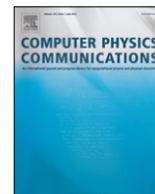




ELSEVIER

Contents lists available at ScienceDirect

Computer Physics Communications

journal homepage: www.elsevier.com/locate/cpc

A generic implementation of replica exchange with solute tempering (REST2) algorithm in NAMD for complex biophysical simulations[☆]

Sunhwan Jo, Wei Jiang^{*}

Leadership Computing Facility, Argonne National Laboratory, 9700 South Cass Avenue, Building 240, Argonne, IL 60439, United States

ARTICLE INFO

Article history:

Received 15 June 2015

Received in revised form

13 August 2015

Accepted 20 August 2015

Available online xxx

Keywords:

REST2

NAMD

Tcl

Free energy calculation

ABSTRACT

Replica Exchange with Solute Tempering (REST2) is a powerful sampling enhancement algorithm of molecular dynamics (MD) in that it needs significantly smaller number of replicas but achieves higher sampling efficiency relative to standard temperature exchange algorithm. In this paper, we extend the applicability of REST2 for quantitative biophysical simulations through a robust and generic implementation in greatly scalable MD software NAMD. The rescaling procedure of force field parameters controlling REST2 “hot region” is implemented into NAMD at the source code level. A user can conveniently select hot region through VMD and write the selection information into a PDB file. The rescaling keyword/parameter is written in NAMD Tcl script interface that enables an on-the-fly simulation parameter change. Our implementation of REST2 is within communication-enabled Tcl script built on top of Charm++, thus communication overhead of an exchange attempt is vanishingly small. Such a generic implementation facilitates seamless cooperation between REST2 and other modules of NAMD to provide enhanced sampling for complex biomolecular simulations. Three challenging applications including native REST2 simulation for peptide folding–unfolding transition, free energy perturbation/REST2 for absolute binding affinity of protein–ligand complex and umbrella sampling/REST2 Hamiltonian exchange for free energy landscape calculation were carried out on IBM Blue Gene/Q supercomputer to demonstrate efficacy of REST2 based on the present implementation.

Program summary

Program title: REST2-NAMD

Catalogue identifier: AEXX_v1_0

Program summary URL: http://cpc.cs.qub.ac.uk/summaries/AEXX_v1_0.html

Program obtainable from: CPC Program Library, Queen's University, Belfast, N. Ireland

Licensing provisions: Standard CPC licence, <http://cpc.cs.qub.ac.uk/licence/licence.html>

No. of lines in distributed program, including test data, etc.: 240886

No. of bytes in distributed program, including test data, etc.: 8474342

Distribution format: tar.gz

Programming language: C/C++, Tcl8.5.

Computer: Not computer specific.

Operating system: Any.

Has the code been vectorized or parallelized?: Yes, MPI and/or PAMI parallelized depending on machine system software; ≥ 8192 cores used on IBM Blue Gene/Q

Classification: 3.

External routines: NAMD 2.10 (<http://www.ks.uiuc.edu/Research/namd/>)

Nature of problem: A generic implementation providing user-friendly API including input file preparation and performing replica exchange, and high frequency exchange attempt frequency with minimal communication overhead.

[☆] This paper and its associated computer program are available via the Computer Physics Communication homepage on ScienceDirect (<http://www.sciencedirect.com/science/journal/00104655>).

^{*} Corresponding author.

E-mail address: wjiang@alcf.anl.gov (W. Jiang).

<http://dx.doi.org/10.1016/j.cpc.2015.08.030>

0010-4655/© 2015 Elsevier B.V. All rights reserved.

Solution method: The rescaling procedure of force field parameters controlling REST2 is implemented into NAMD at the source code level. A user can conveniently select hot region through VMD and write the selection information into a PDB file. The rescaling keyword/parameter is written in NAMD Tcl script interface that enables an on-the-fly simulation parameter change. The implementation of REST2 is within communication-enabled Tcl script built on top of Charm++, thus communication overhead of an exchange attempt is vanishingly small.

Running time: 30 min–60 min

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

A fundamental challenge for molecular dynamics (MD) simulations based on all-atom force field is the bottleneck to sufficiently sample infrequent events across multiple potential barriers. Over two decades enormous efforts have been made in MD community to develop novel sampling enhancement algorithms to overcome the notorious sampling issues in complex biophysical simulations. Among those algorithm progresses, temperature replica exchange molecular dynamics (T-REMD) [1] is one of the most popular methods with its simplicity of implementation and remarkable efficiency. However, with the advancement of computer power and increasing complexity of biophysical simulations, T-REMD exposes severe shortcoming, that is, the number of replicas required is proportional to the square root of number of atoms in the simulation system [2]. As a result, the computing resource required for T-REMD rapidly grows with increasing number of atoms in a simulation. For this reason, it remains critical to continue to seek alternative solutions to improve acceleration efficiency of replica exchange algorithm.

Replica exchange with solute tempering (REST) is proved to be a promising direction that is able to overcome bottleneck of T-REMD [3–5]. In REST algorithm, a “hot region”, comprised of relatively small solute molecules, is defined to accelerate the sampling of the solute conformation while the solvent remains “cold”. Similar approaches that split Hamiltonian into two (or more) partitions have been proposed to study hydrophobic interaction and to enhance conformational sampling efficiency [2,6–9]. Originally, REST was developed for efficient sampling of protein folding simulation in an explicit solvent environment [3,5], but it can be easily generalized to accelerate sampling of a small subset of atoms such as active site of protein while the other atoms become a generalized ‘solvent’. Recently, an improved version of REST, namely REST2, with modified scaling of Hamiltonian demonstrated a significantly improved efficiency of conformational sampling of small peptide [5]. Moreover, REST2 is employed in quantitative biophysical MD simulations, such as lipid diffusion [10], relative binding free energy calculation [11–13], and protein dynamics [14,15].

Relying on feasibility of modifying parameter file with in-house codes, REST2 can be implemented by employing a set of modified force field parameter input files, each of which serves one replica of a unique effective temperature [16]. Periodic exchange attempts were carried out through configuration exchange between neighboring replicas. Such an apparently straightforward scheme can work efficiently only for relatively small and moderate size biological systems on small scale computing clusters. However, the increasing demands for simulating large biomolecular systems with complex structure-function relation on modern large-scale supercomputers requires user friendly and low overhead REST2 implementation. In practice, a single trajectory MD simulation for large system (> 100 K–1 M atoms) needs hundreds to thousands of CPU processors. A configuration exchange between neighboring trajectories creates severe communication overhead including global

collective communication within a replica and point-to-point coordinates/velocities swap. Modern MD software has extensively adopted sophisticated domain decomposition and dynamic load balance algorithms, and therefore a sudden coordinates/velocities exchange often results in an expensive reconstruction of workload balance. Moreover, from the MD simulation point of view, increasing exchange attempts as frequently as possible allows optimal rate of exploration of configurational space, as long as the exchanges are done properly [17,18]. In practice, high frequency exchange attempts, i.e. order of 1/100 or higher, can effectively counteract low acceptance ratio of replica exchange due to rugged energy surfaces [19].

In this article, we report a systematic and generic implementation of REST2 algorithm in NAMD’s source code and Tcl interface, aiming at complex quantitative biophysical simulations on modern massively distributed computing platforms. NAMD is a highly scalable program built onto a hybrid spatial/force decomposition managed by the asynchronous parallel programming system Charm++ [20]. Generally, a user can generate simulation control and PDB file through the popular visualization software VMD. 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 without modifying source code. Recently, the interface was further extended to enable communications among different NAMD instances through Tcl commands built on top of low-level communication functions, while the Charm++ programming system can support concurrent multiple NAMD instances by remapping processing elements [21].

In our implementation, REST2 algorithm is implemented in NAMD source code and the rescaling parameter/keyword is written into Tcl script interface, which allows end users to arbitrarily combine REST2 with other modules for various complex biophysical simulations. Here, the implementation details are reported and three representative applications, native REST2 simulation for conformation sampling of solvated polypeptide, free energy perturbation (FEP)/REST2 for absolute binding affinity calculation of protein–ligand complex, umbrella sampling (US)/REST2 simulation for free energy landscape exploration are carried out on the IBM Blue Gene/Q supercomputer at Argonne National Laboratory. The robustness of the present REST2 implementation is demonstrated and how REST2 assists quantitative MD simulations is discussed in detail.

2. Implementation details

In REST2, all of the replicas are run at the same temperature but the potential energy for each replica is scaled differently:

$$E_m^{\text{REST2}}(X) = \frac{\beta_m}{\beta_0} E_{ss}(X) + \sqrt{\frac{\beta_m}{\beta_0}} E_{sw}(X) + E_{ww}(X) \quad (1)$$

where E_{ss} , E_{sw} , and E_{ww} represent the solute–solute, solute–solvent, and solvent–solvent interaction energies, respectively, and X represents the configuration of the whole system. $\beta_m = 1/k_B T_m$ and T_0

is the target temperature we would like to sample. Essentially the charge and vdW parameter of each atom of hot region is rescaled by a factor of $\sqrt{\beta_m/\beta_0}$ and related bonded terms are rescaled correspondingly. By imposing detailed balance conditions it is easy to show that the acceptance ratio for exchange between replicas m and n is determined by

$$\Delta_{mn}(\text{REST2}) = (\beta_m - \beta_n) \left[(E_{ss}(X_n) - E_{ss}(X_m)) + \frac{\sqrt{\beta_0}}{\sqrt{\beta_m} + \sqrt{\beta_n}} (E_{sw}(X_n) - E_{sw}(X_m)) \right]. \quad (2)$$

The first aim of the present implementation is to provide convenience for end user to generate PDB file that contains hot region information with the popular visualization software VMD. In VMD Tk Console interface, an end user can set occupancy values (column O) 1 for all atoms of hot region and generate a corresponding PDB file. During startup phase, NAMD sets a nonzero flag (e.g., 1) for each atom of hot region and these flags are stored as basic molecular information.

The scaling factor of REST2 is controlled by one of the standard keywords in simulation control parameter class. Further it is written into NAMD's Tcl script interface, offering flexibility for end user to change it on the fly and minimize overhead for replica exchange scheme. The nonbonded force field parameter rescaling is executed in force computing classes for direct nonbonded interaction and Particle Mesh Ewald (PME) [22]. For bonded terms, parameter rescaling is done in bonded tuple class template that defines a family of all bonded terms. The parameter scaling of hot atoms creates a few extra integer operations within the nonbonded (i, j) loop, charge grid filling loop of PME and bonded loop, but fortunately they merely result in $\sim 3\%$ – 4% speed loss. In addition, the scaling factor can be arbitrarily adjusted if necessary, instead of using original REST2 scheme, but performance penalty is more severe.

NAMD has a generic Multiple Copy Algorithm (MCA) interface built on top of Charm++. In MCA, multiple NAMD instances run concurrently and communicate through Tcl interface, which intends to provide maximum flexibility for a high-end user to realize its special needs without touching the source code [21]. Master Tcl script periodically does scaling parameter swap attempts for all replica pairs. In principle a user can arbitrarily modify the exchange pattern among replicas to devise new REST2 based algorithms through basic Tcl scripting. Our implementation is based on NAMD version 2.10 and the source code, test, and examples will be available online. Appendix exhibits an excerpt of Tcl scripting code that controls parameter swaps in a REST2 simulation.

3. Application examples

3.1. REST2 for small solute in explicit solvent

Using REST2, we simulated a small peptide, Ac-(AAQAA)₃-NH₂, conformation in explicit water on Blue Gene/Q. An extended configuration of Ac-(AAQAA)₃-NH₂ peptide was prepared by CHARMM [23] and the same initial structures were used for each replica. The simulation system comprised of $\sim 25,000$ atoms and total of 16 replicas were used that spans 300–600 K. The hot region temperature of i th replica (T_i) was determined by following equation,

$$T_i = T_0 \exp \left[\ln \left(\frac{T_{\max}}{T_0} \right) \left(\frac{i}{N_{\text{rep}} - 1} \right) \right]$$

where T_0 and T_{\max} are the lowest and the highest temperature used in the simulation, and N_{rep} is the total number of replicas. Among the bonded interactions, only the dihedral angle terms are

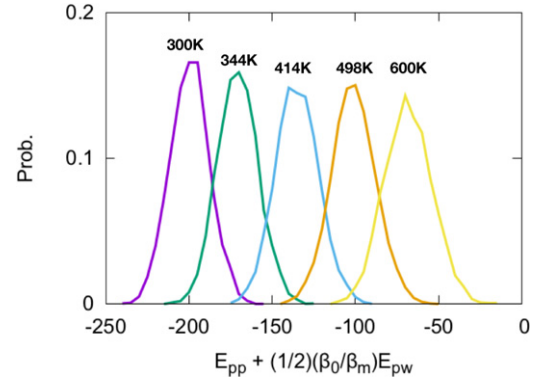


Fig. 1. Distribution of $E_{ss} + (1/2)(\beta_0/\beta_m)E_{sw}$ from Ac-(AAQAA)₃-NH₂ peptide simulation at various effective simulation temperatures. Among 16 effective temperatures (300–600 K), only five of the effective temperatures are shown here for clarity.

subjected to scaling following the suggestion from Wang et al. [5]. Neighboring replicas were subject to exchange at every 500 MD steps. The simulations were performed under NPT (constant pressure at 1 bar and constant temperature) ensemble and periodic boundary condition with CHARMM36 force field [24]. Total 8192 cores (512 cores per replica) of IBM Blue Gene/Q were used to launch the generic parallel/parallel REST2 simulation. Following a 20 ns pre-equilibration, 200 ns production run was carried out.

Ac-(AAQAA)₃-NH₂ peptide has been used widely to understand the secondary structure formation of small peptide [25,26] and optimizing the helical propensity of modern force fields [24]. In the previous computational studies [21,24], T-REMD has been used for efficient sampling of the peptide in explicit water, which requires a large number of replicas that span relatively small temperature range for efficient exchange of replicas [27]. In our previous work, 64 replicas that span temperature range of 278–375 K were used and obtained 45% exchange acceptance ratio for T-REMD of the similar sized system [21]. REST2 algorithm excludes the contribution from solvent molecules, which results dramatically reduced number of replica required for efficient exchange. Fig. 1 shows that potential energy of neighboring replicas overlaps well. As shown in Eq. (2), only the fluctuations of $E_{ss} + (1/2)(\beta_0/\beta_m)E_{sw}$ contribute to the acceptance ratio (when $\beta_m \approx \beta_{m+1} \approx \beta_{m-1}$), which allowed average exchange acceptance ratio over 55% with only 16 replicas spanning temperature range of 300–600 K.

Fig. 2 shows the alpha-helix contents from the simulation trajectory. On average, the residue 6–10, sampled alpha-helical region (defined by $-160^\circ < \phi < -20^\circ$ and $-120^\circ < \psi < 50^\circ$) around 35.7% (averaged over 100–200 ns), which is consistent from previous computational study [24] even though our simulations started from an extended conformation. In addition, we have observed average fraction of helix as 28.3% (averaged over 100–200 ns), defined by a fraction of residues 6–10, which are laid within a stretch of at least three alpha-helical residues. From NMR experiment [26] and circular dichroism experiments of an analogous peptide [25,28], Ac-(AAQAA)₃-NH₂ peptide is found to have around 20% of helical contents in water. Fig. 3 shows some snapshots of the initial peptide conformation and that of the alpha-helical conformations adopted during the simulations. Given the fact that our simulation started from a complete extended conformation instead of folded states, these agreements lay a milestone for testing efficiency of novel sampling enhancement algorithms.

3.2. Free energy perturbation with REST2

In free energy perturbation (FEP) simulations, thermodynamic coupling parameter λ , is used to adjust the potential energy of a

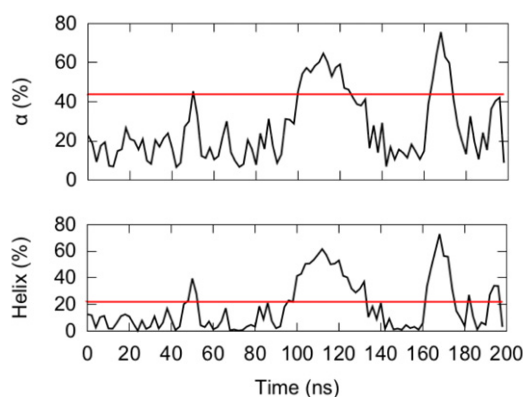


Fig. 2. Time-series of (top) alpha-helical contents and (bottom) fraction of helical residues, defined by a fraction of residues 6–10, which are laid within a stretch of at least three alpha-helical residues, from Ac-(AAQAA)₃-NH₂ simulation system at 300 K.

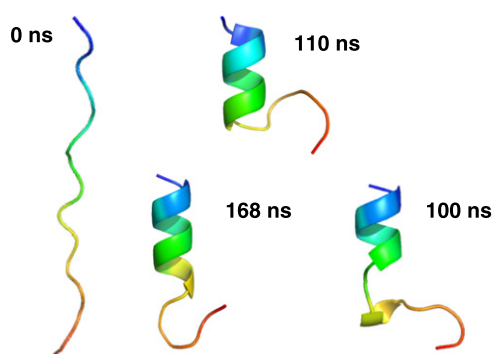


Fig. 3. Snapshots of Ac-(AAQAA)₃-NH₂ peptide from the simulation. The initial structure (0 ns) and some of the folded conformations are arbitrarily selected. The peptide is represented as cartoon and the N- and C-termini are colored as blue and red, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

set of alchemical atoms. For example, binding free energy of receptor–ligand complex can be calculated by slowly turning on the ligand potential energy in multiple stages [29]. The hysteresis issue of FEP simulations can be resolved by using FEP/ λ -REMD scheme [19,30], where the random Monte Carlo move of thermodynamic coupling parameter is allowed. However, effective sampling of orthogonal degrees of freedom along the thermodynamic path is often beyond time scale of a normal MD simulation, and therefore these FEP simulations suffer from so called “Hamiltonian lagging” problems [31], where the surroundings of a ligand are trapped in the initial simulation configuration.

One prominent example of such problem is the binding affinity difference in L99A mutant of T4 lysozyme. For small ligands (e.g., benzene, toluene, and indole) the calculated binding free energy is well converged regardless of whether the simulation started from the *holo* or *apo* state [32–34]. However, for bigger ligands (e.g., *p*-xylene), it has been observed that the calculated binding free energies were inconsistent depending on whether the simulation started from a *holo* or *apo* state [32,33]. In this case, the side chain of Val111, which is in direct contact with the bound ligand, must change its rotameric states to accommodate the ligand, but the energetic barrier for this transition is significantly high (~ 5 kcal/mol) and the reorientation is not well captured in the time scale of FEP simulation [35]. To overcome this problem, Jiang and Roux [35] introduced a “boosting” potential specifically accelerates the sampling of side chain reorganization for the selected residues near the binding pocket, which requires careful construction of boosting potentials for the side chains.

In REST2, the side chain atoms near the binding pocket can be selected for solute tempering and their sampling can be accelerated selectively. To demonstrate the performance of FEP/REST2, we have performed absolute binding free energy calculation with T4 lysozyme and *p*-xylene starting from the *holo* state. We have used previously implemented Weeks–Chandler–Anderson (WCA) decomposition scheme [19,21,36], which decomposes the total potential energy of the system as follows

$$U(\lambda_{\text{rep}}, \lambda_{\text{disp}}, \lambda_{\text{elec}}) = U_0 + U_{\text{rep}}(\lambda_{\text{rep}}) + U_{\text{disp}}(\lambda_{\text{disp}}) + U_{\text{elec}}(\lambda_{\text{elec}})$$

where U_0 is the potential of the system with the non-interacting ligand, $\lambda_{\text{rep}}, \lambda_{\text{dis}}, \lambda_{\text{elec}} \in [0, 1]$ are the thermodynamic coupling parameters, U_{rep} and U_{dis} are the repulsive and dispersive components of the Lennard-Jones potential according to WCA decomposition [36], U_{elec} is the electrostatic contribution. 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.

T4 lysozyme bound to *p*-xylene (PDB:187L) was solvated in explicit water using CHARMM-GUI [37] and the simulation was performed on Blue Gene/Q. The simulation system comprised of $\sim 45,000$ atoms and all simulations were performed under constant pressure and temperature (NPT) ensemble. CHARMM36 force field parameter [24] was used for protein. The force field parameter for *p*-xylene was taken from an earlier work [32]. To compare the performance, binding free energies are calculated with both standard FEP/ λ -REMD [19,30] and FEP/REST2 scheme designed by Wang and co-workers [11]. Total of 64 replicas are evenly located along the alchemical reaction path, employing 36 repulsive windows, 12 dispersion windows and 16 electrostatic windows. For REST2, side chain atoms of Val111 are selected for the accelerated sampling, and the effective temperature for the selected atoms is changed at the same time as the thermodynamic coupling parameter. The effective temperatures at both thermodynamic end-states ($\lambda = 0$ and $\lambda = 1$ states) are kept 300 K and reached maximum of 1200 K at the middle of the thermodynamic perturbation (Fig. 4). The simulations adopt an exchange attempt frequency of 1/100 steps and obtain average acceptance ratio $\sim 60\%$ – 80% . For solvation free energy of *p*-xylene in bulk solvent, the ligand was solvated in a box comprised of 600 TIP3P water molecules and the FEP/ λ -REMD simulations were performed with 64 windows (32 repulsive, 16 dispersive, and 16 electrostatic windows).

Total 2 ns of simulations were performed for both FEP/ λ -REMD and FEP/REST2 methods and simple overlap sampling method [38] was used to analyze the free energy change. Free energy appears to quickly converge within first 100 ps of simulation, thus the last 1.9 ns of trajectories were used for further analysis. Table 1 shows the free energy simulation results from standard FEP/ λ -REMD and FEP/REST2, which we obtained -5.7 ± 0.4 and -4.74 ± 0.7 kcal/mol, respectively. Error bars are calculated by taking standard deviations of free energies obtained with 100 ps trajectory blocks. Both free energies agree well with the experimental binding free energy of -4.67 kcal/mol [39], but there is a noticeable improvement observed in the final binding free energy when FEP/REST2 method was used compared to FEP/ λ -REMD.

Fig. 5 shows the improvement of rotameric state (χ_1 ; C-CA-CB-CG2 dihedral) of Val111 sampling during the FEP/REST2 simulation at different thermodynamic state. At the *holo* state, Val111 favors *gauche* conformation ($\chi_1 \approx -60^\circ$) whereas it favors *trans* conformation ($\chi_1 \approx 180^\circ$) at the *apo* state [35]. The simulation started at the *holo* state, and the χ_1 dihedral angle of Val111 remained at the *gauche* conformation even at the *apo* state during the traditional FEP/ λ -REMD simulation as shown in Fig. 5, while χ_1 dihedral angle quickly moves towards more favorable *trans* conformation at the *apo* state during FEP/REST2 simulations.

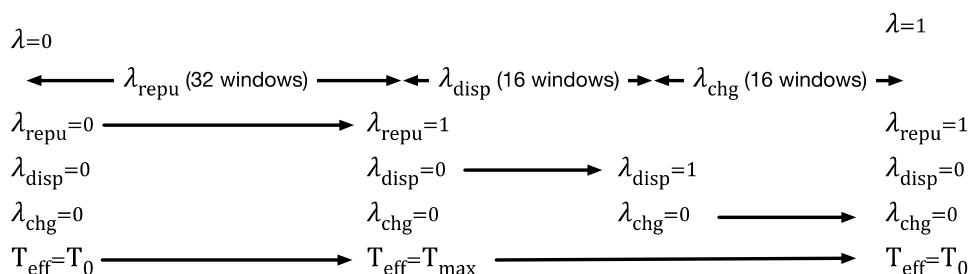


Fig. 4. One-dimensional replica exchange protocol combining REST and FEP. WCA decomposition is adopted during the FEP. Black solid lines represent the change of thermodynamic coupling parameter or effective temperature parameter for the group of windows. Parameters are exchanged by neighbor exchange algorithm.

Table 1

Binding free energy of *p*-xylene to L99A T4 lysozyme (all values in kcal/mol).

	MD	REST2 ^b	REST2 ^c	Exp [39]
ΔG_{site}	-6.3 ± 0.4	-5.3 ± 0.7	-5.3 ± 0.5	
ΔG_{solv}^a	-0.56 ± 0.09	-0.56 ± 0.09	-0.56 ± 0.09	-0.87
ΔG_{bind}	-5.7 ± 0.4	-4.74 ± 0.7	-4.74 ± 0.5	-4.67

^{b,c} Different number of residues were used to define “hot” region.

^a ΔG_{solv} term is calculated once using standard FEP/MD protocol.

^b Only Val111 residue was selected.

^c Total of seven residues including A, B, C, D, E, F, and Val111 were selected.

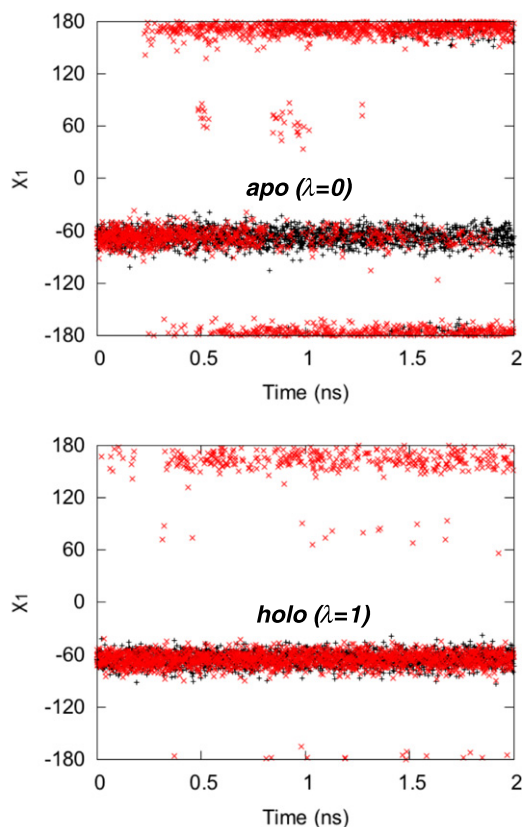


Fig. 5. Enhanced sampling of rotameric states (%1; C-CA-CB-CG2 dihedral) of Val111. Black dots, values obtained with FEP/λ-REMD; red dots, obtained with FEP/REST2. Window #1 turns off all ligand–receptor interactions and therefore corresponds to the *apo* state. Window #64 turns on all ligand–receptor interactions and corresponds to the *holo* state. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Sampling of surrounding reorganization is crucial for reliable binding affinity prediction. REST2 provides a general and convenient route to enhance the sampling of environment around the binding site without *a priori* knowledge of the system. However, the definition of the hot region, both the size of the region and the effective temperatures, remained as an arbitrary choice. As the

current FEP/REST2 scheme thrusts hot region into thermodynamic path to save computing power, it is uncertain that to what extent such a cheap integration affects convergence of FEP simulations without further assumption. In general, it is advised to choose hot region as small as possible to minimize statistical error [11], but larger hot region might be beneficial for enhanced sampling. To quickly test the effect of using larger locally enhanced sampling region, we have performed another 2 ns FEP/REST2 simulation with same simulation conditions while 6 more residues around the binding site were included in the hot region. In this simulation, the final binding free energy was -4.74 ± 0.5 kcal/mol, which is similar to the result obtained with only Val111 in the hot region. Although in this preliminary study we did not observe a noticeable difference between the binding free energies obtained with hot regions of different sizes, we are endeavoring to seek optimal integration of FEP and REST2 for general biomolecular simulations.

3.3. Umbrella sampling with REST2

Umbrella sampling (US) is a powerful method to obtain potential of mean force (PMF) along a chosen reaction coordinates [40–42]. When combined with Hamiltonian replica-exchange scheme (US/ℋ-REMD) [43,44], the convergence can be greatly accelerated. However, when the system is very complex and has hidden orthogonal degrees of freedom that are not targeted by the collective variable (CV), the convergence of the simulation could be slow. Replica-exchange algorithm can mitigate this problem by permitting the biasing parameters of adjacent windows to be exchanged according to a Metropolis energy criterion [45]. Still, the barrier crossing along orthogonal degree of freedom is a rare event and one has to passively wait until such event occurs and propagates to all other windows.

REST2 can selectively accelerate the sampling of atoms in non-specific manner, allowing enhanced sampling along the orthogonal degree of freedom without intimate knowledge of slow degree of freedom. Such scheme is particularly appealing for umbrella sampling of protein–peptide interaction where the conformational samplings of a small subset of residues that are near the binding site are crucial. Here, we combine US/ℋ-REMD and REST2 approaches together and test how sampling efficiency gains. This can be achieved by adopting 2D Hamiltonian replica exchange scheme where umbrella biases are exchanged in one axis and the effective temperatures are exchanged in another axis. A similar approach has been applied by Okamoto and his co-workers [12] for protein and ligand binding.

Here, we have performed binding of a small peptide to the SH3 domain of Abl kinase. The small peptide used in this study has sequence of APSYSPPPPP (hereafter p41) is an artificially designed peptide that has a high affinity to the SH3 domain of Abl kinase [46]. Complete characterization of absolute binding free energy of peptide–protein using molecular dynamics simulation is a still difficult task [47], and is beyond the scope of our manuscript. Instead, here we limit ourselves to computing the PMF

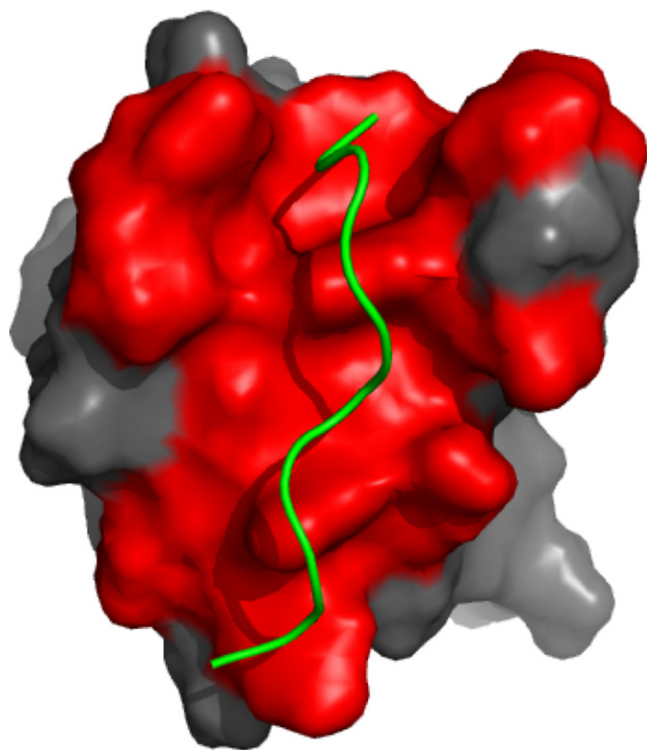


Fig. 6. Crystal structure (PDB:1BBZ) of SH3 domain of Abl kinase (gray surface) bound to p41 peptide (green cartoon). The SH3 domain residues within 4 Å distance from p41 peptide are colored red and selected for selective tempering. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

of separation of the p41 peptide from the SH3 domain of Abl kinase while the translational and orientational freedom is restricted.

The crystal structure of p41 peptide bound to SH3 domain (PDB:1BBZ) was solvated in a water box that consists of 5013 TIP3P water molecules. The solvated system was equilibrated for 1 ns in constant pressure and temperature (NPT) at 1 bar and 300 K, respectively. The orientation and the translational degree of freedom are restricted using the same reference atoms and the reference values used in Gumbart et al. [47] with a force constant of 0.1 kcal/(mol · deg²). The p41 peptide conformation is also restricted to the bound conformation using RMSD restraints with a force constant of 10 kcal/(mol · deg²). The distance between the center of mass of the SH3 domain and the p41 peptide was used as the main CV. To compare the performance, we have performed both US/ \mathcal{H} -REMD and US/REST2 simulations. For US/ \mathcal{H} -REMD, 32 evenly spaced replicas spanning a separation distance from 19.5 to 35 Å were used. For REST2 simulations, four effective temperatures (32 × 4 = 64 replicas) were used that span 300–450 K. The simulations adopt an exchange attempt frequency of 1/100 steps and obtain average acceptance ratio ~20%. SH3 domain side chain atoms that are within 4 Å from the p41 peptide at the bound state were selected for the solute tempering, includes 15 residues and 180 atoms in total (Fig. 6). Weighted histogram analysis method (WHAM) [48–50] was used to produce the result for both simulations.

Fig. 7 shows the final PMF from US/ \mathcal{H} -REMD and US/REST2 simulations after 6 ns of simulations per replica. Both methods produced consistent PMF after 6 ns of simulation, and this also matches well with the previous computational study [47]. To compare which method reaches the convergence faster, we have integrated the PMF as below.

$$I = \int_{\text{site}} dr \exp[-\beta(W(r) - W(r^*))] \quad (3)$$

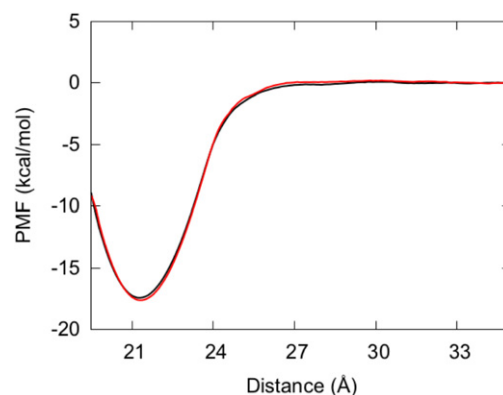


Fig. 7. PMF calculated from US/ \mathcal{H} -REMD (black) and US/REST2 (red) simulations. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

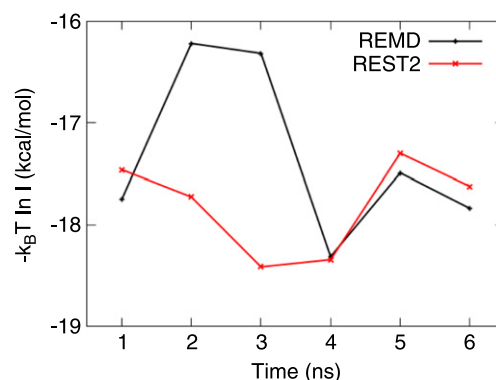


Fig. 8. Convergence of PMF calculated from US/ \mathcal{H} -REMD and US/REST2. PMFs are first calculated by using 1 ns blocks of simulation trajectories and integrated using Eq. (3).

where “site” is defined as the ± 2.5 Å from the minimum PMF distance. $W(r)$ is the distance dependent PMF and r^* is the distance at the bulk state. Fig. 8 shows the integrated PMF using different blocks of simulation trajectories, which demonstrates that the PMFs produced by PMF/REST2 fluctuate less and reach convergence faster than the conventional US/ \mathcal{H} -REMD simulations in the context of separation of two complex biomolecules.

Slow convergence could be attributed to the slow relaxation of receptor and ligand residues near the binding interface. Fig. 9 shows a distribution of distances between the SH3 domain and the p41 peptide and a big gulf in the distributions is observed around the separation distance of 23.5–25 Å, which belong to the windows #11 and #12. Such gap could be explained by the fact that the simulation is initiated from the bound state and the binding interface residues have to adopt a different conformation where no peptide is bound as the peptide leaves the binding interface, which is a slow process. This slow structural reorganization could cause hysteresis in conformations sampled in the corresponding windows. REST2 algorithm helps structural reorganization by scaling the interaction between the binding interface. Although the simulation ran only a short time, the histogram of distance between the two proteins from window #12, which was the collective variable used in the umbrella sampling, became slightly wider (Fig. 9B) and has a better overlap with the neighboring window. Increased overlap between the neighboring windows is generally a good indication of improved sampling.

4. Conclusion

We have implemented REST2 algorithm in the highly scalable program NAMD built upon the parallel programming system

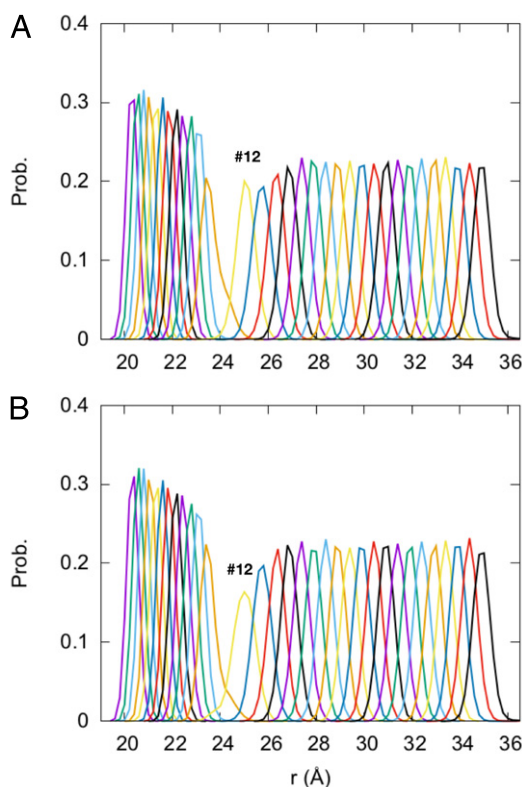


Fig. 9. (A and B) Distribution of the separation distance between SH3 domain and p41 peptide from each umbrella window from (A) US/ℋ-REMD and (B) US/REST2 simulations. A gulf in the distribution is observed between the windows #11 (orange solid line) and #12 (yellow solid line) where the receptor and the ligand are being separated. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Charm++. Our implementation provides a generic and flexible interface for applying REST2 method to variety of biomolecular simulations. Force field parameter rescaling procedures are executed in force computing classes and hot region rescaling parameter is written into Tcl scripting interface. With MCA interface of NAMD a user can perform REST2 based simulation and modify communication patterns of replicas to design novel REST2 based replica exchange scheme without touching NAMD source code. The application of REST2 in NAMD can be straightforwardly combined with other modules for complex biophysical simulations. We have demonstrated the versatility of the present implementation with three REST2 applications, small solute molecule in a large number of explicit solvent, global REST2/ λ -exchange along an entire reaction path in an absolute binding free energy calculation, and umbrella sampling Hamiltonian exchange/REST2. Novel REST2 based algorithm development for free energy calculations and applications to complex biological problems are currently underway.

Acknowledgments

Dr. Sunhwan Jo is a postdoc of Argonne Leadership Computing Facility (ALCF). Dr. Wei Jiang is an ALCF staff scientist. This research used resources of the Argonne Leadership Computing Facility at Argonne National Laboratory, which is supported by the Office of Science of the US Department of Energy under contract DE-AC02-06CH11357. The submitted manuscript has been created by UChicago Argonne, LLC, Operator of ANL. ANL, a US DOE Office of Science laboratory, is operated under Contract No. DE-AC02-06CH11357.

Appendix. Excerpt of Tcl scripting code for REST2

```
if {$replica(index) < $replica(index.$swap)} {
    set pot $POTENTIAL
    set pot2 [replicaRecv $replica(loc.$swap)] ; # Receive instant
    potential value from exchange partner
    sptScaleFactor $replica(factor.$swap) ; # Adjust scaling param-
    eter as of exchange partner
    run 0
    save_array
    set pot_new [expr $saved_array(TOTAL) - $saved_array
    (KINETIC)]
    set pot_new2 [replicaRecv $replica(loc.$swap)] ; # Receive new
    instant potential value from exchange partner
    set delta [expr ($pot_new + $pot_new2 - $pot - $pot2)/($
    BOLTZMAN * $TEMP)]
    set doswap [expr $delta < 0. || exp(-1. * $delta) > rand()]
    replicaSend $doswap $replica(loc.$swap)
}
if {$replica(index) > $replica(index.$swap)} {
    replicaSend $POTENTIAL $replica(loc.$swap)
    sptScaleFactor $replica(factor.$swap) ; # Adjust scaling param-
    eter as of exchange partner
    run 0
    save_array
    set pot_new [expr $saved_array(TOTAL) - $saved_array
    (KINETIC)]
    replicaSend $pot_new $replica(loc.$swap)
    set doswap [replicaRecv $replica(loc.$swap)]
}
set newloc $replica_id
if {$doswap} { # If an exchange attempt is accepted, update the
    physical location of a replica
    set newloc $replica(loc.$swap)
    set replica(loc.$swap) $r
} else {
    sptScaleFactor $replica(factor) # revert scaling factor to original
    value
}
set replica(loc.$other) [replicaSendrecv $newloc $replica(loc.$
other) $replica(loc.$other)] ; # Exchange updated physical location
with next round exchange partner
if {$doswap} {
    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
}
```

References

- [1] Y. Sugita, Y. Okamoto, Replica-exchange molecular dynamics method for protein folding, *Chem. Phys. Lett.* 314 (1999) 141–151.
- [2] H. Fukunishi, O. Watanabe, S. Takada, On the Hamiltonian replica exchange method for efficient sampling of biomolecular systems: Application to protein structure prediction, *J. Chem. Phys.* 116 (2002) 9011–9058.
- [3] P. Liu, B. Kim, R.A. Friesner, B.J. Berne, Replica exchange with solute tempering: a method for sampling biological systems in explicit water, *Proc. Natl. Acad. Sci. USA* 102 (2005) 13749–13754.
- [4] X. Huang, M. Hagen, B. Kim, R.A. Friesner, R. Zhou, B.J. Berne, Replica exchange with solute tempering: ~ efficiency in large scale systems, *J. Phys. Chem. B* 111 (2007) 5405–5410.
- [5] L. Wang, R.A. Friesner, B.J. Berne, Replica exchange with solute scaling: a more efficient version of replica exchange with solute tempering (REST2), *J. Phys. Chem. B* 115 (2011) 9431–9438.
- [6] H.A. Yu, M. Karplus, A thermodynamic analysis of solvation, *J. Chem. Phys.* 89 (1988) 2366–2379.

- [7] A. Roitberg, R. Elber, Modeling side chains in peptides and proteins: Application of the locally enhanced sampling and the simulated annealing methods to find minimum energy conformations, *J. Chem. Phys.* 95 (1991) 9212–9277.
- [8] E. Gallicchio, M.M. Kubo, R.M. Levy, Enthalpy–Entropy and cavity decomposition of alkane hydration free energies: numerical results and implications for theories of hydrophobic solvation, *J. Phys. Chem. B* 104 (2000) 6271–6285.
- [9] H. Li, M. Fajer, W. Yang, Simulated scaling method for localized enhanced sampling and simultaneous “alchemical” free energy simulations: a general method for molecular mechanical, quantum mechanical, and quantum mechanical/molecular mechanical simulations, *J. Chem. Phys.* 126 (2007) 024106.
- [10] K. Huang, A.E. García, Acceleration of lateral equilibration in mixed lipid bilayers using replica exchange with solute tempering, *J. Chem. Theory Comput.* 10 (2014) 4264–4272.
- [11] L. Wang, B.J. Berne, R.A. Friesner, On achieving high accuracy and reliability in the calculation of relative protein–ligand binding affinities, *Proc. Natl. Acad. Sci. USA* 109 (2012) 1937–1942.
- [12] H. Kokubo, T. Tanaka, Y. Okamoto, Two-dimensional replica-exchange method for predicting protein–ligand binding structures, *J. Comput. Chem.* 34 (2013) 2601–2614.
- [13] L. Wang, Y. Wu, Y. Deng, B. Kim, L. Pierce, G. Krilov, D. Lypyan, S. Robinson, M.K. Dahlgren, J. Greenwood, D.L. Romero, C. Masse, J.L. Knight, T. Steinbrecher, T. Beuming, W. Damm, E. Harder, W. Sherman, M. Brewer, R. Wester, M. Murcko, L. Frye, R. Farid, T. Lin, D.L. Mobley, W.L. Jorgensen, B.J. Berne, R.A. Friesner, R. Abel, Accurate and reliable prediction of relative ligand binding potency in prospective drug discovery by way of a modern free-energy calculation protocol and force field, *J. Am. Chem. Soc.* 137 (2015) 2695–2703.
- [14] L.B. Wright, T.R. Walsh, Efficient conformational sampling of peptides adsorbed onto inorganic surfaces: insights from a quartz binding peptide, *Phys. Chem. Chem. Phys.* 15 (2013) 4712–4715.
- [15] J. Huang, A.D. MacKerell Jr., Induction of peptide bond dipoles drives cooperative helix formation in the (AAQAA)₃ peptide, *Biophys. J.* 107 (2014) 991–997.
- [16] T.T. Terakawa, T.T. Kameda, S.S. Takada, On easy implementation of a variant of the replica exchange with solute tempering in GROMACS, *J. Comput. Chem.* 32 (2011) 1228–1234.
- [17] D. Sindhikara, A.D. MacKerell Jr., Exchange frequency in replica exchange molecular dynamics, *J. Chem. Phys.* 128 (2008) 024103–024103.
- [18] D.J. Sindhikara, D.J. Emerson, A.E. Roitberg, Exchange often and properly in replica exchange molecular dynamics, *J. Chem. Theory Comput.* 6 (2010) 2804–2808.
- [19] W. Jiang, M. Hodoseck, B. Roux, Computation of absolute hydration and binding free energy with free energy perturbation distributed replica-exchange molecular dynamics, *J. Chem. Theory Comput.* 5 (2009) 2583–2588.
- [20] J.C. Phillips, R. Braun, W. Wang, J.C. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R.D. Skeel, L. Kalé, K. Schulten, Scalable molecular dynamics with NAMD, *J. Comput. Chem.* 26 (2005) 1802.
- [21] W. Jiang, J.C. Phillips, L. Huang, M. Fajer, Y. Meng, J.C. Gumbart, Y. Luo, K. Schulten, B. Roux, Generalized scalable multiple copy algorithms for molecular dynamics simulations in NAMD, *Comput. Phys. Comm.* 185 (2014) 908–916.
- [22] T. Darden, D.M. York, L.G. Pedersen, Particle mesh Ewald: An $N \log(N)$ method for Ewald sums in large systems, *J. Chem. Phys.* 98 (1993) 10089–10092.
- [23] B.R. Brooks, C.L. Brooks, A.D. MacKerell, L. Nilsson, R.J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caffisch, L. Caves, Q. Cui, A.R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoseck, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R.W. Pastor, C.B. Post, J.Z. Pu, M. Schaefer, B. Tidor, R.M. Venable, H.L. Woodcock, X. Wu, W. Yang, D.M. York, M. Karplus, CHARMM: the biomolecular simulation program, *J. Comput. Chem.* 30 (2009) 1545–1614.
- [24] R.B. Best, X. Zhu, J. Shim, P.E.M. Lopes, J. Mittal, M. Feig, A.D. Mackerell, Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ , ψ and side-chain $\chi(1)$ and $\chi(2)$ dihedral angles, *J. Chem. Theory Comput.* 8 (2012) 3257–3273.
- [25] J.M. Scholtz, H. Qian, E.J. York, J.M. Stewart, R.L. Baldwin, Parameters of helix-coil transition theory for alanine-based peptides of varying chain lengths in water, *Biopolymers* 31 (1991) 1463–1470.
- [26] W. Shalongo, L. Dugad, E. Stellwagen, Distribution of helicity within the model peptide acetyl(aaqa)(3)amide, *J. Am. Chem. Soc.* 116 (1994) 8288–8293.
- [27] H. Fukunishi, O. Watanabe, S. Takada, On the Hamiltonian replica exchange method for efficient sampling of biomolecular systems: Application to protein structure prediction, *J. Chem. Phys.* 116 (2002) 9058–9067.
- [28] V. Munoz, L. Serrano, Local versus nonlocal interactions in protein folding and stability—An experimentalist’s point of view, *Folding & Design* 1 (1996) R71–R77.
- [29] Y. Deng, B. Roux, Computations of standard binding free energies with molecular dynamics simulations, *J. Phys. Chem. B* 113 (2009) 2234–2246.
- [30] H. Li, W. Yang, Sampling enhancement for the quantum mechanical potential based molecular dynamics simulations: A general algorithm and its extension for free energy calculation on rugged energy surface, *J. Chem. Phys.* 126 (2007) 114104–114108.
- [31] D.A. Pearlman, P.A. Kollman, The lag between the Hamiltonian and the system configuration in free energy perturbation calculations, *J. Chem. Phys.* 91 (1989) 7810–7831.
- [32] Y. Deng, B. Roux, Calculation of standard binding free energies: aromatic molecules in the T4 lysozyme L99A mutant, *J. Chem. Theory Comput.* 2 (2006) 1255–1273.
- [33] D.L. Mobley, J.D. Chodera, K.A. Dill, On the use of orientational restraints and symmetry corrections in alchemical free energy calculations, *J. Chem. Phys.* 125 (2006) 084902.
- [34] D.L. Mobley, A.P. Graves, J.D. Chodera, A.C. McReynolds, B.K. Shoichet, K.A. Dill, Predicting absolute ligand binding free energies to a simple model site, *J. Mol. Biol.* 371 (2007) 1118–1134.
- [35] W. Jiang, B. Roux, Free energy perturbation hamiltonian replica-exchange molecular dynamics (FEP/H-REMD) for absolute ligand binding free energy calculations, *J. Chem. Theory Comput.* 6 (2010) 2559–2565.
- [36] J.D. Weeks, D. Chandler, H.C. Anderson, Role of repulsive forces in forming the equilibrium structure of simple liquids, *J. Chem. Phys.* 54 (1971) 5237–5247.
- [37] S. Jo, T. Kim, V.G. Iyer, W. Im, CHARMM-GUI: a web-based graphical user interface for CHARMM, *J. Comput. Chem.* 29 (2008) 1859–1865.
- [38] N. Lu, D.A. Kofke, T.B. Woolf, Improving the efficiency and reliability of free energy perturbation calculations using overlap sampling methods, *J. Comput. Chem.* 25 (2004) 28–40.
- [39] D.A.S. Sitkoff, K. A. B. Honig, Accurate calculation of hydration free-energies using macroscopic solvent models, *J. Physical Chemistry* 98 (1994) 1978–1988.
- [40] J. Kirkwood, Statistical mechanics of fluid mixtures, *J. Chem. Phys.* (1935).
- [41] G.M. Torrie, J.P. Valleau, Nonphysical sampling distributions in Monte-Carlo free-energy estimation—umbrella sampling, *J. Comput. Phys.* 23 (1977) 187–199.
- [42] J. Kačstner, Umbrella sampling, *Wiley Interdisciplinary Reviews: Comput. Mol. Sci.* 1 (2011) 932–942.
- [43] H. Lou, R.I. Cukier, Molecular dynamics of apo-adenylate kinase: A distance replica exchange method for the free energy of conformational fluctuations, *J. Phys. Chem. B* 110 (2006) 24121–24137.
- [44] S. Park, T. Kim, W. Im, Transmembrane helix assembly by window exchange umbrella sampling, *Phys. Rev. Lett.* 108 (2012).
- [45] K. Murata, Y. Sugita, Y. Okamoto, Free energy calculations for DNA base stacking by replica-exchange umbrella sampling, *Chem. Phys. Lett.* 385 (2004) 1–7.
- [46] M.T. Pisabarro, L. Serrano, Rational Design of Specific High-Affinity Peptide Ligands for the Abl-SH3 Domain (2015) 1–7.
- [47] J.C. Gumbart, B. Roux, C. Chipot, Standard binding free energies from computer simulations: What is the best strategy? *J. Chem. Theory Comput.* 9 (2013) 794–802.
- [48] A.M. Ferrenberg, R.H. Swendsen, New Monte-Carlo technique for studying phase-transitions, *Phys. Rev. Lett.* 61 (1988) 2635–2638.
- [49] A.M. Ferrenberg, R.H. Swendsen, Optimized Monte-Carlo data-analysis, *Phys. Rev. Lett.* 63 (1989) 1195–1198.
- [50] S. Kumar, D. Bouzida, R. Swendsen, P.A. Kollman, J. Rosenberg, The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method, *J. Comput. Chem.* 13 (1992) 1011–1021.