

A Benchmark Test on the Leapfrog Integrator and its Middle Alternative

Zhaoxi Sun^{1*}

¹Beijing National Laboratory for Molecular Sciences, College of Chemistry and Molecular Engineering, Institute of Theoretical and Computational Chemistry, College of Chemistry and Molecular Engineering, Peking University, Beijing 100871, China

*To whom correspondence should be addressed: z.sun@pku.edu.cn

Testing the free energy simulation results when using different integrators for the equations of motion: A comparison between the leapfrog (LF) integrator and the leapfrog-middle (LF-middle) scheme.

Benchmark tests are performed on three systems. In the first two systems including the NMA system (ACE-NME) and an AT tract (A₇-T₇), the conformational change of dihedral flipping is simulated. In the last system malonaldehyde, the intramolecular proton transfer is investigated. The LF integrator is the default option used in various molecular dynamics (MD) packages including AMBER¹. We used Langevin dynamics for temperature regulation. In this case, the integrator widely accepted to have better stability and efficiency is the BAOAB method,²⁻⁴ which could be generalized to other thermostats and named as the middle scheme.⁵

⁶ The algorithm is named as LF-middle when the method is coupled to the LF integrator. The simulation is performed at 300 K with a 5 ps⁻¹ friction coefficient for the thermostat. We used isotropic position scaling and the Berendsen barostat to regulate the pressure in NPT simulations. The SHAKE algorithm^{7,8} is an influencing factor of the simulation speed and the time step acceptable. In the current test, the simulation is performed with and without the SHAKE constraints. The SHAKE constraint constrains only the bonds involving hydrogen.

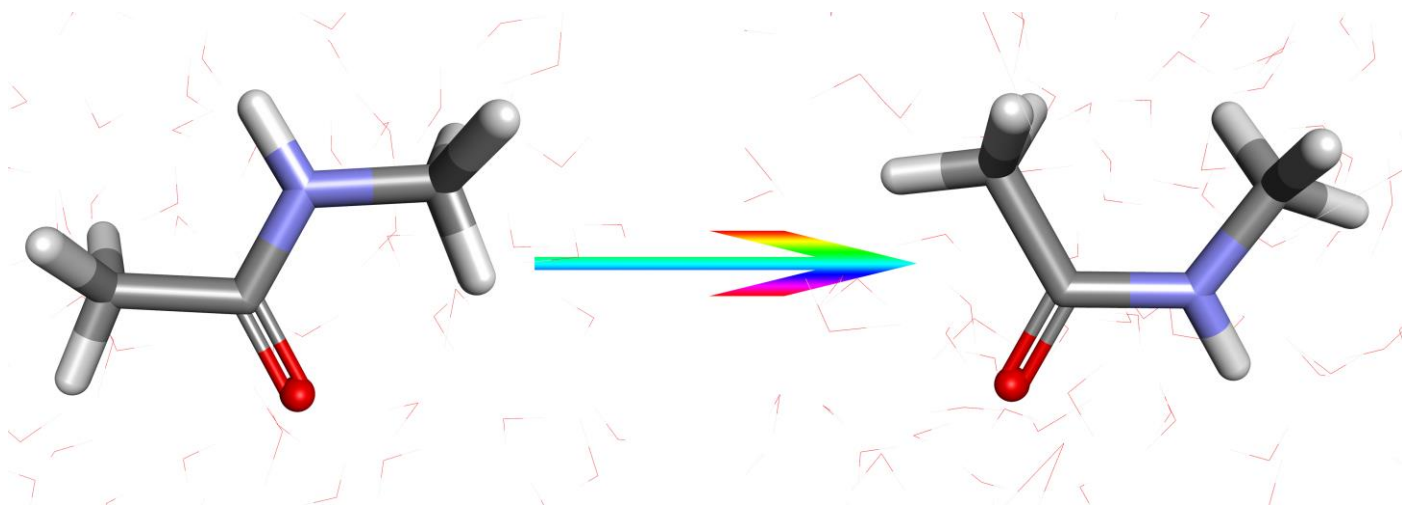


Fig. 1. Model system (ACE-NME). The backbone C-C-N-C dihedral is used as the reaction coordinate describing the dihedral flipping. In multiscale quantum mechanics (QM)/ molecular mechanics (MM) simulations, the atoms depicted with the stick type are described with QM Hamiltonians, while the line-type atoms are described with MM potentials. Namely, the whole solute is included in the QM region, while the water molecules are represented by the MM force field.

1. ACE-NME with all-atom force field

We use the model system (ACE-NME) shown in Fig. 1 as an example to test the time step and integrator dependence of the free energy profile. The NMA system is biologically relevant and simple enough to achieve a sufficient level of sampling convergence with moderate amounts of computational resources.⁹⁻¹¹ The small two-residue peptide is described with AMBER99SB¹² and solvated in TIP3P water^{13, 14} in an octahedron box of 2814 atoms (934 water molecules). The cutoff for non-bonded interactions in the real space is set to 8 Å (AMBER default) and the long-range electrostatics are treated with the PME method.¹⁵ Therefore, there is no finite-size effect influencing the simulation outcome. The umbrella sampling¹⁶⁻¹⁸ method is used to enhance the sampling efficiency due to its simplicity and accuracy in producing highly accurate estimates of the free energy results. The parameters for the umbrella sampling simulation are 72 windows equally spaced between 0 ° and 360 ° and a force constant of 100 kcal/(mol·rad²), which has been successfully employed in various dihedral flipping cases.¹⁹⁻²⁴ The sampling interval of the backbone dihedral is 1 ps, which is longer than the autocorrelation time according to our previous calculations.^{9, 11} Potential of mean force (PMF) are obtained from the perturbation-based variational free energy profile (vFEP)²⁵ reweighting.

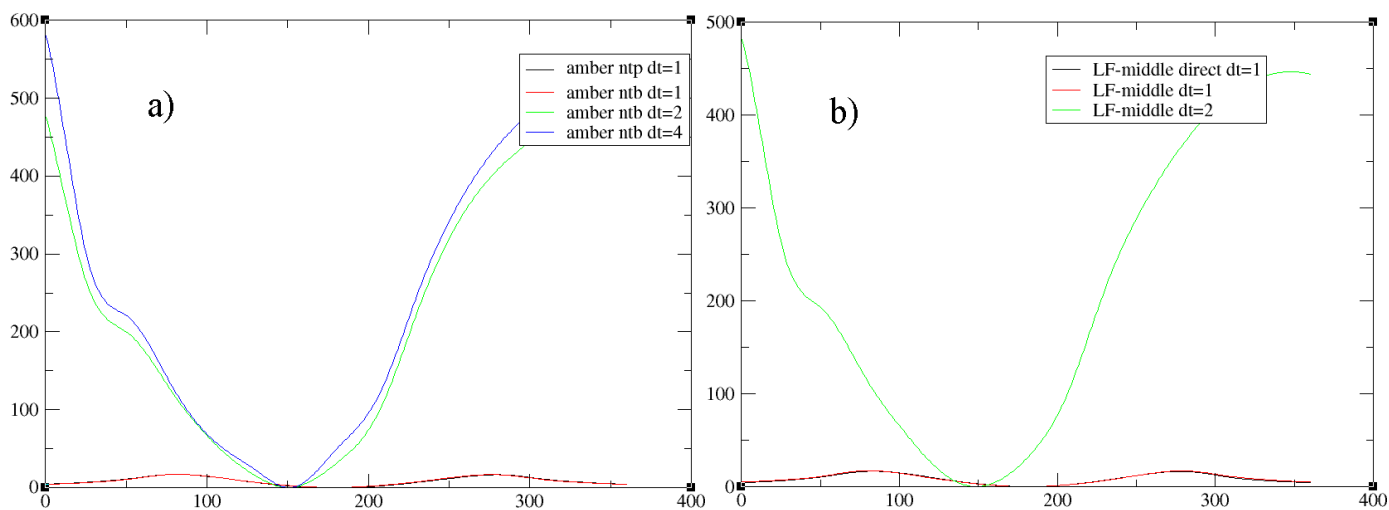


Fig. 2. Free energy profiles in kcal/mol along the dihedral flipping pathway obtained from different simulation protocols with a) the LF integrator and b) the LF-middle scheme without the SHAKE constraint. The time step used is represented by $dt=n$ in the legend. NPT and NVT simulations are represented by ntp and ntb keywords, respectively. When the LF-middle integrator is employed, only NVT simulations are performed. The word ‘direct’ in the legend means that the simulation initiates from newly generated configurations under the target Hamiltonian (i.e., nonequilibrium pulling and then equilibration), while the other simulations are initiated from equilibrated structures from ‘amber ntp $dt=1$ ’.

1.1 No SHAKE.

We first check the results obtained without SHAKE. In each window, 100 ps simulation is performed and the sampling interval is 1 ps. Therefore, there are 100 points in each window. Simulation speeds with the LF integrator and the LF-middle scheme are almost identical, as the force-evaluation step is the most time-consuming calculation limiting the simulation speed.

As the LF integrator is the default option in AMBER, we use amber in the legend to describe the LF results. Also, as the ntb flag is used to define the periodic boundaries and the ntp flag is for pressure regulation, we use them to represent NVT and NPT ensembles, respectively. The PMFs obtained with the LF integrator are shown in Fig. 2a. Simulations with time steps larger than 1 fs (i.e. 2 fs and 4 fs) collapse due to particle exceeding the velocity limit in the equilibration step (i.e. the system blows up). The free energy profiles obtained from 1 fs time step in NVT and NPT simulations agree with each other, but the results obtained with larger time steps are completely wrong. Thus, without SHAKE, with the LF integrator, the time step should not be larger than 1 fs.

Then, we check the results obtained with the LF-middle integrator. The direct simulation from a single structure and those initiated from the equilibrated structure in previous NPT simulations are performed, and the results are shown in Fig. 2b. Obviously, with the LF-middle scheme, without the SHAKE constraints, the time step still cannot be larger than 1 fs.

We compare the PMFs obtained with the LF and LF-middle integrators in Fig. 3. We notice that the PMFs are similar, and the fluctuations originate from the fluctuation in the statistics and insufficient sampling, which is not the focus of the current research. Thus, we conclude that without SHAKE, for both integrators, the time step should be no larger than 1 fs. In another systematic test presented in the following part, we observe that 1.5 fs is also usable for both integrators, but normally we would like to use time steps as integers (e.g., 1 fs or 2 fs).

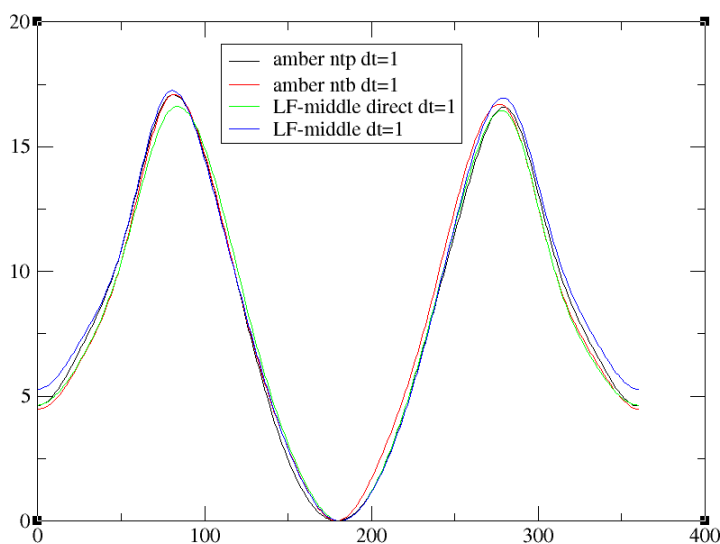


Fig. 3. Comparison between the free energy profiles (kcal/mol) obtained with the LF and LF-middle integrators without the SHAKE constraint. The time step is represented by $dt=n$ in the legend. NPT and NVT simulations are represented by ntp and ntb keywords, respectively. When the LF-middle integrator is employed, only NVT simulations are performed. The word ‘direct’ in the legend means that the simulation initiates from newly generated configurations under the target Hamiltonian, while the other simulations are initiated from equilibrated structures from ‘amber ntp $dt=1$ ’.

1.2 With SHAKE.

In practical MD simulation for complex systems, the phenomena of interest are relevant to large-scale motions of biomolecules, which are often decoupled from the high-frequency bond-stretching movement. Therefore, the SHAKE constraint is favored to enable the use of a larger time step. We then tested the performance with SHAKE to check how long the time step could be in practical simulations. In each window, 1 ns simulation is performed and the sampling interval is 1 ps. Therefore, there are 1000 points in each window.

With the LF integrator, the PMFs from NVT and NPT simulations and different time steps shown in Fig. 4a are very similar, which indicates that the NVT and NPT conditions do not influence the simulation results, and the time step as long as 4 fs could be used with the SHAKE constraints. The results obtained with the LF-middle integrator with different time steps shown in Fig. 4b are also similar, which indicates that the simulation

results with the LF-middle integrator are also independent of the time step. For both integrators, the simulation collapses when dt is set to 8 fs, which indicates that the LF-middle integrator does not really touch the available magnitude of the time step.

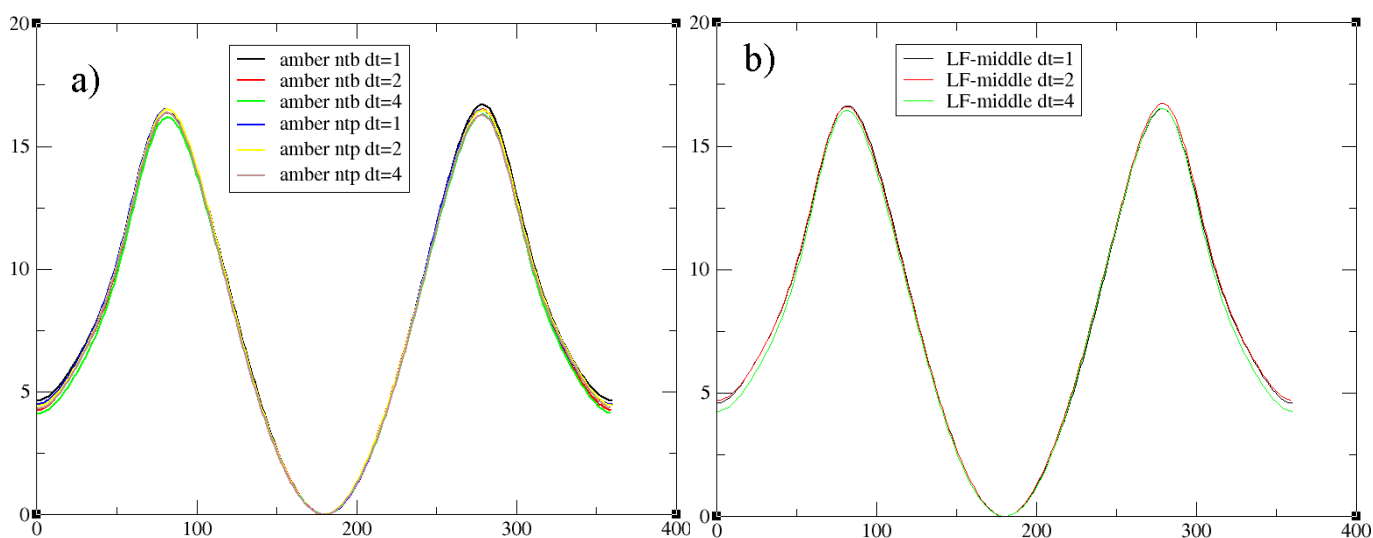


Fig. 4. Free energy profiles in kcal/mol along the dihedral flipping pathway obtained from different simulation protocols with a) the LF integrator and b) the LF-middle scheme with the SHAKE constraint. The time step used is represented by $dt=n$ in the legend. NPT and NVT simulations are represented by ntp and ntb keywords in the legend, respectively. When the LF-middle integrator is employed, only NVT simulations are performed.

We compare the free energy profiles obtained with the LF and LF-middle integrators at different time steps in Fig. 5a and 5b. In Fig. 5a, only the NVT results are shown, as the PMFs from NVT and NPT simulations are extremely similar. We can see that there is no significant difference between the LF PMFs and the LF-middle results under different time steps, which indicates that the LF integrator is already accurate enough and the LF-middle scheme does not really touch the thermodynamic profile. In Fig. 5b, PMFs from all protocols are compared. We can see that the PMFs obtained from all protocols with SHAKE constraints are very similar, which again indicates that the LF-middle scheme neither change the thermodynamics, nor change the acceptable time step.

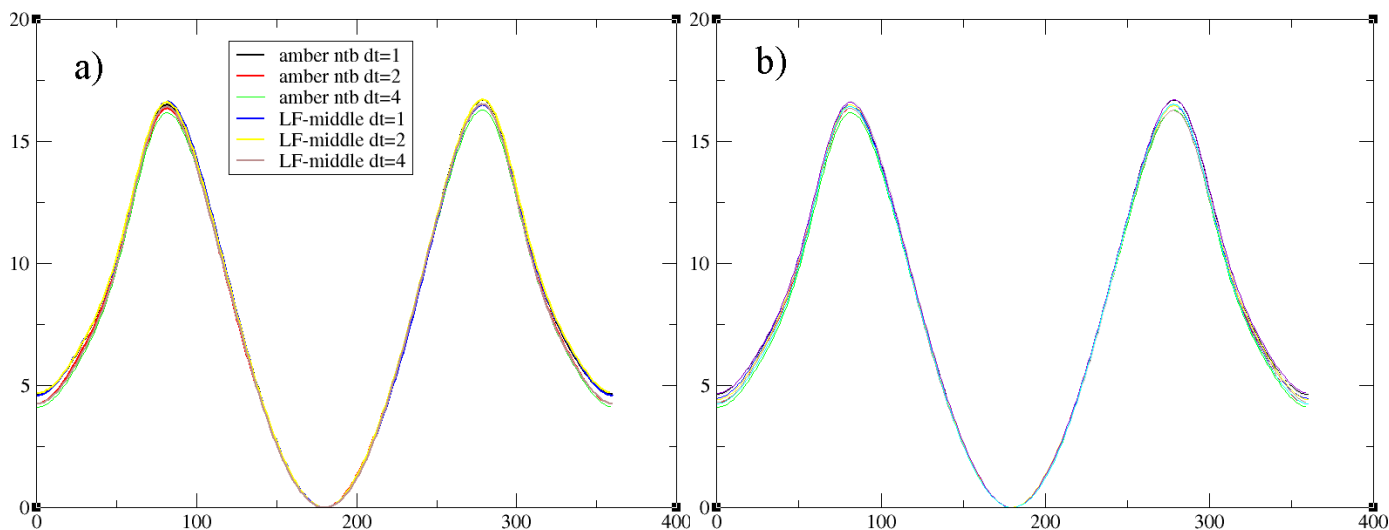


Fig. 5. Comparison between the free energy profiles with the LF integrator and LF-middle with the SHAKE constraints. In a), only results from LF NVT simulations and LF-middle NVT simulations are provided, as there is no obvious difference between the NVT and NPT results, while in b) results from all protocols are compared. The time step used is represented by $dt=n$ in the legend. NVT simulations are represented by the `ntb` keyword in the legend.

Another comparison performed but not shown here is the comparison between the SHAKE-on and SHAKE-off PMFs. The PMFs are very similar, which indicates that the SHAKE option could be preferably on when dealing with dihedral flipping.

We performed longer simulations (6 ns in each window) to achieve a better level of convergence and the results are summarized in Fig. 6. A quick visual inspection of the free energy profiles obtained with the two integrators at different time steps does not really give hints on their difference, which again indicates that the time-step-dependence is minimal. This time-step-induced systematic error is quantified to be about 0.1 kcal/mol for large time steps such as 4 fs with the SHAKE constraint, but the time-step-related error is negligible with the recommended value for time steps (e.g., 2 fs with the SHAKE constraint). This shifted free energy difference relates to the perturbation of the distributions in the simulated ensemble. In the appendix, we performed some analyses on the potential energy terms and related distributions to investigate the contributions from various interaction terms. A main observation from the energy comparison is that the simulation is actually exploring the perturbed ensemble, thus leading to perturbed/shifted results. The perturbed distribution could be corrected by reweighting with the shadow work performed during the system-thermostat coupling,²⁶⁻²⁸ but with the recommended value of time steps (i.e., 1 fs without SHAKE and 2 fs with SHAKE) this seems unnecessary, as the statistical error for complex systems ($\sim k_B T$) is much larger.

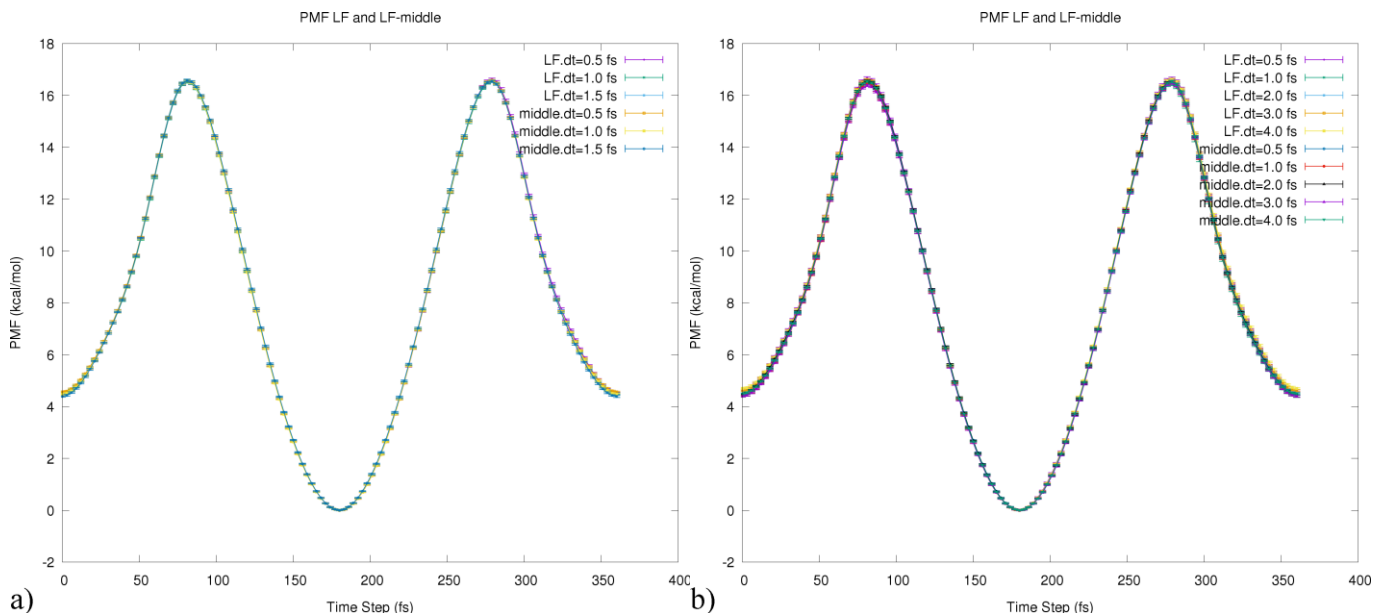


Fig. 6. The free energy profile along the flipping dihedral obtained with the LF and LF-middle integrators at different time steps.

2. ACE-NME with the semi-empirical QM Hamiltonian Parametrized Model number 6 (PM6)

The simulations in section 1 are repeated with the semi-empirical QM Hamiltonian PM6²⁹ to check the Hamiltonian-dependence of the time step (and the effect of the multi-scale treatment). The solute ACE-NME is described with PM6 while the TIP3P water is described with MM. The QM region does not use SHAKE in any simulations, while the SHAKE option could be on or off for MM regions. The results are similar to those in section 1. As the QM region does not use SHAKE under any circumstances, there are some differences but the main conclusions remain unchanged. Thus, we just briefly discuss them.

2.1 No SHAKE.

In Fig. 7, the comparison between PMFs obtained with different integrators in NVT and NPT simulations without the SHAKE constraints is given. Only 1 fs time step is acceptable and longer time steps result in system collapse. The PMFs obtained from different simulation protocols are very similar.

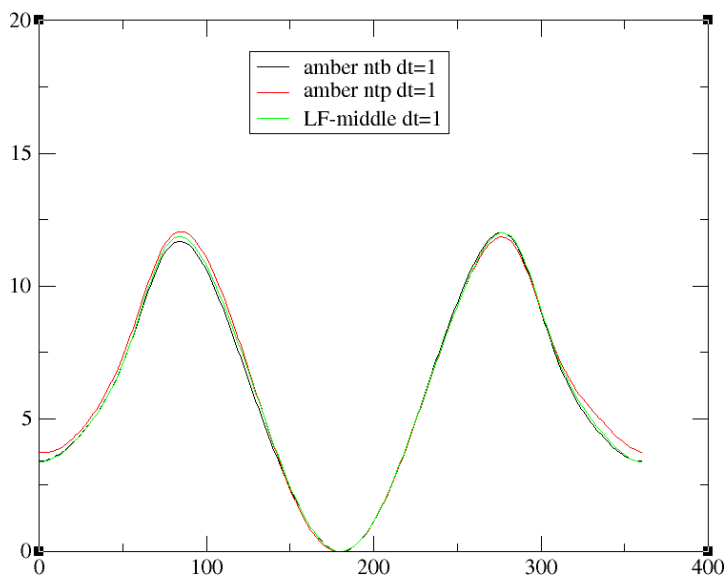


Fig. 7. Under the PM6 Hamiltonian, the comparison between the free energy profiles with the LF and LF-middle integrators without the SHAKE constraints. The time step used is represented by $dt=n$ in the legend. NPT and NVT simulations are represented by ntp and ntb keywords in the legend, respectively. When the LF-middle integrator is employed, only NVT simulations are performed.

2.2 With SHAKE.

We then apply the SHAKE constraints on MM molecules (i.e., water), while the QM region does not include constraints. This leads to a little different behaviors of the system, compared with the previous SHAKE-off section. Specifically, as the time step acceptable for the solute differs from that for the solvent, the minimum of them determines the largest time step usable. The simulations with 1 fs and 2 fs time steps could be run, while 4 fs time step is too large for stable dynamics. Comparing this phenomenon with the previous section, we know that without SHAKE, acceptable time steps for the solute could be as long as 2 fs and 4 fs is a bit large, while the time step for water molecules must be as small as 1 fs. Note that in this multi-scale case, the multiple time step (MTS)³⁰⁻³² treatment of the time evolution of different degrees of freedom (DOFs) could be useful to accelerate the simulation. In each window, 1 ns simulation is performed and the sampling interval is 1 ps. Therefore, there are 1000 points in each window. The PMFs (not shown) are similar for different simulation protocols and thus we do not discuss the results further.

3. AT tracts: A₇-T₇

Genetic codes are deposited in biopolymers as the sequence of nucleotides, specifically in the base pairs.³³ As the most widely observed form of DNA in cell, the duplex structural motif features the inner hydrogen bond interactions connecting the two strands and the outer phosphate and 5-membered sugar rings.³⁴⁻³⁷ The functionality of the DNA systems could be activated by the base flipping event, where the base-pair hydrogen bonds are broken and the base flips outward to be solvent-exposed. As a result, the bases are made accessible to external agents, which enables the occurrence of crucial biological processes such as repair, replication, transcription and recombination,³⁸⁻⁴³ DNA methylation,⁴⁴⁻⁴⁷ melting,⁴⁸⁻⁵⁷ bubbling,^{58, 59} and protein-DNA binding.^{33, 60-65} The model system A₇-T₇ shown in Fig. 8 is a short DNA duplex with sufficient biological relevance and complexity to test the time step dependence of the two integrators. We use the bsc1⁶⁶ force field to describe the DNA duplex. The system is solvated with TIP3P water molecules and Na⁺ cations^{67,68} are added for neutralization. The resulting system contains 10511 particles. The octahedron box is replicated with periodic boundary conditions. The length of each dimension is about 51 Å.

The pseudo-dihedral defined by four centers of mass (COM) is chosen as the collective variable (CV) describing the base flipping event, which has been successfully applied in various computational studies.¹⁹⁻²⁴ The four groups include the flipping base of A4, the sugar moiety of A4, the sugar moiety of A5 (the 3' side of A4) and the bases of the base pair A5-T10, as shown at the bottom of Fig. 8. The window spacing and force constant parameters are the same as the previous backbone dihedral case. The starting configurations are obtained from our previous works on the AT tracts (the last configurations in each umbrella window under the bsc1 force field).¹⁹ Therefore, no long equilibration is performed. In each window, 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.^{19, 21, 22} Therefore, there are 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.¹⁹ PMFs are obtained from vFEP reweighting.

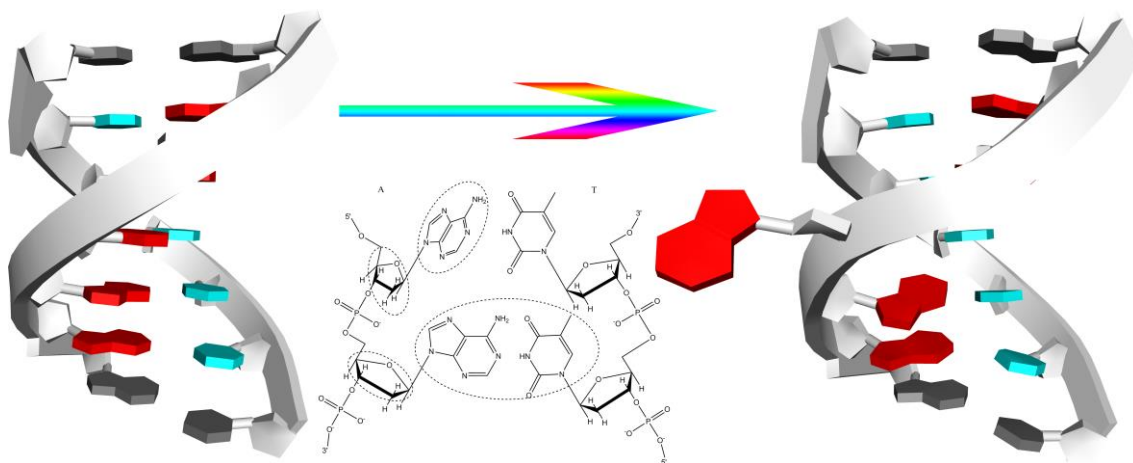


Fig. 8. Illustration of base flipping in the A₇-T₇ duplex. The COM dihedral is used as the reaction coordinate.

3.1. No SHAKE.

Without SHAKE, the bonds are not constrained and all degrees of freedom could be freely relaxed. A time step of 1 fs is used to propagate the dynamics, while larger time steps result in the termination of simulation. In Fig. 9, we compare the PMFs obtained with and without the SHAKE constraints. 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. The results obtained from different integrators (i.e., LF and LF-middle) are also similar, which indicates that the simulation result is insensitive to the integrator used. It is worth noting that although the LF-middle scheme is shown to have higher accuracy in integrating the dynamics, it helps nothing in the size of time steps when the SHAKE option is off. Therefore, using LF-middle in simulations without SHAKE seems unnecessary.

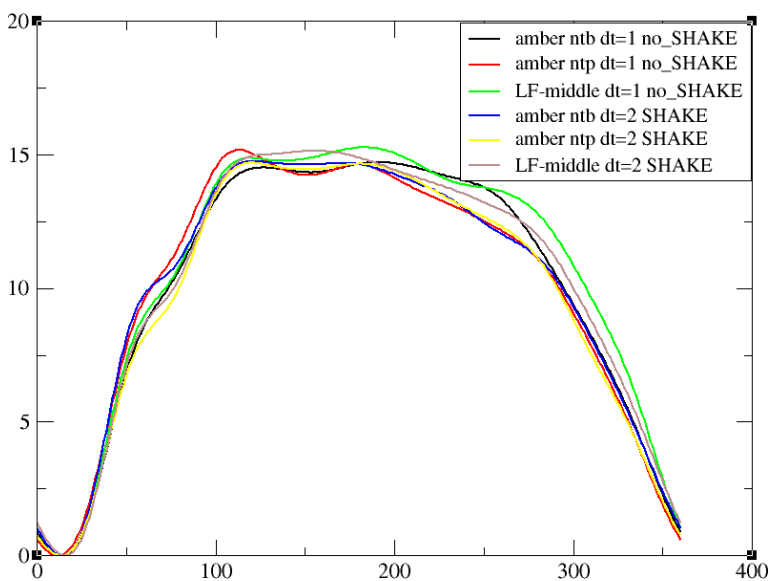


Fig. 9. Free energy profiles in base flipping with and without the SHAKE constraints. For simulation without SHAKE, only 1 fs time step is usable, while larger time steps result in unstable dynamics.

3.2. With SHAKE.

We then turn the SHAKE option on and check the simulation results below.

In Fig. 10a, we compare the 2 fs and 4 fs results obtained with the LF and LF-middle integrators. For the LF integrator, 2 fs results with different ensembles are very similar, while the 4 fs time step in NVT and NPT ensembles provide wrong results, deviating from the true expectation. The simulations in many windows stop due to unstable dynamics. Therefore, for the LF integrator, 2 fs should be used in base flipping simulations. For the LF-middle integrator, 2 fs and 4 fs results are very similar and are consistent with the LF results. However, the simulations with the LF-middle integrator and 4 fs time step sometimes crash due to unstable dynamics. This phenomenon indicates that the simulation could be run under such conditions, but the algorithm is not stable in some phase space regions explored. Therefore, for either integrator, the 2 fs time step is recommended for the stable propagation of dynamics.

In Fig. 10b, we compare the results obtained with the LF-middle integrator and 1 fs time step with those obtained with 2 fs time step and different ensembles. The results are similar, which indicates that the 2 fs time step could be safely used without a sacrifice of accuracy, compared with smaller time steps.

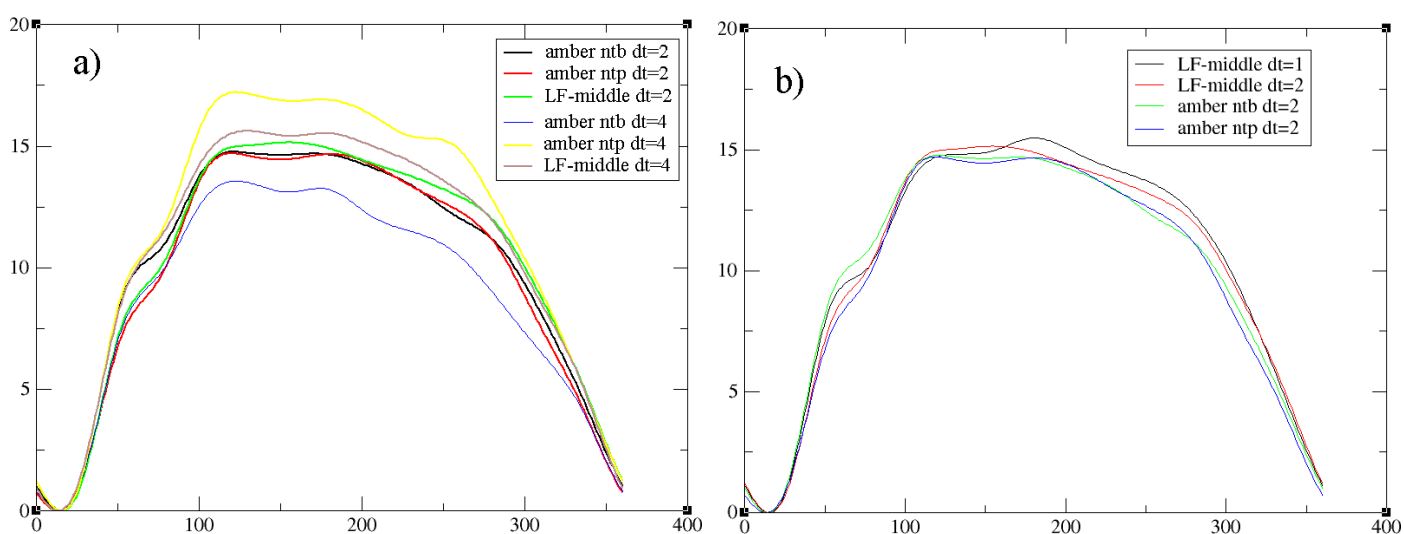


Fig. 10. Free energy profiles along the base flipping pathway obtained with the SHAKE constraints. a) Comparison between 2 fs and 4 fs results with different integrators in different ensembles. b) Comparison between 1 fs and 2 fs time steps obtained from different integrators.

4. Intramolecular proton transfer in malonaldehyde in vacuo

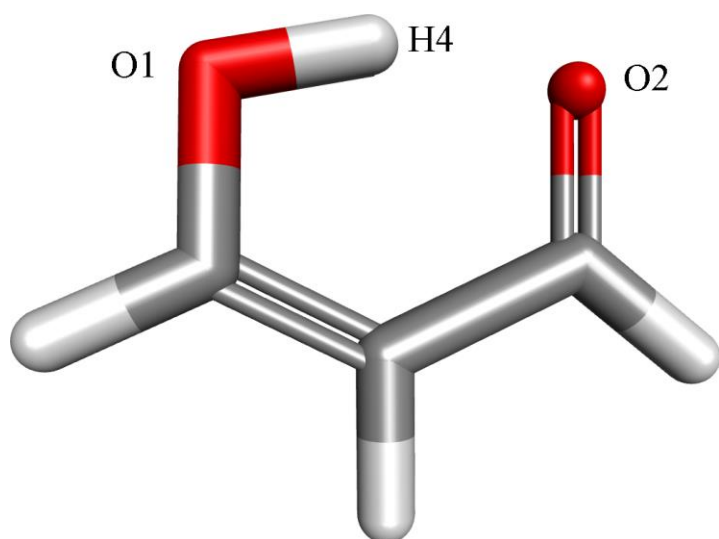


Fig. 11. Illustration of the intramolecular proton transfer case malonaldehyde.

The intramolecular proton transfer case shown in Fig. 11 serves as a good test case for molecular simulation techniques. The intramolecular fluctuation of the proton between the two oxygen atoms is an essentially QM effect, which needs to be treated with *ab initio* or semi-empirical QM methods. Here, due to efficiency consideration, the electronic structure calculation is performed at the semi-empirical PM6²⁹ level and no solvent molecule is added. Two distances including $d(\text{O1-H4})$ and $d(\text{O2-H4})$ are involved in the bond rearrangement, and the one-dimensional CV defined by the difference between these two distances could be used. However, as our aim here is to provide an example for high-dimensional complex processes, to make our simulation protocol more generally applicable to more complex systems, both of the two distances are chosen as the CVs to bias the simulation. Thus, two-dimensional enhanced sampling simulations are performed in this case. As the proton transfer involves the rearrangement of bonds involving hydrogen atoms, the SHAKE constraint is not used. Umbrella sampling is still used to bias the sampling. As for the window spacing scheme, we put windows from 0.8 Å to 1.8 Å with 0.1 Å increments in each dimension, and a force constant of 100 kcal/(mol*Å²) is employed. The sampling time in each window is 900 ps and the sampling interval is set to 0.12 ps. The reweighting procedure is performed with the weighted histogram analysis method (WHAM).^{69, 70} The simulation is performed in vacuo at 300 K.

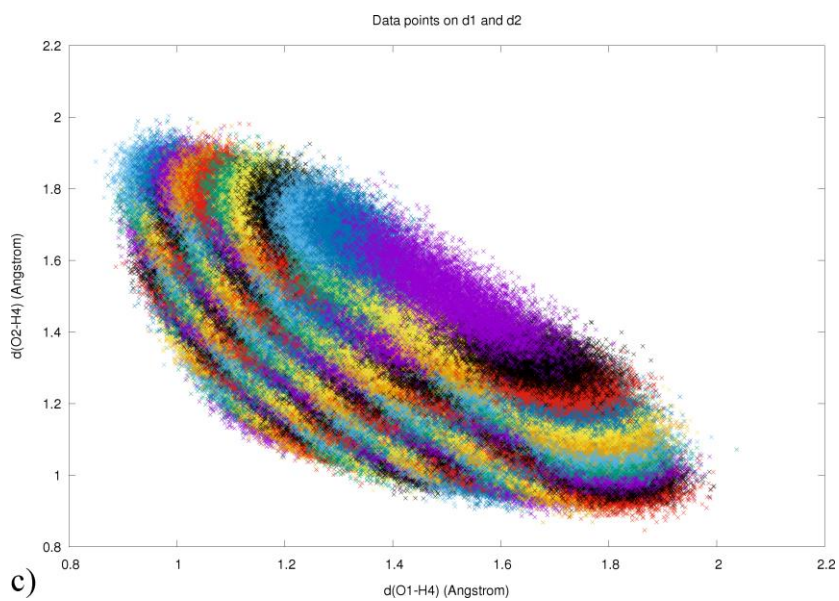
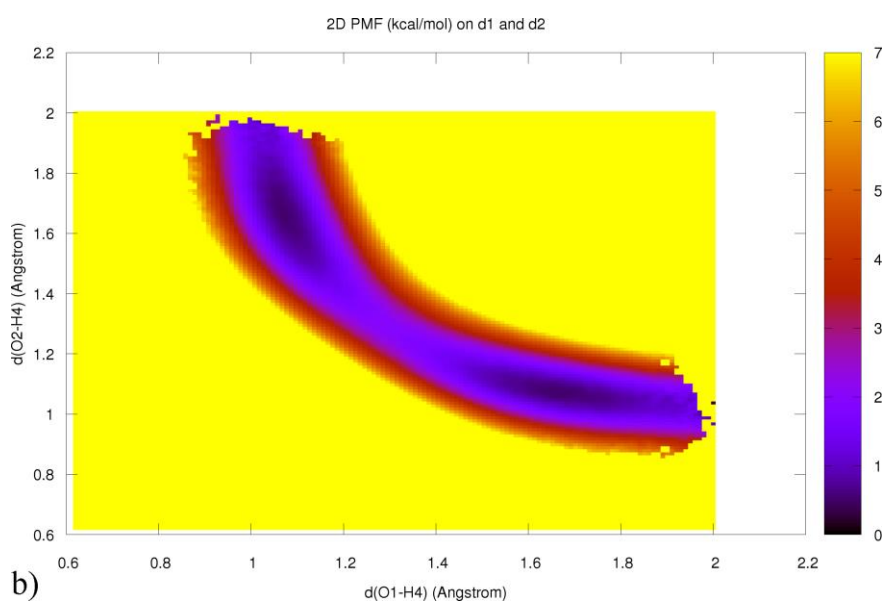
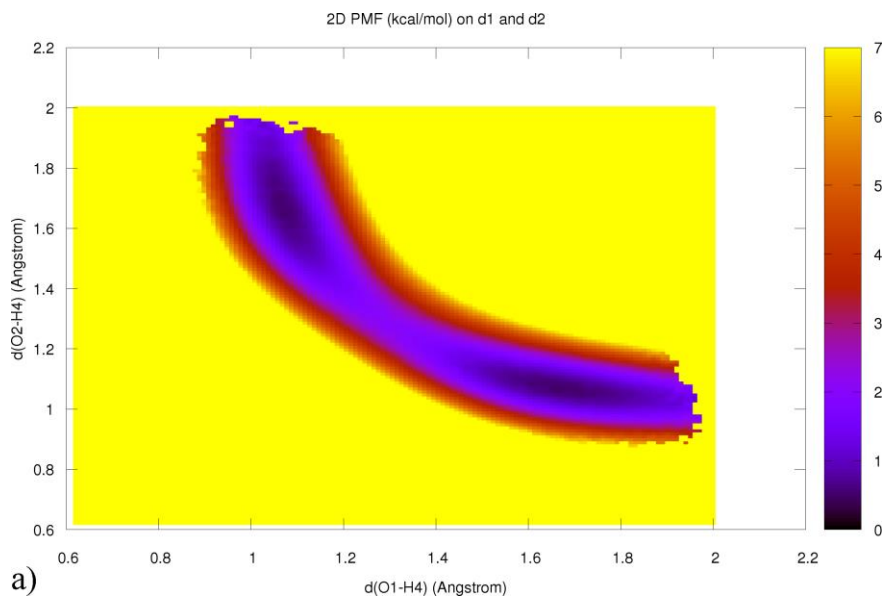


Fig. 12. With 1 fs time step, a) the two-dimensional free energy surface on $d(\text{O1-H4})$ and $d(\text{O2-H4})$ obtained with the LF integrator and b) that obtained with the LF-middle integrator. c) The data points sampled from all umbrella windows.

In Fig. 12, the 2D PMFs projected on the two distance CVs obtained with 1 fs time step and different integrators are shown. As the two ends of the proton transfer reaction are chemically equivalent, their thermodynamic stabilities are identical. A quick visual inspection of the free energy landscapes again indicates that the thermodynamics obtained with different integrators are the same. When checking the statistical data, we locate the free energy minima at (1.061773, 1.707163) and (1.707163, 1.061773). The saddle point satisfies the relationship $y=x$, i.e. $d(\text{O1-H4})=d(\text{O2-H4})$, and is found to be the point (1.290142, 1.290142). Due to the data binning in reweighting, the exact position of free energy minima and the saddle point could differ from the above positions, but the magnitude would be minimal. The data points sampled from all umbrella windows are scattered also on the $d(\text{O1-H4}) - d(\text{O2-H4})$ surface in Fig. 12c, from which we can see that the sampling has covered all of the important regions for the intramolecular proton transfer.

The free energy difference between the global minimum and the saddle point (i.e., the free energy barrier of the proton transfer reaction) could be calculated to quantify this time-step-dependence. A summary of the results is that the time-step-induced perturbation of the free energy barrier is minimal for smaller time steps (e.g., 0.5 fs and 1 fs), but could be statistically non-negligible for larger time steps (e.g., ~ 0.1 kcal/mol for 2 fs), but is still much smaller than the normal range of statistical errors from sampling the configurational space in complex systems (~ 0.6 kcal/mol).

A summary of the conclusions for the time step usable.

We then summarize the conclusion for the choice of time steps used in MD simulations.

In Table 1, the maximum acceptable time steps for the test systems are shown. A general conclusion for simulation without the SHAKE constraints is that a time step of 1 fs should be used. Larger values would result in unstable dynamics.

For simulations with the SHAKE constraints, for the simple ACE-NME system, 4 fs is usable for MM-only simulations, while 2 fs could be used for multiscale QM/MM simulations. In practical QM/MM simulations for complex systems, it is still recommended to use the 1 fs time step, as the SHAKE constraint is often not used for the QM region. For the DNA AT-tract, with the SHAKE constraints, 2 fs time step is usable for the LF integrator and 4 fs is usable for the LF-middle integrator. However, the dynamics with LF-middle and 4 fs time step is not as stable as those with LF and 2 fs time step. Therefore, it is still recommended using 2 fs time step. The choice of integrator seems not a key problem here and can be determined according to personal preference.

Table 1. Acceptable time steps in MD simulations for the test systems.

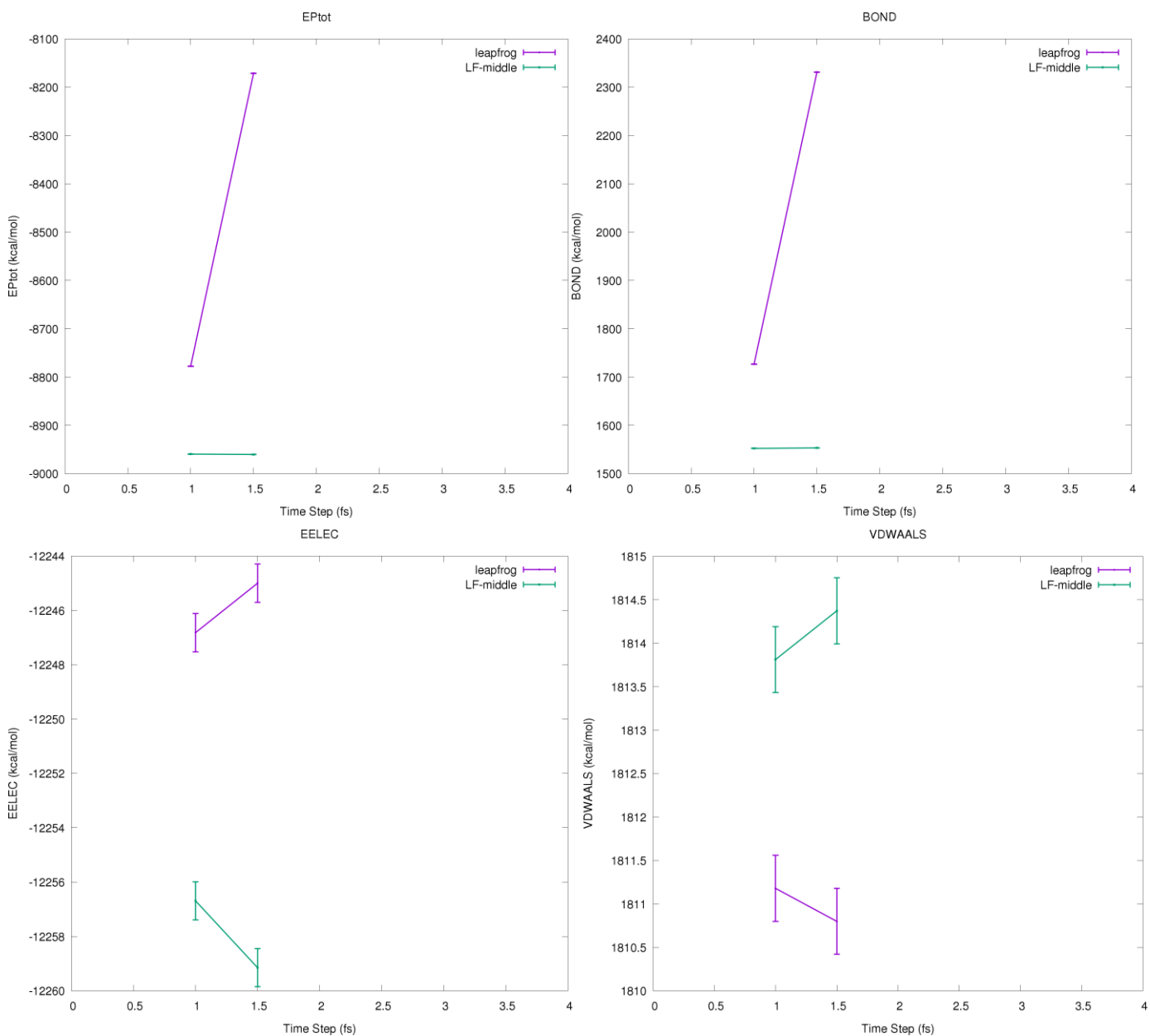
System	SHAKE on hydrogen	Integrator	dt_max
ACE-NME-MM	off	LF	1
		LF-middle	1
	on	LF	4
		LF-middle	4
ACE-NME-PM6/MM	off	LF	1
		LF-middle	1
	on	LF	2
		LF-middle	2
DNA7-bsc1	off	LF	1
		LF-middle	1
	on	LF	2
		LF-middle	4 but still recommend 2

Appendix. A test of the time-step-dependence of the potential energy with the LF and LF-middle integrators for the NMA system.

The increase of the time step leads to perturbations of the energy distributions of the system. As the total energy is decomposable, we then investigate the time-step dependence of different energy components. We performed unbiased simulations for the NMA system with the MM force field and the PM6/MM Hamiltonian (i.e., the multi-scale QM/MM simulation) without the SHAKE constraint to investigate the contributions of various energy terms. In the MM-only simulations, the potential energy could be decomposed into the bond, angle, dihedral, electrostatic and vdW terms. By contrast, in the QM/MM case, the potential energy of the QM region is extracted as a new term named PM6ESCF, which gives the potential energy of the solute. As the TIP3P model is treated as rigid water with only bond-stretching simulated explicitly in AMBER, solvent molecules do not have contributions to the angle and dihedral terms. Note that although these two terms are zero for water molecules, the values of the H-O-H angle at different time steps are similar. These two terms of the solute are included in the QM potential of PM6ESCF. As a result, these two stiff DOFs are both zeros in the QM/MM case. Therefore, under the PM6/MM Hamiltonian, the total potential energy could be decomposed into water bond-stretching, PM6 energy function, and intermolecular electrostatic and vdW interactions, as presented in Fig. A1. We can see that the energy of the solute changes very small with the increase of the time step, the contributions from the intermolecular interactions to the change of the total energy are also small, and the largest contribution to the increase of the total potential energy comes from the intramolecular bond-stretching in water. Due to the large number of solvent molecules, the increase of the bond term in each water molecule is accumulated to be a large number, which leads to the highly time-step-dependent behavior of the bond-stretching potential energy. Aside from the mean of various potential energy terms, a more straightforward view of the time-step-induced perturbation is the distributions of the total potential energy. As shown in the last subplot of Fig. A1, the time-step perturbation could be very large for the distribution of the potential energy when using large time steps (e.g., 1.5 fs in the absence of the SHAKE constraint). The sampling results from these perturbed distributions are thus dependent of the magnitude of the time step. The LF-middle integrator gives much better stability of the potential energy on the variation of the time step.

We performed the unbiased simulations with the SHAKE constraint for the NMA system to mimic the system setup in practical biomolecular simulations. We investigate the behavior of various observables in the MM-only system. Here, we choose the backbone dihedral, the electrostatic and vdW components of solute-solvent interaction energy, and the electrostatic interaction energy between one water molecule and the other particles in the system. In Fig. A2, the distributions of various observables are similar for different setups, i.e., integrator and time step. Therefore, the time-step-induced perturbation is small when investigating the local

interaction patterns, and the merit of using the LF-middle scheme could be limited. This is in agreement with our test on the free energy profile, where the relative probabilities of different states show negligible variations upon the increase of the time step.



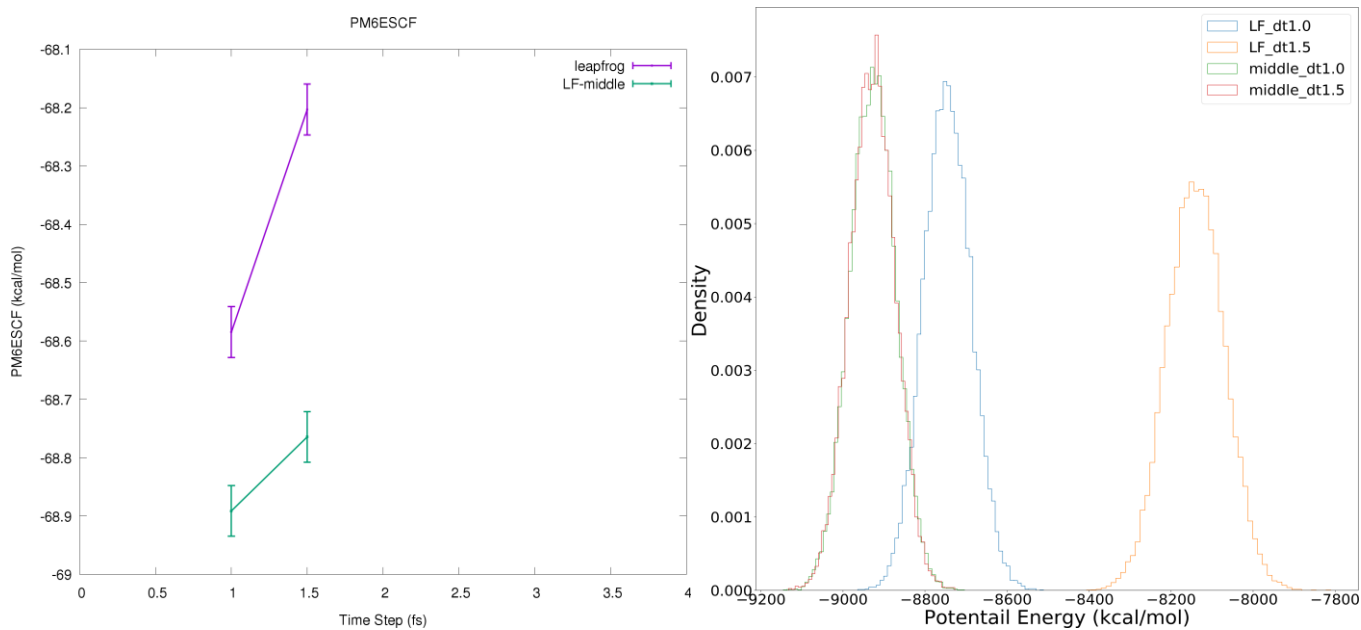


Fig. A1. Under the PM6/MM Hamiltonian and without the SHAKE constraint, the time-averaged potential energies and its various compositions including the electrostatic, vdW and PM6ESCF terms and the distribution of the total potential energy obtained with LF and LF-middle integrators at different time steps.

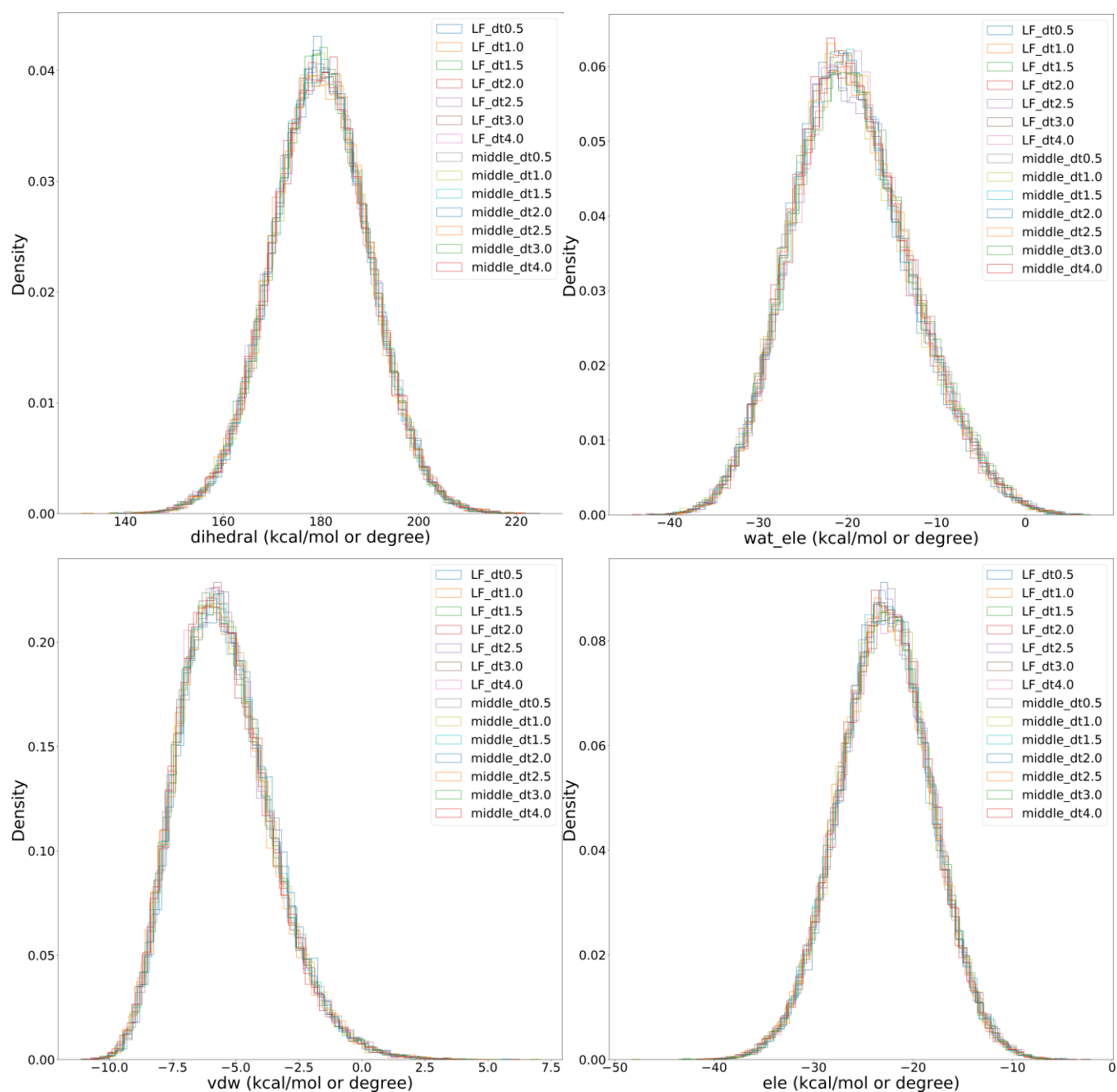


Fig. A2. Under the MM Hamiltonian and with the SHAKE constraint, the distributions of the backbone dihedral, the electrostatic interaction energy between one water molecule and the other in the system, and the electrostatic and vdW components of solute-solvent interaction energy obtained with LF and LF-middle integrators at different time steps.

Version Details

The current note is an altered version of the part A of the technical notes VI. The full manuscript of the technical note could be accessed in personal communication.

Reference

1. Case, D. A.; Cheatham, T. E.; Tom, D.; Holger, G.; Luo, R.; Merz, K. M.; Alexey, O.; Carlos, S.; Bing, W.; Woods, R. J., The Amber Biomolecular Simulation Programs. *J. Comput. Chem.* **2005**, *26*, 1668-1688.
2. Leimkuhler, B.; Matthews, C., Rational construction of stochastic numerical methods for molecular sampling. *Applied Mathematics Research Express* **2012**, *2013*, 34-56.
3. Leimkuhler, B.; Matthews, C., Robust and efficient configurational molecular sampling via Langevin dynamics. *J. Chem. Phys.* **2013**, *138*, 05B601_1.
4. Grønbech-Jensen, N.; Farago, O., A simple and effective Verlet-type algorithm for simulating Langevin dynamics. *Molecular Physics* **2013**, *111*, 983-991.
5. Zhang, Z.; Liu, X.; Yan, K.; Tuckerman, M. E.; Liu, J., A Unified Efficient Thermostat Scheme for the Canonical Ensemble with Holonomic or Isokinetic Constraints via Molecular Dynamics. *J. Phys. Chem. A* **2019**.
6. Zhang, Z.; Liu, X.; Chen, Z.; Zheng, H.; Yan, K.; Liu, J., A unified thermostat scheme for efficient configurational sampling for classical/quantum canonical ensembles via molecular dynamics. *J. Chem. Phys.* **2017**, *147*, 034109.
7. Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C., Numerical Integration of The Cartesian Equations of Motion of A System with Constraints: Molecular Dynamics of n -alkanes. *J. Comput. Phys.* **1977**, *23*, 327-341.
8. Miyamoto, S.; Kollman, P. A., Settle: An Analytical Version of The SHAKE and RATTLE Algorithm for Rigid Water Models. *J. Comput. Chem.* **1992**, *13*, 952-962.
9. Wang, X.; He, Q.; Sun, Z., BAR-Based Multi-Dimensional Nonequilibrium Pulling for Indirect Construction of a QM/MM Free Energy Landscape. *Phys. Chem. Chem. Phys.* **2019**, *21*, 6672-6688
10. Sun, Z., BAR-based multi-dimensional nonequilibrium pulling for indirect construction of QM/MM free energy landscapes: from semi-empirical to ab initio. *Phys. Chem. Chem. Phys.* **2019**, *21*, 21942-21959
11. Wang, X.; Xingzhao, T.; Boming, D.; John Z. H., Z.; Sun, Z., BAR-based Optimum Adaptive Steered MD for Configurational Sampling. *J. Comput. Chem.* **2019**, *40*, 1270-1289.
12. Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C., Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* **2006**, *65*, 712-25.
13. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926-935.
14. Price, D. J.; Brooks III, C. L., A Modified TIP3P Water Potential for Simulation with Ewald Summation. *J. Chem. Phys.* **2004**, *121*, 10096-10103.
15. York, D. M.; Darden, T. A.; Pedersen, L. G., The effect of long -range electrostatic interactions in simulations of macromolecular crystals: A comparison of the Ewald and truncated list methods. *J. Chem. Phys.* **1993**, *99*, 8345-8348.
16. Hoof, R. W.; van Eijck, B. P.; Kroon, J., An Adaptive Umbrella Sampling Procedure in Conformational Analysis using Molecular Dynamics and Its Application to Glycol. *J. Chem. Phys.* **1992**, *97*, 6690-6694.
17. Mezei, M., Adaptive Umbrella Sampling: Self-consistent Determination of the Non-Boltzmann Bias. *J. Comput. Phys.* **1987**, *68*, 237-248.
18. Kästner, J., Umbrella sampling. *Wiley Interdisip. Rev. Comput. Mol. Sci.* **2011**, *1*, 932-942.
19. Wang, X.; Sun, Z., Determination of Base Flipping Free Energy Landscapes from Nonequilibrium Stratification. *J. Chem. Inf. Model.* **2019**, *59*, 2980-2994.
20. Sun, Z.; Wang, X.; Zhang, J. Z. H., Protonation-dependent Base Flipping in The Catalytic Triad of A Small RNA. *Chemical Physics Letters* **2017**, *684*, 239-244.
21. Sun, Z.; Wang, X.; Zhang, J. Z. H.; He, Q., Sulfur-substitution-induced base flipping in the DNA duplex. *Phys. Chem. Chem. Phys.* **2019**, *21*, 14923-14940.
22. Sun, Z.; Zhang, J. Z. H., Thermodynamic Insights of Base Flipping in TNA Duplex: Force Fields, Salt Concentrations, and Free-Energy Simulation Methods. *CCS Chemistry* **2020**, *2*, 1026-1039.
23. Lemkul, J. A.; Savelyev, A.; MacKerell Jr, A. D., Induced Polarization Influences The Fundamental Forces in DNA Base Flipping. *J. Phys. Chem. Lett.* **2014**, *5*, 2077-2083.
24. Zheng, H.; Cai, Y.; Ding, S.; Tang, Y.; Kropachev, K.; Zhou, Y.; Wang, L.; Wang, S.; Geacintov, N. E.; Zhang, Y., Base Flipping Free Energy Profiles for Damaged and Undamaged DNA. *Chemical research in toxicology* **2010**, *23*, 1868-1870.

25. Lee, T. S.; Radak, B. K.; Pabis, A.; York, D. M., A New Maximum Likelihood Approach for Free Energy Profile Construction from Molecular Simulations. *J. Chem. Theory Comput.* **2013**, *9*, 153-164.
26. Fass, J.; Sivak, D. A.; Crooks, G. E.; Beauchamp, K. A.; Leimkuhler, B.; Chodera, J. D., Quantifying configuration-sampling error in Langevin simulations of complex molecular systems. *Entropy* **2018**, *20*, 318.
27. Sivak, D. A.; Chodera, J. D.; Crooks, G. E., Time Step Rescaling Recovers Continuous-Time Dynamical Properties for Discrete-Time Langevin Integration of Nonequilibrium Systems. *J. Phys. Chem. B* **2014**, *118*, 6466-6474.
28. Sivak, D. A.; Chodera, J. D.; Crooks, G. E., Using Nonequilibrium Fluctuation Theorems to Understand and Correct Errors in Equilibrium and Nonequilibrium Simulations of Discrete Langevin Dynamics. *Phys. Rev. X* **2013**, *3*, 011007.
29. Stewart, J. J., Optimization of parameters for semiempirical methods V: modification of NDDO approximations and application to 70 elements. *Journal of Molecular modeling* **2007**, *13*, 1173-1213.
30. Marchi, M.; Procacci, P., Coordinates scaling and multiple time step algorithms for simulation of solvated proteins in the NPT ensemble. *J. Chem. Phys.* **1998**, *109*, 5194-5202.
31. Procacci, P.; Marchi, M., Taming the Ewald sum in molecular dynamics simulations of solvated proteins via a multiple time step algorithm. *J. Chem. Phys.* **1996**, *104*, 3003-3012.
32. Procacci, P.; Berne, B., Computer simulation of solid C60 using multiple time-step algorithms. *J. Chem. Phys.* **1994**, *101*, 2421-2431.
33. Klimasauskas, S.; Kumar, S.; Roberts, R. J.; Cheng, X., HhaI methyltransferase flips its target base out of the DNA helix. *Cell* **1994**, *76*, 357-369.
34. Zhang, Z.-L.; Wu, Y.-Y.; Xi, K.; Sang, J.-P.; Tan, Z.-J., Divalent Ion-mediated DNA-DNA Interactions: A Comparative Study of Triplex and Duplex. *Biophysical journal* **2017**, *113*, 517-528.
35. Špačková, N. A.; Cubero, E.; Šponer, J.; Orozco, M., Theoretical study of the guanine --> 6-thioguanine substitution in duplexes, triplexes, and tetraplexes. *Journal of the American Chemical Society* **2004**, *126*, 14642-50.
36. Wärmländer, S.; Sandström, K.; Leijon, M.; Gräslund, A., Base-pair dynamics in an antiparallel DNA triplex measured by catalyzed imino proton exchange monitored via ¹H NMR spectroscopy. *Biochemistry* **2003**, *42*, 12589-12595.
37. Cheng, Y.-K.; Pettitt, B. M., Stabilities of double- and triple-strand helical nucleic acids. *Prog. Biophys. Mol. Biol.* **1992**, *58*, 225-257.
38. Mees, A.; Klar, T.; Gnau, P.; Hennecke, U.; Eker, A. P.; Carell, T.; Essen, L.-O., Crystal structure of a photolyase bound to a CPD-like DNA lesion after in situ repair. *Science* **2004**, *306*, 1789-1793.
39. C. Clement, F.; Kaczmarek, N.; Mathieu, N.; Tomas, M.; Leitenstorfer, A.; Ferrando-May, E.; Naegeli, H., Dissection of the xeroderma pigmentosum group C protein function by site-directed mutagenesis. *Antioxidants & redox signaling* **2011**, *14*, 2479-2490.
40. Alt, F. W.; Schwer, B., DNA double-strand breaks as drivers of neural genomic change, function, and disease. *DNA Repair* **2018**, *71*, 158-163.
41. Silvestrov, P.; Maier, S. J.; Fang, M.; Cisneros, G. A., DNARcDb: A database of cancer biomarkers in DNA repair genes that includes variants related to multiple cancer phenotypes. *DNA Repair* **2018**, *70*, 10-17.
42. Voloshin, O. N.; Camerini-Otero, R. D., Synaptic complex revisited: a homologous recombinase flips and switches bases. *Mol. Cell* **2004**, *15*, 846-847.
43. Meulenbroek, E. M.; Peron Cane, C.; Jala, I.; Iwai, S.; Moolenaar, G. F.; Goosen, N.; Pannu, N. S., UV damage endonuclease employs a novel dual-dinucleotide flipping mechanism to recognize different DNA lesions. *Nucleic acids research* **2012**, *41*, 1363-1371.
44. O'Gara, M.; Klimasauskas, S.; Roberts, R. J.; Cheng, X., Enzymatic C5-cytosine Methylation of DNA: Mechanistic Implications of New Crystal Structures for HhaI Methyltransferase-DNA-AdoHcy Complexes. *Journal of molecular biology* **1996**, *261*, 634-645.
45. Allan, B. W.; Reich, N. O.; Beechem, J. M., Measurement of the Absolute Temporal Coupling between DNA Binding and Base Flipping. *Biochemistry* **1999**, *38*, 5308-5314.
46. Cal, S.; Connolly, B. A., DNA Distortion and Base Flipping by the EcoRV DNA Methyltransferase A STUDY USING INTERFERENCE AT dA AND T BASES AND MODIFIED DEOXYNUCLEOSIDES. *Journal of Biological Chemistry* **1997**, *272*, 490-496.
47. O'Gara, M.; Roberts, R. J.; Cheng, X., A Structural Basis for the Preferential Binding of Hemimethylated DNA by HhaI

- DNA Methyltransferase. *Journal of molecular biology* **1996**, 263, 597-606.
48. Wildes, A.; Theodorakopoulos, N.; Valle-Orero, J.; Cuesta-López, S.; Garden, J.-L.; Peyrard, M., Structural Correlations and Melting of B-DNA Fibers. *Phys. Rev. E* **2011**, 83, 061923.
49. Velizhanin, K. A.; Chien, C.-C.; Dubi, Y.; Zwolak, M., Driving Denaturation: Nanoscale Thermal Transport as A Probe of DNA Melting. *Phys. Rev. E* **2011**, 83, 050906.
50. Wong, K.; Pettitt, B., The Pathway of Oligomeric DNA Melting investigated by Molecular Dynamics Simulations. *Biophysical Journal* **2008**, 95, 5618.
51. Blake, R.; Bizzaro, J., Jr.; Day, G.; Delcourt, S.; Knowles, J.; Marx, K.; Santalucia, J. J., Statistical Mechanical Simulation of Polymeric DNA Melting with MELTSIM. *Bioinformatics* **1999**, 15, 370-375.
52. Gotoh, O., Prediction of Melting Profiles and Local Helix Stability for Sequenced DNA. *Advances in biophysics* **1983**, 16, 1-52.
53. Lazurkin, Y. S.; Frank-Kamenetskii, M.; Trifonov, E., Perspectives Report: Melting of DNA: Its Study and Application as A Research Method. *Biopolymers* **1970**, 9, 1253-1306.
54. Zimm, B. H., Theory of "Melting" of The Helical Form in Double Chains of The DNA Type. *J. Chem. Phys.* **1960**, 33, 1349-1356.
55. Alexandrov, B. S.; Fukuyo, Y.; Lange, M.; Horikoshi, N.; Gelev, V.; Rasmussen, K. Ø.; Bishop, A. R.; Usheva, A., DNA Breathing Dynamics Distinguish Binding from Nonbinding Consensus Sites for Transcription Factor YY1 in Cells. *Nucleic acids research* **2012**, 40, 10116-10123.
56. Peyrard, M.; Cuesta-Lopez, S.; James, G., Nonlinear Analysis of The Dynamics of DNA Breathing. *J. Biol. Phys.* **2009**, 35, 73.
57. Duduiala, C. I.; Wattis, J. A.; Dryden, I. L.; Laughton, C. A., Nonlinear Breathing Modes at A Defect Site in DNA. *Phys. Rev. E* **2009**, 80, 061906.
58. Travers, A., DNA Dynamics: bubble 'n' flip for DNA Cyclisation? *Curr. Biol.* **2005**, 15, R377-R379.
59. Zeida, A.; Machado, M. R.; Dans, P. D.; Pantano, S., Breathing, Bubbling, and Bending: DNA Flexibility from Multimicrosecond Simulations. *Phys. Rev. E* **2012**, 86, 021903.
60. Huang, N.; Banavali, N. K.; MacKerell, A. D., Protein-facilitated Base Flipping in DNA by Cytosine-5-methyltransferase. *Proc. Natl. Acad. Sci. USA* **2003**, 100, 68-73.
61. Vassilyev, D. G.; Kashiwagi, T.; Mikami, Y.; Ariyoshi, M.; Iwai, S.; Ohtsuka, E.; Morikawa, K., Atomic model of a pyrimidine dimer excision repair enzyme complexed with a DNA substrate: structural basis for damaged DNA recognition. *Cell* **1995**, 83, 773-782.
62. Cheng, X.; Blumenthal, R. M., Finding a basis for flipping bases. *Structure* **1996**, 4, 639-645.
63. Lyakhov, I. G.; Hengen, P. N.; Rubens, D.; Schneider, T. D., The P1 phage replication protein RepA contacts an otherwise inaccessible thymine N3 proton by DNA distortion or base flipping. *Nucleic acids research* **2001**, 29, 4892-4900.
64. Hornby, D. P.; Ford, G. C., Protein-mediated base flipping. *Current opinion in biotechnology* **1998**, 9, 354-358.
65. Goedecke, K.; Pignot, M.; Goody, R. S.; Scheidig, A. J.; Weinhold, E., Structure of the N6-adenine DNA methyltransferase M•TaqI in complex with DNA and a cofactor analog. *Nature Structural and Molecular Biology* **2001**, 8, 121.
66. Ivani, I.; Dans, P. D.; Noy, A.; Pérez, A.; Faustino, I.; Hospital, A.; Walther, J.; Andrio, P.; Goñi, R.; Balaceanu, A., Parmbsc1: A Refined Force Field for DNA Simulations. *Nat. Methods* **2016**, 13, 55.
67. Joung, I. S.; Cheatham III, T. E., Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. *J. Phys. Chem. B* **2008**, 112, 9020-9041.
68. Joung, I. S.; Cheatham, T. E., Molecular Dynamics Simulations of the Dynamic and Energetic Properties of Alkali and Halide Ions Using Water-Model-Specific Ion Parameters. *J. Phys. Chem. B* **2009**, 113, 13279-13290.
69. Hub, J. S.; Groot, B. L. D.; Spoel, D. V. D., g_wham—A Free Weighted Histogram Analysis Implementation Including Robust Error and Autocorrelation Estimates. *J. Chem. Phys.* **2015**, 6, 3713-3720.
70. Ono, S.; Nakajima, N.; Higo, J.; Nakamura, H., The Multicanonical Weighted Histogram Analysis Method for the Free-energy Landscape Along Structural Transition Paths. *Chemical physics letters* **1999**, 312, 247-254.