Sampling Phase Space by a Combined QM/MM ab Initio Car−Parrinello Molecular Dynamics Method with Different (Multiple) Time Steps in the Quantum Mechanical (QM) and Molecular Mechanical (MM) Domains

Tom K. Woo,[†,‡] Peter Margl,[†,§] Peter E. Blöchl,[||] and Tom Ziegler*[,†]

Department of Chemistry, University of Calgary, 2500 University Drive, Northwest, T2N1N4 Calgary, Alberta, Canada, and Institute of Theoretical Physics, Clausthal University of Technology, Leibnizstrasse 10, D-38678 Clausthal-Zellerfeld, Germany

Received: September 20, 2001

This study considers a scheme for sampling phase space in large molecules based on Car−Parrinello ab initio molecular dynamics. The scheme makes use of a combined quantum mechanics and molecular mechanics (QM/MM) method augmented with a multiple-time-step technique. This scheme makes it possible to oversample the computationally less expensive MM region relative to the QM domain. The goal here is to provide better ensemble averaging in the MM region that is usually larger in size and therefore typically has a higher degree of configurational variability. It is shown that the multiple-time-step integrator will generate the same trajectory as a standard molecular dynamics integrator. Moreover, with a gradual rescaling of masses, the energy conservation of a multiple-time-step simulation can be satisfied to the same extent as a standard simulation. Finally, it is demonstrated that the multiple-time-step QM/MM method can accelerate the equilibration and configurational sampling of a molecular dynamics simulation as it is used in thermodynamic integration. The scheme is not intended as a tool for generating trajectories in actual dynamics.

1. Introduction

The Car−Parrinello[1] based ab initio molecular dynamics method (CP-AIMD) has over the past 15 years emerged as a powerful tool in chemical and physical research. The CP-AIMD scheme is still computationally demanding for larger molecules, and there has, as a consequence, been several attempts to improve the way in which the required computational time scales with size. The improvements include linear scaling methods[2,3] and sophisticated techniques for integrating the equations of motion.[4]

Taking a different route toward improved scaling with size, we have in a previous study implemented[5] the combined quantum mechanics and molecular mechanics (QM/MM) method[6] into the Car−Parrinello ab initio molecular dynamics framework. In the present investigation we report on the implementation of a multiple-time-step scheme[4] such that phase space can be sampled at a faster rate than the quantum mechanics domain, thereby providing better ensemble averaging during the calculation of the free energy barriers. The method is intended for sampling phase space, but not for generating trajectories in real dynamics. Section 2 will deal with details of the QM/MM implementation,[5] whereas section 3 describes our multiple-time-step scheme.

2. Outline of the Combined QM/MM Car−Parrinello Ab Initio Molecular Dynamics Implementation

We shall in this section deal briefly with details of the QM/MM implementation[5] not reported previously.

Partitioning of the Real System into QM and MM Domains. In the combined QM/MM method the molecular system is divided into a smaller part described by a quantum mechanical method and a larger part accounted for by molecular mechanics. This partitioning is illustrated in Figure 1 where the complete system is represented by 3-methylhexane (**1a**). This system is divided into a quantum mechanical ethane core (**1b**) and three molecular mechanics regions represented respectively by two methyl and one propyl group (**1b**). The partitioning into the QM and MM domains crosses through three C−C bonds. We denote the carbons that remain in the QM region the C* link atoms (**1c**), whereas the carbons residing in the MM domain are denoted the C# link atoms (**1c**). Finally, hydrogen capping atoms (H′) are added to the QM region in order to satisfy the valence (**1c**) in this domain. The capping atoms are not part of the real system.[5] As described elsewhere,[5b] the bond angle and dihedral angle used to define the link atom are set equal to the bond angle and dihedral angle used to define the capping atom, respectively. Also the bond distance used to define the link atom is equal to the bond distance defining the capping atom except for a constant ΔR. This is required in order not to introduce redundant degrees of freedom.[5b] The present investigation did not consider any electrostatic coupling between the QM and MM atoms. However, such a coupling can be introduced.[5c]

The combined QM/MM total energy in this framework can be defined[5b] as

$$E_{\mathrm{QM/MM}} = E_{\mathrm{QM}} + E_{\mathrm{MM}} \tag{1}$$

where E_{QM} is simply the energy of the QM model system with the capping atom included, and E_{MM} is the sum of all the MM energy terms of the real system that contain at least one MM atom (for example, the bond stretching potential between the QM link atom and the MM link atom is handled by the

[†] University of Calgary.
[‡] Permanent address: Department of Chemistry, University of Western Ontario, London, Ontario, Canada.
[§] Permanent address: Corporate R&D Computing, Modeling and Information Sciences, The Dow Chemical Company, 1702 Bldg, Midland, MI 48674.
[||] Clausthal University of Technology.

10.1021/jp0135860 CCC: $22.00
Published on Web 01/24/2002

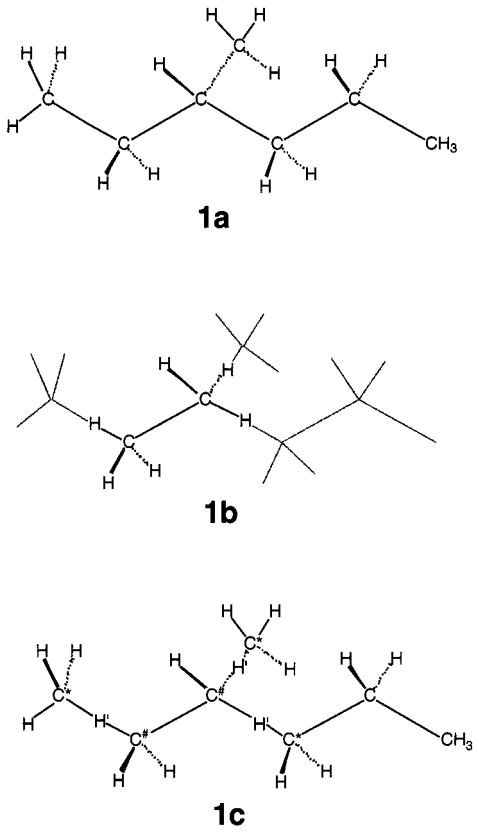

Figure 1. Partitioning scheme used in QM/MM method. (a) Complete system represented by 3-methylhexane. (b) Partitioning into MM part (dotted line) and QM part represented by ethane molecule. (c) Position of link atoms and capping-atoms.

appropriate MM stretching potential). We note that most authors define the total combined QM/MM energy expression with a component that express the interaction energy between QM and MM regions. In eq 1 E_{QM} and E_{MM} include these interaction energies.[5]

Verlet Integration Scheme. In a molecular dynamics simulation, the motion of the nuclei is determined by integrating Newton's second-order differential equation to obtain the position of all atoms as a function of time. Generally, there is no analytical solution to the problem for molecular systems, and numerical methods have to be utilized. In other words, given the velocities and positions at a time t, one determines the same quantities with reasonable accuracy at a later time $t + \Delta t$ using the calculated forces on the nuclei. Consider the motion of an atom along a coordinate x. If one knows the position $x(t)$ at time t, the position at time $t + \Delta t$ is given by a standard Taylor expansion, as shown in eq 2.

$$x_i(t+\Delta t) = x_i(t) + \dot{x}_i(t)\Delta t + \ddot{x}_i(t)\frac{\Delta t^2}{2} + O(\Delta t^3) + \ldots \quad (2)$$

There are a variety of integration schemes for solving the above problem. The most common methods truncate after the quadratic term and therefore are of third-order accuracy in the time step Δt. One such method is the Verlet[7] integration scheme that is commonly used in ab initio molecular dynamics. In Verlet dynamics, the position of the system at the at a time $t + \Delta t$ is given by a function of the position at the current time step, $x_i(t)$, previous time step, $x_i(t-\Delta t)$, and the forces,

$$F_i(t) = -\delta E_{Tot}/\delta x_i \quad (3a)$$

at the current time step, as related in eq 3b.

$$x_i(t+\Delta t) = 2x_i(t) - x_i(t-\Delta t) + \frac{F_i(t)}{m_i}\Delta t^2 \quad (3b)$$

With each new time step, a new geometry is generated and therefore the forces on the nuclei have to be recalculated. To simulate molecular vibrations, the time step Δt must be at least a third smaller than the period of the fastest vibration.

Combined QM/MM Lagrangian. The combined QM/MM methodology has been implemented within the PAW ab initio molecular dynamics package of Blöchl[8] by extending the Car−Parrinello Lagrangian[1] to include the molecular mechanics subsystem:

$$L = \Sigma\mu_i\langle\ddot{\Psi}_i|\ddot{\Psi}_i\rangle + \frac{1}{2}\sum M_{I,QM}\dot{R}_{I,QM}{}^2 - E_{DFT}(|\Psi\rangle,R_{QM}) + \sum_{i,j}\Lambda_{ij}(\langle\Psi_i|\Psi_j\rangle - \delta_{ij}) + \frac{1}{2}\sum M_{I,MM}\dot{R}_{I,MM}{}^2 - E_{MM}(R_{QM},R_{MM}) \quad (4)$$

The first three terms of eq 4 are equivalent to the ordinary quantum mechanical terms in the CP scheme,[1] whereas the last two terms in eq 4 refer to the kinetic energy of the MM nuclei and the potential energy derived from the MM force field. Equation 4 essentially describes the coupled equations of motion of three subsystems: the QM nuclei, the QM wave function and the MM nuclei. In this way, the first three terms of eq 4 refer to the kinetic and potential energy of the "capped" QM model system.

Thermostating. When the nuclei of a molecular simulation follow Newton's equations of motion, the total energy of the system is conserved. Furthermore, when the volume is held constant (cell size fixed), then the simulation will generate a microcanonical (NVE) ensemble. When chemical reactions are studied, this type of dynamics may be undesirable because the excess heat that is dissipated or absorbed during a reaction could alter the temperature of the system to unwanted values. For this reason the temperature of a molecular dynamics simulation is often controlled, or thermostated, such that a canonical or NVT ensemble is generated. A common thermostating procedure is to couple the molecular system to a heat bath through the method of Nosé. [9a] In this method, an extra degree of freedom corresponding to the heat bath is added to the existing degrees of freedom of the molecular system. A kinetic energy and a potential energy term representing the heat bath are added to the Hamiltonian, which allows energy to flow dynamically back and forth between the system and the heat bath. The method was subsequently reformulated by Hoover,[9b] and so this technique is also referred to as a Nosé−Hoover thermostat. The Nosé thermostat effects the nuclear motion via a velocity dependent friction term in the equations of motion as expressed in

$$m_i\ddot{x}_i = F_i - m_i\dot{x}_i\dot{\zeta} \quad (5)$$

The friction term is governed by the variable ζ, which obeys its own equation of motion as given by

$$Q\ddot{\zeta} = 2\left[\sum_i \frac{1}{2}m_i\dot{x}_i{}^2 - \frac{1}{2}gk_BT\right] \quad (6)$$

In this way, the kinetic energy of the nuclei fluctuates about the mean value $(^1/_2)gk_BT$, where g is the number of degrees of freedom of the nuclear system, k_B is the Boltzmann constant, and T is the desired physical temperature of the simulation. Q in eq 6 is an inertial parameter that controls the time scale of the thermal fluctations. It should be noted that a simulation that is Nosé thermostated also conserves energy if the thermostat potential, $gk_BT\zeta$, and kinetic energy, $(^1/_2)Q\dot{\zeta}^2$, are added to the total energy.

It has recently been shown by Blöchl and Parrinello[10] that the Nosé thermostating method for maintaining constant temperature molecular dynamics can be extended to the fictitious kinetic energy of the wave function in the Car−Parrinello methodology. In this context the thermostat acts to prevent the electronic wave function from slowly drifting away from the Born−Oppenheimer potential surface. The thermostating of the wave function in this way allows stable Car−Parrinello molecular dynamics to be performed for long periods of time without the need of perioidically quenching the wave function to the Born−Oppenheimer potential surface.

A separate Nosé thermostat for the molecular mechanics subsystem has been implemented. Since the MM and QM regions are strongly coupled, there will be energy flux between the two subsystems and their thermostats during a simulation. To prevent strong coupling between the thermostats themselves, the inertial parameters Q (eq 6) are chosen to maintain a large disparity in the time scales of the heat bath fluctuations in the QM and MM regions (approximately 5 times difference). A single Nosé thermostat for the whole QM/MM nuclear system was not implemented because of the complications that would arise with the QM/MM multiple-time-step methodology to be described later.

Equilibration. To properly sample the canonical ensemble, the system must be thermally equilibrated. Thermal equilibration is often done by instantly exciting the system to the desired temperature with a random excitation vector. This is followed by thermostated dynamics for at least a few picoseconds, thereby ensuring that all vibrational modes are excited to an equal extent. There are two problems with this approach in the Car−Parrinello scheme. First, an immediate pulse of kinetic energy in order to excite the nuclei to the desired temperature is likely to dislodge the wave function from the Born−Oppenheimer surface. Second, long periods of equilibration are expensive with ab initio molecular dynamics. For these reasons we take a modified approach to equilbration. First, the nuclei are excited by a series of slowly growing pulses. Each of the excitation vectors is chosen to be orthogonal to the already excited modes, thereby ensuring an evenly distributed thermal excitation. This is followed by a short period of thermostated dynamics. We have found this approach more efficiently achieves a thermally equilibrated system than the convention method. Again, the QM and MM regions are strongly coupled, the MM region cannot be instantly heated to a desired simulation temperature without abruptly dislocating the wave function of the QM region from the Born−Oppenheimer surface. For this reason, the slow warm-up procedure described above has been implemented for the MM subsystem as well.

Mass Rescaling. Since the configurational averages in classical molecular dynamics do not depend on the masses of the nuclei,[11] a common technique to increase the sampling rate involves replacing the true masses with more convenient ones. Since nuclear velocities scale with $m^{-1/2}$, smaller masses move faster and therefore potentially sample configuration space faster. As a result, the masses of the heavy atoms can be scaled down in order to increase sampling. For example, we commonly rescale the masses of C, N, and O in our simulations from 12, 14, and 16 amu, respectively, to 2 amu. There is a limit to the mass reduction, because at some point the nuclei move so fast that the simulation time step has to be reduced. At this point there is no gain in reducing the masses further because if the time step has to be shortened, we have to perform more time steps to achieve the same amount of sampling. It is for this reason we generally scale our hydrogen masses up from 1 to 1.5 amu or higher in order to use a larger time step.

Periodicity. The periodicity of the plane basis sets method may be thought to inhibit the practical application of the QM/MM method within the Car−Parrinello framework. It is often assumed that the simulation cell of the Car−Parrinello simulation must be large enough to encapsulate both the QM and MM regions. If this were true, a larger MM region would then require a larger simulation cell and greater computational effort even though the QM model system may remain the same. This is so since the computational effort of the Car−Parrinello method with plane wave basis scales in part with the dimensions of the simulation cell. Fortunately, this is not true and the size of the molecular mechanics region is inconsequential to the choice of the QM cell size. Even when true electrostatic coupling between the QM and MM regions is invoked, the size of the MM system does not effect the cell size of the QM model system as demonstrated in elsewhere.[5] Furthermore, the molecular mechanics region can itself have periodic boundary conditions with a cell size and type that is independent of the cell size of the QM calculation. The addition of periodic boundary conditions in the MM region is necessary for QM/MM solvent simulations.

QM/MM Features. The molecular mechanics code has been completely written in FORTRAN90, and the core components of the code are shared with the QM/MM implementation within the Amsterdam Density Functional (ADF) program package.[5,12] Thus, all of the features of the ADF QM/MM implementation are shared with the PAW QM/MM implementation. Detailed explanation of the features and the usage of the implementation are described in the PAW QM/MM user's manual.[13]

3. QM/MM Multiple Time Step

Interest in the linear scaling of electronic structure calculations has surged with the recent developments from the groups of Yang,[14] Head-Gordon,[15] and Scuseria.[16] Although these techniques allow the time of an electronic structure calculation to scale linearly with the size of the system, they do not address the problem of nonlinear scaling of the geometry optimization or sampling of configuration space. In this section, we address the issue within the framework of the QM/MM ab initio molecular dynamics methodology. Our goal is to increase the configurational sampling of the MM region without significantly increasing the computational effort put into the QM region.

A QM/MM simulation will most often involve a small QM region embedded within a much larger MM domain, such as with the simulation of the active site chemistry within an enzyme. Associated with the larger size of the MM region is a potential energy surface that is more complex and one that is likely to possess a higher degree of configurational variability. This necessitates an increased degree of sampling in the MM region in order to obtain meaningful ensemble averages from a simulation. In principle, increased sampling of the MM region can be achieved by the technique of mass rescaling. In classical molecular dynamics, configurational averages do not depend on the masses of the nuclei, and therefore the true nuclear masses can be replaced with more convenient values.[11,17] Rescaling of

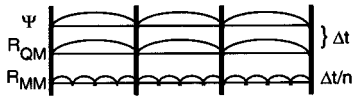

Figure 2. Representation of a simple multiple-time-step scheme within the Car−Parrinello framework. The wave function and the QM subsystems are propagated with a time step Δt, while the MM subsystem is over sampled with a smaller time step $\Delta t/n$.

nuclear masses to smaller values, which allows the particles to move faster, increases the rate of configurational sampling, which is proportional to the inverse square root of the mass, $m^{-1/2}$, of the particles. Therefore, by decreasing the masses of the MM nuclei relative to those in the QM region, one can differentially sample the configuration space of the two regions. However, if the same time step is used for both regions, a limit to the practical over sampling is quickly reached. This is because, as we rescale the masses and increase the speed of the MM nuclei, a smaller time step is required in order to accurately integrate the equations of motion of the fast moving system. Thus, when the QM and MM regions are propagated simultaneously, the "slow" QM region will be propagated with unnecessarily small time steps. This is undesirable because the QM derived forces will be changing slowly and much time will be wasted recalculating these forces at each of the small time steps.

One way to overcome this problem is to propagate the two regions asynchronously with what has been termed multiple-time-step molecular dynamics.[18] The multiple-time-step method in the Car−Parrinello QM/MM framework is represented in Figure 2, where the wave function and the QM nuclei are propagated with a large time step, Δt, while the less computationally demanding MM region is oversampled with a time step of $\Delta t/n$ n times. Thus, the multiple-time-step method in combination with mass rescaling can be applied to the QM/MM molecular dynamics methodology as to differentially sample the QM and MM regions, thereby increasing the sampling of the MM domain without increasing the computational expenditure in the QM region.

The multiple-time-step techniques have been developed since 1978[18] to treat systems with high- and low-frequency motion and/or short- and long-range forces more efficiently. However, the simple algorithm often used (as depicted in Figure 2) has no rigorous basis in theory.[19] Recently, Tuckerman and coworkers[4b,20,21] have pioneered the development of multiple-time-step methods where there is a rigorous separation of time scales. We have adopted the reversible multiple-time-step algorithm of Tuckerman et al.,[4] which allows for numerically stable, energy conserving, multiple-time-step molecular dynamics to be performed. In our implementation, the original formalism, which was based on the velocity Verlet propagation algorithm, has been modified to accommodate the standard Verlet algorithm of eq 3b used in the PAW program.

Displayed in Figure 3 is a schematic representation of the reversible multiple-time-step procedure. Consider the system at a time $T = t_o$, where we are propagating the slow QM subsystem with a long time step Δt and the faster MM subsystem with a short time step Δt/n. First, the faster MM subsystem is propagated $n/2$ times for half of the long interval, $\Delta t/2$, as shown in Figure 3a. Now the slow QM degrees of freedom are propagated for a full long time step, as shown in Figure 3b. It is crucial to point out that the forces used to propagate the slow QM system are the average of the forces evaluated at the half-interval ($T = t_o - \Delta t/2$ and $t_o + \Delta/2$) values of the MM degrees of freedom. Thus, the new positions of the QM degrees of freedom (both nuclear and electronic) are

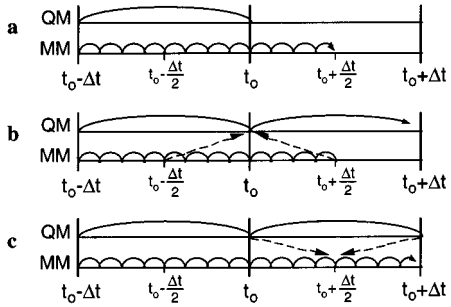

Figure 3. Schematic representation of a Tuckerman's reversible multiple-time-step procedure implemented within the PAW QM/MM method: (a) propagation of the first half of the MM subsystem starting at $T = t_o$; (b) propagation of the slow QM subsystem that occurs at $T = t_o + \Delta t/2$; (c) propagation of the second half of the MM subsystem. The dashed arrows in (b) and (c) illustrate what forces are used to propagate the systems in these intervals (see text for more details).

expressed in

$$x_{QM}(t_o+\Delta t) = 2x_{QM}(t_o) - x_{QM}(t_o-\Delta t) + \frac{\Delta t^2}{M_{QM}}F_{QM} \quad (7)$$

where the forces on the QM nuclei are defined in eq 8, where $c(t)$ denotes the expansion coefficients of the Kohn−Sham wave function. Following the propagation of the slow QM system

$$F_{QM}(t_o) = \frac{1}{2}F_{QM}\left\{x_{QM}(t_o),c(t_o),x_{MM}\left(t_o-\frac{\Delta t}{2}\right)\right\} + F_{QM}\left\{x_{QM}(t_o),c(t_o),x_{MM}\left(t_o+\frac{\Delta t}{2}\right)\right\} \quad (8)$$

with the large time step, the second half of the small time steps involving the MM subsystem is executed. In this interval, the forces on the MM subsystem at time $T = t_o + \Delta t/2$ are evaluated using the updated values of the QM degrees of freedom at $t_o + \Delta t$ averaged with the same values at time t_o. This is expressed in

$$F_{MM}\left(t_o+\frac{\Delta_t}{2}\right) = \frac{1}{2}F_{MM}\left\{x_{QM}(t_o),c(t_o),x_{MM}\left(t_o+\frac{\Delta t}{2}\right)\right\} + F_{MM}\{x_{QM}(t_o+\Delta t),c(t_o+\Delta t),x_{MM}(t)\} \quad (9)$$

The forces on the MM subsystem for the remainder of the interval are given by

$$F_{MM} = F_{MM}\{x_{QM}(t_o+\Delta t),c(t_o+\Delta t),x_{MM}(t)\} \quad (10)$$

We reiterate that the slow QM degrees of freedom are propagated with forces derived from the MM degrees of freedom at $T = t_o + \Delta t/2$ and not $T = t_o$. Conversely, the faster MM degrees of freedom are half propagated with forces derived from the QM degrees of freedom at $T = t_o$ and half with forces derived from the QM degrees of freedom at $T = t_o + \Delta t$. For both the QM degrees of freedom and the MM degrees of freedom, this gives a better or more "averaged" representation of the complementary degrees of freedom over the entire large time step.

Generally, when the "trick" of mass rescaling is used in classical molecular dynamics simulations, the masses of light atoms such as hydrogen are scaled up to make them heavier.[17] This allows for larger time steps to be used during the integration of the equations of motion and therefore effectively increases the simulation times. Conversely, when masses are rescaled to smaller values, the particles move faster and are able to sample configuration space faster. When the masses of the MM atoms

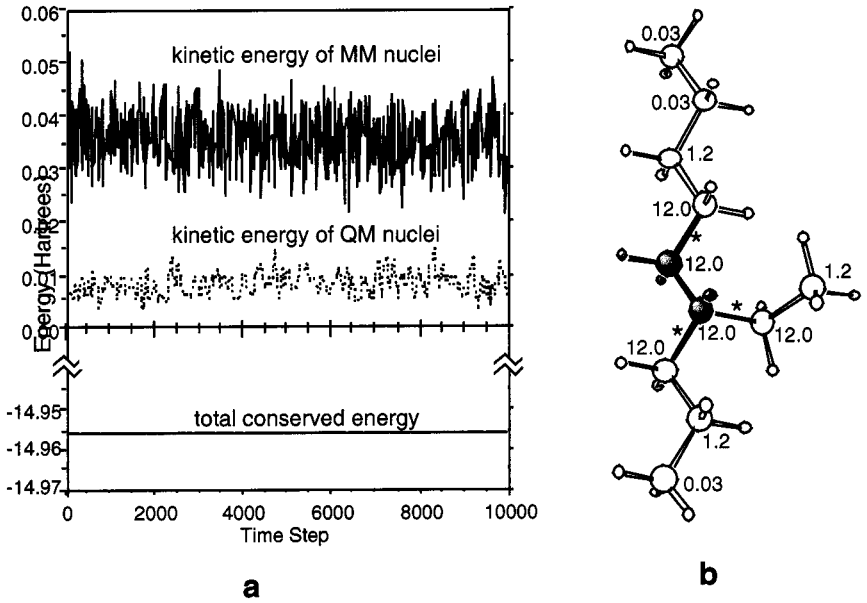

Figure 4. (a) Kinetic energy of the QM and MM nuclei for a multiple-time-step QM/MM dynamics simulation of 4-ethylnonane with the rescaling of the masses depicted in (b). To demonstrate energy conservation of the dynamics, the total energy of the system is shown at the same scale as the kinetic energies. (b) QM/MM partitioning of 4-ethylnonane where the shaded regions represent the QM region. Covalent bonds labeled with asterisks denote the QM/MM link bonds. All of the link bonds have been capped with hydrogen atoms such that the QM model system is ethane. The rescaling of the atomic masses of carbon are depicted in atomic mass units.

are scaled to smaller values, the multiple-time-step procedure allows for the proper integration of the faster moving atoms without increasing the number of force evaluations from the electronic structure calculation of the QM model system. Thus, in the combined QM/MM molecular dynamics framework, this unique combination of techniques allows for oversampling of the MM partition of the system, without increasing the computational expense of the QM subsystem. Since we are the first to develop the multiple-time-step method in this way, the next section will be devoted to testing the validity of the combined QM/MM multiple-time-step dynamics approach.

Test Results of the Multiple-Time-Step QM/MM Approach. Test results of the multiple-time-step QM/MM methodology with mass rescaling will be presented in this section. The method will first be showcased in order to establish how the methodology is intended to be applied. This will be followed by a more systematic validation of the methodology.

Test of Energy Conservation. A combined QM/MM multiple-time-step dynamics simulation of 4-ethylnonane (Figure 4) has been performed. The QM/MM partitioning of the system is shown in Figure 4b, where the link bonds are denoted with asterisks. The calculation involved ethane as the model QM system for which the electronic structure was calculated at the gradient-corrected BP86 DFT level.[22−25] A time step of 3.0 au (∼0.07 fs) was used for the QM system, whereas the MM subsystem was oversampled by a ratio of 20:1 such that a small time step of 3.0/20 = 0.15 au was utilized for the MM region. The simulation involved 10 000 QM time steps (720 fs) and 200 000 MM time steps. The rescaling of the masses is depicted in Figure 4b. Masses of 12.0 and 1.5 amu were used for the C and H atoms, respectively, in the QM system, whereas the masses in the MM region were rescaled 400-fold to enhance the sampling rate by a factor of approximately 20. The rescaling of the masses is detailed in Figure 4b.

Plotted at the top of Figure 4a is the kinetic energy of the MM and QM subsystems. The disparity in the frequency of QM and MM kinetic energies reveals how differential sampling of the two regions is achieved with the method. In this simulation, an average temperature of 300 K is maintained throughout the simulation for both the QM and MM regions. However, as a result of the mass rescaling in the MM region, the MM kinetic energy oscillates much more rapidly than the QM kinetic energy. This is the goal of the methodology since the faster fluctuations in the MM kinetic energy implies a faster motion of the MM subsystem and ultimately a more rapid sampling of the configuration space. This will be further illustrated later.

Figure 4a also displays total energy of the system plotted at the same energy scale as the QM and MM kinetic energies. This illustrates the stability of the multiple-time-step method, which displays no significant drift in the total energy over the period of the whole simulation. There is also no drift in either the kinetic energies of the MM or QM nuclei shown in Figure 4a. This is notable because it points out that there is no significant energy flux between the nuclear kinetic energies and the fictitious kinetic energy of the QM wave function. It is conceivable that the rapidly fluctuating kinetic energy of the MM subsystem might couple with the fictitious kinetic energy of the QM wave function. This would lead to an eventual dislodging of the wave function from the Born−Oppenheimer surface and unphysical dynamics in the QM subsystem. The stability in the nuclear kinetic energies during the course of this simulation reveals that this is not occurring. We note here that no thermostating was applied to any of the subsystems during the dynamics. Drift can also result from net energy flux between the kinetic energies of the QM and MM nuclei, a real physical effect in a nonequilibrium state. However, in this simulation, the dynamics shown in Figure 4a was preequilibrated.

Comparison of Trajectories from Multiple-Time-Step and Standard Verlet Propagators. We now turn our attention to a systematic validation of the method and implementation. First, we intend to demonstrate that starting from the exact same system (structure and masses), the multiple-time-step propagator generates the same trajectory as the standard Verlet propagator. For this we have performed three simulations of 4-ethylnonane, starting from the same structure with no initial velocities. That is, for these simulations the dynamics was initiated from a frozen structure and the resulting temperature of the system was

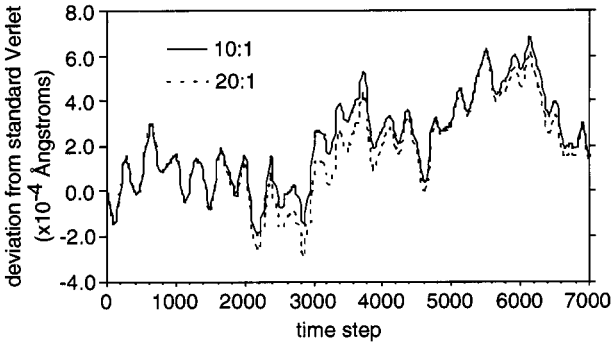

Figure 5. Deviation in the C1−C9 distance of 4-ethylnonane between the standard Verlet trajectory and multiple-time-step trajectories.

approximately 100 K. Acting as our control is a simulation performed with the standard Verlet propagator with a time step of 3.0 au for both the QM and MM subsystems. Two QM/MM multiple-time-step simulations were performed with oversampling ratios of 10:1 and 20:1 such that the time steps in the MM region are 0.3 and 0.15 au, respectively. In all three simulations, the QM/MM partitioning is the same as that depicted in Figure 4b such that ethane constitutes the QM model system. However, unlike the simulation depicted in Figure 4, the masses in these simulations were not rescaled.

Small structural changes in the backbone of the 4-ethylnonane system are likely to be magnified at the extremities of the alkane. Thus, a geometric parameter that is likely to be sensitive to differences in the trajectories is the distance between the two terminal carbons (C1 and C9) of 4-ethylnonane system. Plotted in Figure 5 is the deviation in this distance between the standard Verlet trajectory and the two multiple-time-step trajectories. The difference in this parameter throughout the simulation remains exceptionally small (of the order of 10^{-4} Å).

Stability of QM Energy to Sampling Procedures in MM Region. Another property that is highly sensitive to geometric differences in the trajectory is the electronic energy of the QM model system. Figure 6a plots the total DFT energy of the QM model system for all three trajectories. At the scale of the oscillations in this energy, the three trajectories cannot be distinguished over the course of the simulation (∼500 fs). Shown in Figure 6b is the magnitude at which the electronic energy of the two trajectories generated from the multiple-time-step method (10:1 and 20:1) and the standard Verlet propagator deviate. The deviation in the DFT energy is of the order of 10^{-5} Hartrees with no progressive increase during the 500 fs simulation. We conclude that, well within chemical accuracy, the multiple-time-step propagator generates the same trajectory as the standard Verlet propagator.

Energy Conservation with Both Oversampling and Mass Rescaling in MM Region. We now turn our attention to the energy conservation of the multiple-time-step QM/MM molecular dynamics. To determine the limitations of the multiple-time-step QM/MM method, we will compare the energy conservation of a series of molecular dynamics simulations of 4-ethylnonane where various mass rescaling schemes and multiple-time-step oversampling ratios have been adopted. Acting as our control is the standard Verlet simulation of the system where the masses have not been rescaled. Compared in Figure 7 is the energy conservation of a series of simulations where the masses in the MM region have been rescaled by a factor of $^1/_{400}$ as to oversample the MM region by a ratio of 20:1. The atomic masses of the alkane have been rescaled in an abrupt manner and a gradual manner. The abrupt rescaling of the masses is shown in Figure 8a and corresponds to what is

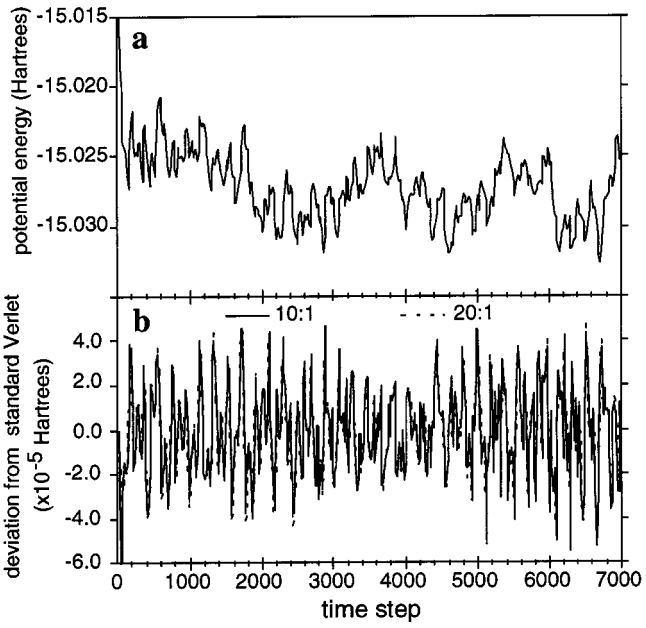

Figure 6. (a) Potential energy of the QM model system for the standard Verlet and the multiple-time-step Verlet 20:1 and 10:1 oversampled. In the scale shown, the potential energy of the various trajectories cannot be differentiated. (b) deviation in the potential energy between the standard Verlet algorithm and the multiple-time-step Verlet algorithm 10:1 (solid) and 20:1 (dashed) oversampled. The plots illustrate the similarity between the standard Verlet and multiple-time-step Verlet trajectories that were initiated from the same structure.

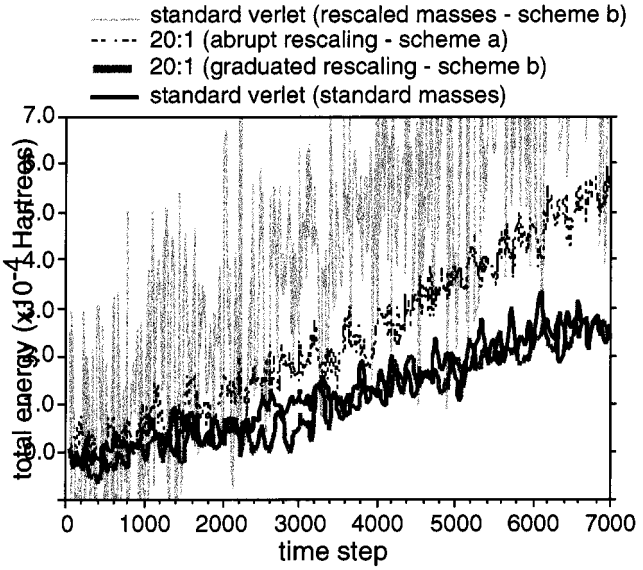

Figure 7. Comparison of the energy conservation during simulations of 4-ethylnonane. Plotted are the total (conserved) energies of the simulations relative to the initial value. The schemes referred to in the plot legend correspond to the mass rescaling schemes depicted in Figure 8. The temperature in these simulations was 300 K.

labeled rescaling scheme a. In this rescaling scheme, the atomic masses are immediately rescaled by a factor of $^1/_{400}$ as we move from the QM region to the MM region. In this way, a carbon atom that has a mass of 12.0 amu will be bonded to a carbon atom with a rescaled mass of 0.03 amu. A more gradual rescaling of the masses is adopted in scheme b, as depicted in Figure 8b. Here the masses of atoms adjacent in connectivity to the MM link atom are first rescaled by a factor of $^1/_{10}$. We have used the modified IMOMM coupling scheme.[5] In this coupling scheme the MM link corresponds to the nuclear degree

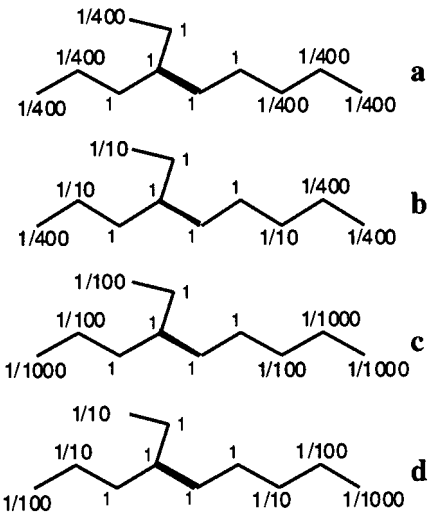

Figure 8. Mass rescaling schemes used for the QM/MM dynamics simulations of 4-ethylnonane shown in Figures 7 and 9. The values next to the backbone carbon atoms reflect the fraction that the standard masses were rescaled. Thus, $^1/_{400}$ means that the 12.0 amu mass of carbon was rescaled to a value of 0.03 amu. Masses of the hydrogen atoms were similarly rescaled.

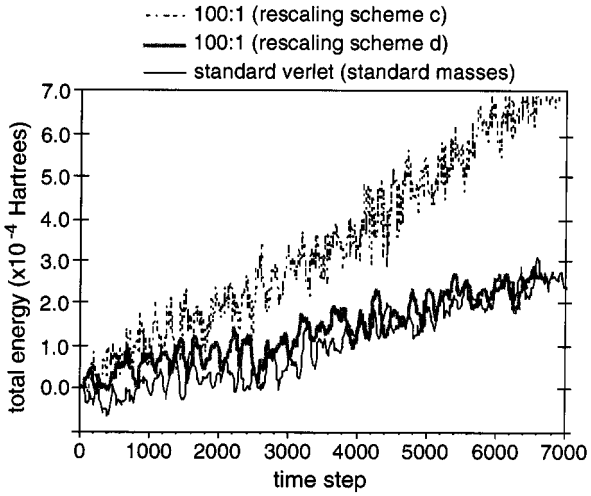

Figure 9. Comparison of the energy conservation of the multiple-time-step simulations of 4-ethylnonane. Oversampling ratios of 100:1 were utilized with rescaling of the masses shown in Figure 8. Plotted are the total (conserved) energies of the simulations relative to the initial value. The energy conservation of the multiple-time-step simulations are compared to that of the standard Verlet simulation with no mass rescaling. The temperature in these simulations was 300 K.

of freedom of the QM dummy atom. Since the QM dummy atom is propagated by the large time step, its mass is not rescaled. All other MM atoms further away in connectivity are rescaled by a factor of $^1/_{400}$. Thus, there is a gradual rescaling of the masses as we pass from the QM region to the MM region.

The total energy of the systems relative to the initial value is plotted in Figure 7. The control system (standard Verlet and no mass rescaling) exhibits a slow linear drift of approximately $+6 \times 10^{-7}$ Hartrees for each femtosecond of simulation time. The energy conservation of the simulation with the gradual mass rescaling scheme matches that of the control simulation. However, with the abrupt rescaling scheme the drift rate is double that of the control, indicating a loss of integration accuracy compared to the control. With an abrupt rescaling, the high-frequency motions of the light MM atoms are propagated into the QM region. This degrades the energy conservation because the larger time step used for the QM subsystem is unable to properly integrate this high-frequency motion. This effect is minimized when the masses are gradually rescaled since the high-frequency motions of the MM subsystem are dampened by the progressively heavier masses. We conclude that in order to apply the multiple-time-step QM/MM method, there must be a graduated rescaling of the masses.

Also shown in Figure 7 is the energy conservation when the masses are rescaled (scheme b of Figure 8b) but the multiple-time-step method is not used. In this simulation, the same large time step is used to integrate both the fast MM and slow QM subsystems. It is apparent from the large fluctuations and the rapid drift of the total energy during this simulation, that the large time step is unable to properly capture the fast motion of the light MM atoms. Thus, to effectively rescale the masses as to increase the sampling in the MM region, a multiple-time-step integrator is necessary.

Energy Conservation for Different Oversampling Schemes in MM Region. Next, we investigate the amount of oversampling that can be effectively applied to our 4-ethylnonane complex. Figure 9 compares the energy conservation of the control to a pair of simulations were the MM masses have been rescaled by a factor of $^1/_{1000}$ and a multiple-time-step ratio of 100:1 has been applied. The rescaling schemes of the two simulations are shown in Figure 8c,d. The rescaling schemes are both gradual, but the step sizes in scheme c are greater than in scheme d. With rescaling scheme c, the energy conservation is diminished from the control as evidenced by the faster drift in the total energy. With scheme d, the more gradual rescaling, the level of energy conservation of the control is matched. Although the integration accuracy with scheme d is excellent, the amount of oversampling at the 100:1 level is minimal since only one carbon atom (terminal C9) is rescaled by the $^1/_{1000}$ factor. The results suggest that, to attain the level of energy conservation of the control, the masses can be graduated by no more than an order of magnitude for each level of connectivity away from the QM subsystem. Therefore, the necessity of using a gradual rescaling of the masses imposes a practical limit to the amount of oversampling that can be achieved with this technique.

Acceleration of the Equilibration Process. Finally, we demonstrate how the multiple-time-step QM/MM technique can accelerate the equilibration process and ultimately configurational sampling compared to a standard simulation in which there is no mass rescaling. Two simulations of normal undecane (C_{11}) have been performed where the terminal methyl groups make up the QM region as illustrated in Figure 10. That is, the calculation of the QM model system involved two methane molecules contained within a 8.5 Å cubic cell. One simulation was performed with the standard Verlet propagator where "standard" masses for the whole system were utilized (12.0 amu for C and 1.5 amu for H). In the other simulation, the standard masses were rescaled in the MM region by a factor of $^1/_{400}$ and the multiple-time-step algorithm was used with an oversampling ratio of 20:1. The masses used in the multiple-time-step simulation are detailed in Figure 10 for the carbon atoms of the backbone (hydrogen atoms were similarly rescaled). A high-energy conformation of the hydrocarbon chain was selected as the initial structure. Thus, with the QM subsystems frozen, the global minimum energy structure of the MM backbone was determined from fully optimizing structures sampled every 1 ps from a 50 ps dynamics simulation run at 800 K. From the same simulation, a high-energy structure of the MM backbone was selected that was conformationally distinct from the global minimum structure. For both simulations, the free, unthermostated dynamics was commenced from this structure with no

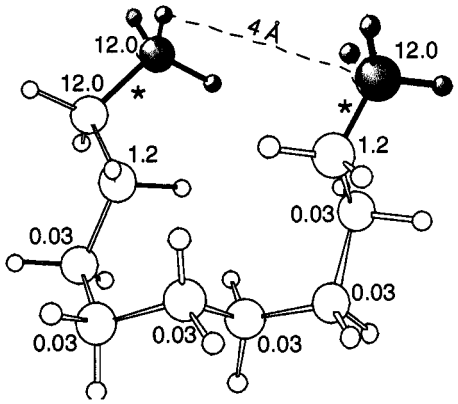

Figure 10. QM/MM partitioning, mass rescaling scheme and constraint definition for the molecular dynamics simulations of *n*-undecane. The shaded regions represent the QM region. The two link bonds, which are labeled with asterisks, have been capped with hydrogen atoms, such that the QM model system consists of two methane molecules. The rescaling of the masses in the multiple-time-step simulation is shown in atomic mass units. Hydrogen atoms are similarly rescaled.

initial velocities. Additionally, the constraint shown in Figure 10 was applied to the system. Here the distance between one of the terminal carbon atoms and a proton of the other terminal carbon atom was constrained to a distance of 4 Å throughout the dynamics (The reason for applying the constraint will be explained later)

The temperature evolution of the QM and MM subsystems for both the standard and the multiple-time-step simulation is shown in Figure 11. Since the dynamics was initiated from a high-energy conformation, there is a steep rise in the temperature of both the QM and MM systems in the first 100 fs of the simulation. When the whole system is properly equilibrated, both the QM and MM systems should be at the same temperature. For the multiple-time-step simulation the temperature evolution plotted in Figure 10b reveals that the QM and MM systems both attain a steady state temperature of approximately 400 K. The temperature of the two systems is well equilibrated at about the 800 fs mark of the simulation. In contrast, for the standard Verlet simulation with no mass rescaling, equilibration does not occur in the first 1400 fs of the simulation. Even at the 1400 fs mark, the fluctuations are large and the disparity between the QM and MM temperatures is still considerable. The continuation of the temperature evolution for the standard time-step simulation is plotted in Figure 12. The plot reveals that equilbration does not begin to occur until the 5500 fs mark, where an equilibrium temperature of 400 K is being established. However, even after 8000 fs temperature equilbration has not been fully established.

It is apparent from Figures 11 and 12 that temperature equilibration of the system occurs much more rapidly with multiple-time-step simulation than it does with the standard simulation. The light masses in the MM region allow this subsystem to equilibrate itself rapidly to the slowly changing QM subsystem. Figure 11b reveals that within the first 20 fs of the simulation a steady temperature of approximately 500 K is attained in the MM region. Following the initial equilibration of the MM subsystem, there is the steady transfer of energy into the QM subsystem. This results in the slow cooling of the MM subsystem as the whole QM/MM system equilibrates to a final temperature of just under 400 K. Thus, the rescaling of the masses in the MM subsystem, which allows for the rapid equilibration of the MM system, facilitates the equilibration of the whole system. In the standard simulation, both subsystems

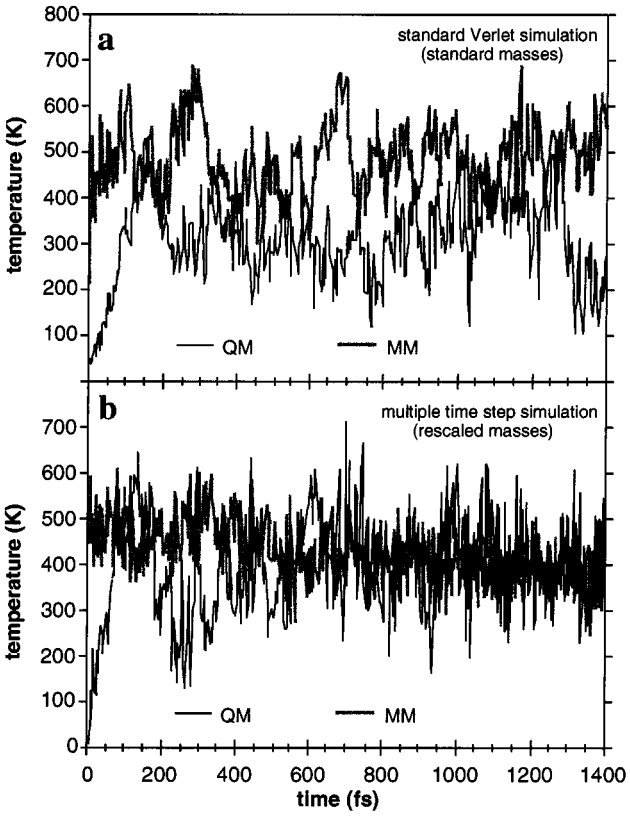

Figure 11. Temperatures of the QM and MM subsystems of the dynamics simulation undecane with (a) standard masses and (b) rescaled masses, as depicted in Figure 9. In simulation (b), where the masses have been rescaled, the multiple-time-step method was used with an oversampling ration of 20:1 (i.e., 20 small time steps for each large one).

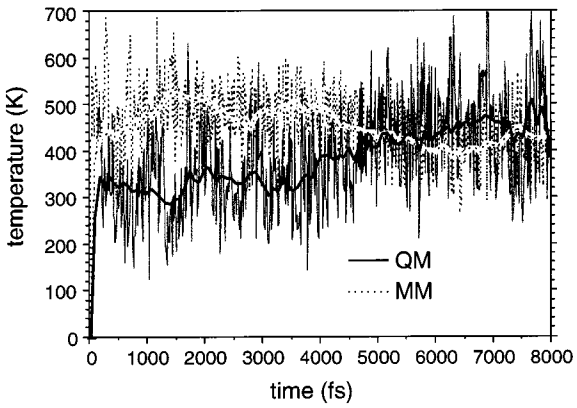

Figure 12. Temperature evolution of the QM and MM subsystems of the dynamics simulation undecane with standard masses and the standard Verlet propagator. This plot is a continuation of Figure 11a. The solid lines are the temperature and its running average (with a window of 100 fs) for the QM subsystem, and the dashed lines are the same quantities for the MM subsystem.

are "heavy" and slow to equilibrate, which delays the process for the whole system.

Convergence of a Constraint Force. The distance constraint was imposed during the dynamics of undecane in order to evaluate if the multiple-time-step method would be effective in accelerating the convergence of the constraint force. The convergence of the force is important since the slow growth simulations we perform to map out reaction free energy profiles involve the integration of the constraint force. A faster convergence of the constraint force would then correspond to

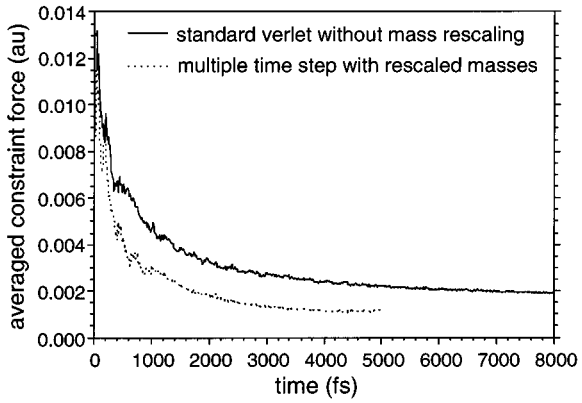

Figure 13. Evolution of the average constraint force during the standard (solid lines) and multiple-time-step (dashed lines) dynamics of undecane. The multiple-time-step simulation was stopped after 5000 fs upon convergence of the average force.

an acceleration in the configurational averaging in these simulations and also smaller errors. Although the undecane system does not correspond to a real reaction, the simulation does imitate the condition where the structure of the MM subsystem strongly influences the force on the reaction coordinate. In the undecane simulation, the constraint force involving the distance between the terminal methyl groups will be regulated by the structure of the hydrocarbon backbone. Thus, the faster the structure of the backbone equilibrates, the faster the constraint should converge. Since most of the backbone resides in the MM partition, we expect that the QM/MM multiple-time-step procedure will accelerate the convergence of the force compared to the standard simulation.

The evolution of the average constraint force, $\langle F \rangle$, during the two undecane simulations is plotted in Figure 13. Since a high-energy starting structure of the backbone was selected, the force on the constraint is initially high. As the backbone structure equilibrates, the force on the constraint should relax to a steady state value. The average constraint force decays much more rapidly in the multiple-time-step simulation than in the standard simulation. This is because the heavier MM subsystem in the standard simulation is slower to settle into an equilibrium conformation. Figure 13 shows that the constraint force in the multiple-time-step simulation acheives a steady state value within approximately 4000 fs. The constraint force in the standard simulation, which should converge to the same value, is slow to decay and convergence is not observed in the first 8000 fs of the simulation. The simulation was stopped after 50 000 time steps, since it was evident that there is an accelerated relaxation of the average force with the multiple-time-step method. Furthermore, since temperature equilibration in the standard simulation does not begin to arise until the end of the 8 ps simulation, the convergence of the constraint force will not occur for some time later.

When the slow growth techique is used to map out free energy surfaces, it is actually the ensemble averaged force on the constraint that is the ideal force used in the integration. Thus, the evolution of the averaged force plotted in Figure 13 emphasis that the multiple-time-step QM/MM method can provide for better sampling and smaller errors. Uncertainty estimates in a slow growth simulation are determined by the amplitude of the fluctuations in the constraint force.[26] The larger the fluctuations, the larger the uncertainies are in the calculated relative free energies. Examination of the unaveraged forces on the constraint (not plotted) show the amplitude of the fluctations is approximately 50% smaller in the multiple-time-step simulation compared to the standard simulation. This corresponds to roughly 20% smaller error bars in the multiple-time-step simulation.

The undecane simulation was designed such that the structure of the MM system strongly influenced the constraint force. Therefore, a significant acceleration in the convergence of the constraint force was observed with the multiple-time-step method. However, most systems are more balanced in that the QM system has a stronger influence on the constraint force. Thus, the benefit of the multiple-time-step method will be less pronounced for most simulations. On the other hand, even if the improvement is small, the additional computational effort is minimal (assuming that the computational expense of the MM system is negligible compared to the QM system). In the Car−Parrinello QM/MM simulation,[27] the calculation of the MM forces amounts typically to less than 0.01% of the total computational effort. Therefore, if the multiple-time-step technique were to be applied to the system with a high oversampling ratio of 100:1, the 100-fold increase in computing the MM forces would still be negligible. Therefore, in these cases where the computational expense of the MM subsystem is negligible and where time-dependent properties are not being investigated, there is no reason not to apply the mutltiple-time-step procedure.

4. Conclusions

A new implementation to carry out Car−Parrinello ab initio molecular dynamics simulations of extended systems using a combined quantum mechanics and molecular mechanics potential is presented. Our implementation allows the QM/MM boundary to cross covalent bonds such that the potential surface of a single molecular system is described by a hybrid potential. Since the potential surface of the molecular mechanics region is usually much less computationally demanding to calculate than that in the QM region, we have implemented a multiple-time-step technique to oversample the MM region relative to the QM region. The goal here is to provide better ensemble averaging in the MM region, which is usually larger in size and therefore usually has a higher degree of configurational variability. We have demonstrated the multiple-time-step integrator will generate the same trajectory as a standard molecular dynamics integrator. Moreover, with a gradual rescaling of masses the energy conservation of a multiple-time-step simulation can be brought to the same level as a standard simulation. Finally, we have demonstrated that the multiple-time-step QM/MM method can accelerate the equilibration and configurational sampling of a molecular dynamics simulation.

Acknowledgment. This investigation was supported by the Natural Science and Engineering Research Council of Canada (NSERC) and by Novacor Research and Technology (NRTC) of Calgary, Alberta, Canada.

References and Notes

(1) Car, R.; Parrinello, M. *Phys. Rev. Lett.* **1985**, *55*, 2471.
(2) Galli, G.; Parrinello, M. *Phys. Rev. Lett.* **1992**, *69*, 3547.
(3) Mauri, F.; Galli, G.; Car, R. *Phys. Rev. B* **1993**, *47*, 9973−9976.
(4) (a) Tuckerman, M. E.; Parrinello, M. *J. Chem. Phys.* **1994**, *101*, 1302. (b) Tuckerman, M. E.; Berne, B. J.; Martyna, G. J. *J. Chem. Phys.* **1992**, *97*, 1990.
(5) (a) Woo, T.K.; Luigi Cavallo, L.; Ziegler, T. *Theor. Chem. Acc.* **1998**, *100* 307−313. (b) Woo, T. K.; Margl, P.; Blöchl, P.; Ziegler, T. *Phys. Chem. B* **1997**, *101*, 7877−7880. (c) Woo, T. K.; Blöchl, P.; Ziegler, T. *THEOCHEM* **2000**, *506*, 313.
(6) (a) Singh, U. C.; Kollman, P. A. *J. Comput. Chem* **1986**, *7*, 718−730. (b) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227. (c) Théry, V.; Rinaldi, D.; Rivail, J.-L.; Maigret, B.; Ferenczy, G. G. *J. Comput. Chem.* **1994**, *15*, 269. (d) Bersuker, I. B.; Leong, M. K.; Boggs, J. E.; Pearlman,

R. S. *J. Comput. Chem.* **1997**, *63*, 1051. (e) Maseras, F.; Morokuma, K. *J. Comput. Chem.* **1995**, *16*, 1170.

(7) (a) Verlet, L. *Phys. Rev.* **1967**, 159, 98. (b) Hockney, R. W. *Methods Comput. Phys.* **1990**, *70*, 921

(8) Blöchl, P. E. *Phy. Rev. B* **1994**, *50*, 17953−17979.

(9) (a) Nosé, S. *Mol. Phys.* **1984**, *52*, 255. (b) Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695.

(10) Blöchl, P. E.; Parrinello, M. *Phys. Rev. B* **1992**, *45*, 9413.

(11) De-Raedt, B.; Sprik, M.; Klein, M. L. *J. Chem. Phys.* **1984**, *80*, 5719.

(12) Fonseca Guerra, C.; Snijders, J G.; te Velde, G.; Baerends E. J. *Theor. Chem. Acc.* **1998**, *99*, 391.

(13) Woo, T. K. "ADF and PAW QM/MM User's Manual," Department of Chemistry, University of Calgary, Calgary, Alberta, Canada, 1998.

(14) Yang, W.; Lee, T.-S. *J. Chem. Phys* **1995**, *1103*, 5674

(15) White, C. A.; Johnson, B. G.; Gill, P. M. W.; Head-Gordan, M. *Chem. Phys. Lett.* **1996**, *253*, 268.

(16) Strain, M. C.; Scuseria, G. E.; Frisch, M. J. *Science* **1996**, *271*, 51.

(17) van-Gunsteren, W. F.; Berendsen, H. J. C. *Angew. Chem., Int. Ed. Engl.* **1990**, *29*, 992−1023.

(18) (a) Streett, W. B.; Tildesley, D. J. *Mol. Phys.* **1978**, *35*, 639−648. (b) Smith, D. E.; Dang, L. X. *J. Chem. Phys.* **1994**, *100*, 3757. (c) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem* **1990**, *11*, 700−733. (d) Lyne, P. D.; Mulholland, A. J.; Richards, W. G. *J. Am. Chem. Soc.* **1995**, *117*, 11345. (e) Bash, P. A.; Field, M. J.; Davenport, R.; Ringe, D.; Petsko, G.; Karplus, M. *Biochemistry* **1991**, *30*, 5826. (f) Hartsough, D. S.; Merz, K. M., Jr. *J. Chem. Phys* **1995**, *99*, 11266.

(19) Gibson, D. A.; Carter, E. A. *J. Phys. Chem.* **1993**, *97*, 13429.

(20) Martyna, G. J.; Tuckerman, M. E.; Tobias, D. J.; Klein, M. L. *Mol. Phys.* **1996**, *87*, 1117−1157.

(21) Tuckerman, M. E.; Parrinello, M. *J. Chem. Phys.* **1994**, *101*, 1316.

(22) Perdew, J. P.; Zunger, A. *Phys. Rev. B* **1981**, *23*, 5048.

(23) Becke, A. *Phys. Rev. A* **1988**, *38*, 3098.

(24) Perdew, J. P. *Phys. Rev. B* **1986**, *33*, 8822−8824.

(25) Perdew, J. P. *Phys. Rev. B* **1986**, *34*, 7406.

(26) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Oxford University Press: Oxford, U.K., 1987.

(27) Woo, T. K.; Blöchl, P. E.; Ziegler, T. *J. Phys. Chem A* **2000**, *104*, 121−129.