

Optimized Hydrogen Mass Repartitioning Scheme Combined with Accurate Temperature/Pressure Evaluations for Thermodynamic and Kinetic Properties of Biological Systems

Jaewoon Jung, Kento Kasahara, Chigusa Kobayashi, Hiraku Oshima, Takaharu Mori, and Yuji Sugita*



Cite This: *J. Chem. Theory Comput.* 2021, 17, 5312–5321



Read Online

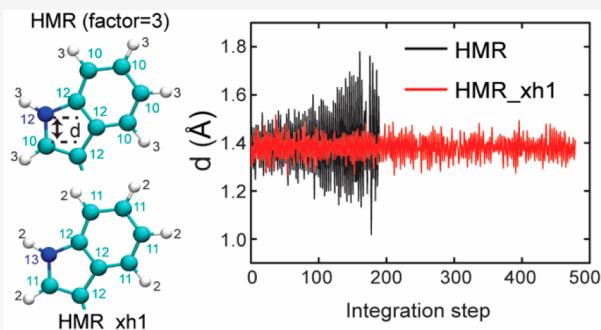
ACCESS |

Metrics & More

Article Recommendations

Supporting Information

ABSTRACT: In recent years, molecular dynamics (MD) simulations with larger time steps have been performed with the hydrogen-mass-repartitioning (HMR) scheme, where the mass of each hydrogen atom is increased to reduce high-frequency motion while the mass of a non-hydrogen atom bonded to a hydrogen atom is decreased to keep the total molecular mass unchanged. Here, we optimize the scaling factors in HMR and combine them with previously developed accurate temperature/pressure evaluations. The heterogeneous HMR scaling factors are useful to avoid the structural instability of amino acid residues having a five- or six-membered ring in MD simulations with larger time steps. It also reproduces kinetic properties, namely translational and rotational diffusions, if the HMR scaling factors are applied to only solute molecules. The integration scheme is tested for biological systems that include soluble/membrane proteins and lipid bilayers for about 200 μ s MD simulations in total and give consistent results in MD simulations with both a small time step of 2.0 fs and a large, multiple time step integration with time steps of 3.5 fs (for fast motions) and 7.0 fs (for slower motions). We also confirm that the multiple time step integration scheme used in this study provides more accurate energy conservations than the RESPA/C1 and is comparable to the RESPA/C2 in NAMD. In summary, the current integration scheme combining the optimized HMR with accurate temperature/pressure evaluations can provide stable and reliable MD trajectories with a larger time step, which are computationally more than 2-fold efficient compared to the conventional methods.



1. INTRODUCTION

Molecular dynamics (MD) simulation has been used extensively in life science and *in silico* drug discovery. It describes the conformational dynamics of proteins, nucleic acids, biological membranes, and other biomolecules at atomic resolution to understand the structure–dynamics–function relationship. Despite the usefulness, short simulation time scales have been pointed out as a major drawback. MD-specialized computers, enhanced conformational sampling algorithms, and other computational methods have been developed to overcome the problem.^{1–7} In all the approaches, a large integration time step in MD simulations is preferred, if stable and accurate computations are realized. In most of the biological MD simulations, SHAKE or RATTLE constraints,^{8,9} which fix the bond lengths involving hydrogen atoms, increase the time step (δt) to 2 or 2.5 fs.

Feenstra et al. suggested two schemes to extend the time step further, namely, the dummy atom scheme and hydrogen mass repartitioning (HMR).¹⁰ In the former, the positions of hydrogen atoms are not explicitly integrated via the equations of motion, while their virtual sites are reconstructed every time step. In the latter, the mass of each hydrogen atom is increased to prevent high-frequency motion, while the mass of a non-

hydrogen atom bonded to a hydrogen atom is decreased to keep the total molecular mass unchanged. These schemes have been examined in MD simulations of proteins or biological membranes with $\delta t = 4$ or 5 fs.^{11,12} Lin and Tuckerman developed an enhanced sampling scheme with a different mass scaling: mass reduction of side chain and solvent atoms.¹³

Another approach to increase the performance with an extended time step is multiple time step (MTS) integration. One of the well-known MTS integrations schemes is to separate time scales with reversible reference system propagator algorithms (RESPA).¹⁴ The MTS integration algorithm with the Ewald electrostatic scheme has been adopted in biological systems with several different choices of separations in large and small time step forces. Batcho et al. assigned the reciprocal-space electrostatic force as the force of the largest time step.¹⁵ Zhou et al. suggested separating the

Received: February 19, 2021

Published: July 17, 2021



small and large time step forces by short-range and long-range forces by rearranging the short-range interaction term in reciprocal-space interactions with additional switch functions.¹⁶ Recently, a better switch function that smoothly distinguishes short- and long-range forces was suggested.¹⁷ Leimkuhler et al. independently suggested a way to separate reciprocal-space interactions into short-range and long-range terms.¹⁸ In MTS integration, one of the main concerns in increasing the time step is to solve the error from resonance, and there have been several suggestions to overcome it.^{18,19}

In MD simulations, accurate T/P evaluations are essential to avoid artificial problems, such as hot-solvent/cold-solute phenomena.²⁰ The optimal temperature, T_{opt} , and half-time step pressure, P_{half} , were previously derived from the generalized equipartition theorem.^{21,22} They are accurate up to the third order of δt , while the conventional definitions have only first-order accuracy. Group pressure or molecular pressure, treating a non-hydrogen atom and its bonded hydrogen atoms as one particle in pressure evaluations, can also increase the accuracy and efficiency by neglecting the high-frequency vibrational motions.^{23,24} Recently, we have proposed an improved time propagator for isothermal–isobaric (*NPT*) simulations based on accurate T/P evaluations with a group-based approach (group T/P).²⁵ The time propagator worked well in MD simulations of hydrated dipalmitoylphosphatidylcholine (DPPC) lipid bilayers with a time step of 5 fs, while we encountered numerical problems in simulations of proteins with the same conditions. Here, the optimized heterogeneous scaling factors in HMR are found to be essential for keeping stable integrations and reproducing thermodynamic and kinetic properties in MD simulations of various biomolecules with a large time step. Note that our integrator used in the study combines the optimized HMR proposed here with the accurate T/P evaluations in the previous work. In contrast, the virtual hydrogen site interaction scheme, one of the dummy atom schemes,^{10,26} where hydrogen atoms are not explicitly integrated in MD simulation, is shown to change the structural properties by fixing local angle terms. We expect our integration scheme could lead to more insight for understanding of various biological phenomena including protein folding or binding.^{27,28}

2. METHODS

2.1. Multiple Time Step (MTS) Integration. In our MTS integration, atomic forces are split into those related to slow and fast motions. Slow motion forces are evaluated less frequently with a larger time step than fast motion forces. Based on the velocity Verlet integration, one MD integration cycle is written as

$$e^{iL_{\text{slow}}\Delta t/2} (e^{iL_{\text{fast}}\delta t})^n e^{iL_{\text{slow}}\Delta t/2} \quad (1)$$

where L_{slow} and L_{fast} are the Liouville operators of slow and fast motion forces, $f_{i,\text{fast}}$ and $f_{i,\text{slow}}$, respectively:

$$\begin{aligned} iL_{\text{slow}} &= \sum_i f_{i,\text{slow}} \cdot \frac{\partial}{\partial p_i} \\ iL_{\text{fast}} &= \sum_i \left(f_{i,\text{fast}} \cdot \frac{\partial}{\partial p_i} + \frac{p_i}{m_i} \cdot \frac{\partial}{\partial r_i} \right) \end{aligned} \quad (2)$$

In the conventional potential functions of biological systems, electrostatic interaction energy is usually described as

$$E_{\text{elec}} = \frac{1}{2} \sum_{\mathbf{n}} \sum_{ij \notin \{\text{bond}\}} (1 - \delta_{ij}^0) \frac{q_i q_j}{|\mathbf{r}_{ij} + \mathbf{n}L|} \quad (3)$$

and this is again decomposed into three terms using the smooth particle mesh Ewald (PME) approximation:²⁹

$$\begin{aligned} E_{\text{elec,real}} &= \frac{1}{2} \sum_{\mathbf{n}} \sum_{ij \notin \{\text{bond}\}} (1 - \delta_{ij}^0) \frac{q_i q_j \operatorname{erfc}(\alpha r_{ij})}{r_{ij}} \\ &\quad - \sum_{ij \in \{\text{bond}\}} \frac{q_i q_j \operatorname{erf}(\alpha r_{ij})}{r_{ij}} \end{aligned} \quad (4a)$$

$$E_{\text{elec,reciprocal}} = \frac{1}{2} \sum_{ij} \sum_{\mathbf{k} \neq 0} \frac{1}{\pi L^3} \frac{4\pi^2}{k^2} q_i q_j e^{-k^2/4\alpha^2} \cos(\mathbf{k} \cdot \mathbf{r}_{ij}) \quad (4b)$$

$$E_{\text{self}} = -\frac{\alpha}{\sqrt{\pi}} \sum_i q_i^2 \quad (4c)$$

In our case, $f_{i,\text{fast}}$ is derived from the negative gradient of energies excluding $E_{\text{elec,reciprocal}}$. $f_{i,\text{slow}}$ is the negative gradient of $E_{\text{elec,reciprocal}}$. The term in eq 4c is constant, and we do not need to consider it in force evaluations. The second term in eq 4a is evaluated to compensate the reciprocal-space energy between bonded pairs. There are also suggestions that separate fast- and slow-motion forces by rearranging eqs 4a and 4b into short- and long-range forces with additional switching functions.^{16,17}

In these schemes, short- and long-range energies are defined as

$$E_{\text{short}} = E_{\text{elec,real}} + \frac{1}{2} \sum_{\mathbf{n}} \sum_{ij, r_{ij} < r_c} (1 - \delta_{ij}^0) \frac{q_i q_j \operatorname{erf}(\alpha r_{ij})}{r_{ij}} \quad (5a)$$

$$E_{\text{long}} = E_{\text{elec,reciprocal}} - \frac{1}{2} \sum_{\mathbf{n}} \sum_{ij, r_{ij} > r_c} (1 - \delta_{ij}^0) \frac{q_i q_j \operatorname{erf}(\alpha r_{ij})}{r_{ij}} \quad (5b)$$

$f_{i,\text{fast}}$ and $f_{i,\text{slow}}$ are obtained from the negative gradients of E_{short} and E_{long} with additional switch functions. According to our tests, as discussed in section 3, evaluating $f_{i,\text{fast}}$ and $f_{i,\text{slow}}$ based on real- and reciprocal-space interactions does not make any significant accuracy error compared to the separation of short- and long-range interactions although the reciprocal-space interactions include short-range interactions.

In isothermal–isobaric conditions, by introducing the thermostat and barostat p_η and p_e , the equation of motions is expressed as³⁰

$$\begin{aligned} \dot{\mathbf{r}}_i &= \frac{\mathbf{p}_i}{m_i} + \frac{p_e}{W}\mathbf{r}_i \\ \dot{\mathbf{p}}_i &= \mathbf{f}_{i,\text{fast}} + \mathbf{f}_{i,\text{slow}} - \frac{p_e}{W}\left(1 + \frac{3}{N_f}\right)\mathbf{p}_i - \frac{p_\eta}{Q}\mathbf{p}_i \\ \dot{V} &= 3V\frac{p_e}{W} \\ \dot{p}_e &= V(P(t) - P_{\text{ext}}) + \frac{2K}{N_f} - \frac{p_\eta}{Q}p_e = F_e + \nu_\eta p_e \\ \dot{\eta} &= \frac{p_\eta}{Q} \\ \dot{p}_\eta &= (N_f + 1)(T - T_{\text{target}}) \end{aligned} \quad (6)$$

where W , V , P_{ext} and K are barostat mass, volume, target pressure, and instantaneous kinetic energy of the system, respectively. In the isothermal–isobaric or isothermal conditions with MTS integration, the update from slow-motion force, thermostat, and barostat can be applied less frequently with a larger time steps using the MTS integration.^{22,23} In this case, one MD integration cycle in the isothermal–isobaric condition is expressed as

$$\begin{aligned} &e^{iL_B\Delta t_{\text{baro}}/2}[e^{iL_T\Delta t_{\text{therm}}/2}\{e^{iL_{\text{slow}}\Delta t/2}(e^{iL_P\Delta t/2}e^{iL_R\delta t}e^{iL_P\Delta t/2})^n e^{iL_{\text{slow}}\Delta t/2}\}^a \\ &\times e^{iL_T\Delta t_{\text{therm}}/2}]^b e^{iL_B\Delta t_{\text{baro}}/2} \end{aligned} \quad (7)$$

where the Liouville operators iL_B , iL_T , iL_{slow} , iL_P , and iL_R are defined as

$$\begin{aligned} iL_B &= F_e \frac{\partial}{\partial p_e} \\ iL_T &= -\frac{p_\eta}{Q} \sum_i \left(\mathbf{p}_i \cdot \frac{\partial}{\partial \mathbf{p}_i} \right) - \frac{p_\eta}{Q} p_e \frac{\partial}{\partial p_e} + (N_f + 1)(T - T_{\text{target}}) \frac{\partial}{\partial p_\eta} \\ iL_{\text{slow}} &= \sum_i \mathbf{f}_{i,\text{slow}} \cdot \frac{\partial}{\partial \mathbf{p}_i} \\ iL_R &= \sum_i \left(\frac{\mathbf{p}_i}{m_i} + \frac{p_e}{W}\mathbf{r}_i \right) \cdot \frac{\partial}{\partial \mathbf{r}_i} + 3V\frac{p_e}{W}\frac{\partial}{\partial V} \\ iL_P &= \sum_i \left(\mathbf{f}_{i,\text{fast}} - \frac{p_e}{W}\left(1 + \frac{3}{N_f}\right)\mathbf{p}_i \right) \cdot \frac{\partial}{\partial \mathbf{p}_i} \end{aligned} \quad (8)$$

Here, time steps for the slow-motion force, Δt , the thermostat evaluation, Δt_{therm} , and the barostat evaluation, Δt_{baro} , are related as

$$\begin{aligned} \Delta t &= n \delta t \\ \Delta t_{\text{therm}} &= a \Delta t \\ \Delta t_{\text{baro}} &= b \Delta t_{\text{therm}} \end{aligned} \quad (9)$$

The integers n , a , and b are greater than or equal to 1. In the study, $a = b$ is assumed and a time step for both thermostat and barostat is defined as $\Delta t_{\text{tb}} = \Delta t_{\text{therm}} = \Delta t_{\text{baro}}$.

The real application of iL_T in MD simulations depends on the thermostat scheme. In the case of the Berendsen thermostat,³¹ particle momentum \mathbf{p}_i is rescaled by

$$\lambda = \left[1 + \frac{\Delta t}{\tau_T} \left(\frac{T_{\text{target}}}{T} - 1 \right) \right]^{1/2} \quad (10)$$

In the Nose–Hoover chain thermostat,³² we prepare additional variables for chains, $p_{\eta 1}$, $p_{\eta 2}$, ..., $p_{\eta M}$, and we rescale particle velocities by $\exp\left(-\frac{p_\eta}{Q}\right)$ after updating p_η from

$$\dot{p}_\eta = N_f(T - T_{\text{target}}) - \frac{p_\eta}{Q_1}p_\eta \quad (11)$$

In the stochastic velocity rescaling thermostat suggested by Bussi et al.,³³ \mathbf{p}_i are scaled to change the temperature change as

$$\Delta T = (T_{\text{target}} - T) \frac{\Delta t}{\tau_T} + 2 \sqrt{\frac{T T_{\text{target}}}{N_f}} \frac{\Delta W}{\sqrt{\tau_T}} \quad (12)$$

where ΔW is the Wiener process. In the case of isothermal–isobaric conditions, p_e and particle momenta are included in the temperature evaluation and N_f is replaced by $N_f + 1$. All of our tests were done by stochastic velocity rescaling thermostat.

2.2. Hydrogen Mass Repartitioning (HMR) with Heterogeneous Scaling Factors. A scaling factor of 3 to each hydrogen atom in HMR, which was suggested by Hopkins et al.,¹¹ was shown to be stable up to a 4 fs time step. However, it causes the SHAKE error for a 5 fs time step that results from the structural instability of amino acid residues having a five- or six-membered ring (Phe, Tyr, Trp, Pro, and His), as shown in section 3.3. A simple solution to avoid the error is to apply a smaller scaling factor to each hydrogen atom in the ring. We suggest a scaling factor of 2 to each hydrogen atom in all the XH_1 groups (X, heavy atom; H, hydrogen atom) (Figure 1). We also found that a scaling factor of 2.5 is

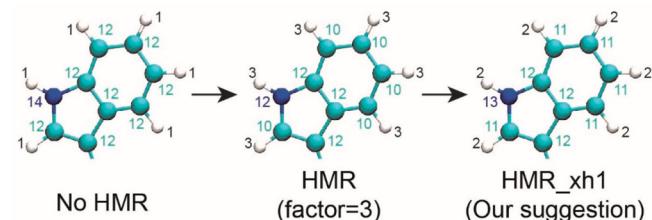


Figure 1. Schematic description of original and modified HMR schemes for the indole ring in tryptophan. The number in each atom indicates the assigned atomic mass.

more suitable to each hydrogen atom in other groups (XH_2 , XH_3 , and XH_4 groups) in simulations using the AMBER force field,³⁴ while a scaling factor of 3 to those atoms reproduces stable trajectories using the CHARMM force field.^{35,36} Hereafter, we call HMR with the heterogeneous scaling factors HMR_xh1.

We also suggest that heterogeneous HMR scaling is not applied to solvent water molecules. Dynamics of a biological system is generally dependent on the solvent viscosity and size of the system according to the Stokes–Einstein equation. From Feenstra's observation, HMR increases the viscosity of water molecules, resulting in a reduction of translational/rotational diffusion coefficients.¹⁰ The error can be reduced by just not applying HMR to water molecules (HMR_solute). If HMR_xh1 is applied only to solute molecules, we denote it as HMR_xh1_solute.

Table 1. Tested Biomolecular Systems and MD Simulation Lengths and Cases

system	no. of atoms	simulation length and cases	total time
80DPPC	18 503	300 ns × 10 cases	3.0 μ s
160DPPC	37 363	300 ns × 35 cases	10.5 μ s
AMPA	55 536	2 μ s × 1 case	2 μ s
SR Ca ²⁺ -ATPase	297 475	1 μ s × 4 cases	4 μ s
SecDF	244 865	1 μ s × 2 cases	2 μ s
chignolin	8 540	1.5 μ s × 12 replicas × 3 sets × 3 cases	162 μ s
PP1	16 321	1 μ s × 3 cases	3 μ s
CI2_1	32 413	10 ns × 40 cases	400 ns
CI2_2	39 565	1 μ s × 3 cases	3 μ s
ApoA1	92 224	40 ps × 36 cases	1.44 ns
STMV	1 066 628	40 ps × 84 cases	3.36 ns

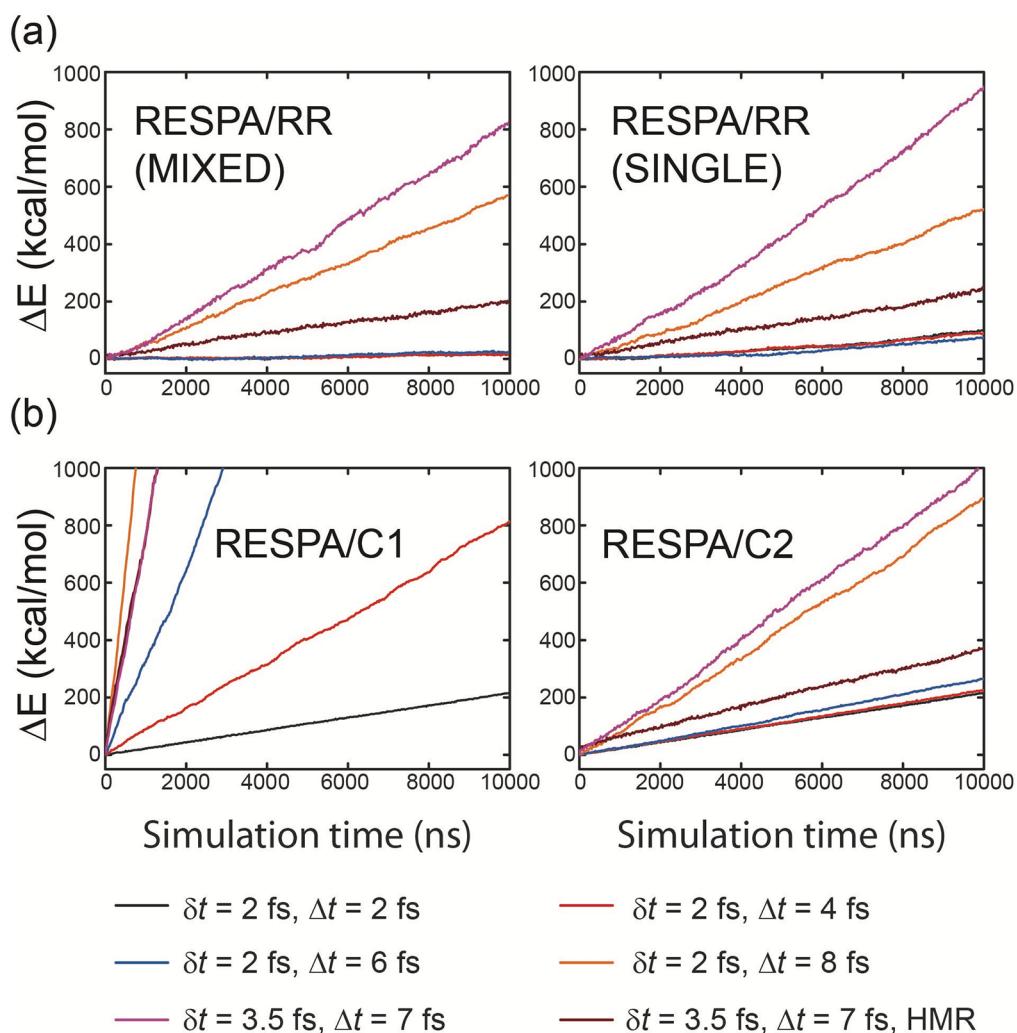


Figure 2. Energy profiles in 10 ns microcanonical ensemble MD simulations of CI2 protein solvated in water molecules using (a) RESPA/RR in GENESIS and (b) RESPA/C1 and RESPA/C2 in NAMD.

3. RESULTS AND DISCUSSION

Tested biomolecular systems are summarized in Table 1. Additional details of the systems and simulation conditions are described in the Supporting Information.

3.1. Validation of Conventional MD Integration. 3.1.1. *Energy Conservation with MTS Integration.* To show the validation of our MTS integration, we first investigate the energy drift of the chymotripsin inhibitor 2 (CI2) protein solvated in water molecules (CI2_1 system in Table 1) using

the microcanonical ensemble with various time steps. The energy drift values are investigated using two programs, GENESIS (version 2.0beta)³⁷ and NAMD (version 2.13),³⁸ because they have different definitions of MTS integration schemes. For all cases, we assign the constraints to all the bonds involving hydrogen atoms. In GENESIS, we test two floating-point precision levels: the single precision where force evaluation and integration are performed with single precision floating points (but constraints and energy evaluation are using

double precision ones) and the mixed precision where force evaluation and integration are carried out using single and double precisions, respectively. Both GENESIS and NAMD use SHAKE/RATTLE/SETTLE^{8,9,39} as constraint algorithms. From Figure 2a, we noticed that the energy drift is highly dependent on the precision of real numbers.

In NAMD, the slow- and fast-motion forces are derived from eqs 5a and 5b with two switch functions with third- and fifth-order polynomials. These two options are denoted as “C1” and “C2” splitting in NAMD’s input, and we here name them RESPA/C1 and RESPA/C2. RESPA/C2 is same as the recent RESPA scheme suggested by Morrone et al.¹⁷ Our MTS scheme implemented in GENESIS is just splitting the force into real- and reciprocal-space interactions based on eqs 4a and 4b, and we name it RESPA/RR (two “R” characters means “real” and “reciprocal”). The energy drift results using MTS integrations are shown in Figure 2b. In the case of $\delta t = 2$ fs, the energy drifts of RESPA/RR and RESPA/C2 are similar to that of velocity Verlet (in Figure 2, it is written as $\delta t = 2$ fs and $\Delta t = 2$ fs) up to $\Delta t = 6$ fs. In other words, RESPA does not change the accuracy up to $\Delta t = 6$ fs. By increasing Δt up to 8 fs, we observe clearly higher energy drift values than those from velocity Verlet integration. RESPA/C1 has much larger energy drift values than RESPA/C2 and RESPA/RR for all time steps. From these observations, the accuracy of RESPA/RR is on the same level as RESPA/C2. The simple separation of real- and reciprocal-space interactions in the MTS integration does not make serious errors of the energy drift up to $\Delta t = 6$ fs. HMR is able to reduce the errors significantly, although it has twice larger energy drift values than the conventional velocity integration case with $\delta t = 2$ fs. There is no consensus about the acceptable energy drift values in MD, but at least we can conclude that both RESPA/RR and RESPA/C2 have almost the same energy drift values as velocity Verlet and HMR reduces the errors for a large time step. For a more quantitative analysis of the energy drift values in each scheme and time step, we investigated the statistics of the 1 ns energy change in Table S1. Both RESPA/RR and RESPA/C2 have average energy drift values within one standard deviation from the velocity Verlet integration up to $\Delta t = 6$ fs. It is also shown that the precision has a more significant effect on energy drift than the multiple time step integration.

3.1.2. Potential Energy Profiles in the Canonical Ensemble. Even though we show the energy drift values in microcanonical ensemble, most MD applications in the life and material sciences are performed in canonical or isothermal–isobaric conditions. Therefore, we also investigate average potential energies, performing 10 ns canonical MD simulations of a CI2 protein solvated in water molecules, using a stochastic velocity rescaling thermostat.³³ According to the results based on Figure 3, we found the following: (1) Single and mixed precisions generate the same potential energy distributions when RESPA/RR is applied. (2) Given $\delta t = 2$ fs, MTS does not change the potential energy distributions up to $\Delta t = 6$ fs. (3) At $\delta t = 3.5$ fs and $\Delta t = 7$ fs the average potential energy is within one standard deviation from the average of $\delta t = 2$ fs and $\Delta t = 2$ fs. We also found that RESPA/C1 has potential energy distributions almost identical with those of velocity Verlet and RESPA/C2 (data not shown). RESPA/RR and RESPA/C2 generate average potential values similar to that obtained with velocity Verlet integration using a time step of 1 fs ($\delta t = 1$ fs and $\Delta t = 1$ fs). As we increase δt , RESPA/C2 has a higher energy change than RESPA/RR. Since temperatures are

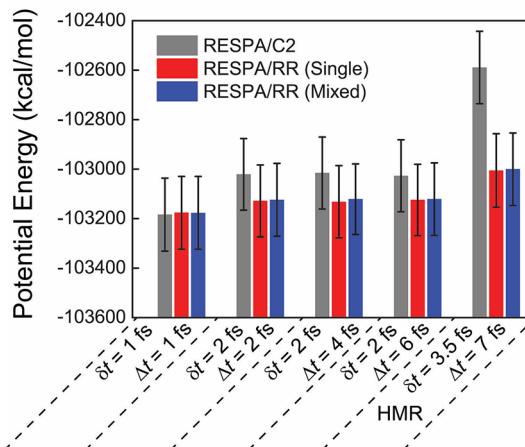


Figure 3. Average potential energy of the CI2 system with isothermal condition.

evaluated from different ways, the temperature evaluation scheme might affect the average potential energy with $\delta t = 3.5$ fs and $\Delta t = 7$ fs. Considering that the average potential energy value in RESPA/RR in the condition $\delta t = 3.5$ fs and $\Delta t = 7$ fs is close to the value obtained by using RESPA/C2 with $\delta t = 2$ fs, and within one standard deviation of the potential energy distributions of $\delta t = \Delta t = 2$ fs, we consider it to be the maximum time step in our MTS integration. In section 3.2, we test this condition in MD simulations of various biological systems.

3.2. Key Conditions in MD Simulations. There are three key conditions in this study. The first one is used as a standard MD simulation condition, while the second and third ones are the proposed time propagators.

(i) The standard condition is the velocity Verlet integrator with $\delta t = \Delta t = 2$ fs without HMR (VV_2fs).

(ii) The second condition is velocity Verlet with a large time step: velocity Verlet integrator with $\delta t = \Delta t = 5$ fs with HMR (VV_5fs_HMR).

(iii) The third condition is MTS integration with a large time step: MTS_3.5fs_HMR. That is, an MTS integrator with $\delta t = 3.5$ fs (fast), $\Delta t = 7.0$ fs (slow), and $\Delta t_{tb} = 21$ fs with HMR (MTS_3.5 fs).

In the cases of ii and iii, we append “xh1” or “solute” when the heterogeneous scaling factor or HMR scaling only for solute molecules is applied, as stated in section 2.2.

3.3. Effect of Heterogeneous HMR Scaling Factor. In MD simulations of the ligand binding domain of AMPA glutamate receptor, VV_5fs_HMR causes a SHAKE error while VV_5fs_HMR_xh1 does not, even starting from the same structure (Figure 4). All the Trp side chains in VV_5fs_HMR_xh1 show local structures similar to those in VV_2fs (Figure S1 in the Supporting Information). This result suggests that our HMR scaling factor reproduces stable MD simulation trajectories even when a large time step is applied.

To examine the effect of HMR in kinetics, we observe the mean square displacement (MSD) and rotational correlation functions ($C(t)$) of a small inhibitor of c-Src kinase, PP1, and a protein, CI2, in water. Here, the correlation function $C(t)$ is evaluated as

$$C(t) = \frac{1}{2} \langle 3(\mathbf{u}(0) \cdot \mathbf{u}(t))^2 - 1 \rangle \quad (13)$$

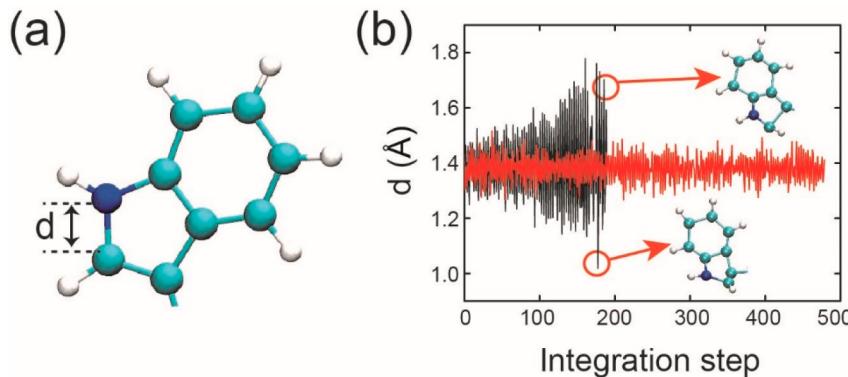


Figure 4. MD trajectories of a bond distance between nitrogen and carbon atoms in the pyrrole ring (“ d ” shown in (a)) of the ligand binding domain of AMPA receptor. The black line starts from 210 steps before the constraint error using HMR. In contrast, HMR_xh1 gives a stable trajectory of the distance, starting from the same structure. (inset) Two structures of distorted indole rings.

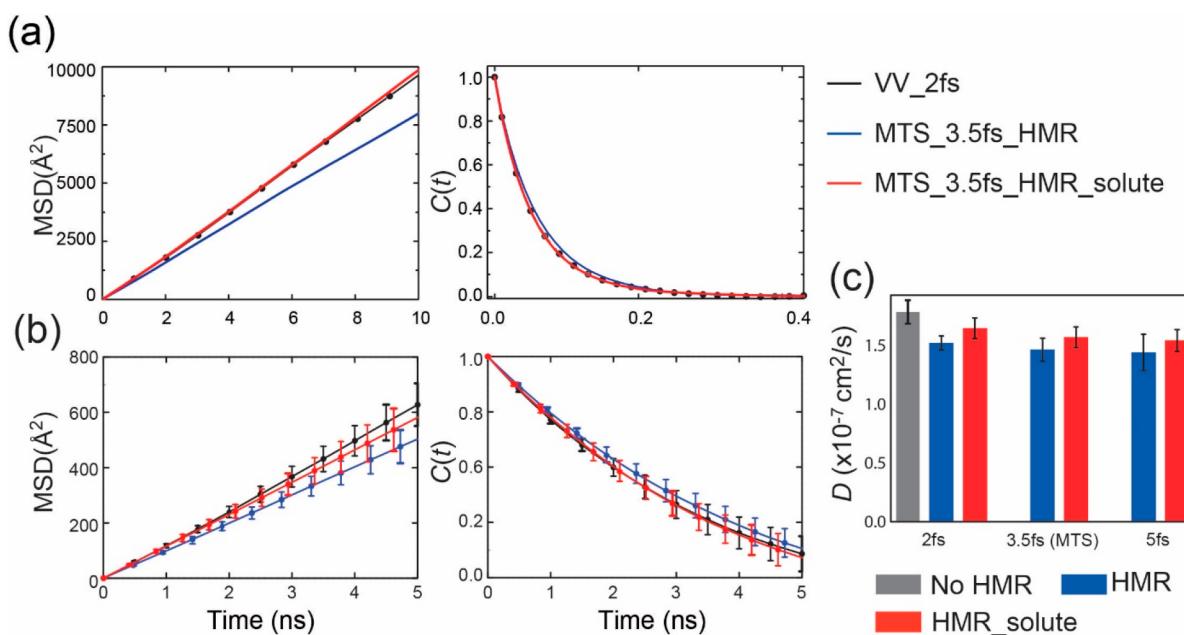


Figure 5. Translational mean squared displacements (left) and rotational correlational function (right) of (a) PP1 and (b) CI2. (c) Lateral diffusion coefficients of DPPC in 160DPPC systems in seven simulations.

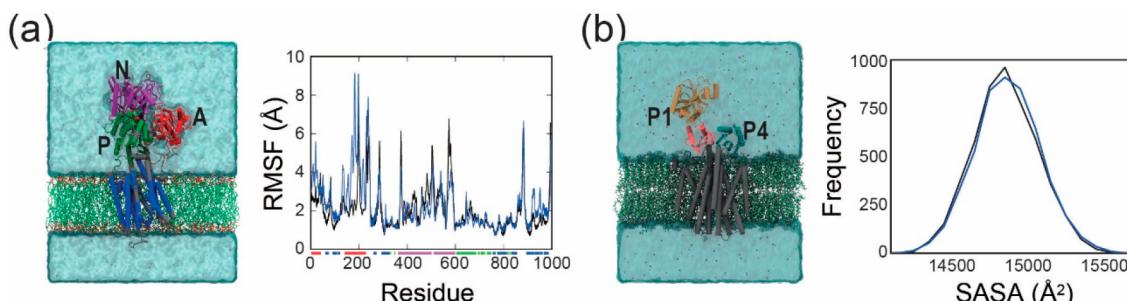


Figure 6. (a) Simulation snapshot of SR Ca^{2+} -ATPase and Ca^α RMSF (black, VV_2fs; blue, MTS_3.5fs_HMR_solute). The underlines in the horizontal axes indicate each domain. (b) Simulation snapshot of SecDF and distributions of total solvent accessible surface area (SASA) of P1 and P4 domains (black, VV_2fs; red, MTS_3.5fs_HMR).

where $\mathbf{u}(t)$ is the unit vector associated with a rigid molecule. For PP1 in water, translational and rotational diffusions obtained with VV_2fs and MTS_3.5fs_HMR_solute agree with each other, while those with MTS_3.5fs_HMR do not (Figure 5a). In Figure 5b, MSD and $C(t)$ of CI2 are slightly

changed within statistical errors between VV_2fs and MTS_3.5fs_HMR_solute. MTS_3.5fs_HMR cannot conserve translational and rotational diffusions of CI2. In Figure 5c, the lateral diffusion coefficients of DPPC in 160DPPC are compared among seven different simulations. The lateral

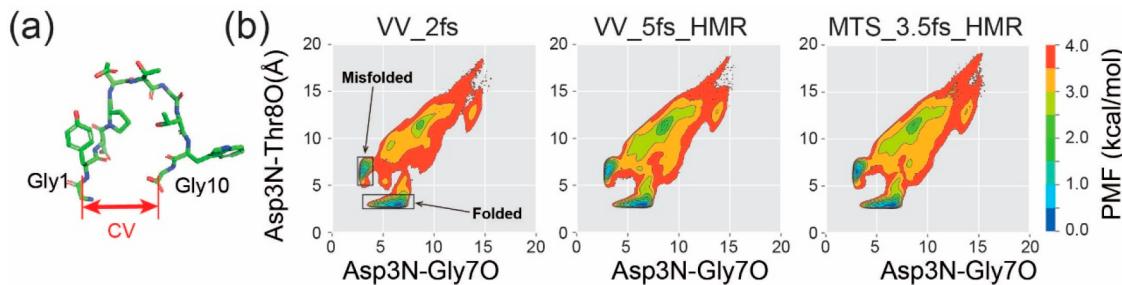


Figure 7. (a) Atomic structure of chignolin and the collective variable (end-to-end distance between C α atoms of Gly1 and Gly10) in REUS simulations. (b) Two-dimensional free-energy landscapes of chignolin folding along two hydrogen-bond distances, Asp3N–Gly7O and Asp3N–Thr8O. Three simulation conditions (VV_2fs, VV_5fs_HMR, and MTS_3.5fs_HMR) are tested. Two squares represent the folded and misfolded states.

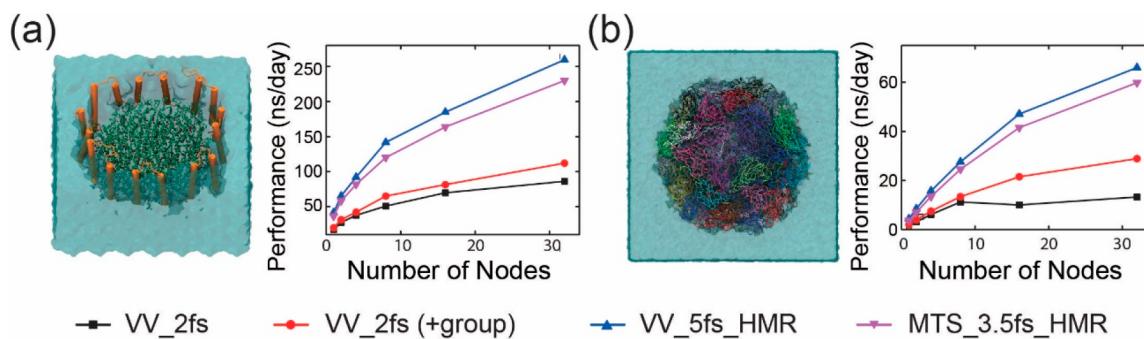


Figure 8. (a) Structure of ApoA1 (92 224 atoms) and benchmark performance in NPT MD simulations. (b) Structure of STMV (1 066 628 atoms) and benchmark performance in NPT conditions. In the benchmarks, each node consists of two Intel Xeon Gold 6142 CPUs (32 cores, 2.60 GHz) and two GeForce GTX 1080 Ti GPU cards connected by InfiniBand.

diffusion is affected by the viscosity of the lipid and water molecules. Therefore, small deviations of diffusion coefficients are observed between simulations with HMR_solute and the standard simulation (VV_2fs). Even in this case, HMR_solute affects the lipid lateral diffusions less than the conventional HMR regardless of a time step (Table S2 in the *Supporting Information*).

3.4. MD Simulations of Various Biomolecular Systems Using the Time Propagator. We examine our propagator by performing multiple 1 μ s MD simulations of sarcoplasmic reticulum (SR) Ca²⁺-ATPase and SecDF with an explicit membrane and solvent. Root-mean-square fluctuations (RMSFs) of the C α atoms in SR Ca²⁺-ATPase (Figure 6a) and the accessible surface area (ASA) of P1/P4 domains of SecDF (Figure 6b) show almost identical results between VV_2fs and MTS_3.5fs_HMR (or MTS_3.5fs_HMR_solute). Other properties, namely the C α atom RMS deviation (RMSD) of whole proteins and their rigid structural domains, fluctuations of membranes in SR Ca²⁺-ATPase, and even side chain dihedral angle distributions in SecDF, agree well between the two different simulations (Figures S2–S4 in the *Supporting Information*).

Next, we examined thermodynamic properties obtained in MD simulations based on the proposed time propagator. The tested molecule is chignolin, which is a 10-residue peptide with sequence GYDPETGTWG, in explicit water (Figure 7a). We apply the replica-exchange umbrella sampling (REUS) method²³ using the end-to-end distance as a collective variable (CV). The one-dimensional (1d) (Figure S5 in the *Supporting Information*) and two-dimensional (2d) free-energy profiles computed with MBAR (Figure 7b) show almost no differences in VV_2fs, VV_5fs_HMR_xh1, and MTS_3.5fs_HMR_xh1.

Small deviations found in the 1d profile at CV > 8 Å are attributed not to the time propagators but to insufficient sampling of the unfolded state structures. In the 2d profiles, two basins are observed: one for the folded state having a hydrogen bond between Asp3N and Thr8O, and the other for the misfolded state with a hydrogen bond between Asp3N and Gly7O (Figure S6 in the *Supporting Information*). This result agrees with previous simulations of the same peptide.

3.5. Computational Performances with the New Integration Scheme. Finally, we perform benchmark calculations of ApoA1 and STMV, which consist of about 92 000 and 1 million atoms, respectively. Figure 8 and Figure S7 in the *Supporting Information* show that NVT and NPT MD simulations are accelerated more than 2-fold by using the proposed time propagator rather than the standard condition (VV_2fs). The time propagator has been introduced in the latest version of GENESIS,²⁴ while it is applicable to any other MD software on standard computational platforms. The scheme proposed in this paper contributes greatly to the investigation of protein dynamics for longer time scales and to the simulation of large-scale biological systems in realistic cellular environments.

4. DISCUSSION

4.1. Comparison between the Dummy Atom Scheme and the Proposed Time Propagator. Another way to perform MD with a large time step is to remove high-frequency motions. One representative is to replace the hydrogen atoms by dummy atoms in virtual interaction sites.¹⁰ This scheme has been implemented in Gromacs software and is suggested to be reliable up to a 5 fs time step.^{6,16} Recently, this scheme and its

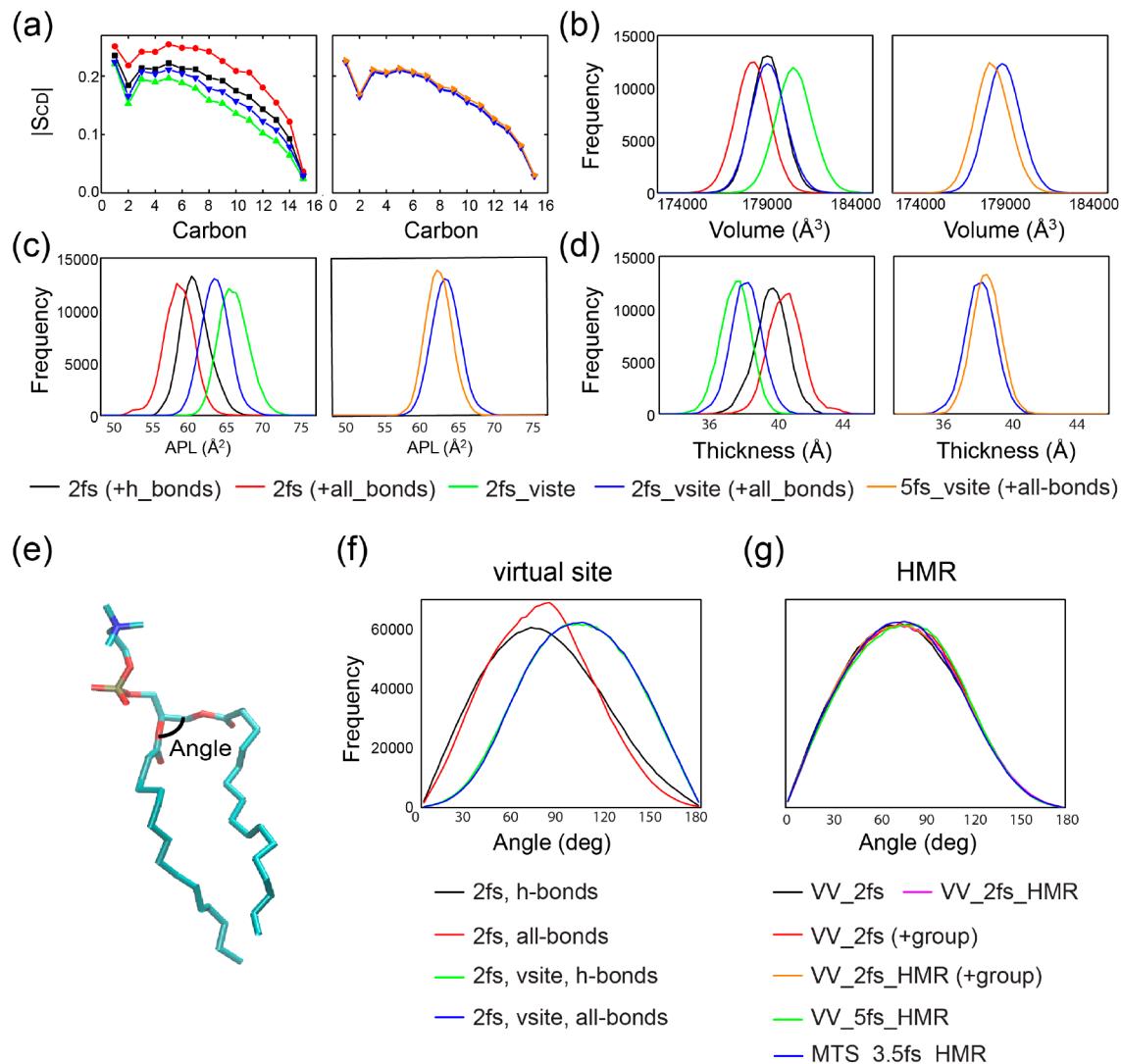


Figure 9. (a) Order parameter of the first chain and distributions of (b) volume distributions, (c) APL, and (d) lipid thickness in *NPT* simulations of 80DPPC using Gromacs MD software. (e) Definition of the target angle in a DPPC. (f, g) Angle distributions in the simulations based on the proposed time propagator using GENESIS software. From (a) to (d), we computed the properties by comparing various working conditions with a time step of 2 fs in the left side. In the right side, we compared the simulations with time steps of 2 and 5 fs for virtual site interaction with all heavy bond constraints.

modification (increasing the hydrogen atom masses at regular intervals in the lipid acyl chains) have been shown to be reliable in lipid bilayer systems by comparing simulation results using time steps of 2 and 5 fs.^{12,26} The scheme is based on the following two working conditions to enable a 5 fs time step: (i) replacing hydrogen atoms by virtual sites to remove the bond/angle vibrations of hydrogen atoms (vsites) and (ii) constraints for bonds involving heavy atoms (all-bonds). Previous research did not observe each effect separately. To understand each more clearly, we observe the structural/physical properties of a DPPC lipid bilayer with the following four conditions: (i) no virtual site and constraints only for bonds involving hydrogen atoms (h-bonds), (ii) no virtual site and constraints for all bonds (all-bonds), (iii) virtual site and constraints only for bonds involving hydrogen atoms (vsites and h-bonds), and (iv) virtual site interaction and constraints for all bonds (vsites and all-bonds). Note that the h-bonds condition is commonly used in many MD programs and parameters in the CHARMM and AMBER force fields have been developed in the condition.

The results of *NPT* MD simulations of the 80DPPC system in Figure 9 and Table S3 in the Supporting Information show that “all-bonds” decreases the area per lipids (APL) of the lipid bilayer and increases lipid thickness, even with the time step of 2 fs. Contrary to “all-bonds”, “vsites” with the same time step increases the APL and decreases the lipid thickness. “Vsites + all-bonds” compensates errors resulting from “all-bonds” and “vsites”, but still a noticeable deviation from the standard condition is observed. With the use of “vsites + all-bonds” and a time step of 5 fs, the amounts of changes in order parameters, volume, APL, and thickness are not observed to be as large. Judging from the results with different conditions, the apparent small errors in the 5 fs integration using “vsites + all-bonds” result from the compensation of two opposite effects. It is difficult to confirm whether the compensation happens in all the other systems.

“Vsites” indeed affects structural properties of DPPC molecules. For instance, we investigate an angle between two lipid tails (Figure 9e–g). “All-bonds” does not change the

angle distributions in the MD simulations from the reference, while “vsite” and “vsite + all-bonds” extend the angle distributions greatly. We observe that the proposed time propagator including HMR does not change the angle distributions (Figures 9e–g) and other properties.²⁵ To leverage the virtual site scheme with a 5 fs time step, force fields should be developed based on the scheme itself. Because the CHARMM or AMBER force fields have been developed using “h-bonds” constraints and not “vsite”, caution is required to make use of the virtual scheme with those force fields.

4.2. Limitation of the Proposed Time Propagator.

Although much improved from the existing integration schemes with large time steps, our new scheme still has some limitations of applicability with a 5 fs time step. We observed that the constraint error happens when we performed temperature replica-exchange MD (REMD) at very high temperatures. For the conditions, the use of a 5 fs time step seems to be dangerous. Also, if the equilibration of complicated biomolecular systems is not sufficient, the use of a 5 fs time step is not recommended. However, the constraint errors in the REMD simulation disappeared if we used MTS_3.5fs_HMR. On the basis of the benchmark tests (Figure 8 and Figure S7 in the Supporting Information), the performance between them is not so different. Judging from the numerical stability and performance benchmark results, the optimal choice for various biological simulations could be MTS_3.5fs_HMR.

5. CONCLUSIONS

We propose a HMR scheme with a heterogeneous scaling factor and combine it with the recently proposed group-based T/P evaluations to use a longer time step in velocity Verlet and multiple time step integrators without losing numerical stability. Extensive MD simulations, ~200 μs in total, of various biomolecules using the time propagator with short and longer time steps show that not only thermodynamic properties but also kinetics of biomolecules are reproduced with the same accuracy as the standard simulation conditions. We expect that our approach can be used for large-scale MD simulations of realistic cellular environments, which will shed more light on the structure–dynamics–function relationships of biomolecules in a living cell.

■ ASSOCIATED CONTENT

SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jctc.1c00185>.

Simulation conditions, simulation systems, evaluation of physical properties; extra analysis on two membrane proteins, chignolin, and DPPC; extra benchmarks of ApoA1 and STMV (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Yuji Sugita — Theoretical Molecular Science Laboratory, RIKEN Cluster for Pioneering Research, Wako, Saitama 351-0198, Japan; Computational Biophysics Research Team, RIKEN Center for Computational Science, Chuo-ku, Kobe, Hyogo 650-0047, Japan; Laboratory for Biomolecular Function Simulation, RIKEN Center for Biosystems Dynamics Research, Chuo-ku, Kobe, Hyogo 650-0047, Japan; pubs.acs.org/doi/10.1021/acs.jctc.1c00185

Japan;  orcid.org/0000-0001-9738-9216; Email: sugita@riken.jp

Authors

Jaewoon Jung — Theoretical Molecular Science Laboratory, RIKEN Cluster for Pioneering Research, Wako, Saitama 351-0198, Japan; Computational Biophysics Research Team, RIKEN Center for Computational Science, Chuo-ku, Kobe, Hyogo 650-0047, Japan;  orcid.org/0000-0002-2285-4432

Kento Kasahara — Laboratory for Biomolecular Function Simulation, RIKEN Center for Biosystems Dynamics Research, Chuo-ku, Kobe, Hyogo 650-0047, Japan

Chigusa Kobayashi — Computational Biophysics Research Team, RIKEN Center for Computational Science, Chuo-ku, Kobe, Hyogo 650-0047, Japan;  orcid.org/0000-0002-5603-4619

Hiraku Oshima — Laboratory for Biomolecular Function Simulation, RIKEN Center for Biosystems Dynamics Research, Chuo-ku, Kobe, Hyogo 650-0047, Japan;  orcid.org/0000-0001-5626-1291

Takaharu Mori — Theoretical Molecular Science Laboratory, RIKEN Cluster for Pioneering Research, Wako, Saitama 351-0198, Japan;  orcid.org/0000-0002-8717-2926

Complete contact information is available at: <https://pubs.acs.org/10.1021/acs.jctc.1c00185>

Notes

The data that support the findings of this study are available from the corresponding author upon reasonable request. The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This research used in-house computer resources and the RIKEN HOKUSAI supercomputer. The computer resources of Oakforest-PACS were also provided through the HPCI System Research project (Project Nos. hp190097, hp200129, hp200135). The research was supported in part by MEXT as the “FLAGSHIP 2020 project”, “Priority Issue on Post-K computer” (Building Innovative Drug Discovery Infrastructure Through Functional Control of Biomolecular Systems), “Program for Promoting Research on the Supercomputer Fugaku” (Biomolecular dynamics in a living cell/MD-driven Precision Medicine), and MEXT/KAKENHI Grant 19H05645 (to Y.S.).

■ REFERENCES

- (1) Brooks, B. R.; Brooks, C. L., 3rd; Mackerell, A. D., Jr.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: the biomolecular simulation program. *J. Comput. Chem.* **2009**, *30* (10), 1545–614.
- (2) Salomon-Ferrer, R.; Case, D. A.; Walker, R. C. An overview of the Amber biomolecular simulation package. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2013**, *3* (2), 198–210.
- (3) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005**, *26* (16), 1781–802.
- (4) Jung, J.; Mori, T.; Kobayashi, C.; Matsunaga, Y.; Yoda, T.; Feig, M.; Sugita, Y. GENESIS: a hybrid-parallel and multi-scale molecular

- dynamics simulator with enhanced sampling algorithms for biomolecular and cellular simulations. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2015**, *5* (4), 310–323.
- (5) Kobayashi, C.; Jung, J.; Matsunaga, Y.; Mori, T.; Ando, T.; Tamura, K.; Kamiya, M.; Sugita, Y. GENESIS 1.1: A hybrid-parallel molecular dynamics simulator with enhanced sampling algorithms on multiple computational platforms. *J. Comput. Chem.* **2017**, *38* (25), 2193–2206.
- (6) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **2008**, *4* (3), 435–447.
- (7) Shaw, D. E.; Grossman, J.; Bank, J. A.; Batson, B.; Butts, J. A.; Chao, J. C.; Deneroff, M. M.; Dror, R. O.; Even, A.; Fenton, C. H.; et al. Anton 2: raising the bar for performance and programmability in a special-purpose molecular dynamics supercomputer. In *SC '14: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*; IEEE Press: 2014; pp 41–53..
- (8) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical-Integration of Cartesian Equations of Motion of a System with Constraints - Molecular-Dynamics of N-Alkanes. *J. Comput. Phys.* **1977**, *23* (3), 327–341.
- (9) Andersen, H. C. Rattle - a Velocity Version of the Shake Algorithm for Molecular-Dynamics Calculations. *J. Comput. Phys.* **1983**, *52* (1), 24–34.
- (10) Feenstra, K. A.; Hess, B.; Berendsen, H. J. C. Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems. *J. Comput. Chem.* **1999**, *20* (8), 786–798.
- (11) Hopkins, C. W.; Le Grand, S.; Walker, R. C.; Roitberg, A. E. Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning. *J. Chem. Theory Comput.* **2015**, *11* (4), 1864–1874.
- (12) Olesen, K.; Awasthi, N.; Bruhn, D. S.; Pezeshkian, W.; Khandelia, H. Faster Simulations with a 5 fs Time Step for Lipids in the CHARMM Force Field. *J. Chem. Theory Comput.* **2018**, *14* (6), 3342–3350.
- (13) Lin, I. C.; Tuckerman, M. E. Enhanced Conformational Sampling of Peptides via Reduced Side-Chain and Solvent Masses. *J. Phys. Chem. B* **2010**, *114* (48), 15935–15940.
- (14) Tuckerman, M.; Berne, B. J.; Martyna, G. J. Reversible Multiple Time Scale Molecular-Dynamics. *J. Chem. Phys.* **1992**, *97* (3), 1990–2001.
- (15) Batcho, P. F.; Case, D. A.; Schlick, T. Optimized particle-mesh Ewald/multiple-time step integration for molecular dynamics simulations. *J. Chem. Phys.* **2001**, *115* (9), 4003–4018.
- (16) Zhou, R. H.; Harder, E.; Xu, H. F.; Berne, B. J. Efficient multiple time step method for use with Ewald and particle mesh Ewald for large biomolecular systems. *J. Chem. Phys.* **2001**, *115* (5), 2348–2358.
- (17) Morrone, J. A.; Zhou, R. H.; Berne, B. J. Molecular Dynamics with Multiple Time Scales: How to Avoid Pitfalls. *J. Chem. Theory Comput.* **2010**, *6* (6), 1798–1804.
- (18) Leimkuhler, B.; Margul, D. T.; Tuckerman, M. E. Stochastic, resonance-free multiple time-step algorithm for molecular dynamics with very large time steps. *Mol. Phys.* **2013**, *111* (22–23), 3579–3594.
- (19) Morrone, J. A.; Markland, T. E.; Ceriotti, M.; Berne, B. J. Efficient multiple time scale molecular dynamics: Using colored noise thermostats to stabilize resonances. *J. Chem. Phys.* **2011**, *134*, 014103.
- (20) Eastwood, M. P.; Stafford, K. A.; Lippert, R. A.; Jensen, M. O.; Maragakis, P.; Predescu, C.; Dror, R. O.; Shaw, D. E. Equipartition and the Calculation of Temperature in Biomolecular Simulations. *J. Chem. Theory Comput.* **2010**, *6* (7), 2045–2058.
- (21) Jung, J.; Kobayashi, C.; Sugita, Y. Optimal Temperature Evaluation in Molecular Dynamics Simulations with a Large Time Step. *J. Chem. Theory Comput.* **2019**, *15* (1), 84–94.
- (22) Jung, J.; Kobayashi, C.; Sugita, Y. Kinetic energy definition in velocity Verlet integration for accurate pressure evaluation. *J. Chem. Phys.* **2018**, *148* (16), 164109.
- (23) Lippert, R. A.; Predescu, C.; Ierardi, D. J.; Mackenzie, K. M.; Eastwood, M. P.; Dror, R. O.; Shaw, D. E. Accurate and efficient integration for molecular dynamics simulations at constant temperature and pressure. *J. Chem. Phys.* **2013**, *139* (16), 164106.
- (24) Di Pierro, M.; Elber, R.; Leimkuhler, B. A Stochastic Algorithm for the Isobaric-Isothermal Ensemble with Ewald Summations for All Long Range Forces. *J. Chem. Theory Comput.* **2015**, *11* (12), 5624–5637.
- (25) Jung, J.; Sugita, Y. Group-based evaluation of temperature and pressure for molecular dynamics simulation with a large time step. *J. Chem. Phys.* **2020**, *153* (23), 234115.
- (26) Bjelkmar, P.; Larsson, P.; Cuendet, M. A.; Hess, B.; Lindahl, E. Implementation of the CHARMM Force Field in GROMACS: Analysis of Protein Stability Effects from Correction Maps, Virtual Interaction Sites, and Water Models. *J. Chem. Theory Comput.* **2010**, *6* (2), 459–466.
- (27) Piana, S.; Lindorff-Larsen, K.; Shaw, D. E. Protein folding kinetics and thermodynamics from atomistic simulation. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (44), 17845–17850.
- (28) Re, S. Y.; Oshima, H.; Kasahara, K.; Kamiya, M.; Sugita, Y. Encounter complexes and hidden poses of kinase-inhibitor binding on the free-energy landscape. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116* (37), 18404–18409.
- (29) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103* (19), 8577–8593.
- (30) Martyna, G. J.; Tuckerman, M. E.; Tobias, D. J.; Klein, M. L. Explicit reversible integrators for extended systems dynamics. *Mol. Phys.* **1996**, *87* (5), 1117–1157.
- (31) Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R. Molecular-Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81* (8), 3684–3690.
- (32) Martyna, G. J.; Klein, M. L.; Tuckerman, M. Nose-Hoover Chains - the Canonical Ensemble Via Continuous Dynamics. *J. Chem. Phys.* **1992**, *97* (4), 2635–2643.
- (33) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126* (1), 014101.
- (34) Wang, J. M.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25* (9), 1157–1174.
- (35) Huang, J.; MacKerell, A. D. CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *J. Comput. Chem.* **2013**, *34* (25), 2135–2145.
- (36) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmuller, H.; MacKerell, A. D. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **2017**, *14* (1), 71–73.
- (37) GENESIS, ver. 2.0beta. <https://github.com/genesis-release-rcs/genesis-2.0>.
- (38) NAMD, ver. 2.13. <http://www.ks.uiuc.edu/Research/namd>.
- (39) Miyamoto, S.; Kollman, P. A. Settle - an Analytical Version of the Shake and Rattle Algorithm for Rigid Water Models. *J. Comput. Chem.* **1992**, *13* (8), 952–962.