



Published in final edited form as:

Comput Phys Commun. 2014 March ; 185(3): 908–916. doi:10.1016/j.cpc.2013.12.014.

Generalized Scalable Multiple Copy Algorithms for Molecular Dynamics Simulations in NAMD

Wei Jiang^{1,*}, James C. Phillips^{4,§}, Lei Huang³, Mikolai Fajer³, Yilin Meng³, James C. Gumbart^{2,6}, Yun Luo^{1,¶}, Klaus Schulten^{4,5,*}, and Benoît Roux^{2,3,*}

¹Argonne Leadership Computing Facility, Argonne National Laboratory, 9700 South Cass Avenue, Building 240, Argonne, Illinois 60439

²Biosciences Division, Argonne National Laboratory, 9700 South Cass Avenue, Building 202, Argonne, Illinois 60439

³Department of Biochemistry and Molecular Biology, Gordon Center for Integrative Science, University of Chicago, 929 57th Street, Chicago, Illinois 60637

⁴Beckman Institute, University of Illinois at Urbana–Champaign, Urbana, Illinois 61801

⁵Department of Physics, University of Illinois at Urbana–Champaign, Urbana, Illinois 61801

⁶School of Physics, Georgia Institute of Technology, Atlanta, Georgia 30332

Abstract

Computational methodologies that couple the dynamical evolution of a set of replicated copies of a system of interest offer powerful and flexible approaches to characterize complex molecular processes. Such multiple copy algorithms (MCAs) can be used to enhance sampling, compute reversible work and free energies, as well as refine transition pathways. Widely used examples of MCAs include temperature and Hamiltonian-tempering replica-exchange molecular dynamics (*T*-REMD and *H*-REMD), alchemical free energy perturbation with lambda replica-exchange (FEP/ λ -REMD), umbrella sampling with Hamiltonian replica exchange (US/*H*-REMD), and string method with swarms-of-trajectories conformational transition pathways. Here, we report a robust and general implementation of MCAs for molecular dynamics (MD) simulations in the highly scalable program NAMD built upon the parallel programming system Charm++. Multiple concurrent NAMD instances are launched with internal partitions of Charm++ and located continuously within a single communication world. Messages between NAMD instances are passed by low-level point-to-point communication functions, which are accessible through NAMD's Tcl scripting interface. The communication-enabled Tcl scripting provides a sustainable application interface for end users to realize generalized MCAs without modifying the source code. Illustrative applications of MCAs with fine-grained inter-copy communication structure, including global lambda exchange in FEP/ λ -REMD, window swapping US/*H*-REMD in multidimensional order parameter space, and string method with swarms-of-trajectories were

*Corresponding authors roux@uchicago.edu, kschulte@ks.uiuc.edu, wjiang@alcf.anl.gov.

¶Current address: Pharmaceutical Sciences Department, College of Pharmacy, Western University of Health Sciences, Pomona, California

§Co-first authors.

carried out on IBM Blue Gene/Q to demonstrate the versatility and massive scalability of the present implementation.

Keywords

MCA; NAMD; Tcl; charm++

1. Introduction

A long-standing challenge for molecular dynamics (MD) simulations based on all-atom models is the difficulty to adequately sample infrequent events, although what has routinely been attainable within typical MD simulations has evolved over the years. With the advancement of novel accelerating computer chips and supercomputers that consist of MD-specific integrated circuits,[1] simple brute-force MD simulations on the order of milliseconds have recently become possible. This is remarkable when compared with what was the state-of-the-art only barely one decade ago. However, the fact remains that a whole host of important biological processes take place on a time scale ranging from milliseconds to seconds, beyond what can be achieved with simple brute-force MD. For this reason, it remains critical to continue to seek alternative solutions to the sampling problem.

One broad computational strategy is offered by algorithms based on multiple copies. Rather than attempting to characterize the system through a single and extremely long trajectory, such multiple copy algorithms (MCAs) adopt a ‘divide-and-conquer’ strategy to carry out the desired computation using massive parallel resources. MCAs can be designed to calculate both equilibrium and non-equilibrium properties.[2–8] For equilibrium properties, the most widely used MCAs include parallel temperature [9–17] and Hamiltonian tempering with replica-exchange MD (REMD) simulations.[18–25] There are also several variants of MCAs focused on non-equilibrium situations, such as the forward-flux sampling,[26] the string method with swarms-of-trajectories,[5] and time-dependent umbrella sampling.[7, 8] MCAs offer some of the most extremely scalable MD methodologies, allowing one to fully exploit the unprecedented computational power of leadership computers such as IBM Blue Gene/Q and Cray XT7. As the most popular applications of MCAs, REMD algorithms have been extensively implemented in several MD engines such as AMBER,[27] GROMACS[28] and Desmond[29] with special application interface (API) as well as post-processing utility. Generally, the REMD algorithms are implemented as regular point-to-point data exchange that is hard coded in source code, or driven by an external master script working with specific machines. As a result, the existing REMD schemes are restricted to a number of plain applications, such as parallel tempering or Hamiltonian exchange with a single biasing parameter, while implementation of any more elaborate algorithm has to touch the low-level MPI programming. However, biological MD simulations often involve several slow degrees of freedom that have to be addressed in multidimensional space of order parameters,[30–33] or multiple thermodynamic coordinates[34–36] to accelerate the sampling efficiency of a reaction path. In addition to the demand of multidimensional REMD algorithms, there can arise an irregular inter-copy communication structure for a complex biophysical problem,

such as case of string method with swarms-of-trajectories,[5, 37] raising the necessity to develop a sustainable API to accommodate various novel MCA solutions.

A prototypical example of a MCA as a high-dimensional Hamiltonian exchange algorithm was recently developed and implemented in the Distributed Replica (REPDSTR) module of CHARMM to calculate the absolute binding free energy of a ligand to a protein.[3, 4] With the 2-dimensional Hamiltonian replica-exchange scheme, an unprecedented global replica exchange through the entire reaction path was performed and accurate free energy information was obtained. Within REPDSTR, coordinates instead of biasing/scaling parameter were swapped during exchange attempts, relying on the legacy parallel structure of CHARMM based on a static task distribution and atomic decomposition.[38] Following a similar strategy, a variety of MCAs could be implemented within the CHARMM/REPDSTR code. However, due to the limitations of the internal parallel structure of CHARMM, this initial implementation of MCAs has been limited to small and moderate size biological systems. To allow scalable simulation of increasingly large systems, the development of contemporary MD software has been directed towards spatial decomposition and dynamic load balance.[28, 39] In spite of its limitations, CHARMM/REPDSTR provides a prototype REMD scheme independent of specific potential terms or parameters and, therefore, can serve to guide further extension in high performance MD software such as NAMD.

NAMD is a highly scalable program built onto a hybrid spatial/force decomposition managed by the asynchronous parallel programming system Charm++.[39] An attractive feature of NAMD is the flexible Tcl interface that provides a user-friendly scripting platform, where one can control simulation parameters on-the-fly. Furthermore, Tcl communication commands can be built on top of low-level communication functions, while the Charm++ programming system can support concurrent multiple NAMD instances by remapping processing elements. With these advanced programming features, NAMD provides an ideal platform to implement a highly flexible API for generalized MCAs.

In this article, we report a systematic implementation of MCAs in NAMD's Tcl interface and Charm++. Implementation details are reported and four representative applications, parallel tempering, free energy perturbation (FEP)/REMD, US/H-REMD with irregular order parameter locations in 2-dimensions, as well as the string method with swarms-of-trajectories are carried out on the IBM Blue Gene/Q computer at Argonne National Laboratory. The superior scalability of the present MCA implementation is demonstrated and discussed in detail.

2. Implementation Details

2.1. MCAs Implementation with Charm++ and NAMD Tcl interface

A primitive MCA can be realized with multiple MPI_subcommunicators. In the MPI machine layer of Charm++, the MPI_Comm_split function splits the default MPI_COMM_WORLD into multiple local sub-communicators, each of which runs an independent Charm++ and NAMD instance. A secondary set of MPI communicators, each spanning like ranks within the local subcommunicators, allow exchanges between the independent NAMD instances. This exchange communication is implemented through new

APIs in both the Charm++ Converse layer and the NAMD Tcl scripting interface to provide a user-friendly interface. As a quick route towards REMD, MPI_subcommunicator is extensively adopted by many MD software, however such a plain application of many MPI_subcommunicators could lose significant network topology information and, therefore, induce further network contention or latency. However, within the Charm++ Converse layer instead of MPI_COMM_WORLD, each Charm++ processing element can be logically mapped onto a designated local partition. When Charm++/NAMD enters the inter-copy communication phase, all localized processing elements are mapped back to global state. In addition, on leadership supercomputers, such internal partitions can obtain further performance gain benefiting from the low-level machine specific communication library, such as Parallel Active Messaging Interface (PAMI)[40] on IBM Blue Gene/Q or user Generic Network Interface (uGNI)[41] on Cray XK7. Tcl interface of NAMD intends to provide maximum flexibility for a high-end user to realize its special needs without touching the source code. For example, by Tcl scripting, variables and expressions used in initially defining options can be changed during a running simulation. After user-friendly Tcl communication commands are built on top of the low-level point-to-point communication functions (either Charm++ Converse or MPI), a user can further design generic MCAs without modifying or adding a single line of C++. Within a MCA Tcl script, inter-copy communications are executed by Tcl replicaSend/Recv/Sendrecv functions on top of Converse communication layer of Charm++, and a user needs to designate communication partners for each copy. Appendix A exhibits an excerpt of Tcl scripting code that controls temperature swaps in a parallel tempering MD simulation. A significant advantage of the present Tcl-based MCA scheme is that a user can realize any type MCA within Tcl scripting, as long as the energy terms or parameters to be biased are registered in the Tcl scripting interface of NAMD. For example, in an absolute free energy calculation, nonbond scaling parameters of different alchemical types can be wrapped into a single parameter unit to be exchanged along the entire alchemical reaction path, or multiple orthogonal order parameters can be alternatively exchanged to form a multidimensional umbrella sampling Hamiltonian exchange (see Section Applications).

2.2. Performance on IBM Blue Gene/Q

The string method with swarms-of-trajectories[5, 37, 42–44] represents a general, but challenging MCA due to the massive number of concurrent trajectories involved and the multi-level communication relations between them. Previously such a sophisticated ensemble task had to be carried out by launching many concurrent NAMD jobs and the communications between jobs were driven by external scripts. At the communication phase of swarms-of-trajectories or images, NAMD jobs have to stop and restart periodically, which results in significant exit/restart overhead. In contrast, a Tcl script wraps all trajectories of an ensemble task into a single large NAMD run, the communications wherein are managed by Charm++ at minimal overhead, such that there is little performance loss compared to a plain single trajectory run. Figure 2 shows the superior scalability of refining the transition path of c-Src kinase with 64 images and 2048 trajectories (see details in sub-section D of Application). It can be seen that the computation strongly scales to 32 Blue Gene/Q racks (524,288 IBM PowerPC A2 cores). In this article, all applications were

carried out on IBM Blue Gene/Q Mira at Argonne National Laboratory using Charm++ 6.5.0 and NAMD 2.10.

3. Applications

3.1. Parallel Tempering (T-REMD)

T-REMD originally introduced in 1999 by Sugita and Okamoto[9] is probably the most widely used MCA. Its aim is to accelerate the configurational space sampling efficiency by overcoming potential barriers with rescaled kinetic energy. Implementation of the T-REMD algorithm in Tcl language is straightforward as its temperature exchange attempts between replicas are simple to state algorithmically. However, efficient exchange of T-REMD require sufficient potential overlap of the energy distributions between neighboring replicas. [45] As a consequence, the number of required replicas grows approximately with the square root of the number of simulated particles. Also high exchange frequency is required to assure enough travelling of a replica through the temperature parameter space.[19, 20] Computationally, the large number of replicas and high frequency exchange attempts requires massively parallel computers with a low-latency network to attain optimal performance. One common approach to avoid the demanding resource requirement is to adopt an implicit solvent model at the expense of significant loss of accuracy.[14, 15] T-REMD simulation of peptide acetyl-(AAQAA)₃-amide[46] in TIP3 solvent was performed on Blue Gene/Q to demonstrate a non-trivial application for a medium size system.

For such an explicitly solvated peptide with helix structure, normal single trajectory MD runs tend to be trapped in the initial conformation. In the present case the simulated system contains ~25,000 atoms and 64 replicas spanning a 278–375 K temperature range were chosen to obtain an average acceptance ratio 45%. The locations of temperature parameters

for the replicas obey a logarithmic spacing $T_i = T_0 * \exp(\ln \frac{T_{max}}{T_0} \div * \frac{i}{nreplicas - 1})$. An exchange frequency of 1/100 steps was adopted to achieve an optimal rate of exploration of configurational space. The T-REMD simulation was performed under constant pressure and periodic boundary condition and employing the CHARMM 36 force field. 32,768 cores were used to launch the generic parallel/parallel T-REMD simulation. Figure 3 demonstrates that with high exchange frequency the 8th replica (close to physical temperature) thoroughly samples the temperature space 5 times within 10ns. A similar sampling efficiency was achieved for higher temperature replicas. Higher sampling efficiency of temperature space can be achieved with higher exchange frequency, but at a considerably increased communication cost. How to balance the exchange frequency and communication cost is an active research direction. The acceleration effect of parallel tempering on conformational sampling of the peptide can be observed from the simulated temperature dependence of helix content. In Figure 4, it can be seen that after 25 ns T-REMD run, the helix contents of different temperatures have reached convergence.

3.2. Free Energy Perturbation/ λ -Exchange Molecular Dynamics (FEP/ λ -REMD)

In our previous REMD implementation in CHARMM, free energy perturbation (FEP) with a staged reversible thermodynamic work protocol designed for the calculation of absolute

ligand binding affinities was combined with a distributed replica exchange MD simulation scheme.[2, 3] It was shown that this FEP/REMD scheme could improve the statistical convergence of FEP calculations by allowing random Monte Carlo moves in an extended ensemble of thermodynamic coupling parameter λ . The staging simulation protocol can be illustrated with the binding free energy simulation: the potential energy is expressed in terms of four coupling (window) parameters[34, 35, 47]

$$U(\lambda_{\text{rep}}, \lambda_{\text{dis}}, \lambda_{\text{elec}}, \lambda_{\text{rstr}}) = U_0 + U_{\text{rep}}(\lambda_{\text{rep}}) + \lambda_{\text{dis}} U_{\text{dis}} + \lambda_{\text{elec}} U_{\text{elec}} + \lambda_{\text{rstr}} U_{\text{rstr}} \quad (1)$$

where U_0 is the potential of the system with the noninteracting ligand, $\lambda_{\text{rep}}, \lambda_{\text{dis}}, \lambda_{\text{elec}}, \lambda_{\text{rstr}} \in [0,1]$ are the thermodynamic coupling parameters, U_{rep} and U_{dis} are the shifted Weeks-Chandler-Anderson (WCA)[48] repulsive and dispersive components of the Lennard-Jones potential, U_{elec} is the electrostatic contribution and U_{rstr} is the restraining potential. The insertion of the molecule into binding pocket is done in three steps, with the help of three thermodynamic coupling parameters, $\lambda_{\text{rep}}, \lambda_{\text{dis}}$, and λ_{elec} , controlling the nonbonded interaction of the molecule with its environment. One additional parameter, λ_{rstr} , is used to control the translational and orientational restraints. As the first step, the WCA decomposition algorithm that finely switches on interaction is implemented into NAMD's alchemical module. As these intermediate thermodynamic states along the alchemical reaction path contains sampling information of different time scales, it is highly desirable to perform a global exchange along the entire path. However, global lambda exchange requires exchange attempts across multiple alchemical stages, which is beyond a naïve REMD implementation that only exchanges a single type of scaling parameter for a specific energy term. In the CHARMM/REPDSTR implementation, coordinate sets of neighboring replicas were swapped and the complexity of parameter swaps was avoided at the expense of higher communication. In the Tcl scripting language of NAMD, each replica can be identified with a unique set of alchemical parameters including lambda value, interaction type and WCA parameters, and therefore a global lambda-exchange is obtained by swapping these parameter sets. When Hamiltonian-exchange is performed, the whole set of alchemical parameters are swapped between neighboring replicas, realized with point-to-point communications of derived data type. The lambda-exchange algorithm follows the conventional Metropolis MC exchange criterion with λ -swap moves. After a FEP/REMD run is done, the shuffled energy outputs due to parameter sets swapping are sorted for the final WHAM[49] post-processing. As a quick test, the absolute binding free energy of benzene to T4/Lysozyme-L99A was calculated with the FEP/REMD scheme. 64 replicas are evenly located along the alchemical reaction path, employing 36 repulsive windows, 12 dispersion windows and 16 electrostatic windows. The average acceptance ratio is > 80% and figure 5 exhibits how frequently a replica traveled through the space of alchemical parameter sets during the REMD run. With an exchange attempt frequency of 1/100 steps and production length of 500 ps (250,000 steps), the free energy simulated is -6.0 ± 0.5 kcal/mol under constant NPT and periodic boundary condition, in agreement with experiment and previous simulations.[35]

3.3. Multiple Dimensional Hamiltonian Replica Exchange MD Umbrella Sampling (US/H-REMD) on Irregular-Shaped Distribution of Umbrella Windows

A commonly used approach for determining a potential of mean force (PMF) along a given reaction coordinate is conventional umbrella sampling (US).[30, 50, 51] Within a stratification strategy, US relies on a series of closely spaced windows along the coordinate, each with the collective variable (CV) of interest harmonically restrained to the window's center. The PMF is then recovered from the biased distributions in each window via the WHAM method.[49] As with any enhanced sampling method, a strict limitation of US lies in the inability of the system to efficiently sample those orthogonal degrees of freedom not targeted by the CV. This limitation can be mitigated, however, by permitting the biasing parameters of adjacent windows to be exchanged according to a Metropolis energy criterion, [52] thus allowing barrier-crossing events along orthogonal degrees of freedom that occur in one window to be progressively communicated and propagated to all other windows. Similar to the T-REMD algorithm, the US/H-REMD along a 1-dimensional order parameter is straightforwardly implemented in NAMD using its Tcl scripting interface and has been applied to binding-free-energy calculations for two systems recently. The first, binding of a small peptide to the SH3 domain of Abl kinase, demonstrated a modest improvement of US/H-REMD over conventional US, with convergence of the latter requiring approximately twice as much simulation time as the former.[53] The second system, the bound state of two similarly sized proteins, barstar and barnase, presented a greater challenge.[54] Even with restraints on interfacial side chains, the complexity of the protein-protein interface made US/H-REMD the most viable approach, reaching convergence already within 2 ns/window, i.e., within 100 ns overall.[54]

In multi-dimensional US/H-REMD implementations, maintaining a regular shape of umbrella windows becomes increasingly computationally inefficient as a regular shaped distribution of windows often covers high free energy regions that are irrelevant. This problem can be circumvented by adopting the self-learning adaptive umbrella sampling protocol[33], which automatically generates the US windows only for the relevant region in the subspace of CVs. However, windows distributed within an irregular shape poses a particular challenge to the implementation of an efficient replica-exchange US scheme, as all the copies must be matched with a swapping partner for efficient window swapping. In this article, a distribution of umbrella windows in CV space is described as having regular shaped if the distribution has 2^N right angles, where N is the dimensionality of the CV space. For example, in the case of two-dimensional US, a rectangular distribution of windows is regular shaped but a triangular distribution is not. This leads to the necessity of developing a general and efficient US/H-REMD algorithm for irregular distribution of umbrella windows. Such an algorithm should aim to keep the Euclidean distance between a pair of exchanging partners as short as possible to maintain a maximal acceptance ratio. To resolve this problem, ordered lists of umbrella windows are constructed to assign neighbors for each replica following the minimal distance rule. A single list such as used in the simple 1D US/H-REMD naturally allows one replica to have at most 2 nearest neighbors. But this is too restrictive in the case of umbrella sampling in multi-dimensions. For an N -dimensional US/H-REMD simulation with M umbrella windows in total, it is possible to generate N distinct ordered lists of the M windows by “weaving” a 1D chain through the

irregular shape (Figure 6). The only difference between the N lists is the order in which the M windows appear in the list. Exchanges are then attempted with nearest neighbors in each 1D list according to the rule odd \leftrightarrow even and even \leftrightarrow odd. By constructing these 1D ordered lists, the number of moves between two different windows is increased. For a set of CVs (x_1, x_2, \dots, x_N), N ordered sets of CVs are created based on the following algorithm: starting from an ordered set $[x_1, x_2, \dots, x_N]$ a new ordered set is generated by popping the left-most CV and pushing it to the right; this process is repeated until $[x_N, x_1, \dots, x_{N-1}]$ is reached. Each ordered list of M windows is created associated with one ordered set of CVs such that a decreasing speed of changing is employed from the left element to the right in the set. An illustration of constructing ordered lists for two-dimensional US/H-REMD with windows distributed as a parallelogram is given in Figure 6.

As an illustration, US/H-REMD simulation was performed in 2D to study the activation/deactivation conformational transition in c-Src kinase domain.[55] The conformational transition in kinase domain primarily involves the α C helix and the activation loop (Figure 7) and, two order parameters that characterize the conformational transition were identified based on the relative motions between relevant atom groups. Two-dimensional self-learning adaptive umbrella sampling[33] calculations are carried out before US/H-REMD calculations in order to define the landscape to be explored by US/H-REMD. 1097 umbrella windows are determined to be essential to characterize the landscape and are further utilized in the US/H-REMD simulation. The two ordered lists utilized in the two-dimensional US/H-REMD calculations are illustrated in Figure 8. Each umbrella window is propagated for 5 ns under constant NPT and periodic boundary condition, and an exchange is attempted every 1 ps. The probability distribution of acceptance ratio in the calculation of c-Src kinase domain conformational transition is given in Figure 9A. The average acceptance ratio is 0.49 and the two extremes are 0.36 and 0.68 for all neighboring pairs. Figure 9B shows the trajectory of replica 435 in the replica space against exchange attempts (time). With the large acceptance ratio, a wide range of replicas (a span of >350 replicas) has been traversed for multiple times, however the entire replica space still is not fully covered. This indicates that either a sampling time of 5 ns per window is not long enough or the exchange frequency is too low, and the PMF could still be refined.

3.4. String Method using Swarms-of-Trajectories

A putative transition path between conformational states can be refined by the swarms-of-trajectories approach.[5, 37] The string is a set of discrete conformations, or images, along the path. A set of collective variables is typically used to reduce the complexity of the system. The free energy gradient at each image can be computed and the images propagated along the component orthogonal to the path, iterating until the string has converged to a local minimum transition path. The swarms-of-trajectories approach utilizes many molecular dynamics trajectories started from a single image that can be combined to yield an average drift. This requires communication between the N copies of a specific image, illustrated by Figure 10. The average drifts for each image compose an update string, and the images are redistributed along this string to remove any drift tangential to the path. The string update process requires communication between the M images that compose the string. After the reparameterization with the current average drifts the new images must be prepared before

the next cycle of drifts can be performed, achieved by performing molecular dynamics simulations with constraints that slowly move the CVs to their reparameterized positions. The multiple copy implementation of this method utilizes MN independent copies that must intercommunicate once per iteration.

In this article, the string method with swarms-of-trajectories was used to refine the activation/deactivation conformational transition in c-Src kinase domain (identical with the sub-section C) in the presence of the regulatory SH2/SH3 domains. The full-length c-Src structure is threaded along the initial string taken from previous work of the kinase domain only. This threading is composed of restraining the CVs of the full-length to the string values and slowly moving those restraints along the string. A total of 64 images are selected from the initial string. Each swarm composes of 32 5-ps trajectories, and the equilibration following the reparameterization is another 5-ps of restrained molecular dynamics.

The discrete Fréchet distance is an order-preserving measure of the similarity between two strings. The distance between each string iteration to the initial string and the distance to the final strings is shown in Figure 11 and are either monotonically increasing or decreasing, respectively. The monotonicity indicates that the swarms-of-trajectories method is performing a steepest descent minimization of the transition path. Furthermore the flattening of the discrete Fréchet distance to the final string around iteration 250 suggests that the transition path is nearly converged. The converged transition path can be studied directly, but is most useful as a starting point for free energy calculation like the US/H-REMD discussed above.

4. Conclusion

We have developed and implemented generalized multiple copy algorithms (MCAs) in the highly scalable program NAMD built upon the parallel programming system Charm++. Inter-copy communication is realized by low-level point-to-point communication functions between independent NAMD instances. At the application level the MCAs are implemented via NAMD's Tcl scripting interface, which provides great flexibility for a user to design novel MCA solutions for complex biophysical problems without modifying the source code. The application of MCAs could be extended to multiple thermodynamic parameters (or reaction coordinates), multi-dimension and irregular communication between copies. We have demonstrated the versatility of the present implementation with three novel MCA applications, global lambda exchange along an entire reaction path in an absolute binding free energy calculation, high-dimensional umbrella sampling with irregular reaction coordinates, and string method with swarms-of-trajectories. Novel applications to complex biological problems are currently underway.

Acknowledgments

We would like to acknowledge Parallel Programming Laboratory, University of Illinois at Urbana-Champaign, for the implementation of Charm++ on IBM Blue Gene/Q. This research is supported by Early Science Program of Argonne Leadership Computing Facility, Department of Energy Office of Science and used resources of the Argonne Leadership Computing Facility at Argonne National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under contract DE-AC02-06CH11357. This work also is supported by the National Institutes of Health through grants 9P41GM104601 (K.S. and J.P.), U54GM087519 (K.S. and B.R.), and K22-AI100927 (J.C.G.) and by MCB-0920261 (B.R.) from the National Science Foundation (NSF).

References

1. Shaw DE, et al. Anton, A Special-Purpose Machine for Molecular Dynamics Simulation. *Communication of the ACM (ACM)*. 2008; 51:91–97.
2. Jiang W, Hodoscek M, Roux B. Computation of absolute hydration and binding free energy with free energy perturbation distributed replica-exchange molecular dynamics. *J Chem Theory Comput*. 2009; 5:2583–2588. [PubMed: 21857812]
3. Jiang W, Roux B. Free Energy Perturbation Hamiltonian Replica-Exchange Molecular Dynamics (FEP/H-REMD) for Absolute Ligand Binding Free Energy Calculations. *J Chem Theory Comput*. 2010; 6(9):2559–2565. [PubMed: 21857813]
4. Jiang W, et al. Calculation of Free Energy Landscape in Multi-Dimensions with Hamiltonian-Exchange Umbrella Sampling on Petascale Supercomputer. *J Chem Theory Comput*. 2012; 8:4672–4680.
5. Pan AC, Sezer D, Roux B. Finding the Transition Pathways Using the String Method with Swarms of Trajectories. *J Phys Chem B*. 2008; 112:3432–3440. [PubMed: 18290641]
6. Bhimalapuram WP, Dinner AR. Nonequilibrium umbrella sampling in spaces of many order parameters. *J Chem Phys*. 2009; 130:074104. [PubMed: 19239281]
7. Dickson A, Dinner AR. Enhanced sampling of nonequilibrium steady states. *Annu Rev Phys Chem*. 2010; 61:441–459. [PubMed: 20367083]
8. Dickson A, et al. Flow-dependent unfolding and refolding of an RNA by nonequilibrium umbrella sampling. *J Chem Theory Comput*. 2011; 7:2710–2720.
9. Sugita Y, Okamoto Y. Replica-exchange molecular dynamics method for protein folding. *Chem Phys Lett*. 1999; 314:141–151.
10. Sugita Y, Kitao A, Okamoto Y. Multidimensional replica-exchange method for free-energy calculations. *J Chem Phys*. 2000; 113(15):6042–6051.
11. Mitsutake A, Sugita Y, Okamoto Y. Generalized-ensemble algorithms for molecular simulations of biopolymers. *Biopolymers*. 2001; 60:96–123. [PubMed: 11455545]
12. Mitsutake A, Okamoto Y. From multidimensional replica-exchange method to multidimensional multicanonical algorithm and simulated tempering. *Phys Rev E*. 2009; 79:047701.
13. Sanbonmatsu K, Garcia A. Structure of Met-enkephalin in explicit aqueous solution using replica exchange molecular dynamics. *Proteins: Struct, Funct, Bioinf*. 2002; 46:225–234.
14. Zhou R, Berne B. Can a continuum solvent model reproduce the free energy landscape of a beta-hairpin folding in water? *Proc Natl Acad Sci USA*. 2002; 99:12777–12782. [PubMed: 12242327]
15. Zhou R. Free energy landscape of protein folding in water: explicit vs. implicit solvent. *Proteins*. 2003; 53:148–161. [PubMed: 14517967]
16. Ostermeir K, Zacharis M. Advanced replica-exchange sampling to study the flexibility and plasticity of peptides and proteins. *Biochim Biophys Acta*. 2013; 1834(5):847–853. [PubMed: 23298543]
17. Kannan S, Zacharis M. Simulation of DNA double-strand dissociation and formation during replica-exchange molecular dynamics simulations. *Phys Chem Chem Phys*. 2009; 11(45):10589–10595. [PubMed: 20145803]
18. Woods CJ, Essex JW, King MA. Enhanced configurational sampling in binding free-energy calculations. *J Phys Chem B*. 2003; 107:13711–13718.
19. Sindhikara D, Meng Y, Roitberg A. Exchange frequency in replica exchange molecular dynamics. *J Chem Phys*. 2008; 128:024103. [PubMed: 18205439]
20. Sindhikara DJ, Emerson DJ, Roitberg AE. Exchange Often and Properly in Replica Exchange Molecular Dynamics. *J Chem Theory Comput*. 2010; 6:2804–2808.
21. Moors S, Michielsens S, Ceulemans A. Improved Replica Exchange Method for NativeState Protein Sampling. *J Chem Theory Comput*. 2010; 7:231–237.
22. Meng Y, Sabri D, Roitberg A. Computing Alchemical Free Energy Differences with Hamiltonian Replica Exchange Molecular Dynamics (H-REMD) Simulations. *J Chem Theory Comput*. 2011; 7(9):2721–2727. [PubMed: 22125475]

23. Matthew AW V, Rohit P. Satisfying the fluctuation theorem in free-energy calculations with Hamiltonian replica exchange. *Phys Rev E*. 2008; 77:026104.
24. Kannan S, Zacharis M. Enhanced sampling of peptide and protein conformations using replica exchange simulations with a peptide backbone biasing-potential. *Proteins: Struct, Funct, Bioinf*. 2007; 66:697–706.
25. Fajer M, Hamelberg D, McCammon JA. Replica-exchange accelerated molecular dynamics (REXAMD) applied to thermodynamic interconversion. *J Chem Theory Comput*. 2008; 4:1565–1569. [PubMed: 19461870]
26. Allen R, Warren P, Wolde P. Sampling Rare Switching Events in Biochemical Networks. *Phys Rev Lett*. 2005; 94:018104. [PubMed: 15698138]
27. Salomon-Ferrer R, Case D, Walker RC. An overview of the Amber biomolecular simulation package. *WIREs Comput Mol Sci*. 2013; 3:198–210.
28. Hess B, et al. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J Chem Theory Comput*. 2008; 4:435–447.
29. Bowers, K., et al. Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters. *Proceedings of the ACM/IEEE Conference on Supercomputing (SC06)*; Tampa, Florida. November 11–17; 2006.
30. Kastner J. Umbrella sampling. *Wiley Interdisciplinary Reviews: Computational Molecular Science*. 2011; 1(6):932.
31. Allen TW, Anderson OS, Roux B. Molecular dynamics — potential of mean force calculations as a tool for understanding ion permeation and selectivity in narrow channels. *Biophys Chem*. 2006; 124:251–267. [PubMed: 16781050]
32. Laio A, Gervasio F. Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. *Rep Prog Phys*. 2008:71.
33. Wojtas-Niziurski W, et al. Self-Learning Adaptive Umbrella Sampling Method for the Determination of Free Energy Landscapes in Multiple Dimensions. *J Chem Theory Comput*. 2013; 9:1885–1895. [PubMed: 23814508]
34. Deng Y, Roux B. Hydration of amino acid side chains: nonpolar and electrostatic contributions calculated from staged molecular dynamics free energy simulations with explicit water molecules. *J Phys Chem*. 2004; 108:16567–16576.
35. Deng Y, Roux B. Calculation of standard binding free energies: aromatic molecules in the T4 lysozyme L99A mutant. *J Chem Theory Comput*. 2006; 2:1255–1273.
36. Deng Y, Roux B. Computation of binding free energy with molecular dynamics and grand canonical monte carlo simulations. *J Chem Phys*. 2008; 128:115103. [PubMed: 18361618]
37. Gan W, Yang S, Roux B. Atomistic view of the conformational activation of Src kinase using the string method with swarms-of-trajectories. *Biophys J*. 2009; 97:L8–L10. [PubMed: 19686639]
38. Brooks BR, et al. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J Comput Chem*. 1983; 4:187–217.
39. Phillips J, et al. Scalable molecular dynamics with NAMD. *J Comput Chem*. 2005; 26:1781–1802. [PubMed: 16222654]
40. Kumar, S.; Sun, Y.; Kale, L. Acceleration of an Asynchronous Message Driven Programming Paradigm on IBM Blue Gene/Q. *IEEE 27th International Symposium on Parallel & Distributed Processing*; 2013. p. 689-699.
41. Sun, Y., et al. Optimizing Fine-grained Communication in a Biomolecular Simulation Application on Cray XK6. *SC12 Proceedings of the international conference on High Performance Computing, Networking, Storage and Analysis*; 2012.
42. EW, Ren W, Vanden-Eijnden E. Transition pathways in complex systems: Reaction coordinates, isocommittor surfaces, and transition tubes. *Chem Phys Lett*. 2005; 413:242–247.
43. EW, Ren W, Vanden-Eijnden E. Finite temperature string method for the study of rare events. *J Phys Chem B*. 2005; 109:6688–6693. [PubMed: 16851751]
44. Ren W, et al. Transition pathways in complex systems: Application of the finite-temperature string method to the alanine dipeptide. *J Chem Phys*. 2005; 123:134109. [PubMed: 16223277]

45. Fukunishi H, Watanabe O, Takada S. On the Hamiltonian replica exchange method for efficient sampling of biomolecular systems: Application to protein structure prediction. *J Chem Phys.* 2002; 116:9058–9067.
46. Shalongo W, Dugad L, Stellwagen E. Distribution of Helicity within the Model Peptide. Acetyl(AAQA)3amide. *J Am Chem Soc.* 1994; 116:8288–8293.
47. Wang J, Deng Y, Roux B. Absolute binding free energy calculations using molecular dynamics simulations with restraining potentials. *Biophys J.* 2006; 91:2798–2814. [PubMed: 16844742]
48. Weeks JD, Chandler D, Anderson HC. Role of repulsive forces in forming the equilibrium structure of simple liquids. *J Chem Phys.* 1971; 54:5237–5247.
49. Kumar S, et al. THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J Comput Chem.* 1992; 13:1011–1021.
50. Kirkwood J. Statistical Mechanics of Fluid Mixtures. *J Chem Phys.* 1935; 3:300.
51. Torrier G, Valleau J. Monte Carlo free energy estimates using non-Boltzmann sampling: Application to the sub-critical Lennard-Jones fluid. *Chem Phys Lett.* 1974; 28:578.
52. Murata K, Sugita Y, Okamoto Y. Free energy calculations for DNA base stacking by replica-exchange umbrella sampling. *Chemical Physics Letters.* 2004; 385:1–7.
53. Gumbart JC, Roux B, Chipot C. Standard binding free energies from computer simulations: What is the best strategy? *J Chem Theory Comput.* 2013; 9:794–802. [PubMed: 23794960]
54. Gumbart J, Roux B, Chipot C. Efficient Determination of Protein–Protein Standard Binding Free Energies from First Principles. *J Chem Theory Comput.* 2013; 9:3789–3798.
55. Xu W. Crystal Structures of c-Src Reveal Features of Its Autoinhibitory Mechanism. *Molecular Cell.* 1999; 3:629–638. [PubMed: 10360179]

Appendix A. Excerpt of Tcl scripting code for parallel tempering

```

if { $replica(index) < $replica(index.$swap) } { # Exchange attempt is
performed by the replica that has
smaller index
set temp $replica(temperature)
set temp2 $replica(temperature.$swap) ; # Temperature parameters of a
replica pair before an exchange
attempt
set dbeta [expr ((1.0/$temp) - (1.0/$temp2)) / $BOLTZMAN]
set pot $POTENTIAL
set pot2 [replicaRecv $replica(loc.$swap)] ; # Receive instant potential
value from exchange partner
set delta [expr $dbeta * ($pot2 - $pot)]
set doswap [expr $delta < 0. || exp(-1. * $delta) > rand()] ; # Metropolis
Monte Carlo algorithm
replicaSend $doswap $replica(loc.$swap) ; # Send swap decision to exchange
partner
}
if { $replica(index) > $replica(index.$swap) } { # Replica that has larger
index sends instant potential
value to exchange partner for exchange attempt and receives swap decision
replicaSend $POTENTIAL $replica(loc.$swap)
set doswap [replicaRecv $replica(loc.$swap)]

```

```

}
set newloc $r
if { $doswap } { # If an exchange attempt is accepted, update the physical
location of a replica
set newloc $replica(loc.$swap)
set replica(loc.$swap) $r
}
set replica(loc.$other) [replicaSendrecv $newloc $replica(loc.$other)
$replica(loc.$other)] ; # Exchange
updated physical location with next round exchange partner
if { $doswap } {
set OLDTEMP $replica(temperature)
array set replica [replicaSendrecv [array get replica] $newloc $newloc] ; #
Update all local attributes
(temperature, exchange partner) for my replica if an exchange attempt is
accepted
set NEWTEMP $replica(temperature) ; # Set new temperature parameter for the
local MD run
rescalelevels [expr sqrt(1.0*$NEWTEMP/$OLDTEMP)]
langevinTemp $NEWTEMP ; # After temperature parameter is updated, velocities
are rescaled and
Langevin temperature is updated as well
}

```

Appendix B

The Tcl scripts for the MCA applications can be downloaded along with NAMD version 2.10 or nightly build source code via <http://www.ks.uiuc.edu/Development/Download/download.cgi?PackageName=NAMD>. In the NAMD source tree, the directory lib/replica contains the relevant MCA scripts and test examples. Replica utility scripts, including replica sorting, visualization and WHAM postprocessing, are also provided.

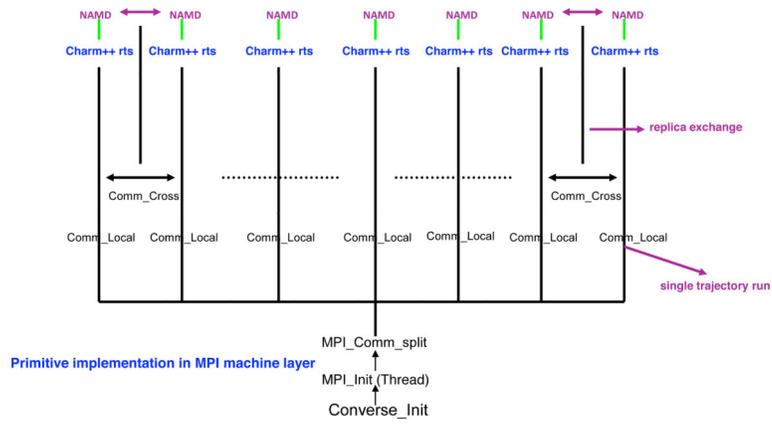


Figure 1.
Generic implementation of MCA in Charm++ run time system/Converse layer

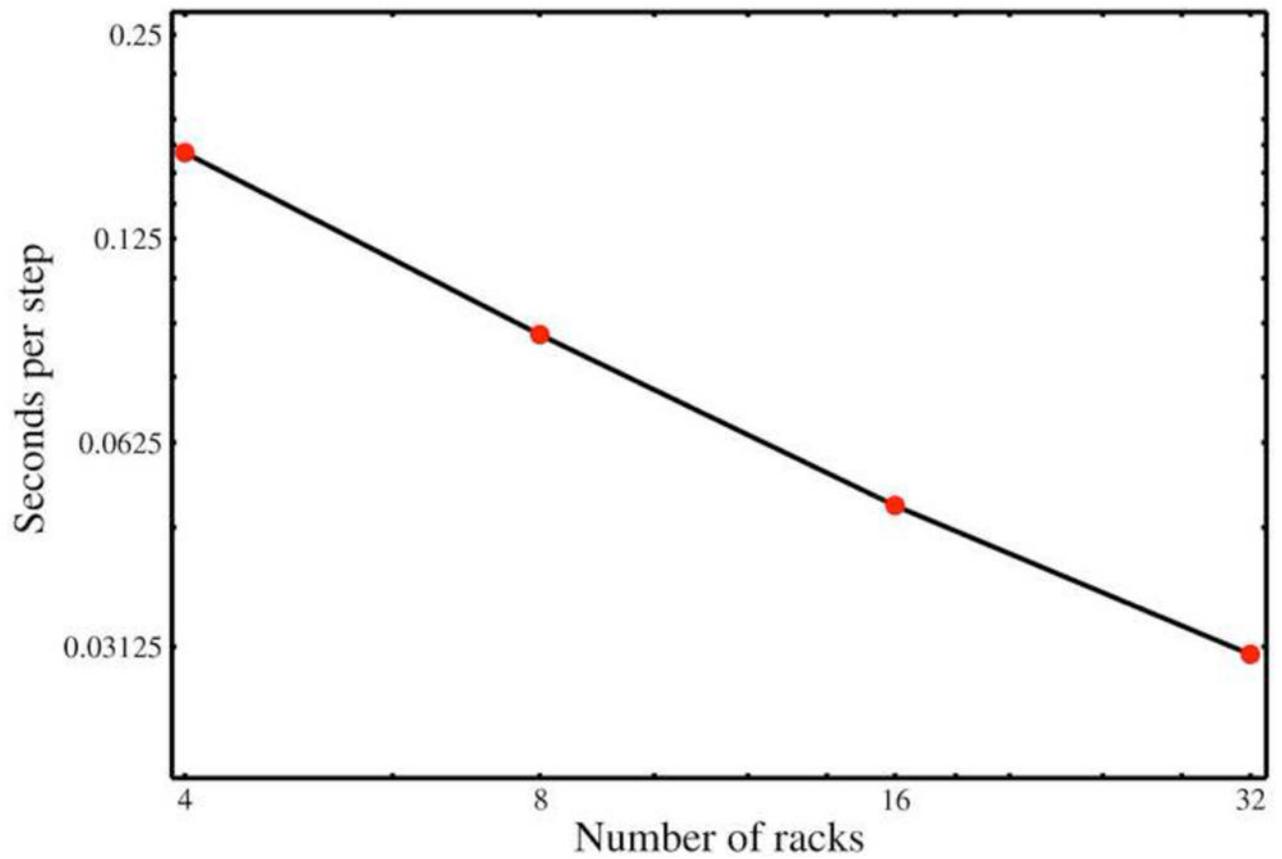


Figure 2.

Strong scaling of the swarms-of-trajectories string method for the full-length c-Src kinase system. The number of cores per trajectory is increased gradually from 32 to 256 with the number of trajectories being fixed. The total number of racks on Blue Gene/Q Mira is 48 and each rack has 16,384 cores.

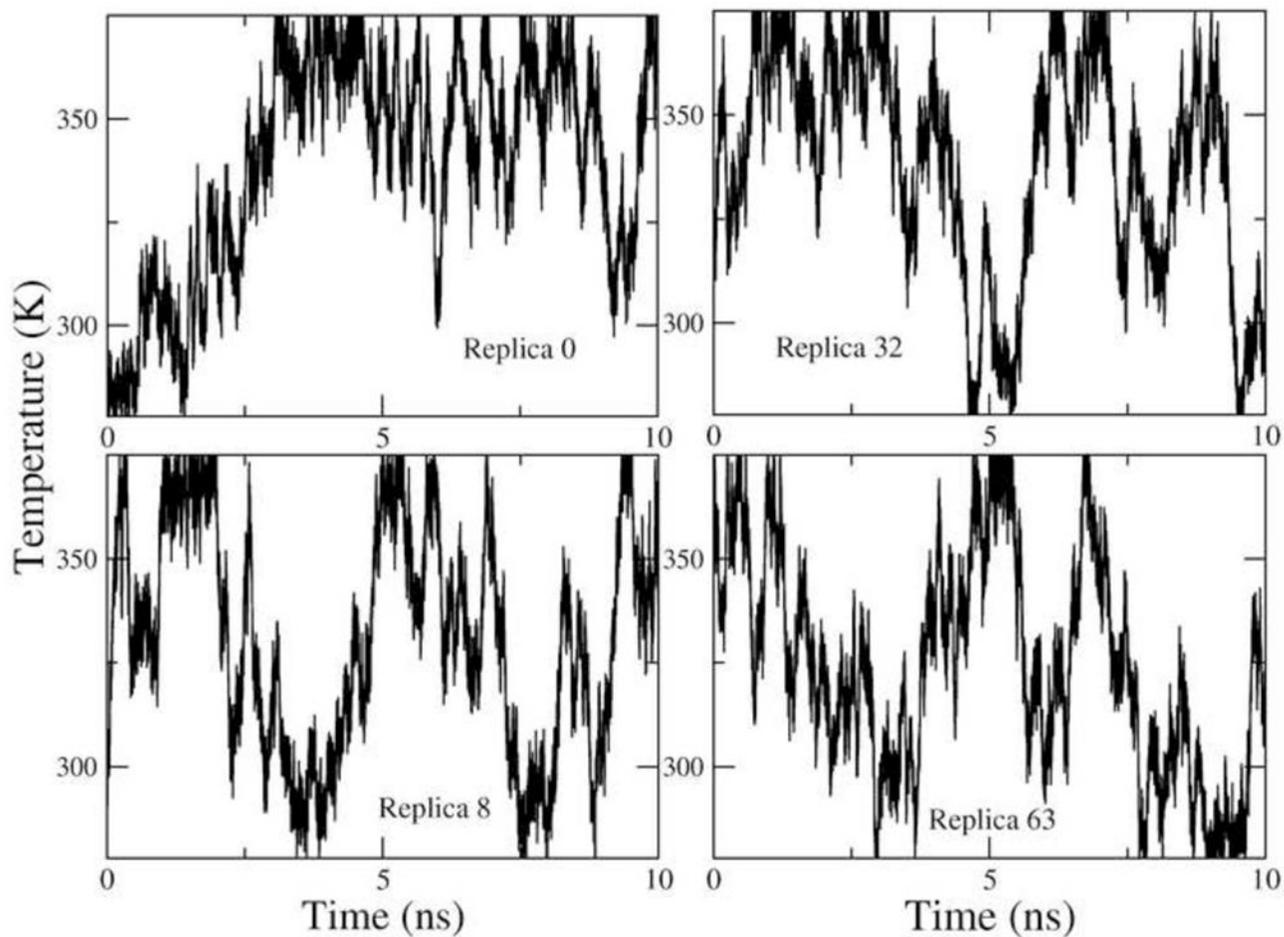


Figure 3. Travelling of selected replicas through the temperature space. All replicas above physical temperature exhibit rapid sampling of temperature space within a timescale of nanoseconds.

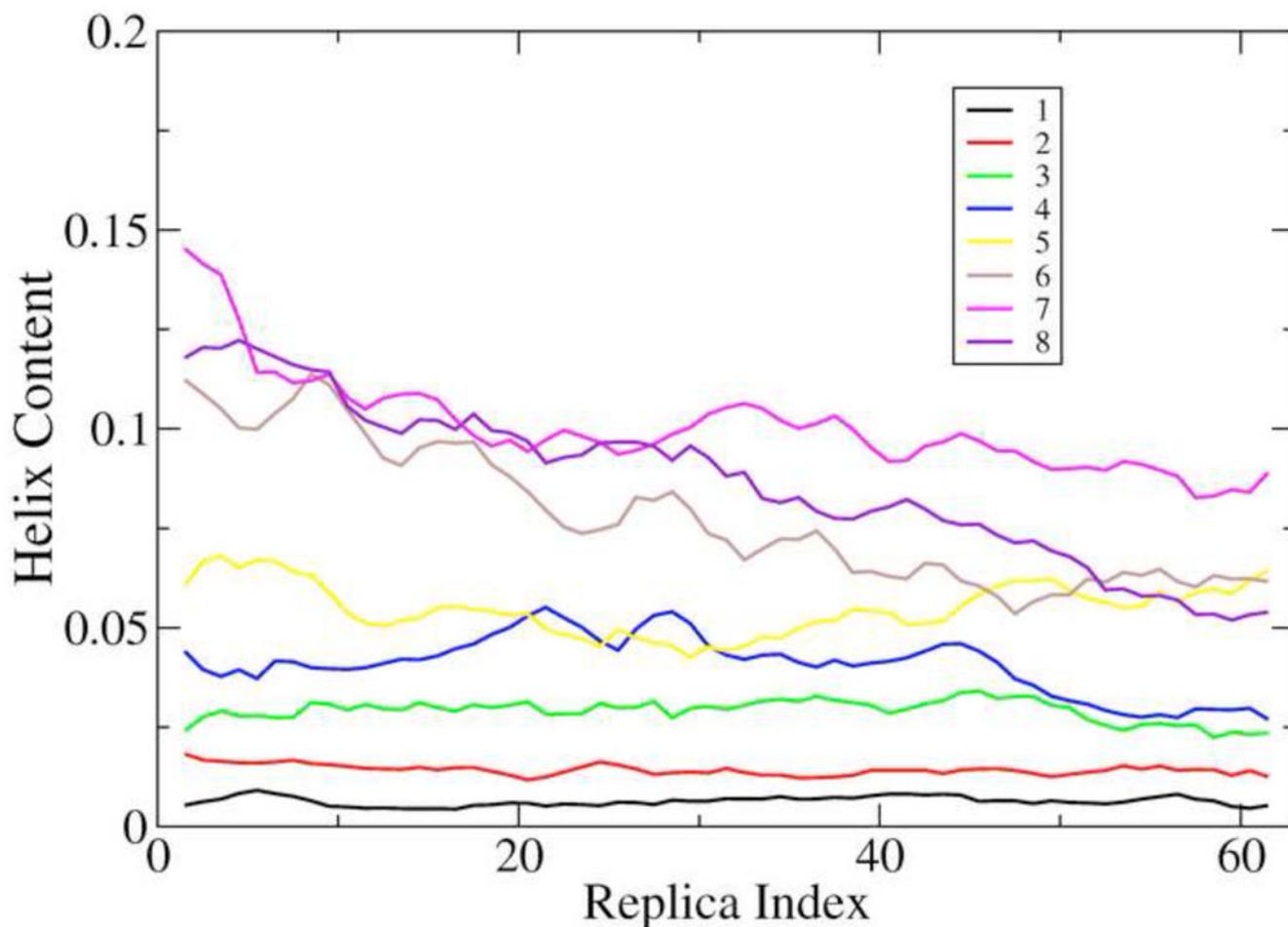


Figure 4. The temperature dependent helix fraction for Baldwin peptides calculated from T-REMD (solid lines) and circular dichroism experimental data (round dots). Fraction of α -helix is determined by at least three residues in the α region of the Ramachandran map ($-100 < \phi < -30$ and $-67 < \psi < -7$). Analysis is done using the last 14 ns of total 25 ns per replica trajectories. Error bars is the standard deviation calculated from the last seven 2 ns blocks.

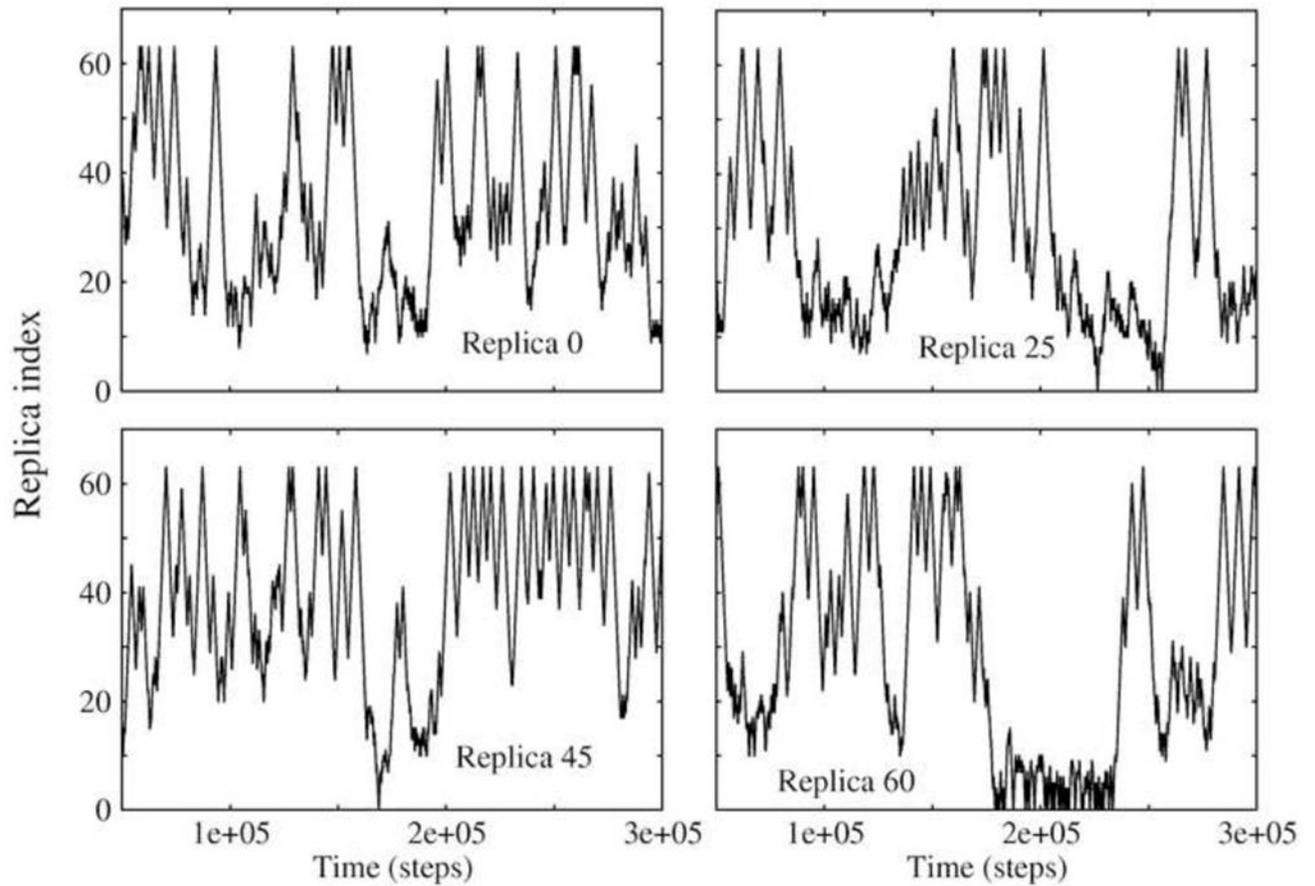
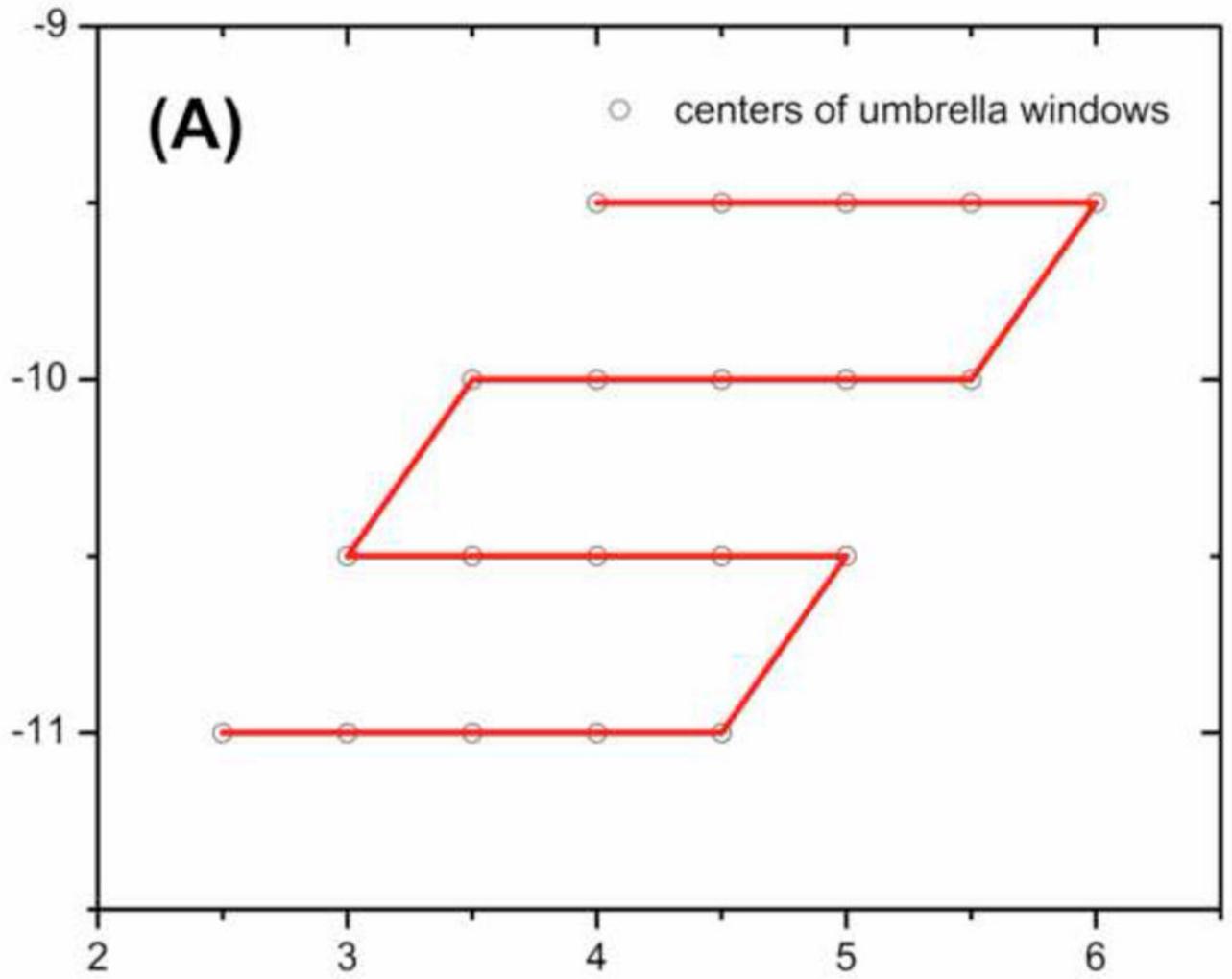


Figure 5.

Travelling of selected replicas through the space of alchemical parameter sets. Each parameter set contains lambda value and alchemical state. The 500 ps production run shown was processed with WHAM to determine the absolute binding free energy.



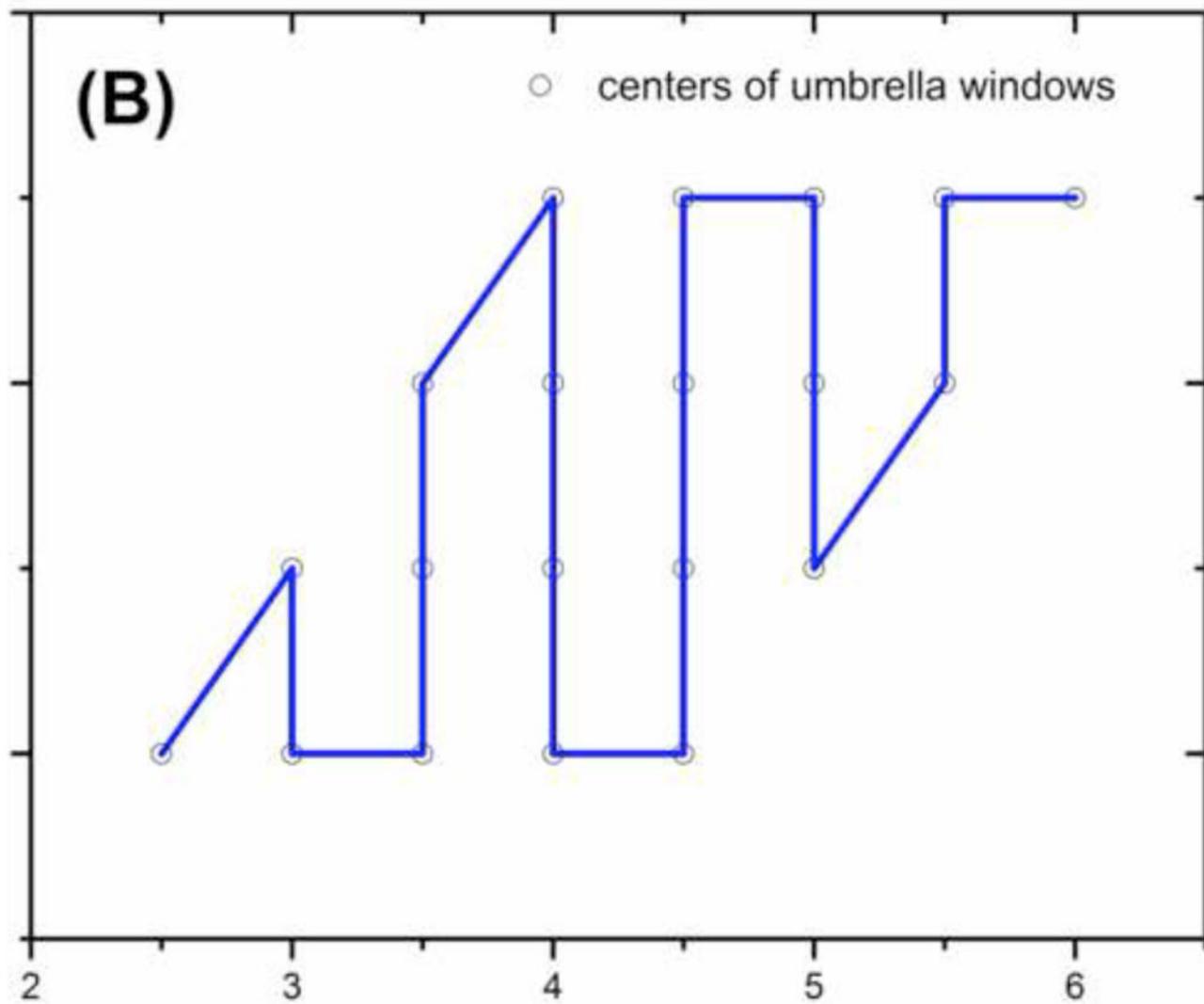


Figure 6.

A graphic demonstration of the two order lists that could be used in 2D US/*H*-REMD. (A) Ordered list when x-axis is the fast-changing coordinate. (B) Ordered list when y-axis is the fast-changing coordinate.

(A)

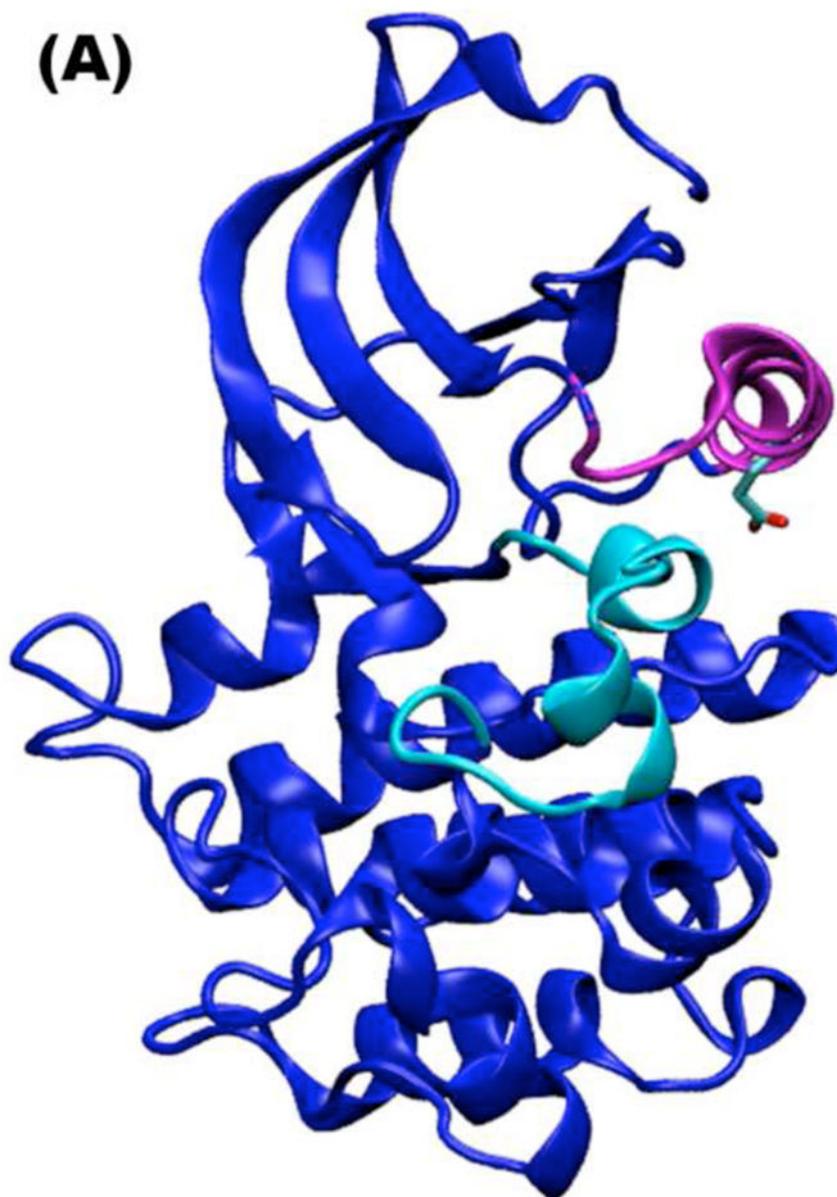
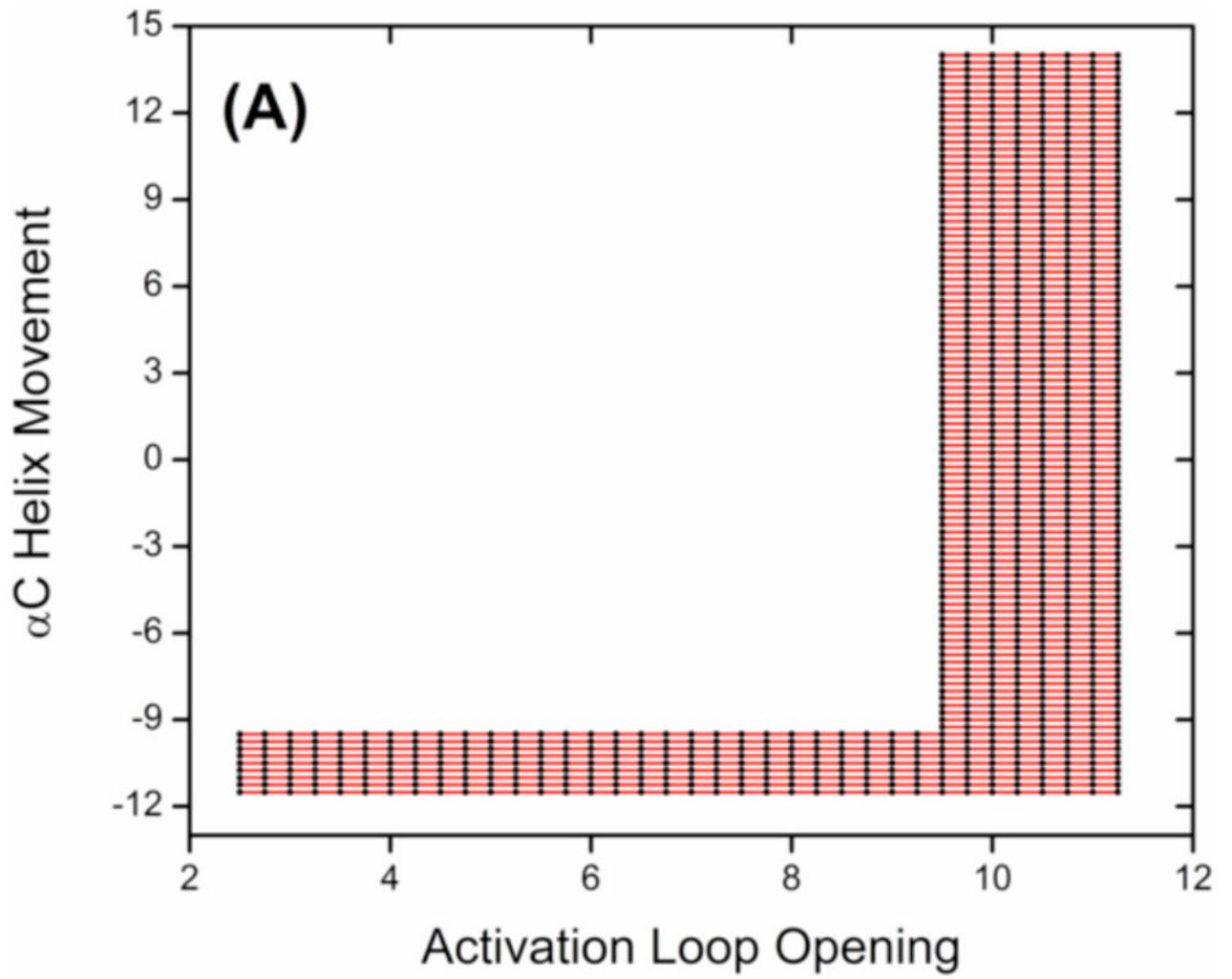
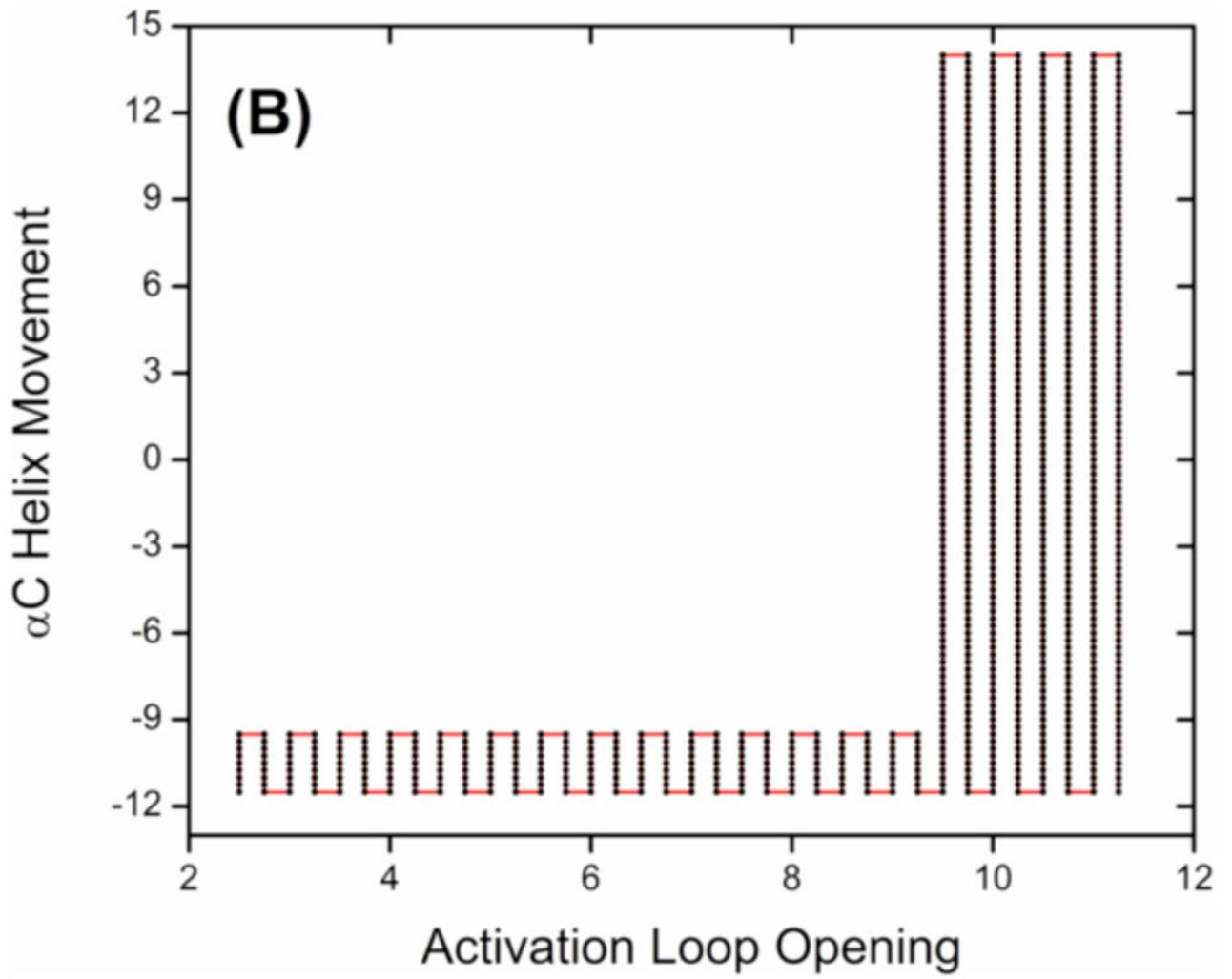




Figure 7. Conformational changes investigated by US/H-REMD calculations. (A) Structure of inactive c-Src kinase domain (adapted from PDB 2SRC). (B) Structure of active-like c-Src kinase domain (adapted from PDB 1Y57). Only the catalytic domain (residue 260 to 521) of c-Src kinase is employed in this study. The α C helix and the A-loop are colored in magenta and cyan, respectively.





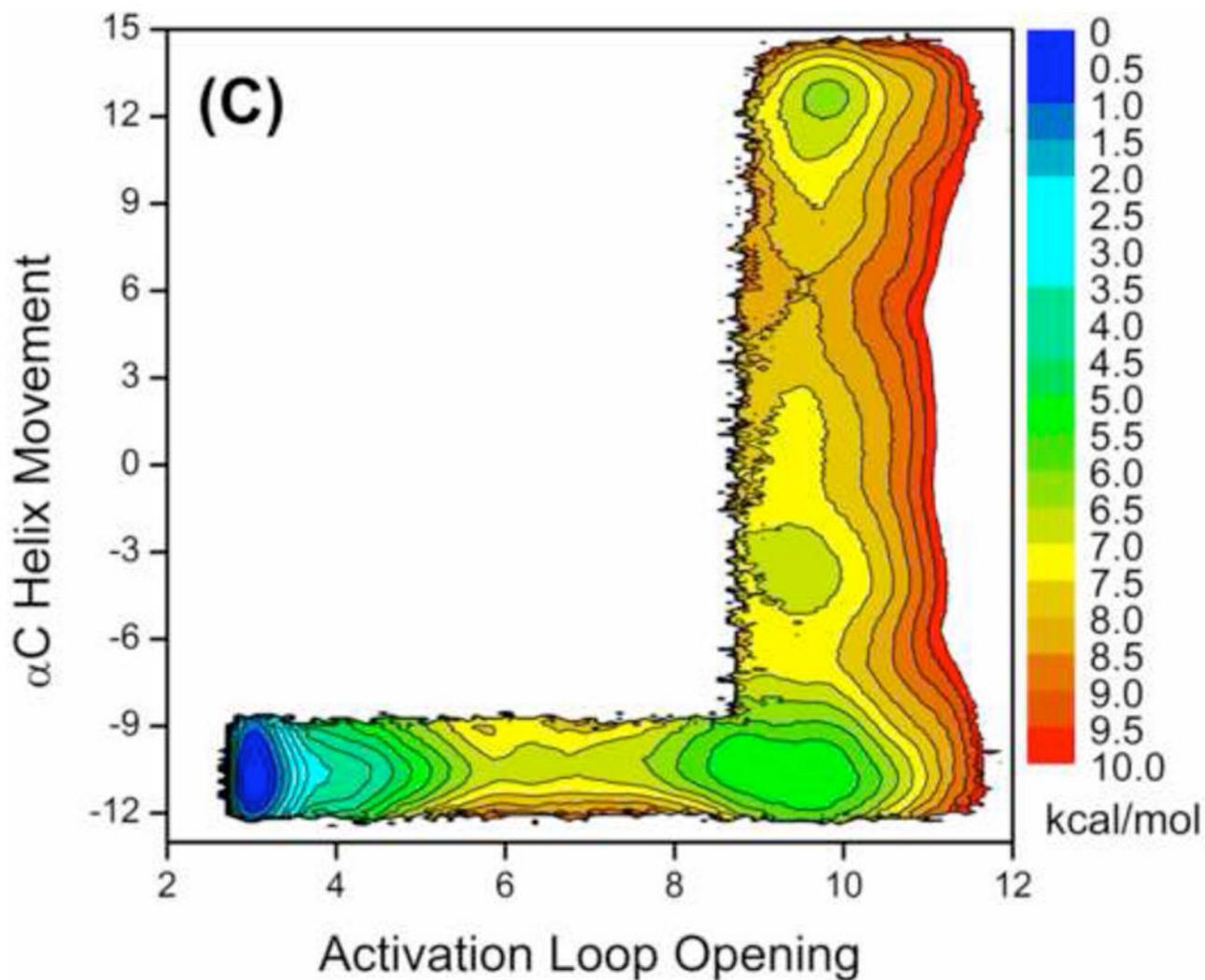
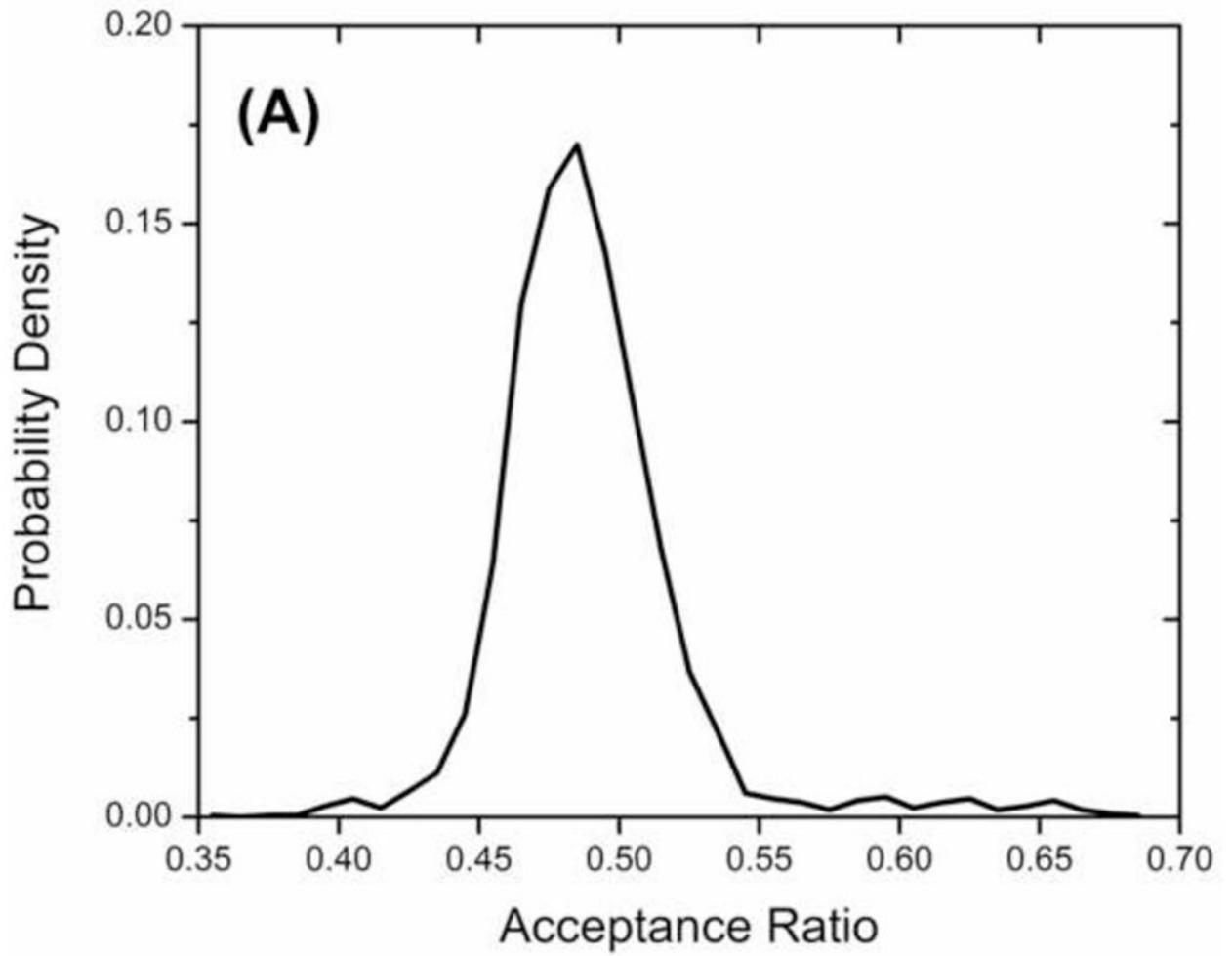


Figure 8.

(A) Ordered list when the x-axis is the fast-changing coordinate. (B) Ordered list when the y-axis is the fast-changing coordinate. (C) PMF determined from US/*H*-REMD simulations using the two ordered lists. The α C helix movement is characterized by the difference between two salt-bridges (d_1-d_2), and the A-loop opening is represented by the average of three distances listed in the text. Both coordinates have the unit of Å.



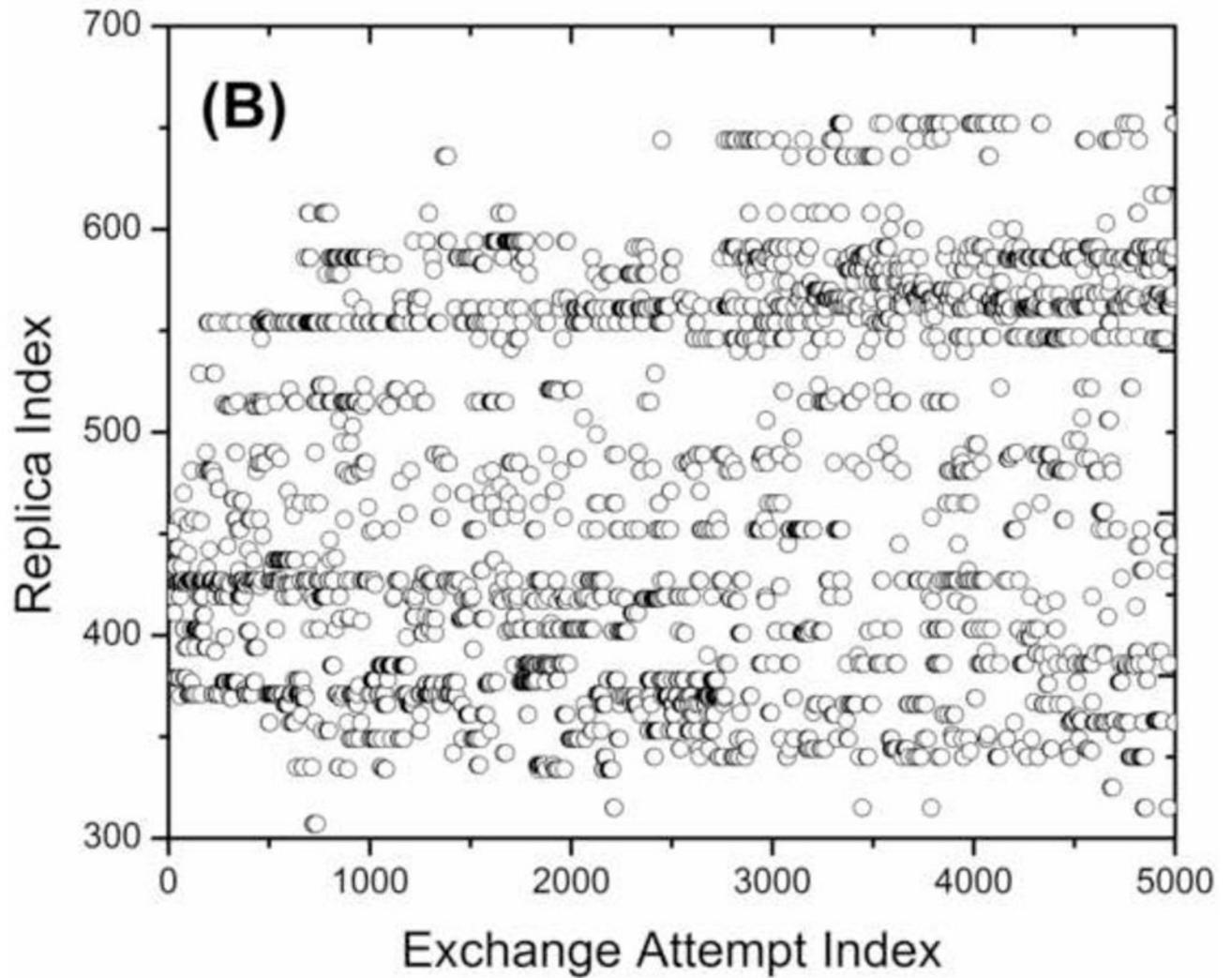


Figure 9. Diagnostics of US/ H -REMD using ordered lists of neighboring windows. (A) Probability distribution of acceptance ratio for all exchanging pairs. (B) The trajectory of replica 435 in the space of replicas.

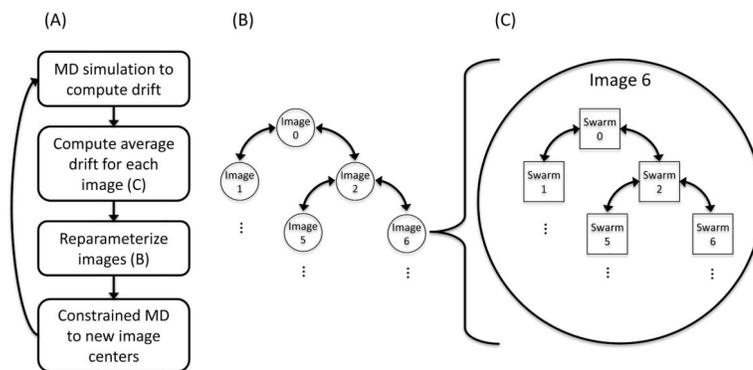


Figure 10.

Multi-level communications are involved in the swarms-of-trajectories string method (Panel A). Each image between the two end states owns a swarm of trajectories and during the refinement iteration of reaction path each trajectory has to talk to each other trajectory in a swarm to compute the position of the next swarm's center (Panel C). Then every swarm must communicate every other swarm to reparameterize the string overall (Panel B).

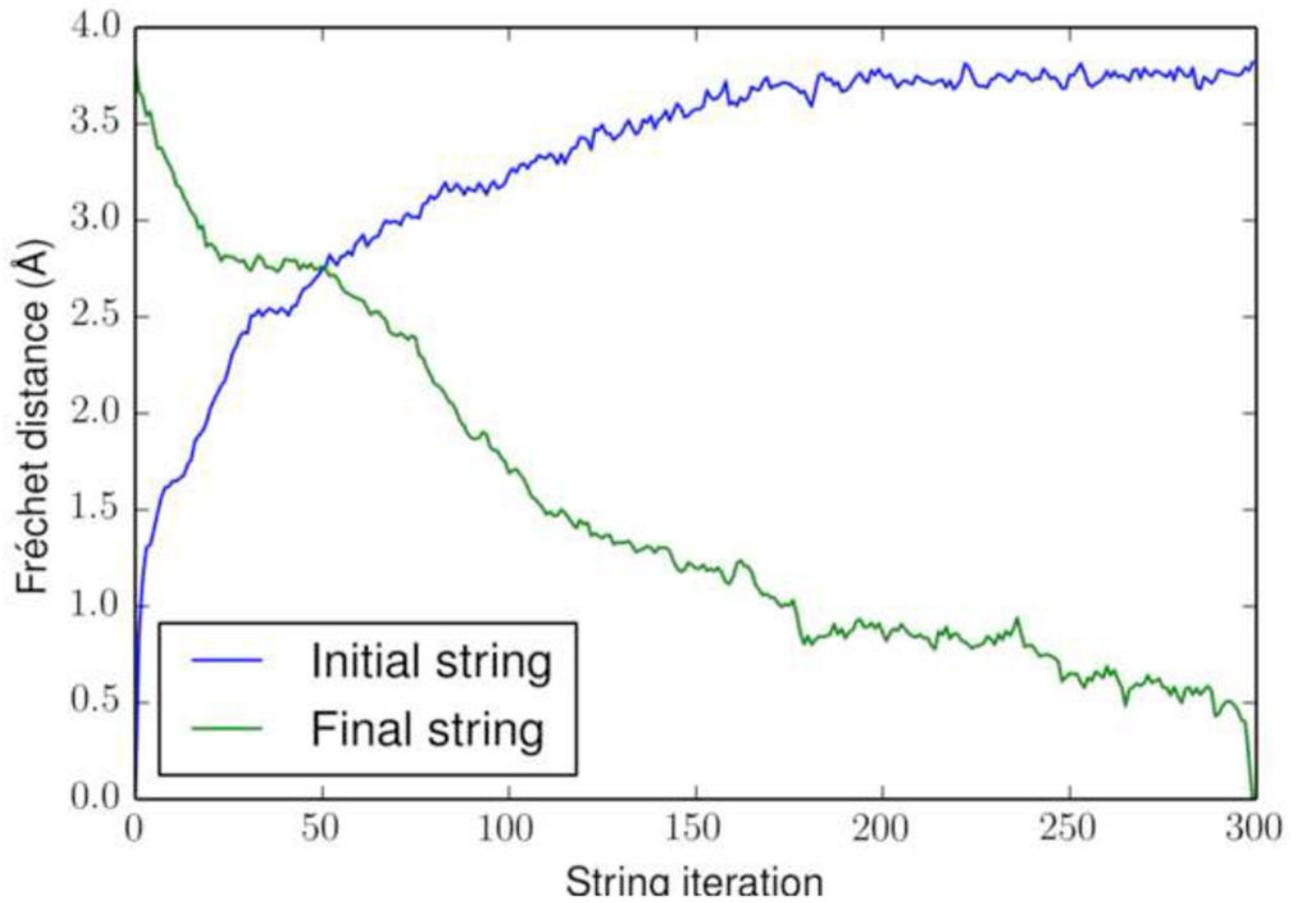


Figure 11.
The discrete Fréchet distance between each string iteration and either the initial string (blue) or the final string (green)