HHS Public Access

Author manuscript

J Chem Theory Comput. Author manuscript; available in PMC 2020 June 12.

Published in final edited form as:

J Chem Theory Comput. 2019 August 13; 15(8): 4699–4707. doi:10.1021/acs.jctc.9b00329.

A Fast Implementation of the Nudged Elastic Band Method in AMBER

Delaram Ghoreishi[†], **David S. Cerutti**[‡], **Zachary Fallon**[¶], **Carlos Simmerling**[¶], **Adrian E. Roitberg**[*,§]

[†]Department of Physics, University of Florida, Gainesville, Florida 32611, United States

[‡]Laboratory for Biomolecular Simulation Research, Department of Chemistry and Chemical Biology, Rutgers University, Piscataway, New Jersey 08854, United States

[¶]Department of Chemistry, and Laufer Center for Physical and Quantitative Biology, Stony Brook University, Stony Brook, New York, 11794, United States

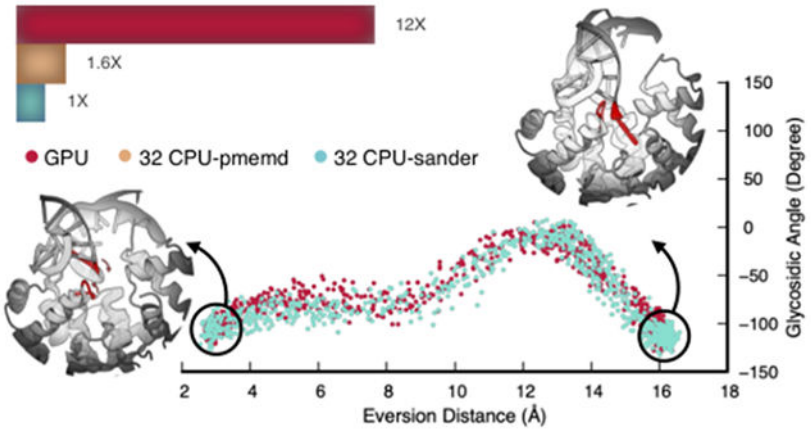
[§]Department of Chemistry, University of Florida, Gainesville, Florida 32611, United States

Abstract

We present a fast implementation of the nudged elastic band (NEB) method into the particle mesh Ewald molecular dynamics (pmemd) module of the Amber software package both for central processing units (CPU) and graphics processing units (GPU). The accuracy of the new implementation has been validated for three cases: a conformational change of alanine dipeptide, the α-helix to β-sheet transition in polyalanine, and a large conformational transition in human 8-oxoguanine–DNA glycosylase with DNA complex (OGG1–DNA). Timing benchmark tests were performed on the explicitly solvated OGG1–DNA system containing ~50k atoms. The GPU-optimized implementation of NEB achieves more than two orders of magnitude speedup compared to the previous CPU implementation performed with a two-core CPU processor. The speed and scalable features of this implementation will enable NEB applications on larger and more complex systems.

Graphical Abstract

[*] roitberg@ufl.edu.



Introduction

The statistical behavior gleaned from simulations of biomolecules yields detailed information about observed biochemical phenomena. Computational biologists can probe the frequency and mechanism of rare transitions that are hard to observe in experiments by locating the minimum free energy pathways. However, traditional molecular dynamics of proteins and biopolymers often fail to sample these important transitions, as the systems are thermally limited to low energy states on a rugged free energy surface. Precision and reproducibility in the result have been limited by the cost of the calculations. The use of graphics processing units (GPUs) significantly accelerates these intensive calculations, offering a base multiplier for enhanced sampling strategies that can be implemented on their advanced architecture, but the multiplier by itself is not enough. The community needs a set of efficient simulation algorithms implemented on vector-accelerated architectures that enhance the exploration of free energy surfaces for detecting the multitude of rare transition pathways.

Different methods have been developed for finding transition pathways.[1–4] Some depend only on the initial structure and follow a minimum ascent path to reach a final structure.[5–8] It is not guaranteed, however, that the desired final structure is reached.[2] Other methods use the second derivatives of the potential energy function to locate the saddle points.[9–11] Once the saddle points are identified, local minimum states can be found using the steepest descent algorithm. But, since calculation and diagonalization of the second derivative matrix at each step of the simulation is expensive, these methods are applicable only to small systems. Other approaches determine the path when both initial and final states are identified.[12,13] Among these methods, chain-of-states algorithms[12–15] are compelling as they adjust and scale with resources to produce the desired precision and efficiency. The nudged elastic band (NEB)[14] method in combination with simulated annealing, for instance, has proven successful in determining minimum energy paths (MEP) for rugged energy surfaces.[16,17] A minimum energy path is a transition passage connecting the initial and final states, for which any point along the path is at the minimum energy value compared to other positions in the hyperplane perpendicular to MEP at that specific point. Hence, the perpendicular component of the gradient of the potential energy at any point along the path

is zero. NEB,[14] first proposed by Jónsson et al., is an evolution of methods such as self penalty walk (SPW),[18] locally updated planes (LUP),[19] and elastic band.[20] NEB results in a continuous representation of MEP by simultaneous energy minimization of a series of connected replicas. This continuity is not guaranteed in LUP, in which the initial choice of the pathway could affect convergence to a connected path. In SPW, the converged path is not the MEP *per se*, but a path along which the averaged potential energy is minimized.[20] Moreover, the elastic band can result in corner cutting and sliding down; problems which are easily excluded in NEB by force decoupling.[14]

A more sophisticated approach to finding MEP is the string method,[15] which results in a smoother path compared to NEB since it uses higher order interpolation schemes. In the string method, it is also possible to change the number of replicas on the fly. Moreover, methods such as temperature-dependent NEB (tNEB)[21] and finite temperature string method[4] account for temperature corrections to MEP. The purpose of this work is to accelerate the current implementation of NEB in Amber suite of molecular dynamics software,[22] in a way that is easily extendable to other chain-of-states methods for searching minimum energy paths such as the string method.

Partial nudged elastic band (PNEB)[16]–which exclusively applies the NEB forces to a user-defined subset of atoms-is the supported implementation of NEB in the Amber18. PNEB allows the use of NEB in systems with explicit solvent molecules, which are left alone to relax and adapt to the conformational changes of the system, free of any additional restraints aside from their physical interactions with other molecules defined by force field parameters. Furthermore, this method decreases the communication overhead between the replicas by incorporating fewer atoms in the NEB calculations, which leads to better scaling of the code. We have extended PNEB by incorporating the routines into the particle mesh Ewald molecular dynamics (PMEMD) module of Amber and further accelerated it by implementing those routines with CUDA to operate on NVIDIA GPUs. In this implementation, a shuttle transfer was developed which minimizes the amount of data that must traverse a message passing interface (MPI) between GPUs. Additional performance enhancement has been made possible through the usage of a flag which enables the users to control the frequency at which NEB forces are computed.

For a system containing ~50k atoms, running in parallel with one NVIDIA Tesla-P100 GPU assigned to each replica, the Amber18 GPU-accelerated PNEB executes simulations more than 60 times faster than two cores of an Intel Xeon Platinum 8160 CPU assigned to each replica. Numerical precision is uncompromised. The new implementation enhances the study of properties of biomolecules due to conformational transitions in a multidimensional configuration space of thousands of atoms–which was otherwise hard to study with current CPU architectures. Our implementation of PNEB for multi-GPU execution within the AMBER software suite could aid computational scientists in developing new drug compounds and novel materials by applying these powerful algorithms on commodity hardware.

Methods

Theory

NEB applies a series of harmonic restraints between replicas (or images), which are first generated along a putative pathway. Replicas are linked to their nearest neighbors via springs, such that the entire system represents a discrete pathway, from reactants to products. The purpose of the springs is to distribute the replicas along the path and prevent them from sliding down to the minimum states.[14] The replicas evolve into a discrete representation of MEP by simultaneous energy minimization of the entire chain. Setting N – 2 replicas between the initial and final states, positions of the discrete points can be denoted by the array $[\boldsymbol{R}_1, \boldsymbol{R}_2, \boldsymbol{R}_3, \ldots, \boldsymbol{R}_N]$, where $\boldsymbol{R}_1$ and $\boldsymbol{R}_N$ are the two endpoints which are kept fixed in the phase space throughout the simulation. Figure 1 shows a mass and spring representation of NEB. Each replica is an atomic representation of the system at a certain position along the pathway that connects the initial and final states.

If no guess for the reaction coordinate is available, this pathway can be constructed by placing half of the replicas on or close to the initial structure and the other half on or close to the final structure. This way of initializing the path requires shorter timesteps and weaker springs at the beginning of the simulation to ensure that the very stretched central spring and other compressed springs will not exert strong forces on the particles of the systems, allowing the ensemble to slowly approach a smooth path.

Translational and rotational differences of the adjacent replicas should be minimized before the calculation of the spring forces. First, the transnational differences are removed by placing the origin of the coordinate system on the corresponding center of mass (COM) coordinates. Then an optimal rotation matrix is applied to the coordinates of the neighboring replicas to minimize the root mean square deviation (RMSD) between the two sets of the atomic coordinates.[23]

A tangent vector at each image position ($\boldsymbol{\tau}_i$) is responsible for decoupling the spring forces and the potential forces to prevent them from interfering. Only the perpendicular component of the forces defined by force field parameters, and the parallel component of the spring forces, are considered in the equations of motion (refer to the inset in figure 2). The total force acting on each atom hence includes two orthogonal components:

$$\boldsymbol{F}_i^{NEB} = \boldsymbol{F}_i^{\parallel} + \boldsymbol{F}_i^{\perp} \tag{1}$$

$$\boldsymbol{F}_i^{\perp} = -\nabla V(\boldsymbol{R}_i) + (\nabla V(\boldsymbol{R}_i) \cdot \boldsymbol{\tau}_i) \cdot \boldsymbol{\tau}_i \tag{2}$$

$$\boldsymbol{F}_i^{\parallel} = (\boldsymbol{F}_i^{s} \cdot \tau_i) \cdot \boldsymbol{\tau}_i \tag{3}$$

where $\boldsymbol{F}_i^{s}$ is the spring force at the position of the ith replica, and $\nabla V(\boldsymbol{R}_i)$ is the potential force described by the force field. The tangents are defined based on the energy of the self and the neighboring replicas as:

$$\boldsymbol{\tau}_i = \begin{cases} \boldsymbol{R}_{i+1} - \boldsymbol{R}_i \ if & V_{i+1} > V_i > V_{i-1} \\ \boldsymbol{R}_i - \boldsymbol{R}_{i-1} \ if & V_{i+1} < V_i < V_{i-1} \end{cases} \quad (4)$$

in which $V_i = V(\boldsymbol{R}_i)$. In this definition, only the position of the neighboring replica with higher energy is considered. If both of the neighboring replicas have either higher or lower energy with respect to replica i, that is, if

$$V_{i+1} > V_i < V_{i-1}$$
$$V_{i+1} < V_i > V_{i-1}$$

then a weighted average will be used to define the tangent estimates:

$$\boldsymbol{\tau}_i = \begin{cases} (\boldsymbol{R}_{i+1} - \boldsymbol{R}_i)\Delta V_i^{max} + (\boldsymbol{R}_i - \boldsymbol{R}_{i-1})\Delta V_i^{min} \ if & V_{i+1} > V_{i-1} \\ (\boldsymbol{R}_{i+1} - \boldsymbol{R}_i)\Delta V_i^{min} + (\boldsymbol{R}_i - \boldsymbol{R}_{i-1})\Delta V_i^{max} \ if & V_{i+1} < V_{i-1} \end{cases} \quad (5)$$

where

$$\Delta V_i^{max} = max(|V_{i+1} - V_i|, |V_{i-1} - V_i|)$$

$$\Delta V_i^{min} = min(|V_{i+1} - V_i|, |V_{i-1} - V_i|)$$

For more detailed information on the definition of tangents, the reader can refer to reference 24. The decoupling ensures a smooth convergence to the path and prevents the images from corner-cutting or sliding down.[14] The force projection decouples the dynamics of the images from the discrete distribution of the images along the path, such that only the true forces are responsible for relaxation of the images while the spring forces keep the images away from the minimum states. Figure 2 shows a schematic representation of the NEB force decoupling on a two-dimensional LEPS harmonic potential energy surface. For more information regarding this potential model refer to appendix A of reference 14.

The minimization step would take the chain towards the local minimum that is most accessible to the initial path. Simulated annealing can increase the chance of NEB simulations converging to the global MEP rather than a local MEP. Supervision may be required to prevent temperature increases from leading to the presence of unphysical structures during the simulation. In the final phase of the simulation, a gradual decrease of the temperature to zero will freeze the replicas along the minimum transition path.

It is possible that the biological system of interest has multiple pathways connecting the two metastable states. This necessitates statistical analysis of multiple independent simulations which result in different pathways–which once again illustrates the importance of fast simulation techniques.[25,26]

Implementation

The PNEB routines have been implemented in the pmemd module of the Amber package and the parallel-MPI simulations can be performed with either CPU or GPU processors (*i.e.*, with pmemd.MPI and pmemd.cuda.MPI). The CPU implementation was straight forward and does not need a detailed explanation. Figure 3 shows the MPI framework of the CPU code. Replicas have to communicate their coordinates to their nearest neighbors for harmonic force calculations. The two arrows connecting the replicas in the figure represent the MPI calls that regulate the data transfer. A higher number of replicas leads to a higher resolution of the path but demands more computational resources. Benchmarking can indicate the best balance between precision and cost. The number of replicas, however, usually goes above what a single cluster node can accommodate. Each replica offloads its computation task to multiple MPI processes. Process zero, which is the master rank, computes the NEB forces and broadcasts them to the other processes. Afterward, an integration step is carried out to update the coordinates.

GPU programming is inevitably susceptible to the latency of the data transfer between the device and the host memory which hinders the performance of the GPU-accelerated code. In the NEB routines, specifically, the coordinate exchange between the replicas at each step of the simulation can aggravate the performance. Hence the GPU implementation required a data transfer optimization other than programming the NEB routines with CUDA. Figure 4 shows the framework of the GPU implementation of PNEB in Amber. Replicas are denoted in this figure by white squares, with a blue polyalanine complex in the middle transitioning from an α-helix to β-sheet conformation. The figure also illustrates our optimized transfer scheme named shuttle transfer. In practice, the information regarding the coordinates of all the atoms could be transferred between the processing units. However, not all coordinates are needed, and an exchange of a smaller set is sufficient. Our shuttle transfer reduces the amount of data that must traverse between the computing units by selectively coalescing the memory sections that only correspond to the two atom masks provided by the user for NEB calculations. These two masks contain the atoms that are included in NEB force calculations and the atoms involved in performing the root mean square fitting to the neighboring structures.

It is possible to increase the performance by applying multiple timesteps to compute NEB forces less frequently. The NEB forces are only updated every n^{th} step of the simulation, while for the other steps, the most recent calculated NEB forces are applied to the replicas. Since for a small number of steps the coordinates do not change drastically, it is safe to skip the spring force calculations on the off-steps and apply the most recent NEB forces instead. The attempt for updating the NEB forces, however, should happen as frequently as possible or else the NEB forces might change too much for two consecutive steps and cause instability in the simulations. The flag that enables this functionality is nebfreq, which has a default value of 1 corresponding to performing NEB on every step.

Computational Details

Three different test cases have been selected for precision and performance examinations. All simulations have been performed using the Amber18 and AmberTools18 suite of

programs.[22] The structures for the first two test cases were built using the leap module of Amber16, with the ff14SB[27] forcefield parameters. The structures for the third test case were obtained from Li et al..[26] The TIP3P water model[28] has been used for explicit solvent simulations. As previously illustrated in the work of Bergonzo et al.[16] and Li et al.,[26] a Langevin thermostat with a high collision frequency is required to control the temperature when performing NEB simulations. High values of collision frequency may not be appropriate for normal MD simulations, but it is recommended for NEB simulations. A strongly coupled thermostat reduces instabilities due to the projection of the potential forces and the addition of the unrealistic springs. Predicting the minimum energy path includes sampling the highly unstable structures around the transition state region. Leading the dynamics of the corresponding replicas towards these specific structures through a high viscous medium is beneficial as it delays the retraction of the structures towards the lower energy states. For this reason, a collision frequency of 1000 ps^{-1} was used to control the temperature for simulations performed in an implicit solvent. For explicit solvent simulations, a lower collision frequency was used (refer to test case 3 for values), since the presence of water molecules contributes to the increase of the viscosity of the system. These values of the collision frequencies are in line with the studies that performed PNEB previously.[16,25,26,29] A strong thermostat, however, would increase the system's temperature too fast. To prevent this from happening, it is recommended to use a slow and a linear change in the temperature.

Test Case 1: Conformational Change of Alanine Dipeptide

A capped alanine dipeptide with a total of 22 atoms was built. In this test case, the pathway between the so-called α_R and the α_L basins on the Ramachandran plot of alanine dipeptide is explored.[16,30] The impose command of the tleap was used to initialize the two structures close to the two basins. GB-Neck2[31] generalized Born implicit solvent model, with sodium chloride salt concentration of 0.2 M, has been used for this test case. The initial structures were then minimized by performing 25000 steps of steepest descent followed by 25000 steps of conjugate gradient. After the minimization, the initial structure had (ϕ, ψ) = (−78.33°, −10.58°), while the final structure had (ϕ, ψ) = (57.50°, 20.55°). NEB calculations have been performed using 16 replicas in implicit solvent. NEB forces have been applied to only the backbone atoms, while all the atoms are included for fitting the neighboring structures to calculate the NEB forces. Simulated annealing was used along with NEB to improve the exploration of the energy landscape. First, 20 ps of simulation with timesteps of 0.5 fs was performed to heat the system to 300 K with spring constants of 10 $kcal \cdot mol^{-1} \cdot Å^{-2}$. Afterward, spring constants were raised to 50 $kcal \cdot mol^{-1} \cdot Å^{-2}$ and 100 ps of simulation was followed with timesteps of 1 fs during which the temperature was held at 300 K. Next, 300 ps of simulated annealing[32] with timesteps of 0.5 fs was performed to gradually increase the temperature up to 500 K and back down to 300 K. Short timesteps and gradual increase of the temperature at this step is critical for the stability of the system. Following the simulated annealing, the temperature of the system was gradually decreased to zero over 120 ps of simulation with timesteps of 1 fs which were followed by quenched MD for 200 ps.

Test Case 2: α-helix to β-sheet Transition in polyalanine

Twelve alanine residues were created from ACE-ALA(12)-NME sequence for a total of 112 atoms. The impose command in the tleap was used to create the initial and final structures in a α-helix and a β-sheet conformation, respectively. Minimization was performed on the endpoint structures for 5000 steps of steepest descent followed by 5000 steps of conjugate gradient. The GBn model[33] with sodium chloride salt concentration of 0.2M has been used to model the implicit solvent. NEB calculations and the selection of the atoms included in the NEB region were performed as described in the previous test case.

Test Case 3: Base Eversion Pathway of the OGG1–DNA Complex

The OGG1–DNA complexes with intrahelical (initial) and extrahelical (final) endpoints together with the additional intermediate structures along the major groove path[26] were generated as described in the supplementary information of reference 26. The parameters for running PNEB were the same as the work of Li et al..[26] The system contains 49534 atoms, 43698 of which belong to the solvent. The NEB simulations were performed with 32 replicas in an explicit solvent. For performing NEB, first, the replicas are equilibrated at 310 K, with spring constants of 1 kcal $\cdot$ mol^{-1} $\cdot$ Å^{-2}, and collision frequency of 100 ps^{-1}, for 100 ps with timesteps of 1 fs. For the rest of the NEB simulations, the spring constants are raised to 20 kcal $\cdot$ mol^{-1} $\cdot$ Å^{-2} and the collision frequency is brought down to 75 ps^{-1}. The system is equilibrated at 310 K for another 500 ps with 1 fs timesteps. The system's temperature is then raised to 380 K during 100 ps with timesteps of 1 fs. Further, the system is equilibrated at 380 K for an extra 200 ps, and finally, the temperature is lowered to 310 K during 100 ps. For the last phase of the NEB simulation, the temperature is kept fixed at 310 K for an extra 500 ps in order to equilibrate the system at that temperature.

Results and Discussion

Accuracy Tests

We use the three test cases to demonstrate that numerical accuracy is not compromised in the new implementations. These tests compare the numerical values of a specific reaction coordinate along the path for different implementations. The choice of the reaction coordinate was based on *a priori* knowledge for alanine dipeptide conformational change along the ϕ and ψ angles, or revelations from prior NEB simulations connecting the initial and final states of the second and third test cases.

Test Case 1: Conformational Change of Alanine Dipeptide

The first test case studies the conformational change of alanine dipeptide. The pathway between two stable conformations corresponding to two minimum regions on the potential energy surface of ϕ and ψ dihedral angles[16,30] is explored. Figure 5 shows the energy landscape of this system along with the positions of the two metastable conformations and one of the possible transition pathways. Another choice for the pathway would be to proceed through the barrier around $\phi = 0°$ and $\psi = 100°$. In this test case, we performed two sets of simulations: one set was performed on sander (previous implementation) and the other on pmemd-CPU (recent CPU implementation). From each set, 20 independent simulations that

proceeded from the pathway shown in figure 5 was selected. When running NEB simulations it is important to perform a set of independent simulations to identify all the possible transition pathways. Free energy profiles along the paths, can then be obtained through various available free energy calculation techniques, such as the umbrella sampling, [34] which can yield more insight into the preferred pathway in terms of the energetics along the path.

Figure 6 shows the potential energies of the NEB replicas along the path. In order to have statistically reliable results, each point in the plot is averaged over the individual simulations. Replicas 1 and 16 correspond to the initial and final minimum energy states, respectively. As we move away from the initial state along the path, the energy of the replicas increases until it reaches a transition state with an energy value of ~6.3 kcal · mol^{-1}. Vertical bars show the standard deviation for each replica. The errors can be represented by dividing the standard deviations by the square root of the ensemble size. The two lines plotted in figure 6 show that the results obtained from pmemd-CPU are in good agreement with the ones obtained from the sander.

Test Case 2: α-helix to β-sheet Transition in polyalanine

The second test case is a small alanine peptide transitioning from α-helix to β-sheet. Figure 7 shows the peptide end to end distance along the path and the initial and the final conformations. The blue and red lines are averaged over 50 individual simulations performed with pmemd-CPU and pmemd-GPU, respectively. Vertical bars represent the standard deviation. This test case demonstrates that the GPU implementation of PNEB provides results that are statistically equivalent to the ones achieved using the pmemd-CPU implementation.

Test Case 3: Base Eversion Pathway of the OGG1–DNA Complex

8-Oxoguanine (8-oxoG) is a result of oxidation of guanine and one of the most common products of oxidative damage in DNA, which can lead to mutations in cells if not excised prior to DNA replication.[25,26,29] Human 8-oxoguanine–DNA glycosylase (OGG1) excises 8-oxoG from damaged DNA in base excision repair. Several studies have been reported that address different aspects of this transition to better understand the base eversion pathways and the preferred binding conformation.[25,26,29,35,36] The two endpoints correspond to the intrahelical and the extrahelical conformations in the base eversion pathway. Figure 8 illustrates the endpoint structures of the region involved in the transition plus the glycosidic angle versus the eversion distance change of this transition. This eversion distance has been reported previously as a reaction coordinate for nucleic acid base eversion.[26,37] The red points are the result of the simulations performed with nebfreq equal to 1. The light blue points are the result of the simulations performed with nebfreq equal to 5 in all the stages of the simulation but the first and the final, at which the value of nebfreq was set to 2. The data points are the averaged values extracted from the trajectories of the last stage of 10 independent NEB simulations. The two sets agree which explains that even when we update the NEB forces every 5 steps the transition pathway still falls into the correct region. A perfect agreement between individual sets of simulations for systems that contain many degrees of freedom is not possible and it is reasonable to expect of a transition region rather

than a single pathway. Once the region is identified, free energy profiles of the region surrounding the path can be obtained.[26]

Timing Benchmarks

Figures 9 and 10 show the performance of the different implementations of PNEB in Amber. All the benchmarks were performed using the third test case consisting of 49534 atoms. The GPU simulations were performed on P100 NVIDIA GPUs and Intel Xeon E5-2680v3 CPU processors linked via Mellanox FDR InfiniBand interconnects. The CPU simulations were performed on Intel Xeon Platinum 8160 processors.

Figure 9 shows the scalability of sander and pmemd-CPU implementations using different numbers of CPU processors per NEB replica. Porting the code from sander to pmemd provides a performance gain of about 1.6X, commensurate with the speed of the CPU pmemd engine relative to the sander.

Figure 10 shows the average increase in the performance of the CPU and the GPU code (with different GPU precision models[38]). CPU benchmarks were performed for sander with nebfreq equal to 1, and for pmemd with nebfreq equal to 1, 2, 5, and 10. GPU benchmarks were performed for a full transfer with nebfreq equal to 1, and a shuttle transfer with nebfreq equal to 1, 2, 5, and 10. In this test case, the simulations were performed with timesteps of 1 fs. As a result, setting nebfreq=n corresponds to updating the NEB forces every n fs. For shuttle transfers, the coordinates of 2630 out of the 49534 atoms traveled between the CPU host and GPU device memory every nebfreq steps. More than 2X performance gain has been achieved through the shuttle transfer scheme for the case of SPFP. Compared to the initial implementation (sander), the GPU code results in more than 10X performance enhancement over 32 CPU processors. For this test case, a further two-fold acceleration was observed by increasing nebfreq to 5.

Another set of benchmarks has been provided in Figure 11, which compares the speed of the GPU code for various numbers of atoms included in the shuttle transfer. For each replica, during each transfer, the coordinate values of a certain number of atoms are transferred, first from the device to the host of that replica, second through an MPI routine between the hosts of the neighboring replicas, and third from the host of the neighboring replicas to their corresponding devices. Data transfer speed depends on a lot of factors including the bandwidth of the hardware, and the type and bandwidth of the interconnect linking the hardware, as well as the latency of each transfer call. But as a general rule, minimizing the size of the transferred data results in a better performance. As shown in figure 11, the performance of the PNEB code is also dependent on the data transfer size. The use of peer to peer communication and interconnect architectures such as NVLink is likely to improve the performance further.

Conclusion

A fast and accurate study of conformational transitions in biological systems can lead to advancements in computational drug discoveries by assisting the rational design of novel drug molecules. Efficient sampling algorithms are pivotal in identifying transition pathways

on multidimensional rugged energy surfaces with numerous degrees of freedom. Chain-of-states methods, like NEB, can adjust to the available computational resources to efficiently predict the transition pathways with the required precision. The CPU and GPU implementation of PNEB has been made available in the pmemd module of the Amber molecular dynamics suite. Our new GPU-PNEB implementation offers 60 times faster sampling than a two-core CPU processor and 7 times greater than a 32-core CPU node. The usage of the nebfreq flag would allow for even faster simulations. The new implementation takes advantage of the efficient vectorization in commodity GPUs and is easily extendable to other chain-of-states methods for transition path sampling. Future generations of GPU architectures and libraries such as NVIDIA Collective Communications Library (NCCL) can potentially accelerate multi-GPU data transfers through higher bandwidth MPI point-to-point communication calls.

Acknowledgement

This research is part of the Blue Waters sustained-petascale computing project and the Extreme Science and Engineering Discovery Environment (XSEDE). Blue Waters is supported by the National Science Foundation (Award Nos.OCI-0725070 and ACI-1238993) and the state of Illinois. Blue Waters is a joint effort of the University of Illinois at Urbana-Champaign and its National Center for Supercomputing Applications. XSEDE is supported by National Science Foundation grant number ACI-1548562. This work was partially supported by the National Institutes of Health R01 GM107104 grant to C.S.

References

(1). Mckee ML; Page M Computing Reaction Pathways on Molecular Potential Energy Surfaces; John Wiley & Sons, Ltd, 2007; pp 35–65.

(2). Henkelman G; Jóhannesson G; Jónsson H Theoretical Methods in Condensed Phase Chemistry; Kluwer Academic Publishers: Dordrecht, 2002; pp 269–302.

(3). E. W; Vanden-Eijnden E Journal of Statistical Physics 2006, 123, 503–523.

(4). E W; Ren W; Vanden-Eijnden E Chemical Physics Letters 2005, 413, 242–247.

(5). Cerjan CJ; Miller WH The Journal of Chemical Physics 1981, 75, 2800–2806.

(6). Nguyen DT; Case DA The Journal of Physical Chemistry 1985, 89, 4020–4026.

(7). Quapp W Chemical Physics Letters 1996, 253, 286–292.

(8). Henkelman G; Jónsson H The Journal of Chemical Physics 1999, 111, 7010–7022.

(9). Malek R; Mousseau N Physical Review E 2000, 62, 7723–7728.

(10). Taylor H; Simons J The Journal of Physical Chemistry 1985, 89, 684–688.

(11). Baker J Journal of Computational Chemistry 1986, 7, 385–395.

(12). Sevick EM; Bell AT; Theodorou DN The Journal of Chemical Physics 1993, 98, 3196–3212.

(13). Gillilan RE; Wilson KR The Journal of Chemical Physics 1992, 97, 1757–1772.

(14). Jónsson H; Mills G; Jacobsen KW Nudged elastic band method for finding minimum energy paths of transitions. Classical and Quantum Dynamics in Condensed Phase Simulations. 1998; pp 385–404.

(15). E W; Ren W; Vanden-Eijnden E Physical Review B 2002, 66, 052301.

(16). Bergonzo C; Campbell AJ; Walker RC; Simmerling C International Journal of Quantum Chemistry 2009, 109, 3781–3790. [PubMed: 20148191]

(17). Herbol HC; Stevenson J; Clancy P Journal of Chemical Theory and Computation 2017, 13, 3250–3259. [PubMed: 28621935]

(18). Czerminski R; Elber R International Journal of Quantum Chemistry 1990, 38, 167–185.

(19). Choi C; Elber R The Journal of Chemical Physics 1991, 94, 751–760.

(20). Elber R; Karplus M Chemical Physics Letters 1987, 139, 375–380.

(21). Crehuet R; Field MJ The Journal of Chemical Physics 2003, 118, 9563–9571.
(22). Case D et al. University of California, San Francisco 2018,
(23). Kabsch W Acta Crystallographica Section A 1976, 32, 922–923.
(24). Henkelman G; Jónsson H The Journal of Chemical Physics 2000, 113, 9978–9985.
(25). Bergonzo C; Campbell AJ; de los Santos C; Grollman AP; Simmerling C Journal of the American Chemical Society 2011, 133, 14504–14506. [PubMed: 21848286]
(26). Li H; Endutkin AV; Bergonzo C; Fu L; Grollman A; Zharkov DO; Simmerling C Journal of the American Chemical Society 2017, 139, 2682–2692. [PubMed: 28098999]
(27). Maier JA; Martinez C; Kasavajhala K; Wickstrom L; Hauser KE; Simmerling C Journal of Chemical Theory and Computation 2015, 11, 3696–3713. [PubMed: 26574453]
(28). Jorgensen WL; Chandrasekhar J; Madura JD; Impey RW; Klein ML The Journal of Chemical Physics 1983, 79, 926–935.
(29). Li H; Endutkin AV; Bergonzo C; Campbell AJ; de los Santos C; Grollman A; Zharkov DO; Simmerling C Nucleic Acids Research 2016, 44, 683–694. [PubMed: 26553802]
(30). Chekmarev Dmitriy S.,; ; Tateki Ishida,; Levy* RM The Journal of Physical Chemistry B 2004, 108, 19487–19495.
(31). Nguyen H; Roe DR; Simmerling C Journal of Chemical Theory and Computation 2013, 9, 2020–2034. [PubMed: 25788871]
(32). Mathews DH; Case DA Journal of molecular biology 2006, 357, 1683–93. [PubMed: 16487974]
(33). Mongan J; Simmerling C; McCammon JA; Case DA; Onufriev A 2006,
(34). Torrie G; Valleau J Journal of Computational Physics 1977, 23, 187–199.
(35). Cheng X; Kelso C; Hornak V; de los Santos C; Grollman AP; Simmerling C 2005,
(36). Song K; Hornak V; de los Santos C; Grollman AP; Simmerling C 2006,
(37). Song K; Campbell AJ; Bergonzo C; de Los Santos C; Grollman AP; Simmerling C Journal of chemical theory and computation 2009, 5, 3105–13. [PubMed: 26609990]
(38). Le Grand S; Götz AW; Walker RC Computer Physics Communications 2013, 184, 374–380.

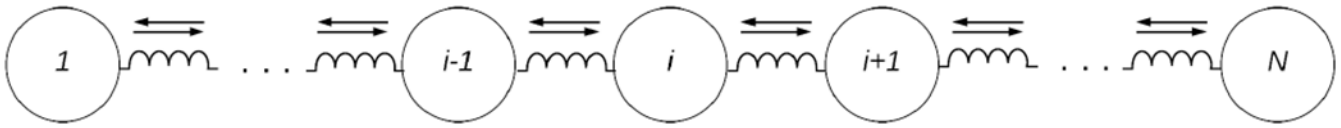


Figure 1:
Mass and spring representation of the nudged elastic band method. Each circle represents an individual simulation called replica which is bound to its nearest neighbors by harmonic potentials modeled as springs in the figure.

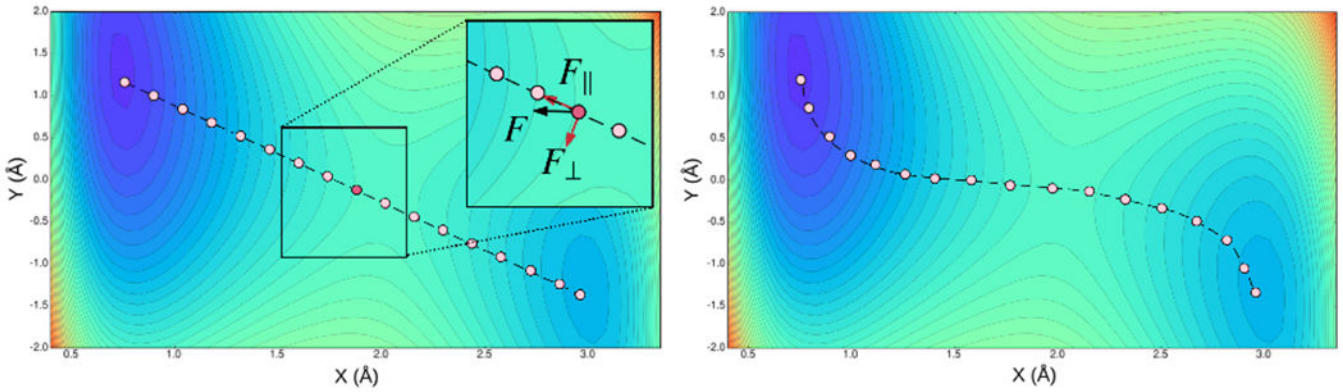


Figure 2:

Two-dimensional potential energy surface of LEPS potential coupled to a harmonic oscillator.[14] LEPS harmonic oscillator potential represents the energy of a system of four atoms. Atoms A and C are fixed. Atom B is allowed to move on the line connecting A and C. Another degree of freedom in terms of a harmonic oscillator is introduced by adding atom D that is coupled to B (refer to appendix A of reference 14). The X and Y axes in the figure represent the distance between atoms A and B and atoms B and D, respectively. The figure on the left shows a schematic representation of an initial path (the dashed straight line) and the replicas along that path (the circles). The force decoupling of one of the replicas is shown. The figure on the right represents a schematic MEP.

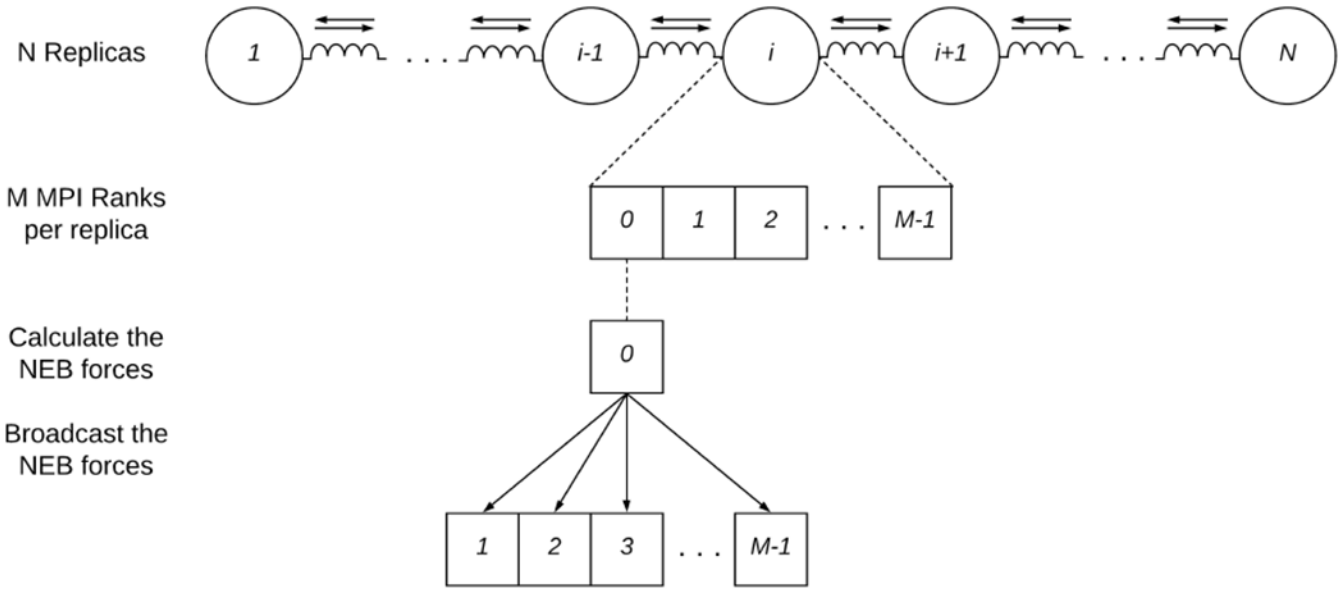


Figure 3:
Each NEB replica offloads its computation to M processes. The MPI process with rank zero calculates the NEB forces and broadcasts them to all other processes.

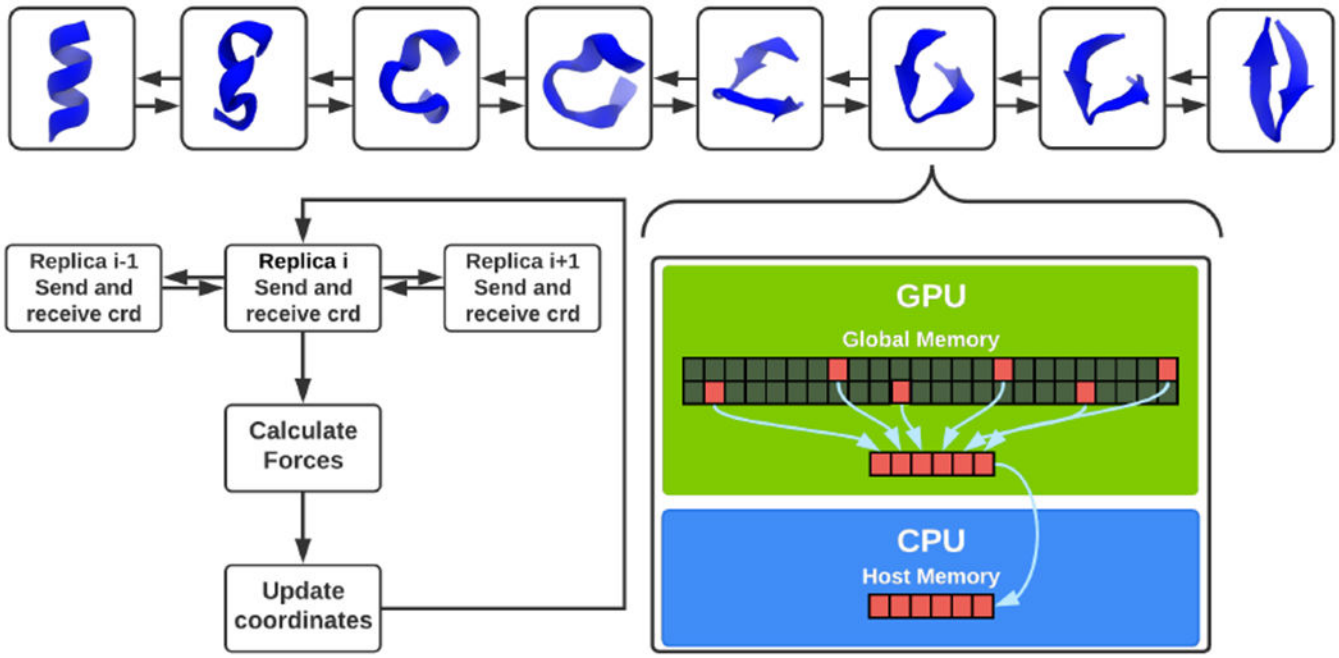


Figure 4:

The NEB replicas, which are shown with polyalanine complex transitioning from α-helix to β-sheet, communicate their coordinates (crd) with their nearest neighbors for harmonic force calculations. In the shuttle transfer scheme, only the coordinates of atoms that are needed in the NEB force calculations are being transferred between the host and the device.

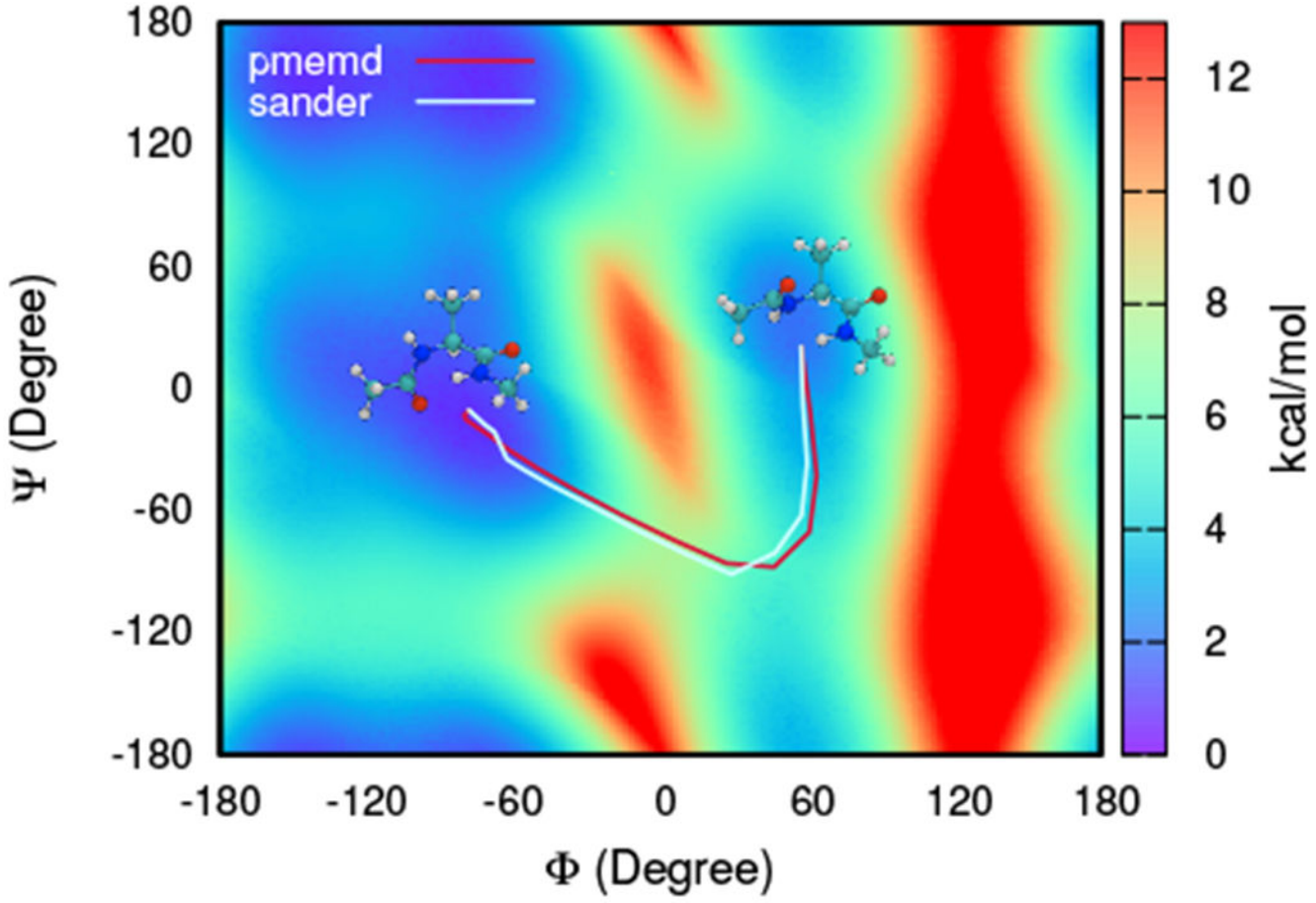


Figure 5:

Two-dimensional potential energy landscape of alanine dipeptide in the ϕ and ψ dihedral angles space. The initial and final conformations are displayed on top of the energy surface, positioned close to their corresponding minimum states. The bottom pathway indicates the averaged results from two independent sets of simulations performed with the pmemd (red) and the sander (light blue). The two transition pathways are lined up close to each other, indicating that the two implementations agree.

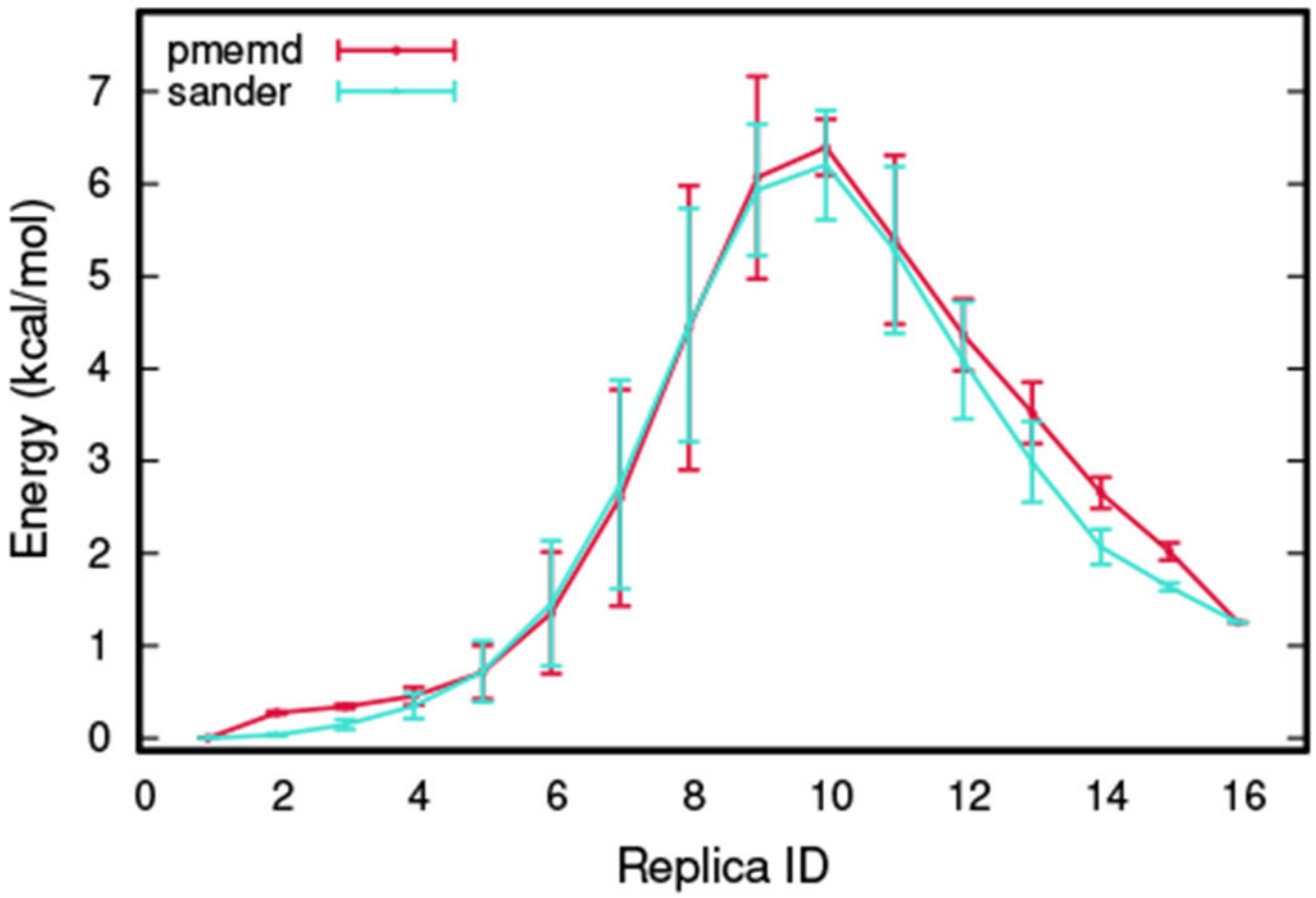


Figure 6:
The potential energy of the replicas, with the end replicas representing minimum states and replicas 9-11 representing the transition region. The light blue line represents simulations done with the sander and the red line represents simulations done with the pmemd-CPU. The lines in the plot are averaged over individual simulations and the vertical bars represent the standard deviations.

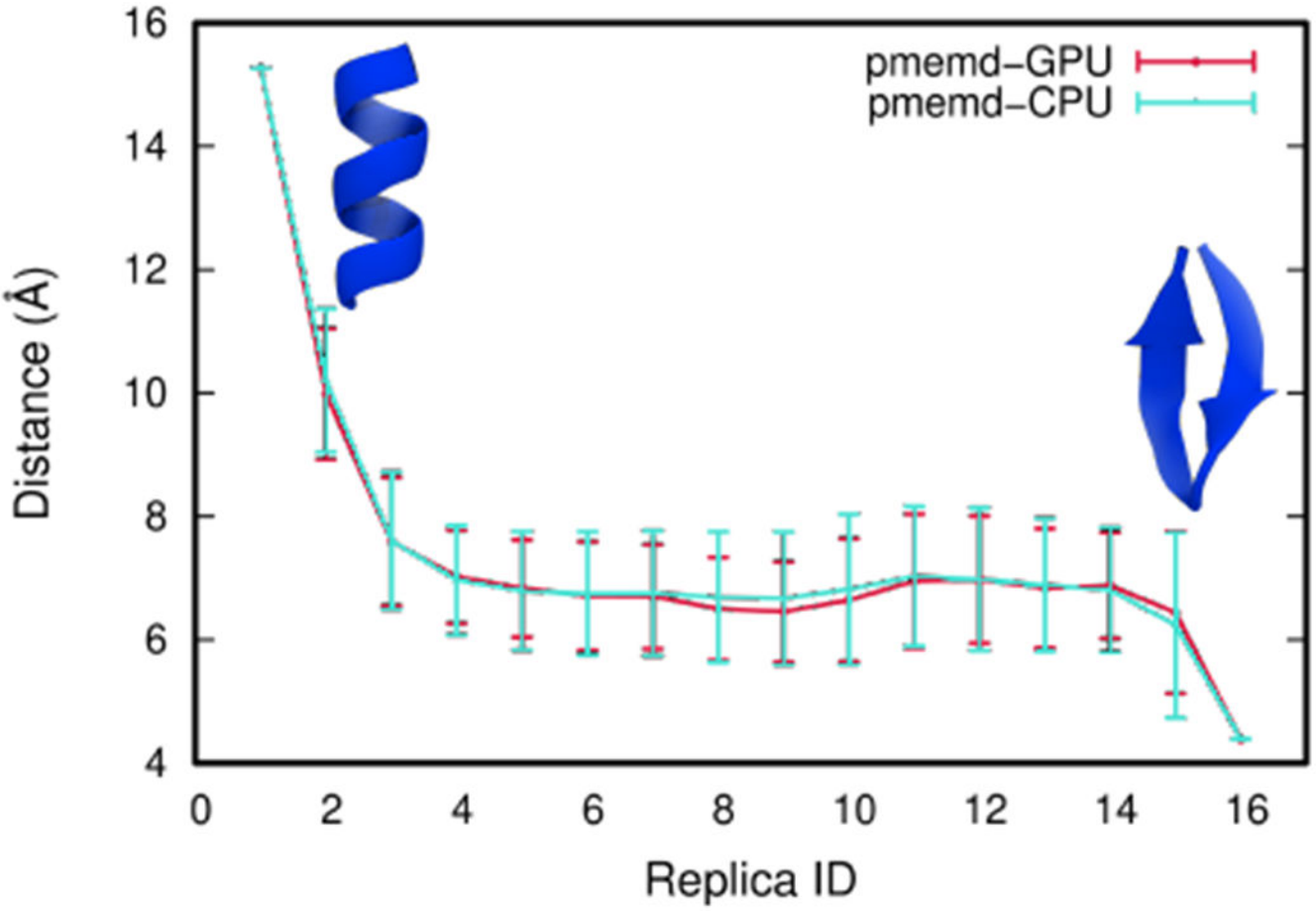


Figure 7:
End to end distance in the 12 alanine residue complex transitioning from α-helix to β-sheet. The blue line represents simulations done with pmemd-CPU while the red line shows simulations done with pmemd-GPU.

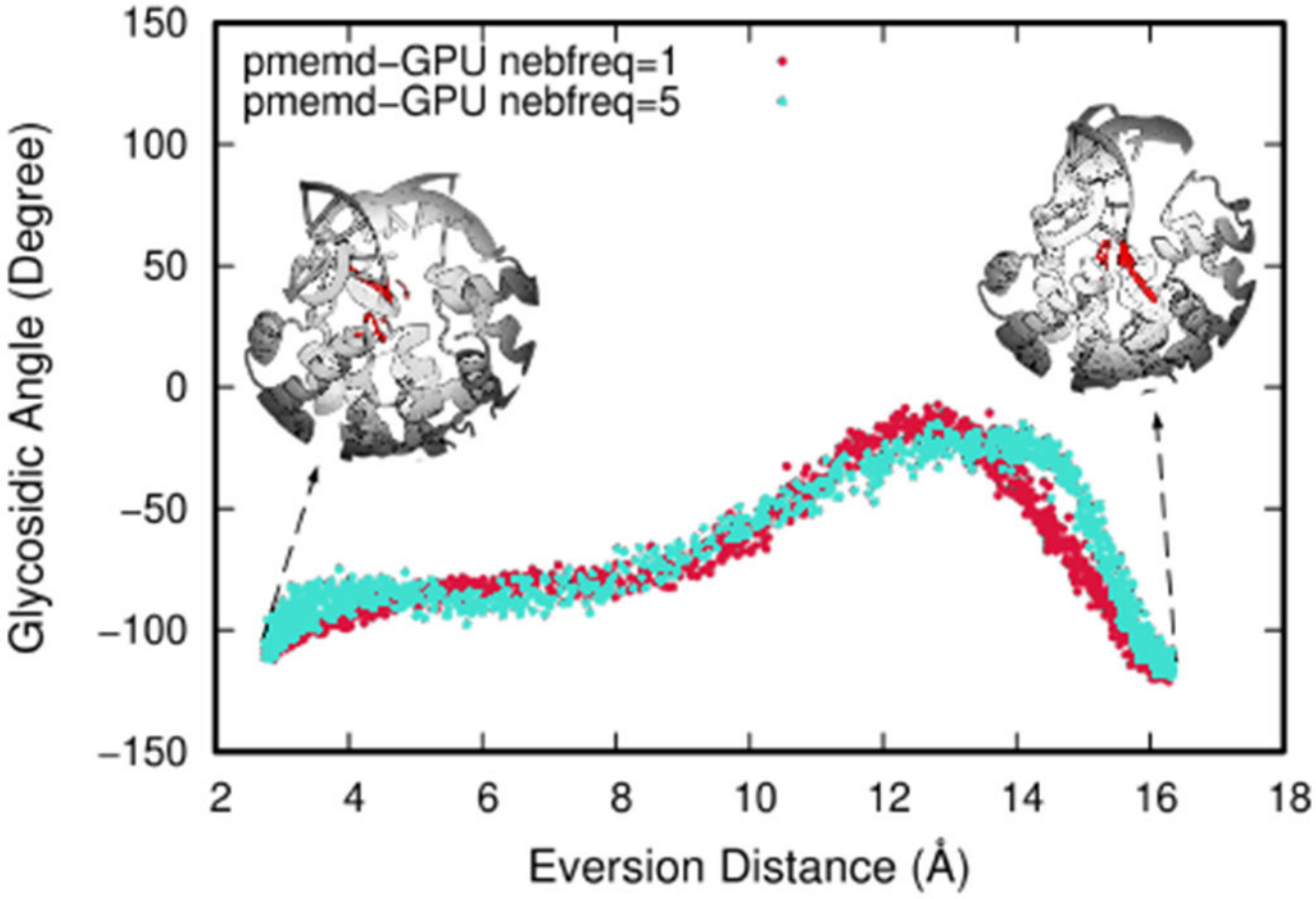


Figure 8:

Glycosidic angle vs. eversion distance involved in transitioning from intrahelical to extrahelical conformations in the OGG1–DNA complex. The blue points represent simulations performed with nebfreq=5 for all the stages but the first and the last, which had nebfreq=2. The red points represent simulations performed with nebfreq=1.

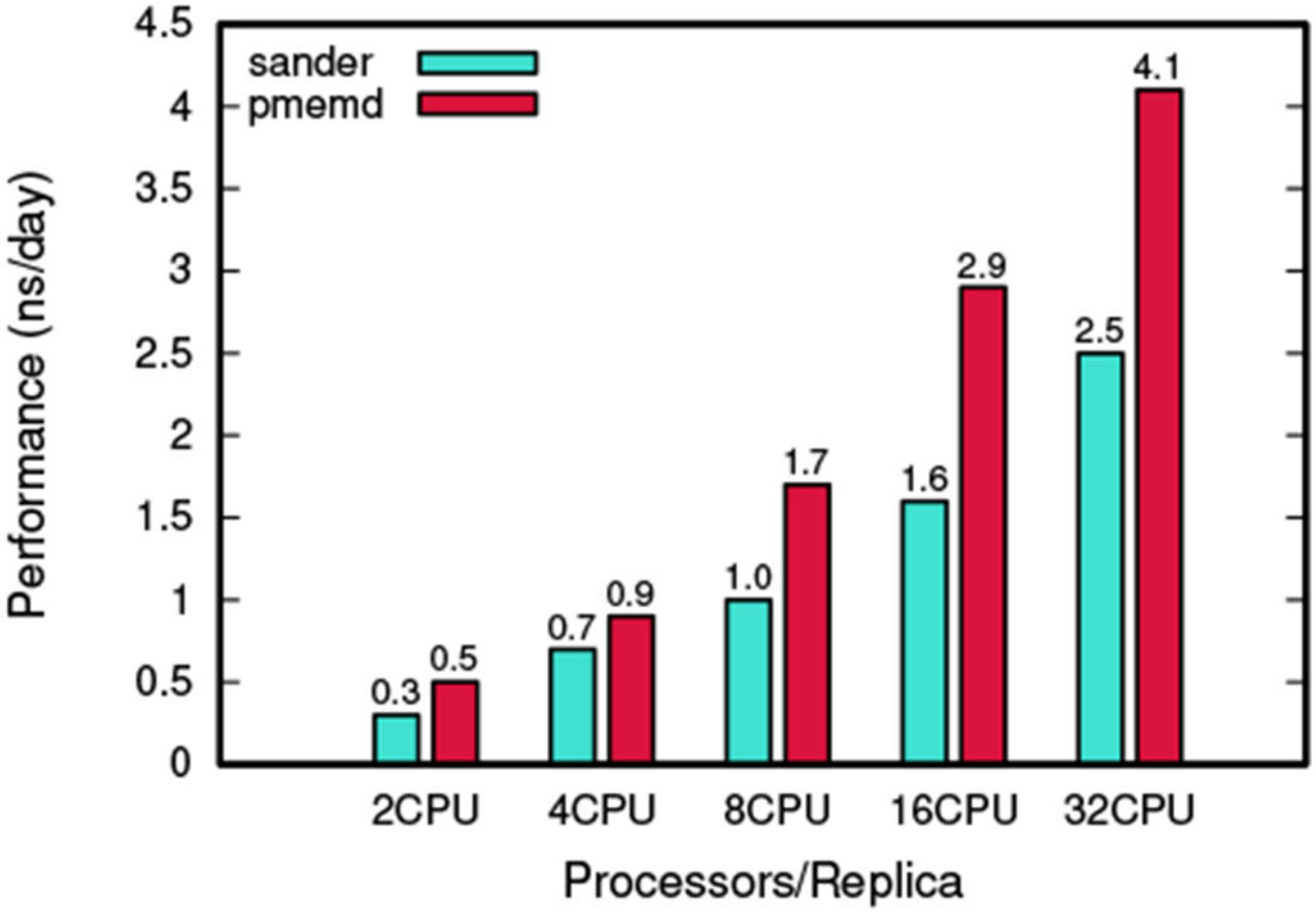


Figure 9:
Performance comparison between sander and pmemd. All simulations have been performed on CPUs.

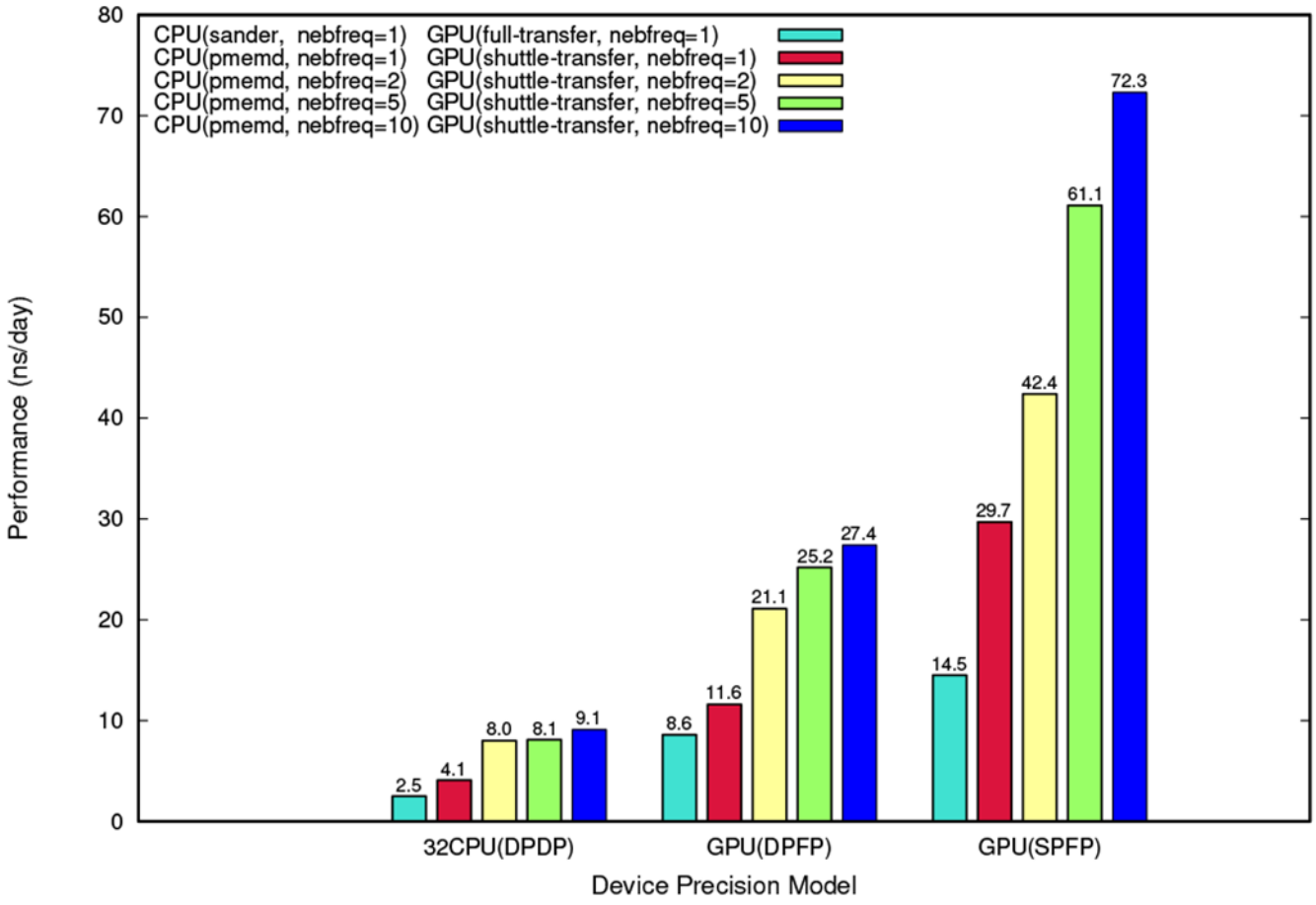


Figure 10:
Performance comparison for different nebfreq values for various CPU and GPU implementations.

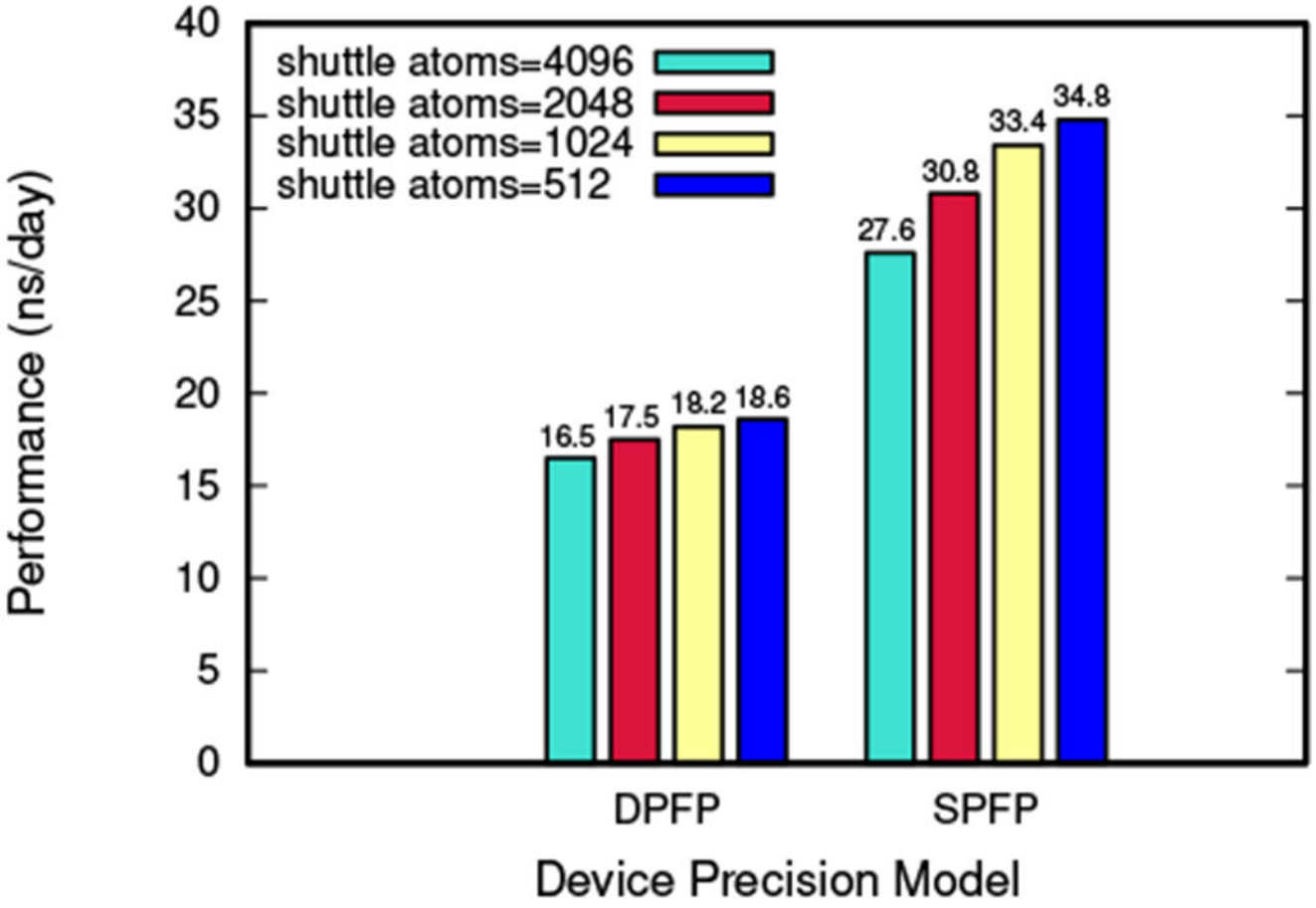


Figure 11:

Performance dependence of different GPU precision models on the size of the data transfers. Shuttle atoms specify the number of atoms for which their coordinates are transferred between the neighboring replicas.