (endothermic) reaction. In T-REMD simulations, the forward (exothermic) curve converges slowly and the backward (endothermic) curve shows no signs of convergence in 4 ns simulation. However, the two curves converge to the theoretical value in about 3 ns in H-REMD simulations.

4 Conclusions

Using a serial reaction models of different activation energies (up to 50 kcal·mol⁻¹) and reaction energies (0 to 10 kcal·mol⁻¹), we examined the robustness of T-REMD and H-REMD methods in simulation of complex chemical reaction.

Two critical parameters, the sequence of controlling variable and the exchange frequency, were examined for each of the methods. The upper limit of the temperature sequence used in T-REMD simulation was set by considering the activation energy that regulates the critical temperature, while the lower limit was determined by the temperature of interest (the experimental temperature). The potential sequence used in H-REMD was made by varying the activation energies in the potential functions. The low-end was determined so that the energy barrier could be overcome in the (short) simulation time at temperature of simulation, while the upper limit was the real potential energy barrier. In both cases, the sequences were dynamically adjusted and determined by aiming a constant acceptance ratio among any adjacent replicas. Twenty-four (24) replicas for T-REMD and twelve (12) replicas for H-REMD were required for a reaction model with activation energy of 40 kcal·mol⁻¹. The exchange attempted frequency was determined empirically in this work. Although high EAF helps improving sampling efficiency in general, excess EAF is unnecessary. The low limit of EAF is determined by the maximum correlation time, which is located at the critical temperature and is independent of the activation and reaction energies. Therefore, the same EAF is applicable for all models of difference reaction energies and activation energies.

With the optimized parameters T-REMD predicts equilibrium constants in close agreement with the normal molecular dynamics and the theoretical predictions for exothermic reactions of modest activation energy (< 20 kcal·mol⁻¹) or low

reaction energy (< 3 kcal·mol⁻¹). However, a combination of high reaction energy and activation energy severely hamper the efficiency of T-REMD. The problem is worsened for endothermic reactions because both temperature and potential energy change in the same direction that reduce the exchange acceptance ratio. Fundamentally, the problem is due to insufficient sampling of the phase space using temperature as the controlling variable. Although the replicas diffuse smoothly in the temperature space, they are limited in a subdomain of the phase space.

The H-REMD simulations with replicas running on potential surfaces having different activation energies demonstrated to be much more efficient than the T-REMD for all reactive models studied. Using 15 kcal·mol⁻¹ as the low boundary in the Hamiltonian space, we carried out H-REMD simulations at 1000 K for reactions of high activation energies up to 40 kcal·mol⁻¹. Using half number of replicas, the H-REMD simulation converges quickly for both exothermic and endothermic reactions.

Acknowledgment: The authors gratefully acknowledge computational resources from Center for High Performance Computing at Shanghai Jiao Tong University.

References

- (1) Steinfeld, J. I.; Francisco, J. S.; Hase, W. L. *Chemical Kinetics and Dynamics*; Prentice Hall Englewood Cliffs: Upper Saddle River, NJ, USA, 1989; Vol. 3.
- (2) Érdi, P.; Tóth, J. *Mathematical Models of Chemical Reactions: Theory and Applications of Deterministic and Stochastic Models*; Manchester University Press: Oxford Road, Manchester M13 9PL, UK, 1989.
- (3) Kresse, G.; Hafner, J. *Phys. Rev. B* **1993**, *47*, 558.
doi: 10.1103/PhysRevB.47.558
- (4) Van Duin, A. C. T.; Dasgupta, S.; Lorant, F.; Goddard, W. A., III. *J. Phys. Chem. A* **2001**, *105*, 9396. doi: 10.1021/jp004368u
- (5) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141.
doi: 10.1016/S0009-2614(99)01123-9
- (6) Hukushima, K.; Nemoto, K. *J. Phys. Soc. Jpn.* **1996**, *65*, 1604.
doi: 10.1143/JPSJ.65.1604
- (7) Fukunishi, H.; Watanabe, O.; Takada, S. *J. Chem. Phys.* **2002**, *116*, 9058. doi: 10.1063/1.1472510
- (8) Itoh, S. G.; Okumura, H. *J. Comput. Chem.* **2013**, *34*, 622.
doi: 10.1002/jcc.23167
- (9) Itoh, S. G.; Okumura, H.; Okamoto, Y. *J. Chem. Phys.* **2010**, *132*, 134105. doi: 10.1063/1.3372767
- (10) Swails, J. M.; Roitberg, A. E. *J. Chem. Theory Comput.* **2012**, *8*, 4393. doi: 10.1021/ct300512h
- (11) Mori, T.; Jung, J.; Sugita, Y. *J. Chem. Theory Comput.* **2013**, *9*, 5629.
doi: 10.1021/ct400445k

- (12) Leahy, C. T.; Kells, A.; Hummer, G.; Buchete, N. V.; Rosta, E. *J. Chem. Phys.* **2017**, *147*, 152725. doi: 10.1063/1.5004774
- (13) Stelzl, L. S.; Hummer, G. *J. Chem. Theory Comput.* **2017**, *13*, 3927. doi: 10.1021/acs.jctc.7b00372
- (14) Wallace, A. F. Replica Exchange Methods in Biomineral Simulations. In *Methods in Enzymology*, De Yoreo, J. J., Ed.; Academic Press: New York, NY, USA, 2013; Vol. 532, Chapter 4, pp. 71–93.
- (15) Zhang, W.; Chen, J. *J. Chem. Theory Comput.* **2013**, *9*, 2849. doi: 10.1021/ct400191b
- (16) Bergonzo, C.; Henriksen, N. M.; Roe, D. R.; Swails, J. M.; Roitberg, A. E.; Cheatham, T. E. *J. Chem. Theory Comput.* **2014**, *10*, 492. doi: 10.1021/ct400862k
- (17) Mori, Y.; Okamoto, Y. *Phys. Rev. E* **2013**, *87*, 023301. doi: 10.1103/PhysRevE.87.023301
- (18) Petraglia, R.; Nicolai, A.; Wodrich, M. D.; Ceriotti, M.; Corminboeuf, C. *J. Comput. Chem.* **2016**, *37*, 83. doi: 10.1002/jcc.24025
- (19) Sato, S. *J. Chem. Phys.* **1955**, *23*, 592. doi: 10.1063/1.1742043
- (20) Plimpton, S. *J. Comput. Phys.* **1995**, *117*, 1. doi: 10.1006/jcph.1995.1039
- (21) Shinoda, W.; Shiga, M.; Mikami, M. *Phys. Rev. B* **2004**, *69*, 134103. doi: 10.1103/PhysRevB.69.134103
- (22) Denschlag, R.; Lingenheil, M.; Tavan, P. *Chem. Phys. Lett.* **2009**, *473*, 193. doi: 10.1016/j.cplett.2009.03.053
- (23) Bittner, E.; Nußbaumer, A.; Janke, W. *Phys. Rev. Lett.* **2008**, *101*, 130603. doi: 10.1103/PhysRevLett.101.130603
- (24) Plattner, N.; Doll, J. D.; Meuwly, M. *J. Chem. Theory Comput.* **2013**, *9*, 4215. doi: 10.1021/ct400355g
- (25) Dupuis, P.; Liu, Y.; Plattner, N.; Doll, J. D. *Multiscale Model. Simul.* **2012**, *10*, 986. doi: 10.1137/110853145
- (26) Wolff, U. *Comput. Phys. Commun.* **2004**, *156*, 143. doi: 10.1016/S0010-4655(03)00467-3
- (27) Flyvbjerg, H.; Petersen, H. G. *J. Chem. Phys.* **1989**, *91*, 461. doi: 10.1063/1.457480
- (28) Katzgraber, H. G.; Trebst, S.; Huse, D. A.; Troyer, M. *J. Stat. Mech. Theory Exp.* **2006**, *03*, P03018. doi: 10.1088/1742-5468/2006/03/P03018