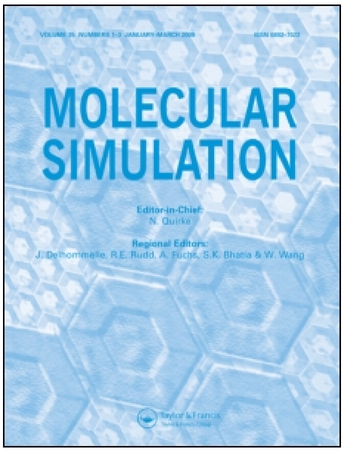

Molecular Simulation
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713644482

Activation Energies from Transition Path Sampling Simulations

Christoph Dellago[a]; Peter G. Bolhuis[b]
[a] Institute for Experimental Physics, University of Vienna, Vienna, Austria [b] Department of Chemical Engineering, University of Amsterdam, Amsterdam, The Netherlands

To cite this Article Dellago, Christoph and Bolhuis, Peter G.(2004) 'Activation Energies from Transition Path Sampling Simulations', Molecular Simulation, 30: 11, 795 — 799
To link to this Article: DOI: 10.1080/08927020412331294869
URL: http://dx.doi.org/10.1080/08927020412331294869

Molecular Simulation, VOL. 30 (11–12), 15 September–15 October 2004, pp. 795–799

Activation Energies from Transition Path Sampling Simulations

CHRISTOPH DELLAGO[a] and PETER G. BOLHUIS[b,*]

[a]*Institute for Experimental Physics, University of Vienna, Boltzmanngasse 5, 1090 Vienna, Austria;* [b]*Department of Chemical Engineering, University of Amsterdam, Nieuwe Achtergracht 166, 1018 WV, Amsterdam, The Netherlands*

(Received November 2003; In final form March 2004)

We derive an new expression for the calculation of activation energies within the framework of transition path sampling. Using this expression one can determine activation energies without knowledge of the reaction mechanism, which is often unavailable for processes occurring in complex systems. Since in this method activation energies are calculated directly from path averages, no computationally expensive calculation of reaction rate constants is necessary. As an illustrative example, we determine the activation energy for the isomerization of a model diatomic immersed in a bath of repulsive soft particles.

Keywords: Activation energy; Computer simulation; Transition path sampling; Chemical kinetics

INTRODUCTION

The temperature dependence of many rate constants for chemical reactions follows the Arrhenius law [1],

$$k = \nu \exp(-\beta E_a), \tag{1}$$

where ν is the so-called pre-exponential factor, E_a is the activation energy, and $\beta = 1/k_B T$ is the inverse temperature. Both the activation energy E_a and the pre-exponential factor ν can be determined from experimental data by plotting the logarithm of the reaction rate constant k as a function of β. The negative slope of this line is the activation energy:

$$E_a = -\frac{\partial}{\partial \beta} \ln k. \tag{2}$$

The pre-exponential factor ν, on the other hand, can be determined from the intersect of the ln k vs. β curve with the y-axis. In the framework of transition state theory the activation energy E_a is the potential energy difference between reactants and the transition state, usually identified with a saddle point on the potential energy barrier separating reactants from products. The pre-exponential factor ν depends on the entropy difference between reactants and transition state and it also includes dynamical corrections.

Since the Arrhenius parameters E_a and ν are accessible experimentally, their theoretical prediction is of great interest. In principle, these parameters can be determined computationally by calculating the reaction rate constant as a function of temperature and then proceeding in a manner analogous to the analysis of the experimental data. The calculation of rate constants, however, is computationally expensive. For this reason, a direct calculation of kinetic observables, such as the activation energy, is highly desirable. In this paper, we present a procedure to accomplish exactly that. Based on the transition path sampling methodology, our approach to determine activation energies does not require full rate constants calculations. Instead, the derivative of the reaction rate constant with respect to the inverse temperature β is expressed in terms of transition path averages which can be evaluated directly in a transition path sampling simulation. Since the transition path sampling method does not require any knowledge of mechanisms or transition states, our approach can be used to determine activation energies in complex systems where such knowledge is usually unavailable. We demonstrate the validity and practicality of the method by applying it to the isomerization of

---

*Corresponding author. E-mail: bolhuis@science.uva.nl

ISSN 0892-7022 print/ISSN 1029-0435 online © 2004 Taylor & Francis Ltd
DOI: 10.1080/08927020412331294869

a diatomic molecule immersed in a fluid of purely repulsive particles.

The remainder of this paper is organized as follows. To introduce the formalism and the terminology we give a brief outline of the transition path sampling methodology in the second section. In the third section, we derive the expressions necessary to calculate the activation energy E_a with transition path sampling techniques. An example calculation for the solvated diatomic is presented in the fourth section and conclusions are drawn in the fifth section.

TRANSITION PATH SAMPLING

Transition path sampling, developed to simulate rare processes in complex systems, is a computational methodology based on a statistical mechanics of trajectories [2]. Recently, this methodology has been applied to the study of a variety of important systems [3–6]. The reader is referred to Refs. [7,8] for in-depth reviews of the transition path sampling approach. In this section, we will give a brief account of the transition path sampling method in order to introduce the concepts and the notation necessary to follow the derivations in the following sections. Although transition path sampling is not restricted to a particular type of dynamics, here, for simplicity, we limit ourselves to deterministic dynamics where entire trajectories are determined completely by their initial conditions.

Consider a molecular system with Hamiltonian $\mathcal{H}(x)$ evolving according to Newton's equations of motion and described by a phase space point $x = \{q, p\}$ which includes positions q and momenta p of all particles. Assume further that initial conditions x_0 are distributed canonically, i.e. $\rho(x_0) = \exp[-\beta\mathcal{H}(x_0)]/Q$. Here, Q is a factor normalizing the phase space distribution. The state x_t of the system at time t is obtained by solving Newton's equations of motion starting from initial conditions x_0 at time 0. Since our main purpose is to study activated processes connecting long-lived stable states, we focus on trajectories starting in a specific region of configuration space, A, and ending in a different region, B. For instance, A and B would be reactants and products in the case of a chemical reaction, while for a phase transition A and B may denote two different phases of a condensed material. We can now define the *transition path ensemble*, i.e. the distribution of trajectories (represented by their initial conditions x_0) originating in region A and ending in region B at a certain time t:

$$f_{AB}(x_0, t) = \rho(x_0)h_A(x_0)h_B[x_t(x_0)]/Q_{AB}(t). \tag{3}$$

Here, the indicator functions $h_A(x)$ and $h_B(x)$ are unity if x belongs to the respective stable region and they vanish otherwise. The "partition function" $Q_{AB}(t)$ normalizes the transition path ensemble.

The transition path ensemble defined in Eq. (3) is a statistical description of *all* trajectories connecting A and B within time t. This ensemble can be sampled with efficient Monte Carlo methods [2,7]. In this approach, one proceeds by generating a new path from an old one and then rejecting or accepting it with a probability depending on how the weight of the new path compares to that of the old path. Various algorithms are available to generate a new path from an old one. In the shooting algorithm [2], for instance, a new path is generated from an old one by selecting a point along the path at random. Then, the momenta belonging to this point are changed by a small, random amount and the equations of motion are integrated forward and backward starting from the state with changed momenta to complete the new trajectory. The new trajectory is then accepted using the Metropolis rule designed to satisfy detailed balance between the forward and backward path move. By repeating this procedure one generates a sequence of pathways, each of which is reactive, i.e. it connects region A with region B, and occurs with a probability proportional to its weight in the transition path ensemble. Analysis of the harvested pathways can then yield detailed mechanistic insight into the reaction mechanism.

The transition path sampling formalism can also be used to calculate reaction rate constants for processes involving rare events. For this purpose, the time correlation functions connecting the microscopic dynamics of the system with the phenomenological description of the kinetics are most conveniently cast in form of averages over the transition path ensemble [7,10]. In this case, however, it is advantageous to consider a transition path ensemble differing slightly from the ensemble defined in Eq. (3):

$$f^*_{AB}(x_0, t) = \rho(x_0)h_A(x_0)H_B(x_0, t)/Q^*_{AB}(t), \tag{4}$$

where $Q^*_{AB}(t)$ is a normalizing factor. Note that in Eq. (4) the function $H_B(x_0, t)$ replaces the indicator function $h_B(x_t)$. In contrast to $h_B(x_t)$, which depends only on a particular state of the system, the function $H_B(x_0, t)$ depends on all states along the trajectory of length t starting at x_0. If in the time interval between 0 and t the trajectory visits region B at least once, $H_B(x_0, t)$ is unity and it vanishes otherwise, or, in other words, $H_B(x_0, t) \equiv \max_{0\leq\tau\leq t} h_B[x_\tau(x_0)]$. Thus, pathways with a non-zero weight in the path ensemble $f^*_{AB}(x_0, t)$ are required to begin in A and visit region B before t but do not need to end in B.

Writing time correlation functions in terms of path averages, the reaction rate constant can be calculated in two steps [7]. The first step consists in determining the conditional probability $C_{AB}(t)$ of observing the system in state B at a particular time t provided the system was located in A at time 0. This probability can be written in terms of a "free energy" difference between different ensembles of pathways and its numerical evaluation is computationally expensive. In a second, less expensive step a single transition path sampling simulation can be used to obtain $C_{AB}(t)$ for a wide range of times t. The forward reaction rate constant is then given by the plateau value of the time derivative of $C_{AB}(t)$. A detailed description of the procedure sketched above can be found in Ref. [7].

The calculation of the free energy, or reversible work, to change between ensembles of trajectories is often computationally very expensive and, for some systems, beyond the capabilities of current computing technology. For this reason, alternative methods for calculating kinetic observables, such as the activation energy E_a, are highly attractive. In the following section, we describe a way to determine the derivative of the rate constant with respect to β, which is related to the activation energy E_a for activated processes. This method does not require to carry out demanding free energy computations.

CALCULATING ACTIVATION ENERGIES

The correlation function

$$C_{AB}(t) = \frac{\langle h_A(x_0) h_B(x_t)\rangle}{\langle h_A\rangle} \tag{5}$$

relates the microscopic dynamics of a system to its phenomenological description in terms of reaction rate constants. In the case of simple activated dynamics this correlation function approaches its asymptotic value exponentially, $C_{AB}(t) \approx \langle h_B\rangle[1 - \exp(-t/\tau_{\rm rxn})]$, where $\tau_{\rm rxn} = (k_{AB} + k_{BA})^{-1}$ is the reaction time, and k_{AB} and k_{BA} are the forward and backward reaction rates, respectively. This exponential behavior breaks down for times shorter than the molecular time scale $\tau_{\rm mol}$, roughly the time required to cross the barrier. If there is a separation of time scales, i.e. if $\tau_{\rm mol} \ll \tau_{\rm rxn}$, the correlation function $C(t)$ grows linearly in the intermediate time regime $\tau_{\rm mol} < t \ll \tau_{\rm rxn}$. The plateau value of its time derivative,

$$k(t) \equiv \frac{\mathrm{d}}{\mathrm{d}t} C_{AB}(t), \tag{6}$$

in this time regime is the forward reaction rate constant k_{AB} [9]. In analogy to Eq. (2) we now define a time dependent activation energy $E_a(t)$ whose plateau value in the intermediate time regime is the activation energy:

$$E_a(t) \equiv -\frac{\partial}{\partial\beta} \ln k(t). \tag{7}$$

For Newtonian dynamics and a canonical distribution of initial conditions we rewrite the above equation as

$$E_a(t) = -\frac{\partial}{\partial\beta} \ln \frac{\int \mathrm{d}x_0 \exp[-\beta\mathcal{H}(x_0)] h_A(x_0) \dot{h}_B(x_t)}{\int \mathrm{d}x_0 \exp[-\beta\mathcal{H}(x_0)] h_A(x_0)}, \tag{8}$$

where the dot denotes a time derivative. Evaluating the derivative with respect to β we obtain

$$E_a(t) = \frac{\int \mathrm{d}x_0 \exp[-\beta\mathcal{H}(x_0)] h_A(x_0) \dot{h}_B(x_t) \mathcal{H}(x_0)}{\int \mathrm{d}x_0 \exp[-\beta\mathcal{H}(x_0)] h_A(x_0) \dot{h}_B(x_t)} - \frac{\int \mathrm{d}x_0 \exp[-\beta\mathcal{H}(x_0)] h_A(x_0) \mathcal{H}(x_0)}{\int \mathrm{d}x_0 \exp[-\beta\mathcal{H}(x_0)] h_A(x_0)}. \tag{9}$$

The second term on the right hand side of the above equation is simply the average $\langle\mathcal{H}\rangle_A$ of the energy $\mathcal{H}(x_0)$ restricted to region A. The first term involves more complicated averages over pathways. To make this term accessible to a transition path sampling simulation, we insert the function $H_B(x_0, \mathcal{T})$ into the integrand of numerator and denominator. Here, $\mathcal{T}$ is the maximum time for which we desire to evaluate $E_a(t)$. The reason that $H_B(x_0, \mathcal{T})$ can be inserted into the integrals without changing them is that $H_B(x_0, \mathcal{T})$ vanishes only if $h_B(x_t) = 0$ for all $0 \le t \le \mathcal{T}$, in which case the integrand vanishes anyways. If, on the other hand, $h_B(x_t) = 1$ for some t between 0 and $\mathcal{T}$, $H_B(x_0, \mathcal{T}) = 1$ and the integrands remain unchanged also in this case. At the same time we multiply and divide that term with the normalization constant of the ensemble of pathways starting in A in visiting B before time $\mathcal{T}$:

$$\begin{aligned} E_a(t) = {} & \frac{\int \mathrm{d}x_0 \exp[-\beta\mathcal{H}(x_0)] h_A(x_0) H_B(x_0,\mathcal{T}) \dot{h}_B(x_t) \mathcal{H}(x_0)}{\int \mathrm{d}x_0 \exp[-\beta\mathcal{H}(x_0)] h_A(x_0) H_B(x_0,\mathcal{T})} \\ & \times \frac{\int \mathrm{d}x_0 \exp[-\beta\mathcal{H}(x_0)] h_A(x_0) H_B(x_0,\mathcal{T})}{\int \mathrm{d}x_0 \exp[-\beta\mathcal{H}(x_0)] h_A(x_0) H_B(x_0,\mathcal{T}) \dot{h}_B(x_t)} \\ & - \langle\mathcal{H}\rangle_A. \end{aligned} \tag{10}$$

Thus, we can write the time dependent activation energy $E_a(t)$ in terms of path averages:

$$E_a(t) = \frac{\langle \dot{h}_B(x_t) \mathcal{H}(x_0)\rangle^*_{AB}}{\langle \dot{h}_B(x_t)\rangle^*_{AB}} - \langle\mathcal{H}\rangle_A. \tag{11}$$

The paths over which one averages have length $\mathcal{T}$ and t is in the interval $[0, \mathcal{T}]$.

Equation (11), which is the central result of this paper, suggests the following procedure for calculating activation energies in the transition path sampling formalism. First, one carries out a transition path simulation for pathways starting in A and visiting B before time $\mathcal{T}$. The two functions $\langle h_B(x_t)\mathcal{H}(x_0)\rangle^*_{AB}$ and $\langle h_B(x_t)\rangle^*_{AB}$ are calculated as averages in this ensemble.

Then, one takes the time derivative of both expressions and divides one by the other for all times $0 \le t \le \mathcal{T}$. Finally, the average total energy of state A, calculated in a separate molecular dynamics or Monte Carlo simulation is subtracted from this ratio. The activation energy E_a is then given by the plateau value of $E_a(t)$. Note that in this procedure no expensive free energy calculation in path space is required.

ISOMERIZATION OF A DIATOMIC MOLECULE

To demonstrate the practicality of this procedure we have applied it to the isomerization of a model diatomic immersed in a fluid of purely repulsive soft spheres [10,11]. This model, which has been used earlier to validate transition path sampling algorithms for the calculation of reaction rate constants [10], consists of N point particles in two dimensions interacting with each other according to the Weeks–Chandler–Andersen potential [12],

$$V_{\mathrm{WCA}}(r) = \begin{cases} 4\epsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] + \epsilon & \text{for } r \le r_{\mathrm{WCA}} \equiv 2^{1/6}\sigma, \\ 0 & \text{for } r > r_{\mathrm{WCA}}. \end{cases} \tag{12}$$

Here, r is the interparticle distance, and ϵ and σ specify the strength and the interaction radius of the potential, respectively. In addition, two of the N particles experience a double well potential binding them to each other,

$$V_{\mathrm{dw}}(r) = h\left[1 - \frac{(r - r_{\mathrm{WCA}} - w)^2}{w^2}\right]^2. \tag{13}$$

Here, h denotes the height of the barrier separating the two potential energy wells located at $r_{\mathrm{WCA}} = 2^{1/6}\sigma$ and $r_{\mathrm{WCA}} + 2w$, respectively. The diatomic can exist in either of these two states differing in the interatomic distance r. If the barrier height h is large compared to k_BT, transition between these states are rare.

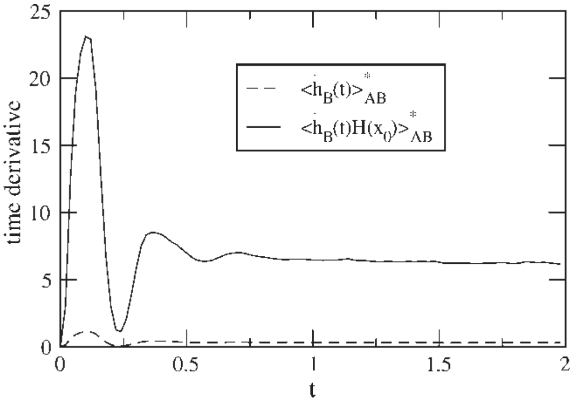

FIGURE 1 Path averages $\langle \dot{h}_B(x_t)\mathcal{H}(x_0)\rangle^*_{AB}$ and $\langle \dot{h}_B(x_t)\rangle^*_{AB}$ as a function of time t for isomerizations of the model diatomic described in the main text.

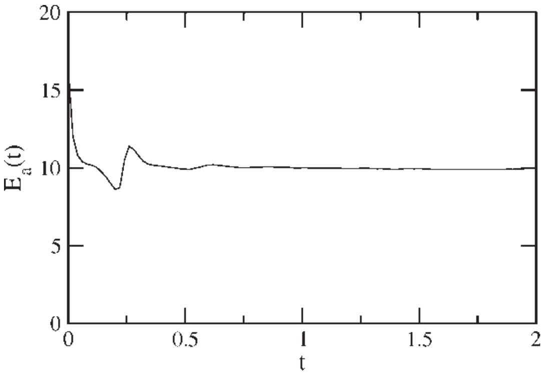

FIGURE 2 Activation energy $E_a(t)$ as a function of time t obtained from the path averages presented in Fig. 1. The plateau value of this function is the activation energy E_a.

We have calculated $E_a(t)$ for the transition from the contracted to the extended state using the approach described above for a barrier height of $10\,\epsilon$. In this example calculation, the total particle number was $N = 9$ (2 particles belonging to the diatomic and 7 solvent particles), the temperature was $k_BT = \epsilon$, the barrier width was $\omega = 0.25\sigma$ and the density was $\rho = 0.6\sigma^{-2}$. The equations of motion were integrated with the velocity Verlet algorithm with a time step of $\Delta t = 0.002(m\sigma^2/\epsilon)^{1/2}$. A total of 2×10^6 pathways of length $2.0(m\sigma^2/\epsilon)^{1/2}$ were harvested with the shooting and shifting algorithms (60% shooting moves and 40% shifting moves). Momentum displacements drawn for each degree of freedom from a Gaussian distribution with width $\delta p = 0.04(m\epsilon)^{1/2}$ yielded and acceptance rate of 50% for shooting moves. Shifting moves with time shifts drawn from a uniform distribution with maximum shift $\pm 0.4(m\sigma^2/\epsilon)^{1/2}$ were accepted with a probability of 80%. The system was defined to be in state A if the interatomic distance $r < 1.35\sigma$ and in state B if $r > 1.45\sigma$. The time derivatives of the path averages $\langle h_B(x_t)\mathcal{H}(x_0)\rangle^*_{AB}$ and $\langle h_B(x_t)\rangle^*_{AB}$ obtained from this transition path sampling simulation are depicted in Fig. 1. The resulting time dependent activation energy $E_a(t)$ is shown in Fig. 2. The plateau value of $E_a(t)$ is about $10\,\epsilon$, nicely coinciding with the potential energy barrier separating the two states of the diatomic.

CONCLUSION

We have derived and demonstrated a practical transition path sampling procedure to determine the activation energy of chemical reactions in complex systems, where knowledge of transition states or reaction coordinates is unavailable. In this method one calculates the activation energy directly without the need to carry out computationally demanding calculations of reaction rate constants. While in this paper we have focused on activated processes with an Arrhenius temperature dependence of the reaction rate constant, our approach is useful also for non-Arrhenius processes. In this case, the derivative of ln k with respect to β computed via transition path sampling simulations cannot be identified with an activation energy, but it can still be compared to its experimental counterpart.

Our method is also of interest in parallel tempering transition path sampling simulations [13]. In this case, one can determine the reaction rate constant using traditional method for one single temperature. Then, reaction rate constants at all other temperatures can be obtained at no extra cost by integrating the temperature derivative of k calculated using the approach described in this paper.

While our approach is exact and its application to small systems is straightforward, large energy fluctuations may complicate the calculation of activation energies for processes taking place in large systems. To simulate configurational transitions of proteins, for instance, thousands of particles are usually necessary. Since energy fluctuations grow with the square root of the particle number, it may be computationally difficult to determine the two right hand side terms of Eq. (11) with a statistical accuracy sufficient for an accurate calculation of the activation energy. Current work in our groups is directed towards overcoming this complication. We note that the approach to calculate the activation energy advocated in this paper is very generally applicable to other path sampling techniques. For instance, for the transition interface sampling method [14] a similar result can be derived [15].

Acknowledgements

We thank Rodolphe Vuillemieur for suggesting the calculation of activation energies in the transition path sampling method. The calculations were carried out in part on the *Schrödinger* Linux cluster of the Vienna University Computer Center.

References

[1] Atkins, P.W. (2002) *Physical Chemistry*, 7th ed. (Oxford University Press, Oxford).

[2] Dellago, C., Bolhuis, P.G., Csajka, F.S. and Chandler, D. (1998) "Transition path sampling and the calculation of rate constants", *J. Chem. Phys.* **108**, 1964.

[3] Geissler, P.L., Dellago, C., Chandler, D., Hutter, J. and Parrinello, M. (2001) "Autodissociation of a water molecule in liquid water", *Science* **291**, 2121.

[4] Bolhuis, P.G., Dellago, C. and Chandler, D. (2000) "Reaction coordinates of bimolecular isomerization", *Proc. Natl Acad. Sci. USA* **97**, 5877.

[5] Ensing, B. and Baerends, E.J. (2002), "Reaction path sampling of the reaction between iron(II) and hydrogen peroxide in aqueous solution", *J. Phys. Chem.* **106**, 7902.

[6] Bolhuis, P.G. (2003) "Transition-path sampling of beta hairpin folding", *Proc. Natl Acad. Sci. USA* **100**, 12129.

[7] Dellago, C., Bolhuis, P.G. and Geissler, P.L. (2002) "Transition path sampling", *Adv. Chem. Phys.* **123**, 1.

[8] Bolhuis, P.G., Chandler, D., Dellago, C. and Geissler, P.L. (2002) "Transition path sampling: throwing ropes over rough mountain passes, in the dark", *Ann. Rev. Phys. Chem.* **53**, 291.

[9] Chandler, D. (1978), "Statistical mechanics of isomerization dynamics in liquids and the transition state approximation", *J. Chem. Phys.* **68**, 2959.

[10] Dellago, C., Bolhuis, P.G. and Chandler, D. (1999), "On the calculation of rate constants in the transition path ensemble", *J. Chem. Phys.* **110**, 6617.

[11] Straub, J.E., Borkovec, M. and Berne, B.J. (1988) "Molecular dynamics study of an isomerising diatomic in a Lennard-Jones fluid", *J. Chem. Phys.* **89**, 4833.

[12] Weeks, J.D., Chandler, D. and Andersen, H.C. (1971) "Role of repulsive forces in determining the equilibrium structure of simple liquids", *J. Chem. Phys.* **54**, 5237.

[13] Vlugt, T.J.H. and Smit, B. (2001) "On the efficient sampling of pathways in the transition path ensemble", *Phys. Chem. Commun.* **2**, 1.

[14] van Erp, T.S., Moroni, D. and Bolhuis, P.G. (2003) "A novel path sampling method for the calculation of rate constants", *J. Chem. Phys.* **118**, 7762.

[15] van Erp, T.S. and Bolhuis, P.G. (2004) *J. Comp. Phys.* (in press).