ARTICLE

Extensive Numerical Tests of Leapfrog Integrator in Middle Thermostat Scheme in Molecular Simulations[†]

Zhaoxi Sun, Payam Kalhor, Yang Xu, Jian Liu*

Beijing National Laboratory for Molecular Sciences, College of Chemistry and Molecular Engineering, Institute of Theoretical and Computational Chemistry, Peking University, Beijing 100871, China

Accurate and efficient integration of the equations of motion is indispensable for molecular dynamics (MD) simulations. Despite the massive use of the conventional leapfrog (LF) integrator in modern computational tools within the framework of MD propagation, further development for better performance is still possible. The alternative version of LF in the middle thermostat scheme (LF-middle) achieves a higher order of accuracy and efficiency and maintains stable dynamics even with the integration time stepsize extended by several folds. In this work, we perform a benchmark test of the two integrators (LF and LF-middle) in extensive conventional and enhanced sampling simulations, aiming at quantifying the time-stepsize-induced variations of global properties (*e.g.*, detailed potential energy terms) as well as of local observables (*e.g.*, free energy changes or bond-lengths) in practical simulations of complex systems. The test set is composed of six chemically and biologically relevant systems, including the conformational change of dihedral flipping in the *N*-methylacetamide and an AT (Adenine-Thymine) tract, the intra-molecular proton transfer inside malonaldehyde, the binding free energy calculations of benzene and phenol targeting T4 lysozyme L99A, the hydroxyl bond variations in ethaline deep eutectic solvent, and the potential energy of the blue-light using flavin photoreceptor. It is observed that the time-step-induced error is smaller for the LF-middle scheme. The outperformance of LF-middle over the conventional LF integrator is much more significant for global properties than local observables. Overall, the current work demonstrates that the LF-middle scheme should be preferably applied to obtain accurate thermodynamics in the simulation of practical chemical and biological systems.

Key words: Molecular dynamics, Leapfrog integrator, Middle thermostat scheme
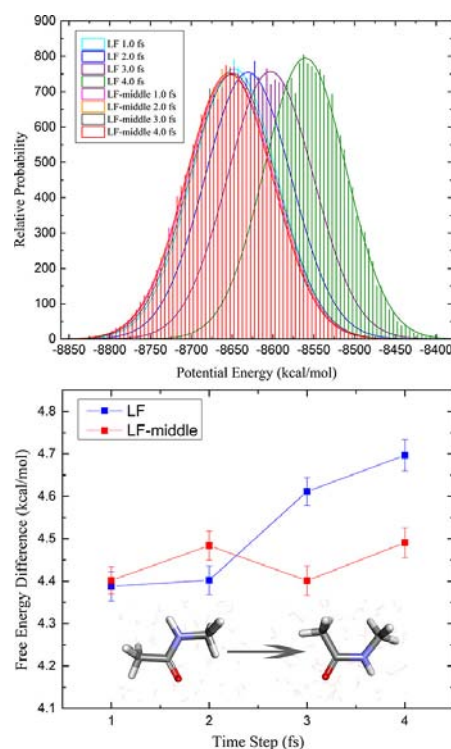
I. INTRODUCTION

Since the pioneering work [1, 2] in 1950's, molecular dynamics simulations have become a feasible route to access the detailed atomic-scale motions of realistic complex systems in chemistry, physics, biology, materials, and environmental science [3–10]. Provided that the ergodicity assumption is valid, the converged simulation outcome leads to the ensemble average. It is, however, often difficult to achieve converged sampling due to the large number of degrees of freedom (DOFs) of realistic processes, the high (free) energy barrier between different states of interest, and the gap between the intrinsic time scale of the process and the integration time stepsize. For example, biologically relevant processes often occur at μs, ms and even longer time scales [11, 12], while the maximum time step for integrating the equations of motion is on the order of 1 fs. As a result, to obtain sufficient statistics for post-process analysis, a huge number of integrations are required. This poses a challenging problem with even modern computational resources. To obtain better convergence behaviors, we often employ enhanced sampling techniques such as the equilibrium replica exchange [13], umbrella sampling [14, 15], and nonequilibrium steered molecular dynamics [16–18]. Because the system Hamiltonian is modified upon the addition of some bias potentials, post-process reweighting is utilized to recover the statistics of the original unperturbed ensemble, which is often achieved by free energy perturbation [19–21]. Another direction to accelerate the exploration of the configurational space is enlarging the integration time stepsize. The leapfrog (LF) integrator is the widely used algorithms for integrating the equations of motion. The LF integrator is set as default in many software packages for molecular dynamics (MD) simulations, e.g., AMBER [22]. Recent development suggests that a special form of the splitting operator formulation could lead to higher accuracy in the configurational distribution produced by Langevin dynamics [8, 23–27]. We have recently pointed out that the essential key idea is constructing the thermostatting algorithms where the thermostat part is updated in the middle of the updation of coordinate variables, which leads to the unified middle scheme where more efficient and robust algorithms can also be developed for other stochastic/deterministic thermostats [9, 10, 26–28]. When the LF integrator is used in the middle thermostat scheme (LF-middle), it provides better performance than the original LF algo-

rithm [10, 28]. As the frequencies of different motions differ significantly, the constraint or multiple time step techniques could be useful to accelerate the simulation [29–31]. Their algorithms with the middle thermostat scheme have been demonstrated to be more efficient [10].

In our previous study, the middle scheme has been performed in a few representative but rather simple systems (e.g., model systems with harmonic and quadratic potentials, Lennard-Jones clusters, liquid water, and small peptide in implicit solvent) [8–10, 26–28, 32]. In the work, extensive numerical benchmarks on the performance of LF and LF-middle integrators will be presented in several more complex systems with chemical and biological significance. Both conventional and enhanced sampling simulations are considered, aiming at providing a practical and general evaluation of the two regimes.

II. RESULTS AND DISCUSSION

Consider the canonical ensemble. The LF integrator (default in such as AMBER) reads

$$\mathbf{p}(0) \leftarrow \mathbf{p}\left(-\frac{\Delta t}{2}\right) - \frac{\partial U}{\partial \mathbf{x}(0)}\frac{\Delta t}{2} \qquad (1)$$

Thermostatic for a time step Δt, in which $\mathbf{p}(0)$ is updated.

$$\mathbf{p}\left(\frac{\Delta t}{2}\right) \leftarrow \mathbf{p}(0) - \frac{\partial U}{\partial \mathbf{x}(0)}\frac{\Delta t}{2} \qquad (2)$$

$$\mathbf{x}(\Delta t) \leftarrow \mathbf{x}(0) + \mathbf{M}^{-1}\mathbf{p}\left(\frac{\Delta t}{2}\right)\Delta t \qquad (3)$$

In comparison, the LF-middle integrator is

$$\mathbf{p}\left(\frac{\Delta t}{2}\right) \leftarrow \mathbf{p}\left(-\frac{\Delta t}{2}\right) - \frac{\partial U}{\partial \mathbf{x}(0)}\Delta t \qquad (4)$$

$$\mathbf{x}\left(\frac{\Delta t}{2}\right) \leftarrow \mathbf{x}(0) + \mathbf{M}^{-1}\mathbf{p}\left(\frac{\Delta t}{2}\right)\frac{\Delta t}{2} \qquad (5)$$

Thermostat for a time step Δt, in which $\mathbf{p}\left(\frac{\Delta t}{2}\right)$ is updated.

$$\mathbf{x}(\Delta t) \leftarrow \mathbf{x}\left(\frac{\Delta t}{2}\right) + \mathbf{M}^{-1}\mathbf{p}\left(\frac{\Delta t}{2}\right)\frac{\Delta t}{2} \qquad (6)$$

In Eq.(1)−Eq.(6), $\mathbf{x}$, $\mathbf{p}$, $\mathbf{M}$, and $U(\mathbf{x})$ are the coordinates, momenta, (diagonal) mass matrix, and potential energy surface of the system.

A. Dihedral flipping in solvated ACE-NME

The first example is the *N*-methylacetamide (NMA) system formed by ACE-NME (caps of protein) shown in FIG. 1. The NMA system is biologically relevant and simple enough to achieve a sufficient level of sampling convergence with moderate amounts of computational resources [33, 34]. The small two-residue peptide is described with AMBER99SB [35] and solvated in TIP3P water [36, 37] in an octahedron box of 2814 atoms (including 934 water molecules). A cutoff of 8 Å for non-bonded interactions in the real space is applied and the Particle-mesh Ewald [38] method is implemented for the long-range electrostatics. The simulation is performed at 300 K with 5 ps^{-1} as the friction coefficient for Langevin dynamics.

Conventional and enhanced sampling simulations are performed to probe the dynamic behavior of the system with the LF and LF-middle integrators. In the enhanced sampling simulations, the flipping of the backbone dihedral is studied with umbrella sampling [39–41] due to its simplicity and accuracy in producing highly accurate estimates of the free energy. The selected collective variable (CV), *i.e.*, the backbone dihedral, is not coupled to other slow DOFs, ensuring fast convergence. Therefore, it has been used in the test of various enhanced sampling methods [33, 34, 42]. The conformational change under investigation does not incur significant rearrangement of the whole solvated system. Thus, all simulations are performed in the NVT ensemble. As the constraint on bond length is an influencing factor of the simulation speed and the acceptable time stepsize, we performed the simulations with and without the SHAKE [43, 44] constraints on the bond-lengths involving hydrogen atoms.

1. Conventional MD

We first discuss the simulation results obtained with unbiased simulations. Twenty-four ns simulations starting from the 180° conformation (*i.e.*, the left one in FIG. 1) are performed at three time stepsizes including 0.5 fs, 1 fs, and 1.5 fs, respectively, and the sampling interval is set to 1.2 ps, which is sufficiently long to decorrelate successive samples according to autocorrelation analysis. As the same sample size is extracted at different time stepsizes, the statistical errors should be similar. Without the SHAKE constraints, the usable time stepsize is limited to smaller values such as 1 fs due

FIG. 1 Model system (ACE-NME) in water. The backbone C−C−N−C dihedral is used as the reaction coordinate for describing the dihedral flipping.

to the high-frequency vibrations of bonds involving hydrogen atoms. Here, we focus on the potential energy and its various sub-terms forming the AMBER force field, including the bond, angle, dihedral, electrostatic and vdW terms. As shown in FIG. 2, with the increase of the time stepsize, all potential energy terms deviate from the 0.5 fs results. The behavior of various energy terms with the LF and LF-middle integrators are different. When the LF integrator is used, the increase of the bond-stretching term due to the increase of the time stepsize is the largest one among these terms because of the high frequency or the large force constant of bond stretching. The other two stiff DOFs, *i.e.*, the angle and dihedral terms, increase relatively small. As for the soft DOFs, the electrostatic term is more sensitive to the time stepsize than the vdW one. The LF-middle integrator behaves differently. It suppresses the deviations of the stiff DOFs. The net change of the total potential energy is small, and the electrostatic term is the part showing the largest time-stepsize-dependence. It is evident that the LF-middle integrator enables the use of a larger time stepsize than LF without the SHAKE constraint.

The application of the SHAKE constraint removes the highest-frequency bond-stretching movement, thus enabling the use of a larger time stepsize. The most widely used technique is constraining bonds involving hydrogen atoms, which is the bond-constraint regime used in this work. We again extract the potential energy terms in this case in FIG. 3. The SHAKE constraint effectively suppresses the fluctuation of the stiff DOFs, and time stepsizes as long as 4 fs can be used. When the LF integrator is employed, similar to the previous SHAKE-off case, all energy terms exhibit significant deviations from the 0.5 fs reference. However, the contributions from the stiff DOFs (*i.e.*, bond, angle and dihedral) are much smaller than the soft DOFs, which is expected due to the application of hydrogen-involved
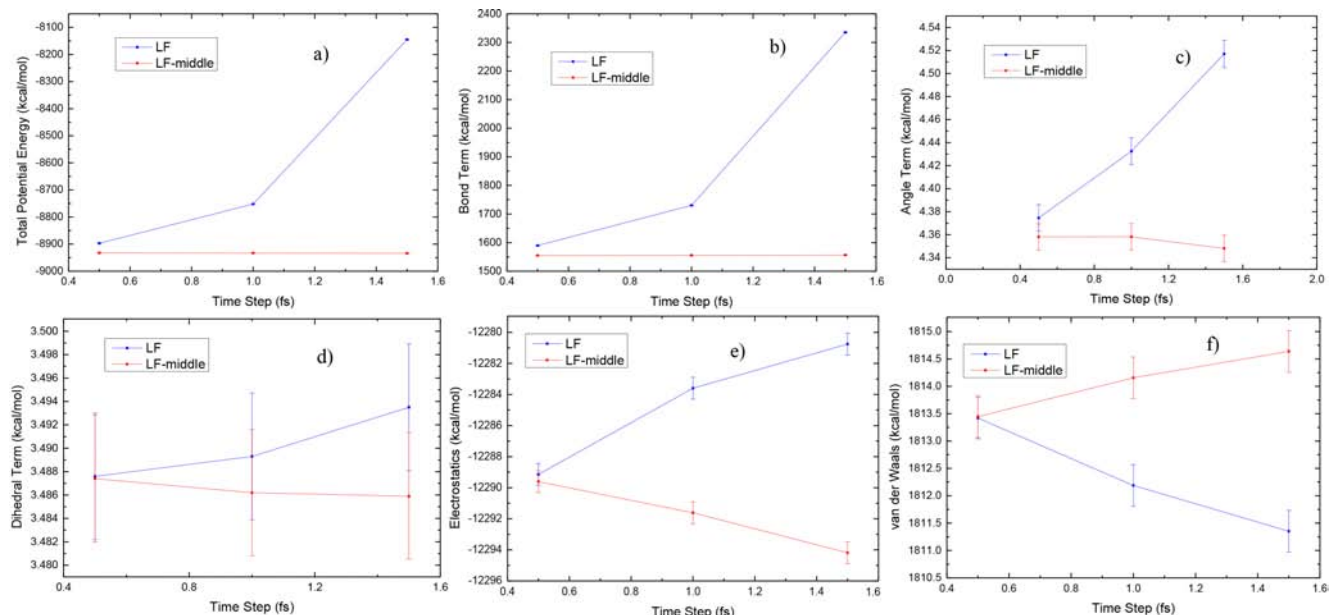
FIG. 2 (a) Time-averaged potential energies and its decomposition into (b) bond, (c) angle, (d) dihedral, (e) electrostatic, and (f) vdW terms obtained with LF and LF-middle integrators at different time steps without the SHAKE constraint.
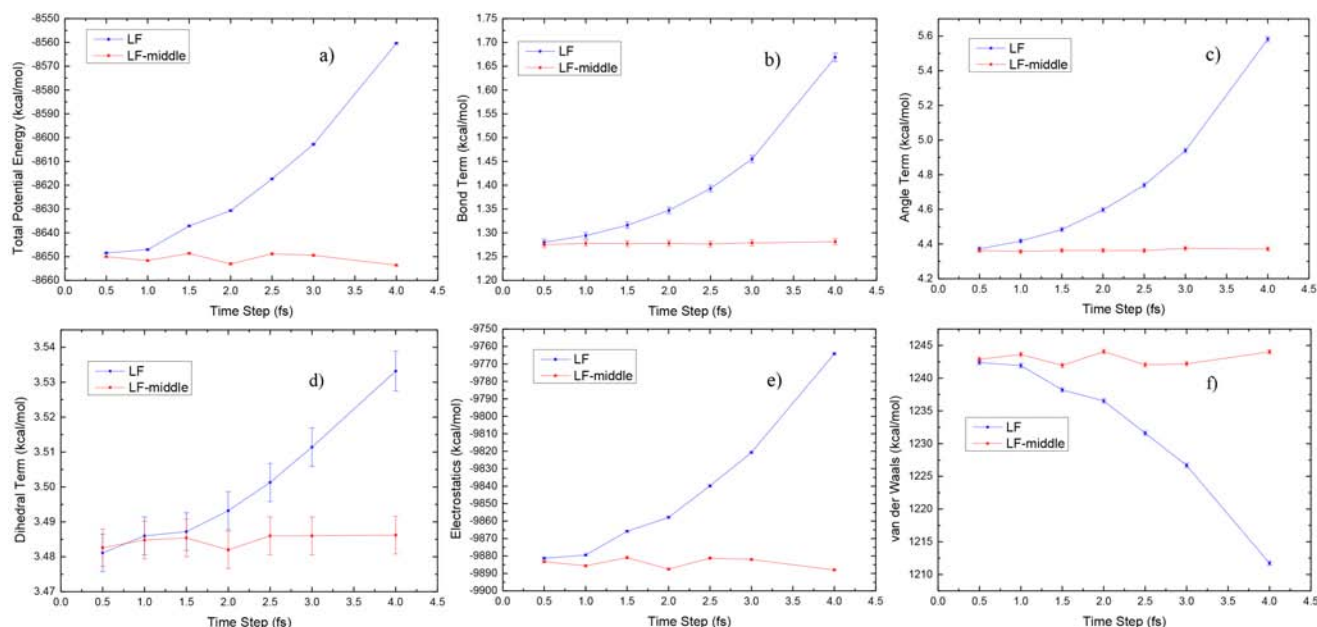

FIG. 3 (a) Time-averaged potential energies and its decomposition into (b) bond, (c) angle, (d) dihedral, (e) electrostatic, and (f) vdW terms obtained with LF and LF-middle integrators at different time steps with the SHAKE constraint on bonds involving hydrogen atoms.

bond-length constraints. The electrostatic term is still more sensitive to the time step than the vdW term. Therefore, the increase of the total potential energy, in the SHAKE-on LF case, is triggered mainly by the increase of the electrostatics interactions. As for the LF-middle case, the deviations of all energy terms from the 0.5 fs reference are much smaller. Similar to the previous SHAKE-off case, the deviations of the stiff DOFs

are suppressed, and the main contributions to the deviations of the total potential energy are from the soft DOFs. Therefore, the LF-middle integrator enables the use of a larger time stepsize than LF with the SHAKE constraint.

The above observation of the time-averaged estimates of various potential energy terms indicates that choosing a large time stepsize introduces errors in the integra-
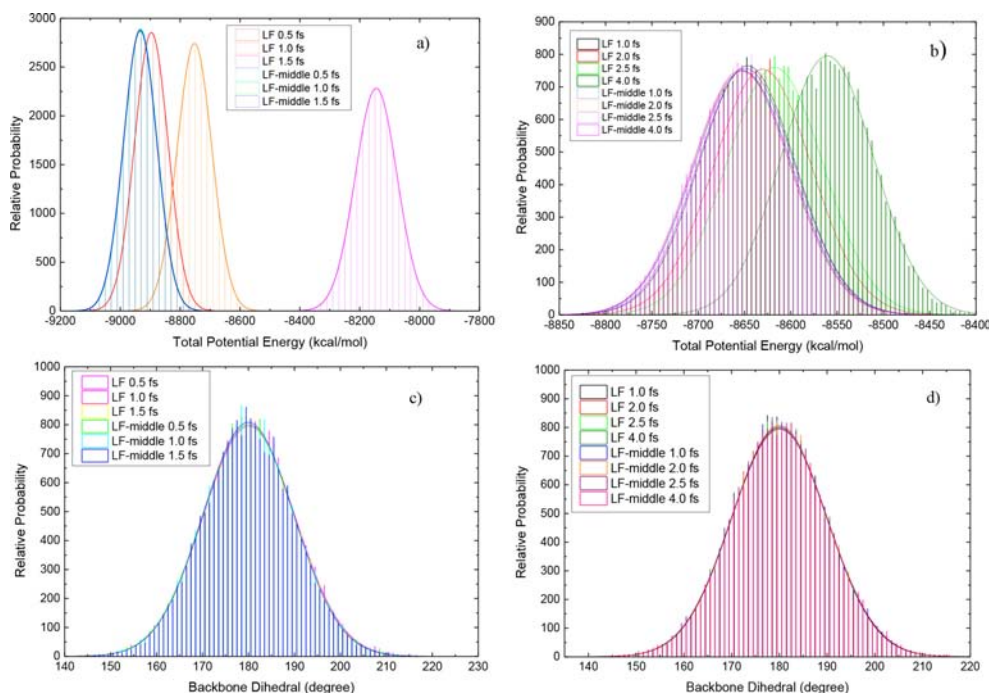
FIG. 4 The distributions of the total potential energy with LF and LF-middle integrators at different time steps in (a) the absence and (b) the presence of the SHAKE constraint on bonds involving hydrogen atoms. The distribution of the backbone dihedral in (c) the absence, and (d) presence of the SHAKE constraint. For the SHAKE on simulations, only the results obtained with four time steps are presented for clarity.

tion of the equations of motion. A more straightforward view of these errors is checking the distributions of the potential energy, which are presented in FIG. 4(a, b) for the SHAKE-off and SHAKE-on cases, respectively. According to these panels, it is clear that the simulation is actually sampling the perturbed distributions, which ultimately lead to discrepancies between estimates obtained with different time stepsizes. The LF-middle integrator still shows significantly higher tolerance of the magnitude of the time stepsize. This is consistent with our previous investigations on: the water molecule where only the intramolecular bond and angle terms exist, $(Ne)_{13}$ that has only the vdW term, and liquid water that has all interactions except the dihedral term [9, 10, 28].

We then check local fluctuations. We present the backbone dihedral, *i.e.*, the biasing CV to be investigated in the later enhanced sampling simulations as an illustrative example. As shown in FIG. 4(c, d), for the local observable of the backbone dihedral, the time-stepsize-induced error is not significant at all for both LF and LF-middle. Although LF-middle does provide better integration accuracy than LF, it takes much more numerical efforts to considerably decrease the statistical error bar to show the difference between the distribu-

tions of the local observables. Although the merit of using LF-middle may not be too significant when only local fluctuations are under investigation, LF-middle is still preferably used because it is in principle more accurate and adds no more computational resources than LF.

2. Enhanced sampling simulations

The previous unbiased simulations provide a comparison of the two integrators when the system is only exploring the neighborhood of the initial low-energy conformation. During conformational changes or chemical reactions, the system needs to overcome high free energy barriers. We compare LF-middle and LF when exploring high-energy regions in this sub-section. Umbrella sampling simulations are used to improve the sampling efficiency. Along the flipping pathway, we put 72 equally spaced windows between $0°$ and $360°$ and use a force constant of 100 kcal·mol^{-1}·rad^{-2}. This window spacing scheme has been employed in various dihedral flipping studies [45–47]. It ensures a sufficient level of phase space overlap between different windows and thus the reliability of perturbation-based reweighting. The extensive sampling in each umbrella window
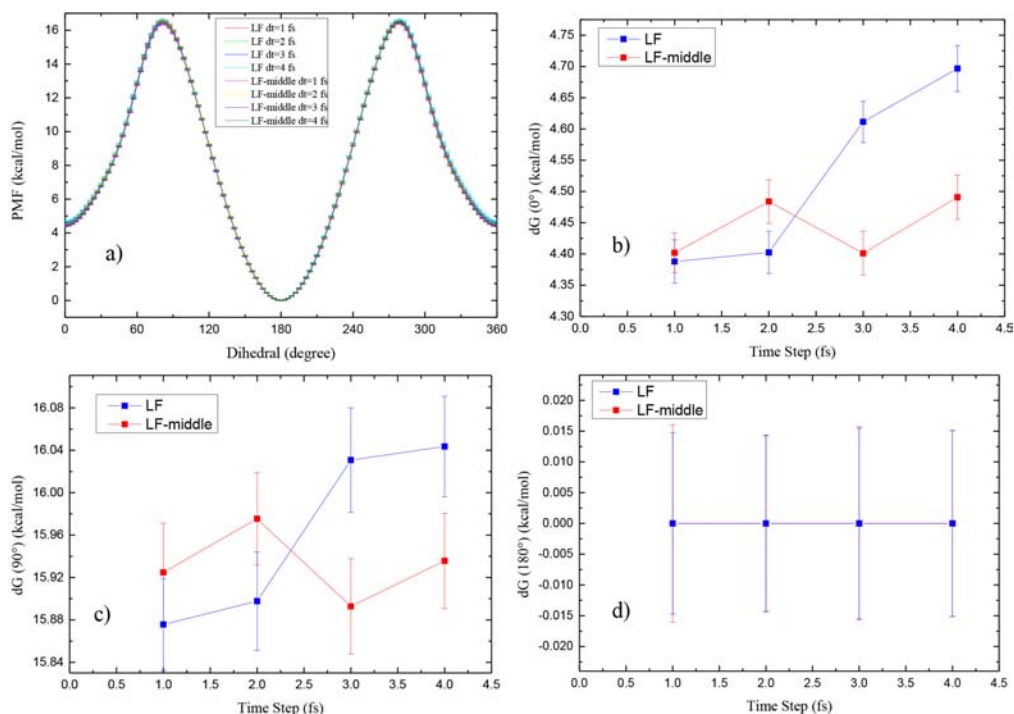
FIG. 5 With the LF and LF-middle integrators, at different time steps: (a) the free energy profile along the flipping dihedral, (b) the relative free energy at the 0° local free energy minimum, (c) the relative free energy at the 90° free energy barrier, (d) the relative free energy at the 180° global free energy minimum.

reaches 6 ns, and the sampling interval of the dihedral is set to 1.2 ps according to autocorrelation analysis. The potential of mean force (PMF) is obtained from the variational free energy profile (vFEP) method [48] and bootstrap resampling is used to obtain the numerical statistical error. In this case, only simulations with the SHAKE constraint are performed. The time step-size employed is set to 1 fs, 2 fs, 3 fs, and 4 fs.

The free energy profiles obtained with all time step-sizes are shown in FIG. 5(a). Visual inspection of the free energy profiles obtained with the two integrators at different time steps does not give hints on their difference. Considering the scale of the axis, the time-stepsize-induced bias is expected to be insignificant. As in practical simulations we want the simulation setup to introduce no/negligible artifacts to the outcome, we calculate the free energy difference between selected points/states to quantify the magnitude of the bias introduced by the integration time stepsize. Three points of the free energy profile including 0°, 90°, and 180° are of chemical/biological interests. The interconversion between the two minima at 0° and 180° needs to overcome the free energy barrier at 90°. Thus, the free energy differences between these states quantify the barrier height and the relative thermodynamic stability

of different conformational states. The time-stepsize-dependence of these free energy differences are shown in FIG. 5(b−d). The global minimum is used as a reference with a free energy of zero and the relative free energies of the other points are presented. For the 0° free energy minimum, its relative free energy changes with the time stepsize, and the situation is more severe for the LF integrator. The situation is similar for the free energy barrier at 90°. The 90° results with LF-middle at different time stepsizes agree with each other within statistical error, while those with LF change monotonically. With the LF integrator, the results obtained at 1 fs and 2 fs are very similar, which is expected due to the application of the SHAKE constraints on high-frequency bond DOFs involving hydrogen atoms, but for larger time steps (*i.e.*, 3 fs and 4 fs), the free energy change upon the time step variation achieves about 0.2 kcal/mol, indicating the unsafety of using these large time stepsizes. The relative free energy in the last point 180° remains zero regardless of the time step and the integration scheme, which indicates that the integrator and the magnitude of the time stepsize do not perturb much the position of the global minimum at 180°. Because the LF-middle integrator provides more accurate energetics (*e.g.* total potential energy) than

LF, the improvement for local observables involved in biologically relevant processes is noticeable when the longer time stepsize (*e.g.*, 4 fs with SHAKE) is used. It is expected that when more sampling data are available to decrease the error bar, the trend will be even more significant. This is entirely consistent with the observation for the case that the alanine dipeptide solvated in an implicit solvent model at 300 K, where replica exchange is used as the enhanced sampling method instead [10].

B. Intramolecular proton transfer in malonaldehyde in vacuo

The first dihedral flipping case illustrates the performance of the two integrators in biologically relevant conformational changes. In the second test case, the intramolecular proton transfer in malonaldehyde shown in FIG. 6(a) serves as a good example of chemical reactions. The fluctuation of the bonding feature is an essentially quantum mechanical effect. Thus, electronic structure calculations with *ab initio* or semi-empirical quantum mechanics methods are needed to account for the bond rearrangement. We used the Parametrized Model number 6 (PM6) [49] semi-empirical Hamiltonian and no solvent molecule is added due to efficiency consideration. Two distances including the $d(O1-H4)$ and $d(O2-H4)$ are involved in the bond rearrangement reaction and one-dimensional CV defined by the difference between these two distances could be used to describe the reaction. However, to make the simulation protocol more generally applicable to complex systems, we use the two distances as CVs to bias the simulation and perform two-dimensional umbrella sampling simulations. Umbrella windows are put from 0.8 Å to 1.8 Å with 0.1 Å increments in each dimension, and a force constant of 100 kcal·mol^{-1}·Å^{-2} is employed. The sampling time in each window is 27 ns and the sampling interval is set to 0.12 ps, which is sufficiently long to decorrelate successive samples according to autocorrelation analysis. The reweighting procedure is performed with WHAM and the statistical error is obtained with bootstrap resampling. The simulation is performed in vacuo at 300 K. The time steps used include 0.2 fs, 0.5 fs, 1 fs, and 2 fs, while higher values lead to unstable dynamics and the ns-length sampling cannot be successfully finished.

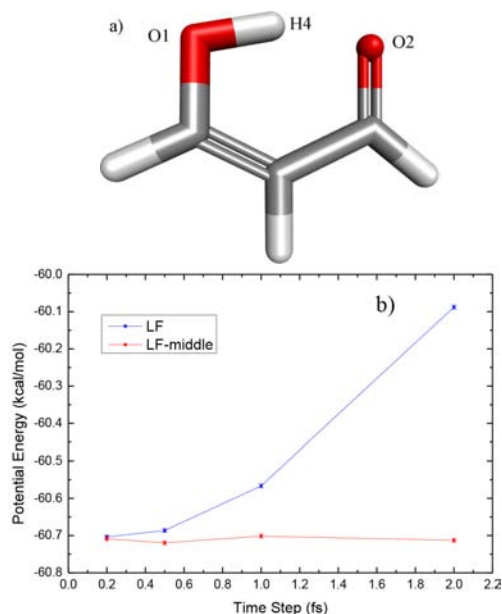A quick comparison between the LF and LF-middle results is shown in FIG. 6(b). The potential energy

FIG. 6 (a) Illustration of the intramolecular proton transfer case malonaldehyde, and (b) the time-averaged potential energy in the umbrella window centered at (1.3, 1.3).

of the system shows obvious time-stepsize-dependence when using the LF integrator, while the potential energy obtained with the LF-middle integrator only demonstrates very small changes with the increase of the time stepsize. We then turn to the free energy landscape of the proton transfer process.

In FIG. 7(a), the 2D PMF projected on the two distance CVs obtained with the 0.5 fs time step is shown. As the two ends of the proton transfer reaction are chemically equivalent, their thermodynamic stabilities are identical. The free energy minima from our projection locate at (1.061773, 1.707163) and (1.707163, 1.061773). This symmetric numerical result agrees with the theoretical expectation. From the free energy surface, we find that the saddle point locates at (1.290142, 1.290142), which lies on the $y=x$ line (*i.e.*, $d(O1-H4)=d(O2-H4)$) and again agrees with the theoretical expectation. Note that due to the details of data binning in reweighting, the exact positions of free energy minima and the saddle point could differ from the above positions, but the magnitude would be minimal. To quantify the time-stepsize-dependence of the free energy landscape, we calculate the free energy difference between the global minimum and the saddle point (*i.e.*, the free energy barrier of the proton transfer reaction) at different time stepsizes in FIG. 7(b), where the free energy barrier changes as the time step-
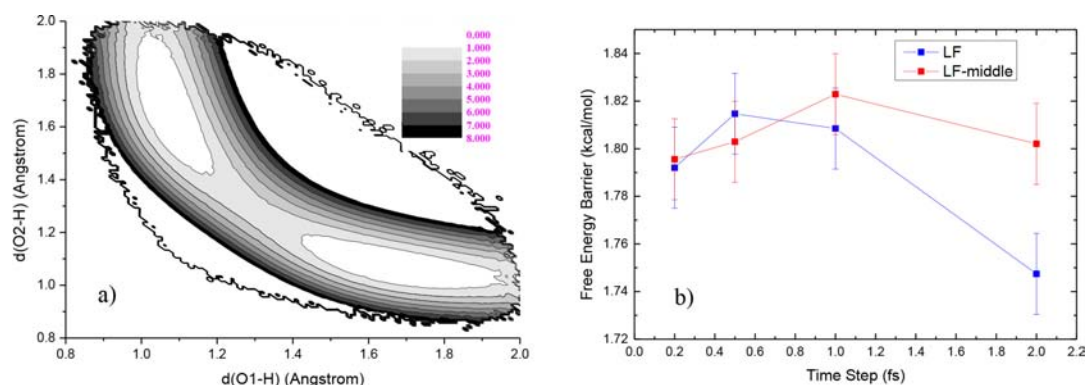
FIG. 7 (a) The two-dimensional free energy surface on $d(O1-H4)$ and $d(O2-H4)$ obtained with the 0.5 fs time step and the leapfrog integrator. The time-step-dependence of (b) the free energy difference between the minimum and the saddle point (*i.e.*, the barrier height) with different integrators.
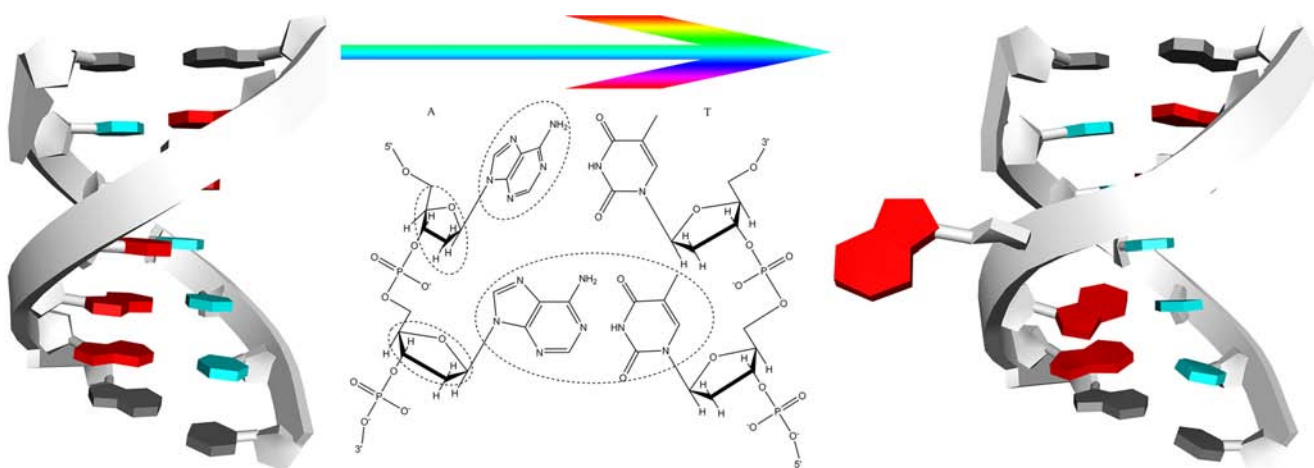

FIG. 8 Illustration of base flipping in the A_7-T_7 duplex. The pseudo-dihedral in the middle is used as the reaction coordinate.

size varies. In the LF-middle case, the free energy barriers obtained at different time steps agree within their statistical errors. By contrast, in the LF case, the free energy barriers show statistically significant deviations when using 2 fs time step. It is worth noting that the time-stepsize-induced bias in free energy difference is relatively small, in comparison to the potential energy deviations in FIG. 6(b). Considering the higher accuracy and more stable dynamics provided by LF-middle, this algorithm could be preferably used in molecular simulations of chemical reactions.

C. DNA AT-tract

The fundamental role of nucleic acid polymers in genetic coding has been widely recognized [50, 51]. The genetic codes are deposited by the sequence of the nucleotide (*i.e.*, the base pair), which hides deeply inside the duplex, triplex and various other structural motifs.

Despite the diversity of the structural features of DNA systems, the most widely observed one in DNA is the duplex structure [52]. Nucleotide systems incorporate a variety of crucial biological processes such as DNA methylation [53], melting [54, 55], bubbling [56, 57], breathing [58, 59], and protein-DNA interaction [60]. As the previous caps of protein and gas-phase intramolecular proton transfer are relatively simple compared with real biopolymers, we select a DNA duplex as the third case in the current integrator comparison. The A_7-T_7 duplex shown in FIG. 8 has been used as model systems in various benchmark tests of nucleotide force fields and simulation techniques [61–63]. The solute DNA is described with the bsc1 [64] force field and solvated in TIP3P water molecules in an octahedron box. Na^+ cations [65, 66] are added for neutralization. The resulting simulation box contains 10511 atoms with the periodic boundary condition applied. A cutoff of 9 Å for non-bonded interactions in the real space is ap-

plied and the long-range electrostatics are treated with the Particle-mesh Ewald [38] method. The simulation is performed at 300 K with 5 ps^{-1} as the friction coefficient for Langevin dynamics.

1. Unbiased Simulations

In conventional MD simulations, as the bond stretching has negligible influences on the conformational dynamics of biological systems, the SHAKE [43, 44] constraint is applied to the bonds involving hydrogen atoms. This setup enables the use of larger time steps and we tested 1 fs, 2 fs, 3 fs and 4 fs time steps. For each setup, we performed 30 ns unbiased simulations and extracted one sample per 1.2 ps, which according to autocorrelation analysis is sufficiently long to decorrelate successive samples. For the LF integrator, the unbiased MD simulation with 4 fs is unstable and collapses at about 25 ns. In comparison, when LF-middle is used, all of the four time stepsizes maintain stable dynamics during the whole 30 ns simulations. It indicates that LF-middle is more robust than LF.

The case of potential energy is similar to the previous cases. The various terms of the potential energy show obvious dependence on the time stepsize, as shown in FIG. 9(a−c). The distributions of the total potential energy are shown in FIG. 9(d). We can see that the distributions deviate from the converge one when using larger time stepsizes with the LF integrator, while the LF-middle integrator effectively maintains the correct distribution even with a time stepsize as large as 4 fs. Therefore, the LF-middle integrator could provide more accurate results for potential energies. Two other biologically relevant observables are also checked. The first one is the number of base-pair hydrogen bonds due to their importance in stabilizing the duplex. Due to the fraying of terminal bases, only the central 5 base pairs are included in the calculation. Namely, No.2−6 and No.9−13 bases are under consideration. In FIG. 9(e), the average number of base-pair hydrogen bonds shows some dependence on the time stepsize. The results with 1−3 fs time steps are similar, while the 4 fs time step leads to obvious deviations. This phenomenon indicates that although LF-middle could provide good potential energies and enlarge the time stepsize usable, the accuracy of some observables could still be not satisfactory at large time stepsizes. The second observable we calculate is the flipping dihedral, which reflects the local fluctuations of base-pair dynamics and characterizes the

flipping of the central base [67]. The pseudo-dihedral is defined by the centers of masses of the flipping base of A_4, the sugar moiety of A_4, the sugar moiety of A_5 (the 3′ side of A_4), and the bases of the base pair $A_5 - T_{10}$. The distributions of the flipping dihedral obtained with the two integrators at different time steps are presented in FIG. 9(f). Still, although the potential energies are perturbed significantly, the distribution of the pseudo-dihedral shows small variations as the time stepsize changes. Overall, the results suggest that although the LF-middle integrator is much more accurate for many configuration-dependent properties such as potential energies, its improvement for some observables with biological significance over LF is not so significant unless high accuracy and low statistical error are requested.

2. Enhanced sampling

Enhanced sampling simulations explore high-energy regions and the system is driven out of the most stable base-paired canonical state. To get a thorough view of the performance of the LF and LF-middle integrators in conformational sampling in practical biological processes, we simulate the system with and without the SHAKE constraint in umbrella sampling simulations. As discussed previously, the biasing coordinate is the pseudo-dihedral, which describes the flipping of the central base relative to the duplex backbone. As this conformational change involves large-scale motions of various components in the simulated system, we simulate the system in both the NVT and NPT ensemble in the LF case to investigate the impact of box-volume fluctuations, while for LF-middle only NVT simulations are performed. In NPT simulations, isotropic position scaling and the Berendsen barostat are applied for pressure regulation. The window spacing and force constant parameters are the same as the previous backbone dihedral case.

A long equilibration step is required to obtain well-equilibrated simulation boxes along the whole base flipping pathway, *i.e.*, in all 72 umbrella windows. The length of equilibration is about tens of ns according to our previous experience on this system [61]. To avoid unnecessary waste of computation time, the starting configurations are obtained from our previous work on the AT tracts, *i.e.*, the last configurations in each umbrella window under the bsc1 force field [61]. As these configurations are well-equilibrated in the sampled en-
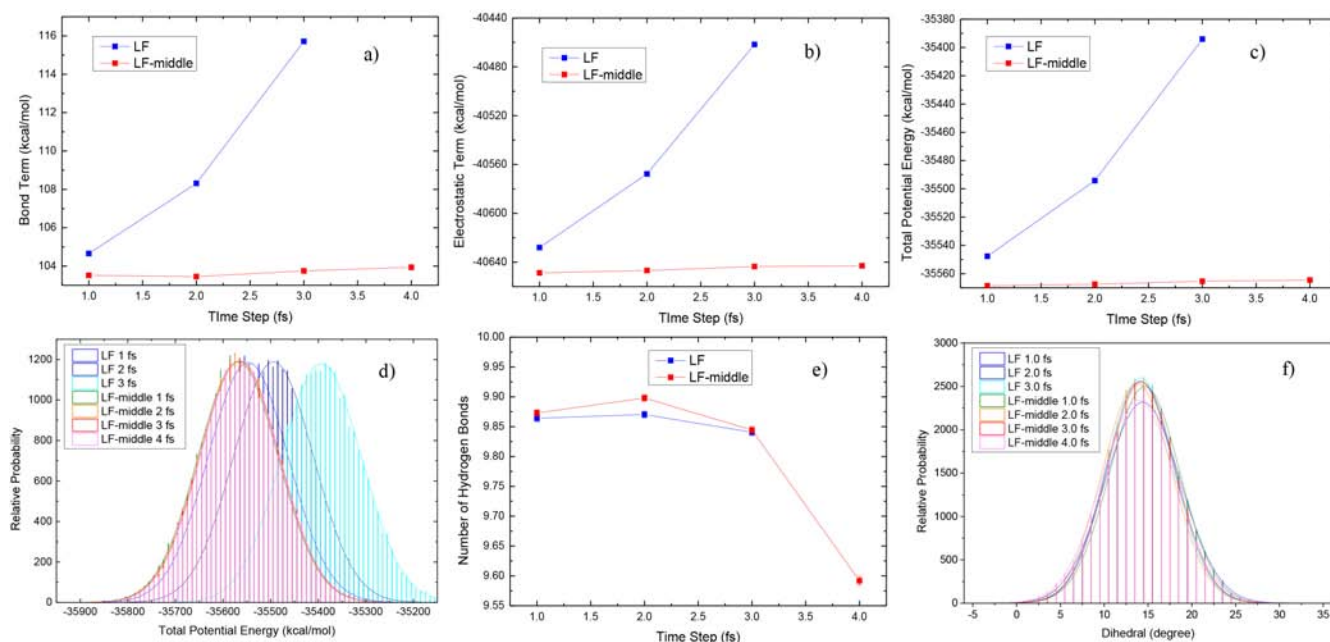
FIG. 9 (a) The bond term of the potential energies, (b) the electrostatic part, (c) the total potential energy, (d) the distribution of the total potential energy, (e) the number of base-paired hydrogen bonds for the central 5 base pairs (*i.e.*, residue No.2−6 and No.9−13), and (f) obtained with LF and LF-middle integrators at different time steps with the SHAKE constraint on bonds involving hydrogen atoms.

semble, no equilibration is performed. In each umbrella window, a 4 ns simulation is performed and the sampling interval is 2 ps, which is similar to the autocorrelation time of the flipping dihedral according to our previous autocorrelation analyses [61, 67, 68]. Therefore, there are approximately 2000 points in each window. According to our previous benchmark test on the convergence time of base flipping simulations, this length of sampling time is already long enough for converged estimates of the free energy profiles along the base flipping pathway. The potentials of mean force (PMFs) are obtained from vFEP reweighting.

Without SHAKE, the bonds are not constrained and all DOFs could be freely relaxed. A time stepsize of 1 fs is used to propagate the dynamics, while larger time stepsizes result in the termination of the simulation. In FIG. 10(a), we compare the LF and LF-middle PMFs. The results obtained from different integrators (*i.e.*, LF and LF-middle) are also similar, which indicates that the simulation result is relatively insensitive to the integrator used. As LF-middle introduces smaller artifacts than LF, we can conclude here that both integrators are almost bias-free at this time step. Thus, without the SHAKE constraint, the merit of LF-middle in free energy simulations is not significant. We also compare the PMFs obtained with the 2 fs time step and the SHAKE

constraints in FIG. 10(a). This combination is the normal setup used in biomolecular simulations. Obviously, the PMFs obtained with the 1 fs time step and without SHAKE constraints are very similar to those obtained with the 2 fs time step and SHAKE constraints. Therefore, in base flipping simulations, it is safe to turn the SHAKE option on and use 2 fs time steps.

We then investigate the simulation outcomes with the SHAKE constraints in detail. In FIG. 10(b), we compare the 2 fs and 4 fs results obtained with the LF and LF-middle integrators. For the LF integrator, the 2 fs results with different integrators are very similar, while the 4 fs time stepsize setup in the NVT and NPT ensembles provides wrong results, deviating from the true expectation. Many windows cannot finish the ns-length sampling due to unstable dynamics. The 2 fs PMFs obtained in NVT and NPT ensembles are very similar, which suggests that the simulation box obtained from previous work is sufficiently equilibrated and the box volume in NVT simulations is properly set. Therefore, for the LF integrator, the usable time step should not be larger than 2 fs in base flipping simulations. When the LF-middle integrator is employed, 2 fs and 4 fs PMFs are very similar and are consistent with the 2 fs LF results. However, the simulations with a 4 fs time step crash in some windows due to unstable dynamics, which
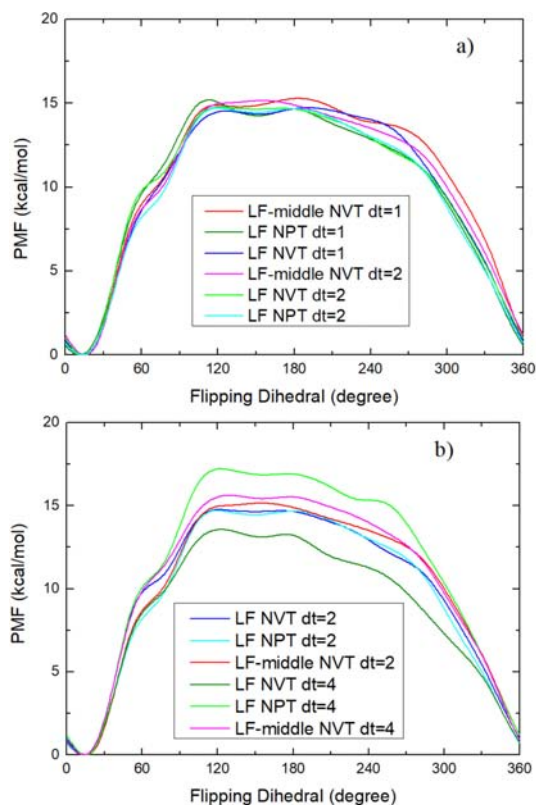
FIG. 10 Free energy profiles along the base flipping pathway obtained in (a) the absence or (b) the presence of the SHAKE constraints. For simulation without SHAKE, only 1 fs time stepsize is usable, while larger time steps result in unstable dynamics. The last three PMFs in the first panel are obtained with the SHAKE constraints to check whether SHAKE would change the thermodynamic profiles.

suggests unstable dynamics in some explored regions. It is expected the acceptable time stepsize for obtaining converged results is 3 fs for LF-middle, while that is 2 fs for LF.

D. Binding free energy calculations: T_4 lysozyme with benzene and phenol

The above comparisons between the LF and LF-middle integrators provide hints on their time-stepsize-induced errors in the configurational space. Within the range of usable time stepsizes employed in MD simulations, both integrators lead to similar results for some local observables when the statistical error is not requested to be significantly small, although LF-middle performs much more robust than LF for some global properties. As the above examples focus on the conformational changes or chemical reactions along physical pathways, in the last case we present an example in another direction, along the non-physical alchemical

pathway. The alchemical method constructs artificial transformation pathways connecting states of interest, which enables feasible calculations of the free energy difference between them [69–71]. In drug industry, the method has been widely applied in the calculations of solvation free energies, protein-ligand binding affinities, and pK_a shifts [72–77]. Due to the biological significance and the complexity of the protein-ligand problem, we consider this case in the integrator comparison. The protein system is T_4-lysozyme L99A [78, 79], and two ligands targeting it include benzene and phenol. This model system has been widely applied in the benchmarks of various free energy methods and thus serves as a nice illustrative example for the current integrator comparison.

To enable a fully reproducible benchmark here, the starting configuration and the parameter setup are obtained from the A9 tutorial of AMBER, which is available at https://ambermd.org/tutorials/advanced/tutorial9/index.html. The calculation of the relative binding free energy of the two ligands is decomposed into the calculation of the free energy changes along the two legs (protein-ligand complex and ligands) of the thermodynamic cycle, as shown in FIG. 11. The difference between the binding affinities of protein-benzene and protein-phenol complexes is thus equivalently estimated by the difference between the free energy changes when mutating benzene to phenol in the presence and absence of the protein. The transformation along each leg is performed in a three-step fashion. The decharge, vdW-bonded and recharge transformations are further stratified into 11 equally spaced windows (*i.e.* $\Delta\lambda=0.1$) to improve the convergence behavior. The alchemical intermediate states are described with linearly mixed end-state Hamiltonians. The softcore potential [80–84] is employed to avoid the singularity in the vdW-bonded transformation and the whole ligand is included in the softcore region. As our benchmark test aims at providing insights with the most widely applied settings in alchemical free energy calculations, the SHAKE [43, 44] constraint is applied on the bonds involving hydrogen atoms, and the integration time steps tested include 1 fs, 2 fs, and 4 fs. In each alchemical window, the system experiences 500 cycles minimization with 5 kcal·mol^{-1}·Å^{-2} restraints on non-hydrogen atoms of solute, 30 ps NVT heating from 0 K to 300 K and 120 ps NPT equilibration, after which a 420 ps production run with a sampling interval of 0.24 ps is performed to accu-
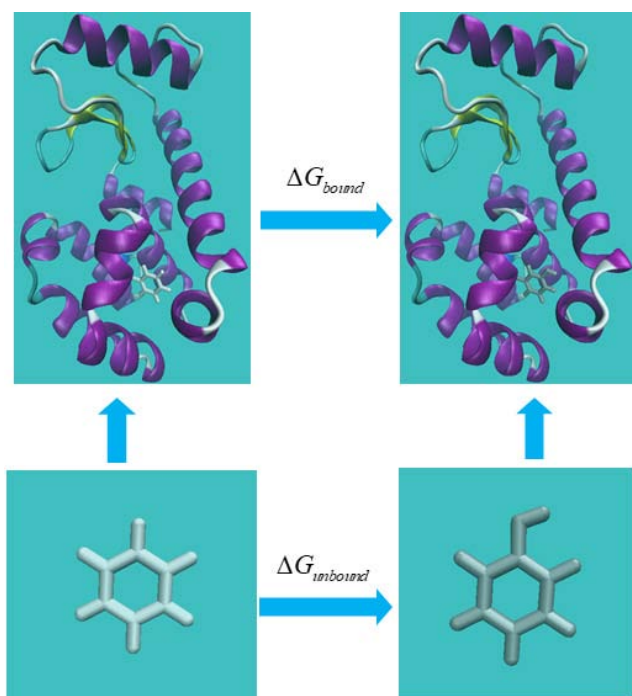
FIG. 11 An illustration of the thermodynamic cycle used to calculate the relative binding free energy of the two ligands with the T_4 lysozyme. The arrows marked with free energy changes are directly simulated, the top arrow represents the complex leg, while the bottom one is for the ligands leg.
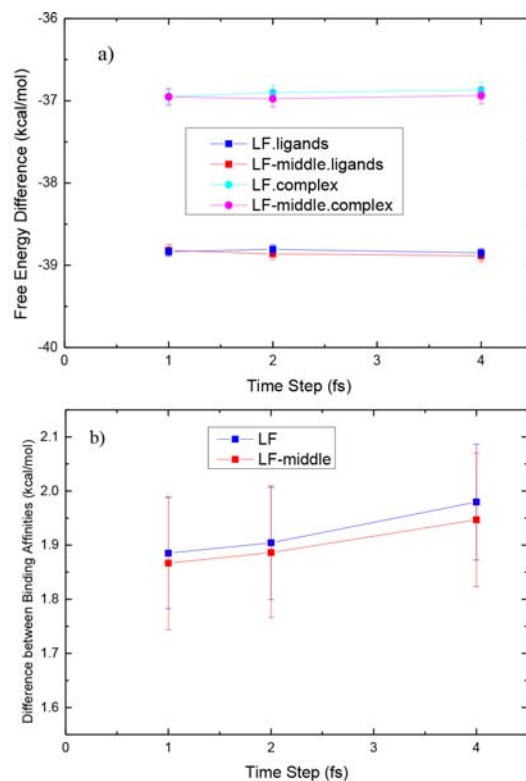

FIG. 12 (a) Free energy differences along the two legs of alchemical transformations and (b) the overall free energy difference that estimates the relative binding free energy of the two ligands.

mulate the time-series data. The free energy differences are extracted with the trapezoid rule thermodynamic integration (TI) method [85–87], and the corresponding statistical uncertainties are obtained with the analytical formula of the estimator. Note that before free energy analysis, the whole dataset is sub-sampled according to the autocorrelation of the partial derivative of the alchemical Hamiltonian $\frac{\partial U}{\partial \lambda}\big|_{\lambda_i}$.

We then check the simulation outcome with different integrators. As the core of alchemical free energy calculations is the free energy difference, our comparison between LF and LF-middle focuses on this statistical quantity. The two alchemical legs are checked separately in FIG. 12(a). For both the ligands-only and the protein-ligand legs, the free energy differences under different time steps agree quite well, which suggests that the time-stepsize-induced bias is relatively small for the two integrators even with very large time steps (4 fs). Thus, large time steps (*e.g.* 4 fs) can be safely used in alchemical transformations in protein-ligand and solvated ligand systems. However, we should note that this time-step tolerance is system-dependent and the exact maximum value for other complex systems requires

more numerical tests. Another major observation in FIG. 12(a) is the great accordance between the two integrators, which indicates that the improvement when shifting LF to LF-middle is not very significant. We then combine the results along the two alchemical legs to estimate the target value in the thermodynamic cycle, *i.e.*, the difference between the binding affinities of the two ligands. The time-stepsize-invariant behavior is also observed. The systematic bias introduced when using a large time stepsize such as 4 fs is very small, which again supports the use of large time stepsizes in alchemical free energy calculations in complex systems. The time-stepsize-induced error with LF-middle is smaller than the LF one.

E. Ethaline deep eutectic solvent

Ethaline deep eutectic solvent (DES) was considered as a major breakthrough in green chemistry [88]. DESs are often defined as binary mixtures of properly mixed components capable of developing multiple hydrogen bonds [89]. The freezing point of the resulting DES is
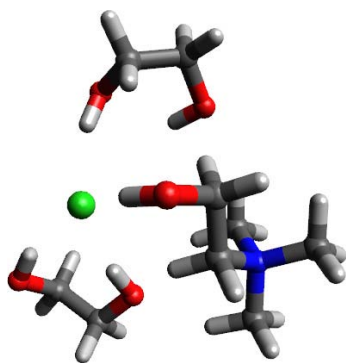
FIG. 13 The chemical structure of ethaline.

usually highly depressed compared to those of the constituents [90]. Despite the extensive studies on DESs, there still exists a scarcity of understanding on the most major principles behind the formation and some extraordinary properties of DESs [91]. On the basis of their constituents, DESs are categorized to five classes out of which type III DESs, often composed of choline chloride (ChCl) as the hydrogen bond acceptor (HBA) and a variety of hydrogen bond donors (HBDs) such as (poly)alcohols, amides and carboxylic acids, are the most studied ones in literatures [92–94]. As a representative type III DES, ethaline in FIG. 13 is formed from a 1:2 molar ratio of ChCl and ethylene glycol (EG) with a freezing point of $-66\ ^{\circ}$C, which is remarkably lower than that of ChCl ($303\ ^{\circ}$C) and EG ($-12.9\ ^{\circ}$C) [92, 95].

In our recent works, we found that the decrease of the $Cl^{-}\rightarrow H-O_{Ch}^{+}$ charge transfer towards $Cl^{-}\rightarrow H-O_{EG}$ upon ChCl and EG mixing is a major contributor to the DES formation [93, 94]. Consequently, the $O-H$ bonds of Ch^{+} and EG which are directly H-bonded to Cl^{-} tolerate changes, influencing the spectral properties of the mixture. In this regard, the changes of the bond lengths of the $O-H$ of Ch^{+} and EG would be of chemical significance in molecular simulations of the ethaline DES. Therefore, as the fifth test case we investigate the time-stepsize-dependence of simulation results in the ethaline DES, with a special focus on bond lengths, the total potential energy of the simulated system and the bond stretching term. The system is composed of 512 ChCl and 1024 EG molecules with the periodic boundary condition. The atomic charges for each component are derived with the restrained electrostatic potential (RESP) scheme at the HF/6-31G* level of theory and the other parameters are obtained from the generalized amber force field (GAFF). MD simulations were performed at 298.15 K and 1 atm. The time

stepsizes employed in the LF and LF-middle integrators include 0.15, 0.5, 1, 1.5, 2, and 2.5 fs. After the energy minimization and heating processes, the system was equilibrated in the NPT and NVT ensembles for 23 and 15 ns, respectively. Then, the NVT simulation with a sampling interval of 10 ps is initiated for data accumulation. For the short-range non-bonded interactions, a cutoff of 12 Å was applied, while the long-range electrostatics interactions were treated by the Particle-mesh Ewald method.

Shown in FIG. 14 (a) and (b) are the changes of the time-averaged $O-H$ bond length of Ch^{+} and EG upon increasing the time stepsize, respectively. It is seen that as the time stepsize increases using both LF and LF-middle integrators, the bond lengths do not change. It implies that the $O-H$ bond lengths are not heavily influenced by the employed integrator. However, with the increase of the time stepsize, the total potential energy (FIG. 14(c)) and its bond term (FIG. 14(d)) deviate abruptly from the 0.5 fs results when using the LF integrator. It clearly shows that the LF-middle integrator enables the use of larger time stepsizes than LF and that LF-middle is more accurate than LF. We also see that the outperformance of LF-middle over LF is more significant in global properties (potential energies) than in local ones (bond-lengths).

F. AppA blue-light using flavin receptor

The last system is the blue-light using flavin (BLUF) photoreceptor involved in many light-activated biological processes [96–99]. The outstanding modular and long-range signaling capabilities of BLUF make them promising candidates for optogenetics. Here, we focus on the BLUF domain of AppA from Rhodobacter sphaeroides, the structure of which is obtained from the PDB bank with the ID of 1YRX [100]. The cofactor FMN presented in the structure file is included in the modelling. We use the AMBER03 force field to describe the protein system. Solvation is performed with 8431 TIP3P water molecules and counter ions (2 Na^{+} cations, to be specific) are added for neutralization. The core region for this BLUF is around the cofactor FMN and two protein residues of Tyr9 and Gln51, an illustration of which is presented in FIG. 15(a). For this reactive region, we employ the QM/MM treatment to enable accurate description of the energy. Minimization and equilibration at 300 K are performed, after which
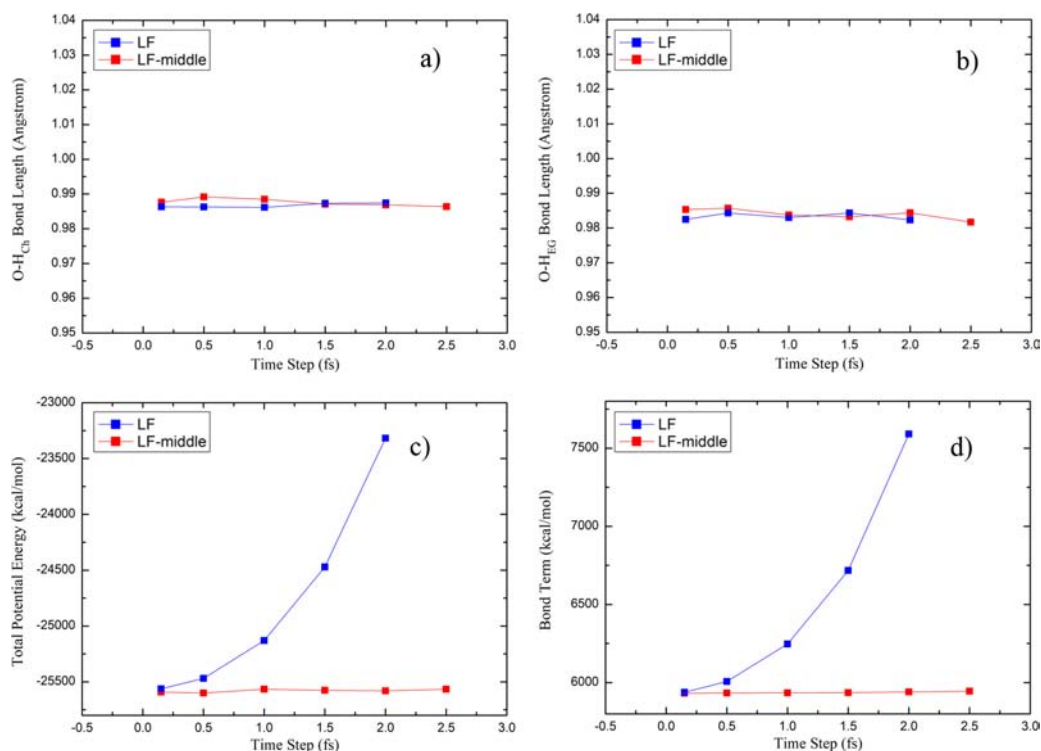
FIG. 14 Time-averaged (a) O−H bond length of Ch^+, (b) O−H bond length of EG, (c) total potential energy and (d) bond term of the total potential energy. All quantities are obtained with LF and LF-middle integrators at different time stepsizes without the SHAKE constraint.
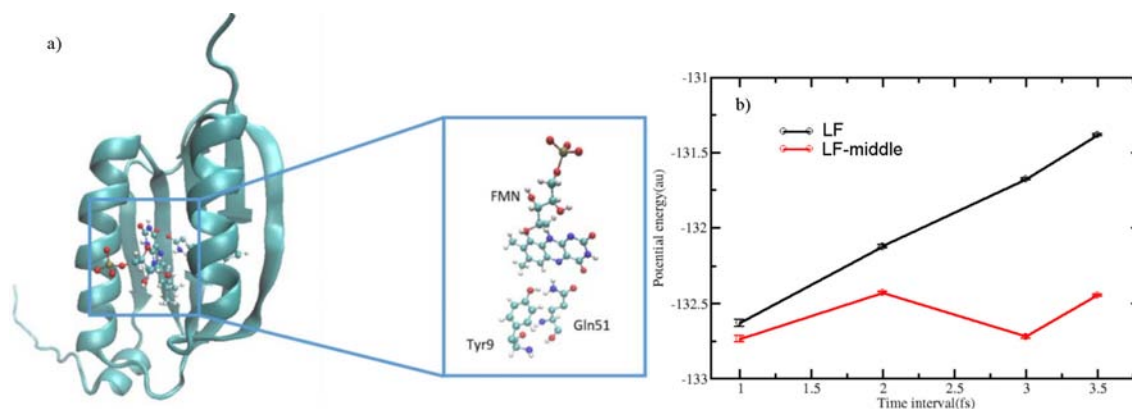

FIG. 15 (a) An illustration of the core region of the BLUF and (b) the potential energy obtained with LF and LF-middle.

production runs are initiated. For the test of the two integrators, we employed 4 different time steps including 1 fs, 2 fs, 3 fs, and 3.5 fs, and the total sampling time was 20 ns. The resulting potential energy in atomic units are given in FIG. 15(b), from which it is clear that the accuracy of the potential energy is successfully maintained as the time stepsize changes when using the LF-middle integrator. In comparison, the LF integrator produces significant bias. Thus, for accurate calculation of energetics, the LF-middle integrator should be

preferably employed.

III. CONCLUSION

Although current computational tools enable the investigation of many physical, chemical and biological processes, the extension of the accessible time scale lies at the center of molecular simulations. The significant gap between the time scales of biological and chemical interests and the time stepsize acceptable in molecu-

lar dynamics simulation often involves a large number of time steps, and methods for improving the sampling efficiency are always preferably applied in modern simulation studies.

A way to solve the problem is enlarging the integration time stepsize. On this aspect, the middle thermostat scheme is capable of maintaining the same accuracy with a larger time stepsize (about 4 times or more) than conventional MD algorithms [8–10, 26–28]. In the work, we focus on the LF algorithm, which is a typical and representative integrator for the equations of motion and is widely applied and set default in many computational packages, *e.g.*, AMBER. We compare LF-middle (the version of LF algorithm of the middle thermostat scheme) and the original/default LF integrator for several complex systems with biological and chemical significance. The extensive numerical tests suggest that LF-middle generally provides better statistics and yields more accurate results than the default LF integrator. The situation is quite significant for various terms of potential energies of the whole system, especially the electrostatic part. The outperformance of LF-middle over the default LF integrator is much more significant for global properties (energies) than local observables such as the bond-length, backbone or pseudo dihedral, or difference between PMFs. Because LF-middle causes no computational overhead on the conventional LF integrator, our results suggest that LF-middle effectively suppresses the time-stepsize-induced error and thus ensures accurate calculation of thermodynamics of complex systems. We expect that the extensive tests that we present in the work will encourage others to use LF-middle and other algorithms of the "middle" thermostat scheme to study more systems with biological and chemical significance with MD [9, 10, 28] or path integral MD [8, 9].

IV. ACKNOWLEDGMENTS

[1] B. J. Alder and T. E. Wainwright, J. Chem. Phys. **27**, 1208 (1957).

[2] E. Fermi, J. Pasta, and S. Ulam, *Studies of Non Linear Problems*, Los Alamos Report LA-1940, (1955).

[3] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids*, Oxford: Clarendon Press, 385 (1987).

[4] D. Frenkel and B. Smit, *Understanding Molecular Simulation: from Algorithms to Applications*, 2nd Edn, San Diego: Academic Press, (2002).

[5] Y. Xiang, L. L. Duan, and J. Z. H. Zhang, Phys. Chem. Chem. Phys. **12**, 15681 (2010).

[6] B. Wang, Y. F. Qi, Y. Gao, and J. Z. H. Zhang, Phys. Chem. Chem. Phys. **22**, 8461 (2020).

[7] J. Z. Chen, L. X. Pang, W. Wang, L. F. Wang, J. Z. H. Zhang, and T. Zhu, J. Biomol. Struct. Dyn. **38**, 985 (2020).

[8] J. Liu, D. Z. Li, and X. J. Liu, J. Chem. Phys. **145**, 024103 (2016).

[9] Z. J. Zhang, X. Z. J. Liu, Z. F. Chen, H. F. Zheng, K. Y. Yan, and J. Liu, J. Chem. Phys. **147**, 034109 (2017).

[10] Z. J. Zhang, X. Z. J. Liu, K. Y. Yan, M. E. Tuckerman, and J. Liu, J. Phys. Chem. A **123**, 6056 (2019).

[11] J. Z. Chen, S. L. Zhang, W. Wang, L. X. Pang, Q. G. Zhang, and X. G. Liu, J. Chem. Inf. Model. **61**, 1954 (2021).

[12] Z. X. Sun, Z. H. Gong, F. Xia, and X. He, Chem. Phys. **548**, 111245 (2021).

[13] H. Fukunishi, O. Watanabe, and S. Takada, J. Chem. Phys. **116**, 9058 (2002).

[14] R. W. W. Hooft, B. P. van Eijck, and J. Kroon, J. Chem. Phys. **97**, 6690 (1992).

[15] G. H. Paine and H. A. Scheraga, Biopolymers **24**, 1391 (1985).

[16] C. Jarzynski, Phys. Rev. Lett. **78**, 2690 (1997).

[17] R. Chelli, C. Gellini, G. Pietraperzia, E. Giovannelli, and G. Cardini, J. Chem. Phys. **138**, 214109 (2013).

[18] E. Giovannelli, P. Procacci, G. Cardini, M. Pagliai, V. Volkov, and R. Chelli, J. Chem. Theory Comput. **13**, 5874 (2017).

[19] V. Lindahl, J. Lidmar, and B. Hess, Phys. Rev. E **98**, 023312 (2018).

[20] Z. X. Sun, Q. L. He, X. Li, and Z. D. Zhu, J. Comput. Aided Mol. Des. **34**, 589 (2020).

[21] Z. X. Sun, J. Comput. Aided Mol. Des. **35**, 105 (2021).

[22] D. A. Case, T. E. Cheatham III, T. Darden, H. Gohlke, R. Luo, K. M. Merz Jr., A. Onufriev, C. Simmerling, B. Wang, and R. J. Woods, J. Comput. Chem. **26**, 1668 (2005).

[23] B. Leimkuhler and C. Matthews, Appl. Math. Res. eXpress **2013**, 34 (2012).

[24] B. Leimkuhler and C. Matthews, J. Chem. Phys. **138**, 174102 (2013).

[25] N. Grønbech-Jensen and O. Farago, Mol. Phys. **111**, 983 (2013).

[26] D. Z. Li, X. Han, Y. C. Chai, C. Wang, Z. J. Zhang, Z. F. Chen, J. Liu, and J. S. Shao, J. Chem. Phys. **147**, 184104 (2017).

[27] D. Z. Li, Z. F. Chen, Z. J. Zhang, and J. Liu, Chin. J. Chem. Phys. **30**, 735 (2017).

[28] Z. J. Zhang, K. Y. Yan, X. Z. J. Liu, and J. Liu, Chin. Sci. Bull. **63**, 3467 (2018).

[29] M. Marchi and P. Procacci, J. Chem. Phys. **109**, 5194 (1998).

[30] P. Procacci and M. Marchi, J. Chem. Phys. **104**, 3003 (1996).

[31] P. Procacci and B. J. Berne, J. Chem. Phys. **101**, 2421 (1994).

[32] B. Leimkuhler and C. Matthews, Proc. Roy. Soc. A: Math., Phys. Eng. Sci. **472**, 20160138 (2016).

[33] X. H. Wang, Q. L. He, and Z. X. Sun, Phys. Chem. Chem. Phys. **21**, 6672 (2019).

[34] X. H. Wang, X. Z. Tu, B. M. Deng, J. Z. H. Zhang, and Z. X. Sun, J. Comput. Chem. **40**, 1270 (2019).

[35] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling, Proteins **65**, 712 (2006).

[36] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, J. Chem. Phys. **79**, 926 (1983).

[37] D. J. Price and C. L. Brooks III, J. Chem. Phys. **121**, 10096 (2004).

[38] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, J. Chem. Phys. **103**, 8577 (1995).

[39] R. W. W. Hooft, B. P. van Eijck, and J. Kroon, J. Chem. Phys. **97**, 6690 (1992).

[40] M. Mezei, J. Comput. Phys. **68**, 237 (1987).

[41] J. Kästner, WIREs: Comput. Mol. Sci. **1**, 932 (2011).

[42] Z. X. Sun, Phys. Chem. Chem. Phys. **21**, 21942 (2019).

[43] J. P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen, J. Comput. Phys. **23**, 327 (1977).

[44] S. Miyamoto and P. A. Kollman, J. Comput. Chem. **13**, 952 (1992).

[45] Z. X. Sun, X. H. Wang, and J. Z. H. Zhang, Chem. Phys. Lett. **684**, 239 (2017).

[46] J. A. Lemkul, A. Savelyev, and A. D. MacKerell Jr., J. Phys. Chem. Lett. **5**, 2077 (2014).

[47] H. Zheng, Y. Q. Cai, S. Ding, Y. J. Tang, K. Kropachev, Y. Z. Zhou, L. H. Wang, S. L. Wang, N. E. Geacintov, Y. K. Zhang, and S. Broyde, Chem. Res. Toxicol. **23**, 1868 (2010).

[48] T. S. Lee, B. K. Radak, A. Pabis, and D. M. York, J. Chem. Theory Comput. **9**, 153 (2013).

[49] J. J. P. Stewart, J. Mol. Model. **13**, 1173 (2007).

[50] N. K. Banavali and A. D. MacKerell Jr., J. Mol. Biol. **319**, 141 (2002).

[51] Y. S. Lazurkin, M. D. Frank-Kamenetskii, and E. N. Trifonov, Biopolymers **9**, 1253 (1970).

[52] N. A. Špačková, E. Cubero, J. Šponer, and M. Orozco, J. Am. Chem. Soc. **126**, 14642 (2004).

[53] M. O'Gara, R. J. Roberts, and X. D. Cheng, J. Mol. Biol. **263**, 597 (1996).

[54] K. A. Velizhanin, C. C. Chien, Y. Dubi, and M. Zwolak, Phys. Rev. E **83**, 050906(R) (2011).

[55] B. H. Zimm, J. Chem. Phys. **33**, 1349 (1960).

[56] A. Travers, Curr. Biol. **15**, R377 (2005).

[57] A. Zeida, M. R. Machado, P. D. Dans, and S. Pantano, Phys. Rev. E **86**, 021903 (2012).

[58] M. Peyrard, S. Cuesta-López, and G. James, J. Biol. Phys. **35**, 73 (2009).

[59] C. I. Duduiala, J. A. D. Wattis, I. L. Dryden, and C. A. Laughton, Phys. Rev. E **80**, 061906 (2009).

[60] N. Huang, N. K. Banavali, and A. D. MacKerell Jr., Proc. Natl. Acad. Sci. USA **100**, 68 (2003).

[61] X. H. Wang and Z. X. Sun, J. Chem. Inf. Model. **59**, 2980 (2019).

[62] R. Galindo-Murillo, J. C. Robertson, M. Zgarbová, J. Šponer, M. Otyepka, P. Jurečka, and T. E. Cheatham III, J. Chem. Theory Comput. **12**, 4114 (2016).

[63] P. D. Dans, I. Ivani, A. Hospital, G. Portella, C. González, and M. Orozco, Nucl. Acids Res. **45**, 4217 (2017).

[64] I. Ivani, P. D. Dans, A. Noy, A. Pérez, I. Faustino, A. Hospital, J. Walther, P. Andrio, R. Goñi, A. Balaceanu, G. Portella, F. Battistini, J. L. Gelpí, C. González, M. Vendruscolo, C. A. Laughton, S. A. Harris, D. A. Case, and M. Orozco, Nat. Methods **13**, 55 (2016).

[65] I. S. Joung and T. E. Cheatham III, J. Phys. Chem. B **112**, 9020 (2008).

[66] I. S. Joung and T. E. Cheatham III, J. Phys. Chem. B **113**, 13279 (2009).

[67] Z. X. Sun and J. Z. H. Zhang, CCS Chem. **3**, 1026 (2021).

[68] Z. X. Sun, X. H. Wang, J. Z. H. Zhang, and Q. L. He, Phys. Chem. Chem. Phys. **21**, 14923 (2019).

[69] X. H. Wang, X. Z. Tu, J. Z. H. Zhang, and Z. X. Sun, Phys. Chem. Chem. Phys. **20**, 2009 (2018).

[70] M. Muñoz and C. Cárdenas, Phys. Chem. Chem. Phys. **19**, 16003 (2017).

[71] F. Nerattini, R. Chelli, and P. Procacci, Phys. Chem. Chem. Phys. **18**, 15005 (2016).

[72] Z. X. Sun, X. H. Wang, and J. N. Song, J. Chem. Inf. Model. **57**, 1621 (2017).

[73] P. Mikulskis, D. Cioloboc, M. Andrejić, S. Khare, J. Brorsson, S. Genheden, R. A. Mata, P. Söderhjelm, and U. Ryde, J. Comput. Aided Mol. Des. **28**, 375

(2014).

[74] J. Michel and J. W. Essex, J. Comput. Aided Mol. Des. **24**, 639 (2010).

[75] Z. X. Sun, X. H. Wang, and J. Z. H. Zhang, Phys. Chem. Chem. Phys. **19**, 15005 (2017).

[76] Z. Huai, Z. X. Shen, and Z. X. Sun, J. Chem. Inf. Model. **61**, 284 (2021).

[77] Z. Huai, H. Y. Yang, X. Li, and Z. X. Sun, J. Comput. Aided Mol. Des. **35**, 117 (2021).

[78] A. E. Eriksson, W. A. Baase, J. A. Wozniak, and B. W. Matthews, Nature **355**, 371 (1992).

[79] A. Morton and B. W. Matthews, Biochemistry **34**, 8576 (1995).

[80] T. Steinbrecher, D. L. Mobley, and D. A. Case, J. Chem. Phys. **127**, 214108 (2007).

[81] M. Zacharias, T. P. Straatsma, and J. A. McCammon, J. Chem. Phys. **100**, 9025 (1994).

[82] T. C. Beutler, A. E. Mark, R. C. van Schaik, P. R. Gerber, and W. F. van Gunsteren, Chem. Phys. Lett. **222**, 529 (1994).

[83] D. A. Pearlman and P. A. Kollman, J. Chem. Phys. **91**, 7831 (1989).

[84] M. Levitt, J. Mol. Biol. **170**, 723 (1983).

[85] Bruckner and S. Boresch, J. Comput. Chem. **32**, 1320 (2011).

[86] H. Resat and M. Mezei, J. Chem. Phys. **99**, 6052 (1993).

[87] H. Resat and M. Mezei, J. Chem. Phys. **101**, 6126

(1994).

[88] A. P. Abbott, G. Capper, D. L. Davies, R. K. Rasheed, and V. Tambyrajah, Chem. Commun. **1**, 70 (2003).

[89] P. Kalhor and K. Ghandi, Catalysts **11**, 178 (2021).

[90] P. Kalhor and K. Ghandi, Molecules **24**, 4012 (2019).

[91] T. El Achkar, H. Greige-Gerges, and S. Fourmentin, Environ. Chem. Lett. **19**, 3397 (2021).

[92] E. L. Smith, A. P. Abbott, and K. S. Ryder, Chem. Rev. **114**, 11060 (2014).

[93] P. Kalhor, J. Xu, H. Ashraf, B. B. Cao, and Z. W. Yu, J. Phys. Chem. B **124**, 1229 (2020).

[94] P. Kalhor, Y. Z. Zheng, H. Ashraf, B. B. Cao, and Z. W. Yu, ChemPhysChem **21**, 995 (2020).

[95] K. Shahbaz, F. S. Mjalli, M. A. Hashim, and I. M. ALNashef, J. Appl. Sci. **10**, 3349 (2010).

[96] S. Y. Park and J. R. H. Tame, Biophys. Rev. **9**, 169 (2017).

[97] J. J. Goings and S. Hammes-Schiffer, J. Am. Chem. Soc. **141**, 20470 (2019).

[98] A. Losi and W. Gärtner, Photochem. Photobiol. **93**, 141 (2017).

[99] J. J. Goings, P. F. Li, Q. W. Zhu, and S. Hammes-Schiffer, Proc. Natl. Acad. Sci. USA **117**, 26626 (2020).

[100] S. Anderson, V. Dragnea, S. Masuda, J. Ybe, K. Moffat, and C. Bauer, Biochemistry **44**, 7998 (2005).