JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

Computational Chemistry

Massive-Scale Binding Free Energy Simulations of HIV Integrase Complexes Using Asynchronous Replica Exchange Framework Implemented on the IBM WCG Distributed Network

Junchao Xia, William Flynn, Emilio Gallicchio, Keith Uplinger, Jonathan D Armstrong, Stefano Forli, Arthur J. Olson, and Ronald M. Levy

J. Chem. Inf. Model., **Just Accepted Manuscript** • DOI: 10.1021/acs.jcim.8b00817 • Publication Date (Web): 13 Feb 2019

ACS Publications

is published by the American Chemical Society. 1155 Sixteenth Street N.W., Washington, DC 20036

Massive-Scale Binding Free Energy Simulations of HIV Integrase Complexes Using Asynchronous Replica Exchange Framework Implemented on the IBM WCG Distributed Network

Junchao Xia, [†] [,#] William Flynn, [†,@] Emilio Gallicchio, [‡] Keith Uplinger, [¶] Jonathan D. Armstrong, [¶] Stefano Forli, [§] Arthur J. Olson, [§] and Ronald M. Levy [*,†]

[†] *Center for Biophysics and Computational Biology and Department of Physics, Temple University, Philadelphia, PA 19122*

[†] *Center for Biophysics and Computational Biology and Department of Chemistry, Temple University, Philadelphia, PA 19122*

[‡] *Department of Chemistry, CUNY Brooklyn College, Brooklyn, NY 11210*

[¶] *IBM WCG team, 1177 S Belt Line Road, Coppell, TX 75019*

[¶] *IBM WCG team, 11400 Burnet Road, 0453B129, Austin, TX 78758*

[§] *Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, CA 92037-1000*

[#] *Present address, Department of Research Computing and PACM, Department of Mathematics, Princeton University, Princetion, NJ 08544*

[@] *Present address, Jackson laboratory for Genomic Medicine, 10 Discovery Drive, Farmington, CT 06032*

E-mail: junchao.xia@temple.edu; ronlevy@temple.edu

Abstract

To perform massive-scale replica exchange molecular dynamics (REMD) simulations for calculating binding free energies of protein-ligand complexes, we implemented the asynchronous replica exchange (AsyncRE) framework of the binding energy distribution analysis method (BEDAM) in implicit solvent on the IBM World Community Grid (WCG) and optimized the simulation parameters to reduce the overhead and improve the prediction power of the WCG AsyncRE simulations. We also performed the rst massive-scale binding free energy calculations using the WCG distributed computing gird and 301 ligands from the SAMPL4 challenge for large-scale binding free energy predictions of HIV-1 integrase complexes. In total there are 10 thousand simulated complexes, 1 million replicas, and 2000 microseconds of aggregated MD simulations. Running AsyncRE MD simulations on the WCG requires accepting a tradeo between the number of replicas that can be run (breadth) and the number of full RE cycles that can be completed per replica (depth). As compared with synchronous Replica Exchange (SyncRE) running on tightly coupled clusters like XSEDE, on the WCG many more replicas can be launched simultaneously on heterogeneous distributed hardware, but each full RE cycle requires more overhead. We compared the WCG results with that from AutoDock and more advanced RE simulations including the use of attening potentials to accelerate sampling of selected degrees of freedom of ligands and/or receptors related to slow dynamics due to high energy barriers. We propose a suitable strategy of RE simulations to re ne high throughput docking results which can be matched to corresponding computing resources: from HPC clusters, to small or median-size distributed campus grids, and nally to massive-scale computing networks including millions of CPUs like the resources available on the WCG.

Introduction

There are three critical components for accurate binding free energy prediction using molecular dynamics simulations which are of importance for structure-based drug design[1 7] in the early stage of computer aided drug discovery: the statistical theory and computational approximations for binding free energy calculations; the force eld functions and parameters for describing the physical systems involved such as the receptor, the ligand, and the solvent; the sampling methods for exploring relevant conformational space. In this report we target the last component, how to combine the replica exchange sampling methods with our binding energy distribution analysis method (BEDAM)[8] for calculating absolute binding free energy and optimize the simulation strategy for a distributed resource like the World Community Grid.

Computational methods for calculating binding free energy[1 7] generally perform individual molecular dynamics (MD) simulations at many intermediate thermodynamic states besides the free and fully coupled states. The MD aggregate times, however, are typically limited to the order of microseconds[9 12] even using high performance computing (HPC) resources from XSEDE or specialized CPU/GPU computing units[13 15]. Developing more advanced conformational sampling methods in the context of generalized ensembles[16 31] such as parallel replica exchange (RE) or parallel tempering methods is one way to accelerate the conformational sampling and overcome the timescale challenge due to high free energy barriers resulting in slow dynamics of biomolecular complexes. However, conventional RE methods are implemented with synchronous exchanges[18 22,32,33] and are designed for homogeneous environments such as HPC clusters that require the allocation and maintenance of necessary resources for all replicas during the entire simulation and are intolerant to the failure of any individual replica simulation. Those limitations prevent the traditional SyncRE approach from being a feasible solution for new RE application simulations requiring hundreds to thousands of replicas.[34 36]

On the other hand the available computing units are not limited to high-end HPC clusters, there exist massively distributed computing units such as the IBM World Com-

munity Grid (WCG), a volunteer grid consists of more than 0.7 million members distributed all over the world and 3.0 million heterogeneous computing units including personal or public workstations, laptops and mobile devices, installed with di erent operating systems such as Linux, Mac OS, and Windows. Those distributed computing grids are highly dynamic and heterogeneous due to the volunteer nature of joined members, the diversity of the computing units, and random pause or termination of running jobs. The implementation of conventional RE methods for those distributed computing grids is di cult and also much less e cient since the slowest computing unit determines the e ciency of the whole RE simulation. There exists previous attempts to develop algorithms better suited for heterogeneous computing grids. Serial tempering (ST)[37 39] or simulated tempering only carries out a single thread of an MC/MD simulation in position space, and updates of the thermodynamic state (such as temperature) of the system are performed periodically. ST methods can perform simulations on a single computing unit but requires a pre-estimation of free energy weights at di erent thermodynamics states and their values are iteratively adjusted to equalize state populations visited.[37 39] Similarly serial replica exchange[40] also performs periods of MD simulations in a single replica but the selections of jumping to other thermodynamic states need corresponding estimated potential energy distribution functions at those states accumulated from time series of previous simulations. Other varieties of serial replica exchange such as virtual replica exchange[41] distributed replica sampling[42] and simulated tempering distributed replica sampling[41] also need to estimate similar potential distribution functions in other states during the simulations, which make the massive-scale simulations of complex systems less applicable.

In recent work we proposed an asynchronous parallel replica exchange (AsyncRE) methodology[43] and corresponding python software package[44] to utilize massive heterogeneous computing grids without pre-estimation of those free energy weights or potential distribution functions. This AsyncRE framework removes the synchronizing feature of the standard implementations of parallel replica exchange. This asynchronization feature make it possible to optimize the usage of heterogeneous HPC clusters and dy-

namically distributed computing networks. The main idea of AsyncRE is to use a client-master communication architecture: the local master server assigns all replicas to one of two states, running or swapping in a dynamic way; running replicas are submitted for execution of a predetermined amount of simulation time (an MD cycle or period) on a remote client through ssh or BOINC transportation mechanisms and thermodynamic parameters of swapping replicas are periodically exchanged on the local server independently of the running replicas; a running replica becomes a swapping replica once its remote MD cycle is completed and simulation results are returned to the server. This client-master communication mechanism does not require a direct network link between compute nodes due to the fact that all exchanges are managed and conducted by the master process, and the asynchronous exchange feature does not require a static homogeneous set of compute nodes. This AsyncRE framework allows the loss of compute nodes or replica simulation failures and is able to scale to very large numbers of replicas and take advantage of heterogeneous and dynamically distributed computational resources, including XSEDE high performance clusters, university grid networks, and world-wide distributed networks contributed by volunteer computing units such as IBM WCG. Our previously developed Python package[44] (`https://github.com/ComputationalBiophysicsCollaborative/AsyncRE`) combines the AsyncRE framework for job preparations and exchanges with the BOINC server for job assignments and collections from clients using the Python interface/wrap to separate the interactions between the MD engine and the BOINC client/server. In this manuscript, we report the C++ implementation which allows more direct communications between the IMPACT MD engine, the BOINC Client/Server and the AsyncRE framework for massive scale simulations involving hundreds of thousands of replicas.

The binding energy distribution analysis method (BEDAM)[8] developed in our group computes absolute binding free energies for receptor-ligand systems in implicit solvent using a single alchemical decoupling path. It is among several implementations[1 7] of a statistical mechanics theory of molecular association equilibria based on atomistic molecular dynamics simulations[1,6,45] that are performed to calculate the free energy di erences

between neighboring thermodynamic states along prede ned paths connecting the fully decoupled (apo-protein + free ligand) state to the fully coupled (binding complex) state. Those paths can be realized as in a physical way such as the potential of mean force (PMF) method[46] constructing the free energy pro le along the distance between the ligand and the receptor, or in an alchemical way such as double decoupling methods (decouple the ligand respectively from the receptor site and the solution environments to vacuum[47] from explicit solvent) and single decoupling as in the BEDAM method (decoupling the ligand directly from the receptor site to the solution environment in implicit solvent[8]) using a parameter ($_b$) to de ne intermediate thermodynamic states and rescale (or weakening) the binding energy (including the intermolecular interactions and implicit solvation energies) between the ligand and the receptor. In the early implementation of BEDAM method in the IMPACT[48] molecular simulation package, parallel Hamiltonian replica exchange molecular dynamics (HREMD) sampling (hopping)[26,49] in the synchronous nearest-neighbor replica exchange scheme (SyncRE) was employed to realize the conformational di usion along the alchemical path connecting the bound ($_b = 1$) to unbound states ($_b = 0$) and accelerate the sampling of the intermolecular (external) degrees of freedom between the ligand and the receptor by weakening the binding energies of intermediate states. Recently we have implemented the BEDAM method in the AsyncRE framework[43,44] to enable asynchronous replica exchange simulations using heterogeneous HPC clusters and campus distributed computing grids. In this report, we focus on the combination of BEDAM and AsyncRE framework for massive-scale simulations of protein-ligand complexes on the IBM WCG.

Due to the rescaling (weakening) feature of binding energy using the parameter $_b$ the BEDAM method is capable of carrying out extensive intermolecular conformational sampling of the relative position and orientation of the ligand with respect to the receptor, an advantage of BEDAM[8,50–53] over existing free energy perturbation (FEP) and absolute binding free energies protocols in explicit solvent. BEDAM is particularly well suited to re ne the prediction results from docking methods and to prioritize the hits for further lead optimization using more expensive free energy simulation methods. Namely

docking methods perform the initial search for strong binders using simpli ed conformational sampling and energy scoring functions in order to perform high throughput virtual screening of large libraries of small molecules (on the order of millions). Then BEDAM simulations can be started from the initial complexes of the top 10^2 to 10^3 candidates predicted by docking methods and BEDAM can be used to re ne binding poses and provide more accurate estimations of binding free energies which can be used to increase enrichment factors. The top predicted results of BEDAM (on the order of 10^1 to 10^2) can be passed to carry out explicit solvent FEP simulations with more accurate atomistic simulations of protein-ligand complexes and more detailed physical potentials and solvation environments.[1 7] *Hence BEDAM targets a niche between low resolution docking methods and high resolution FEP methods in explicit solvent.*[1,6,45] *with the major goal of refining the predicted results from docking methods and significantly reducing the number of candidates to perform further lead optimization using FEP methods in explicit solvent.*

Although post docking analysis using pharmacophore model lters and visual checks can improve initial docking predictions,[54] in this report we focus on an alternative BEDAM re nement procedure starting from initial AutoDock predictions using the SAMPL4 library for the HIV-1 integrase (HIV-1 IN).[53 55] The SAMPL4 library is a low diversity library of 300 ligands designed to target the LEDGF allosteric site of HIV-1 integrase, which was the result of a prior lead identi cation campaign.[55] The protocols described here have been implemented for the FightAIDS@Home Phase 2 project, using the physics based BEDAM model and AsyncRE computational method running on the IBM WCG volunteer grid to score the top candidates from the Phase 1 screening using the AutoDock protocol. Massive binding free energy calculations using replica exchange on heterogeneous distributed volunteer computing networks like WCG is challenging in part because of the need for optimization of the parameters which control how the AsyncRE simulation runs on the massive grid to utilize e ciently the resources in a highly dynamic and heterogeneous environment due to the fact that the wall clock time to nish a MD simulation at client side is very diverse. This diversity

is not only because of the diversity from computing units (in CPU speeds and device types such as desktop, laptop and mobile) but also based on the fact that the simulation can be paused, stopped and aborted in unpredictable ways. Furthermore, there also exist di erent kinds of time overheads including the delay of returning results on the client side and others on the server side such as generating and transferring of input les, queuing of individual jobs, and exchanges of waiting replicas. Previous attempts to utilize distributed computing networks to calculate binding free energies were limited to a few examples[56 59] using short MD simulations without exchanges to build Markov models utilizing cloud computing. These studies have mostly been focused on protein folding and large conformational changes.[41,60 63] To our knowledge no systematic studies for optimizing the simulation parameters and designing general procedures for massive-scale simulations of protein-ligand binding have been reported. We ll this gap using as an example the SAMPL4 library for the HIV-1 integrase, one of several important viral enzymes[64] during the life cycle of the HIV-1 virus. Integrase is responsible for integrating the viral genome into the host genome with the help of the human LEDGF protein (a transcription factor) linking the HIV-1 integrase to the human chromosome.[65] The SAMPL4 library is a focused library developed for lead optimization targeting the LEDGF binding site of HIV-1 IN and preventing the LEDGF protein from binding to stop the insertion of the HIV genome into the human genome, consisting of 300 ligands sharing several common molecular sca olds such as benzoic acid, benzodioxole acid, and benzodioxine acid.[53,55] Since the SAMPL4 library is not very diverse this presents a particular challenge to binding a nity predictions based on docking alone.

Methods

Binding energy distribution analysis method (BEDAM)

BEDAM is based on the statistical mechanics theory of molecular association[6] and the Widom potential distribution theorem.[66] The ligand is decoupled from the complex environment with full intermolecular interactions (bound state) to an uncoupled (unbound)

state in the implicit solvent environment through a prede ned alchemical path which is realized by introducing intermediate states with a parameter ($_b$) rescaling (or weakening) the binding energy (including the intermolecular interactions and implicit solvation energies) between the ligand and the receptor. The BEDAM method[8] calculates the absolute binding free energy G_b^0 between a receptor P and a ligand L by employing a $_b$-dependent e ective potential energy function as follows:

$$V_{\ b}(r) = V_0(r) + \ _b u(r) \tag{1}$$

where $V_0(r) = V(r_P) + V(r_L)$ is the total potential energy of the complex including both the receptor P and the ligand L at the uncoupled (unbound) state in the implicit solvent, and $u(r) = u(r_P\ r_L) = V(r)\quad V_0(r) = V(r_P\ r_L)\quad V(r_P)\quad V(r_L)$ is de ned as the binding energy for each complex conformation ($r = (r_P\ r_L)$), corresponding to the di erence between the total e ective potential energies $V(r)$ with implicit solvation e ects[67 69] of the fully coupled (bound) and decoupled (unbound) states of the complex with the same xed internal conformations. The standard free energy of binding for this system can be calculated as[6,8]

$$G_b^0 = \ k_B T \ln C^0 V_{site} + \ G_b = \ k_B T \ln C^0 V_{site} \ k_B T \int du p_0(u) e^{-\ u} \tag{2}$$

with the rst ideal term on the right $k_B T \ln C^0 V_{site} = k_B T \ln \frac{V_{site}}{V^0}$ representing the entropic change when moving the free ligand in the volume (V^0) of standard concentration into the volume of the binding site (V_{site}) and the second one $G_b = \ k_B T \int du p_0(u) e^{-\ u}$ denotes the total free energy change when turning on the interaction energies between the receptor and the ligand. Moreover k_B is the Boltzmann constant, T is temperature, C^0 (=1M) is the the standard concentration of ligand molecules. and $p_0(u)$ is the probability distribution of binding energy ($u(r)$) collected in an appropriate decoupled ensemble of conformations in which the ligand is present in the binding site but the interactions between the receptor and the ligand are turned o

but both molecules are present in the solvent continuum. To obtain a good estimation of $p_0(u)$ for biomolecules with slow dynamics, a series of discretized $_b$ values between 0 and 1 are introduced to represent intermediate thermodynamic states and depict an alchemical path decoupling the ligand from the complex environment with full interactions in the receptor site ($_b = 1$) to a pure solvent state without protein interactions ($_b = 0$). In practice the G_b can be calculated by reweighting analysis of all snapshot binding energies of all thermodynamic states using the unbinned UWHAM[70] or similar methods.[71–73]

Comparing with the docking methods with empirical docking score functions, one of the advantages of the BEDAM method is the ability to capture the entropic changes for binding processes[8,50,51]. In the BEDAM method, the binding process can be decomposed into two separate steps: 1) the ligand and the receptor reorganizes its conformational (internal and external) ensembles in the unbound state to match corresponding distributions in the bound complex; 2) the receptor-ligand interactions are created in the binding site with no entropic change involved due to the fact of no change in the con g-urational ensemble of the binding partners. The free energy change for the second stage can be calculated as the average binding energy at the fully coupled state ($_b = 1$) , represented by $G^o_{II} =< u >_{RL}$. We should emphasize that G^o_{II} not only includes the interaction energies between the ligand and the receptor but also includes the solvation contribution (the di erence of implicit solvation energy between the bound and unbound state). In contrast, there are both entropic and enthalpic changes in the rst stage due to the reorganization of conformational ensembles, denoted as the reorganization free energy $G^o_I = G^o_{reorg}$. Hence the standard binding free energy can be described as

$$G^0_b = G^o_I + G^o_{II} = G^o_{reorg} + < u >_{RL} \tag{3}$$

We should mention that using of implicit solvation model is crucial to achieve the decomposition of binding free energy and the single-path decoupling of the BEDAM method. Namely the ligand is only decoupled from the coupled complex environment

to the free state in the solution and the additional decoupling from the solution to the vacuum as in the double decoupling method is not necessary since the solvation energies can be evaluated from the ligand and receptor conformations through the AGBNP2 solvation model.[67 69] Implicit solvent models could also lead to faster convergence of free energies than explicit models due to lack of solvent uctuations although simulations with implicit solvation are more di cult to parallelize due to the relatively small number of atomic sites and the complexity of algorithms to evaluate the Born radii.[67 69] For some systems which require explicit representation of water mediated interactions around the binding site, simulations using implicit models may not fully capture the ligand binding e ects and become less accurate than that using explicit solvation models as in the double decouple method.[47]

WCG AsyncRE implementation of BEDAM method

The early implementation of the BEDAM method was accomplished in the SyncRE scheme through the MPI implementation in the IMPACT[48] simulation package targeting homogeneous HPC clusters where all CPUs involved in an RE simulation have the same speed but with a limited number (tens) of replicas due to the necessity of synchronization of replica exchange. To overcome this synchronization limitation and enable large-scale simulations (thousands) using heterogeneous HPC clusters and distributed computing resources, recently we proposed the AsyncRE framework[43]and developed a corresponding python package[44] where only the replicas in the local waiting list can participate in the exchange process using an algorithm for the sampling of the state permutation space which does not require the prior identi cation of neighboring states and attempts to exchange two replicas are randomly picked but follow the same Metropolis criterion and repeated many (M^3 to M^5 where M is the total number of replicas on the exchange queue) times to reach the in nite swap limit.[43,74,75] This Python package (`https://github.com/ComputationalBiophysicsCollaborative/AsyncRE`) has been installed to perform AsyncRE simulations on XSEDE high performance resources, and on BOINC distributed computing networks at Temple University, Brooklyn College at

CUNY.[43,44,76] The Python interface/wrap separate the direct communication between the MD engine and the BOINC server and there is also no direct data transferring between the MD engine and the BOINC client for the screensaver visualization. In this way we can minimize the tasks for code development of new implementations using di erent MD engines, which is more suitable for setting up small or median size computing grids up to 10^3 computing units. For massive computational grids such as the IBM WCG grid involving hundreds of thousands of computing units, in this report we introduce the C++ implementation (See Fig. 1 for the corresponding work ow for FightAIDS@Home Phase 2 re nement) which allows more directly communications between the IMPACT MD engine and the BOINC Client, and between the BOINC Server and the AsyncRE framework for massive scale simulations involving hundreds of thousands of replicas, although it shares the same framework of AsyncRE algorithm with the Python implementation for HPC clusters and university grids. We have summarized the major di erences and provided more implementation details in Supporting Information.

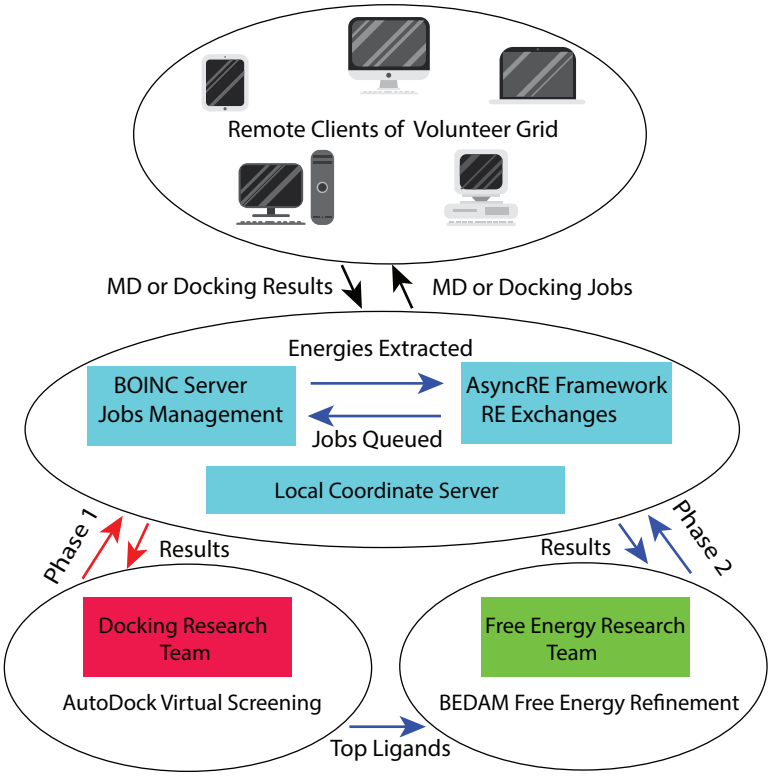

Figure 1: Schematic work ow of IBM implementation of the AsyncRE framework.

Beside the di erent implementations of BEDAM targeting various computing re-

sources, as an alternative way to accelerate conformational sampling of slow dynamics from receptor or ligand, we have also introduced attening potentials on selected bonded and nonbonded intramolecular interactions to lower corresponding energy barriers and accelerate the conformational sampling of internal degrees of freedom of ligands and or receptors (See a short summary on this methodology in Supporting Information and more details in Ref.[76,77]). Further improvement of sampling for WCG simulations can be achieved by combining this attening feature with the WCG AsyncRE framework which is under development.

Metrics to evaluate the e ciency of the REMD algorithm

There is no universally accepted metric for evaluating the sampling e ciency in REMD simulations but there are several methods described in previous work such as the replica relaxation correlation time and end-to-end transit time of the replica state index,[74] variances of the estimated means of relevant observables,[78 80] slowest eigenvalue of the Markov chain from analyzing the state transition matrix,[81] root-mean-square deviation (RMSD) of related observables of test simulations from the corresponding reference simulation,[82,83] and our previous statistical ine ciency analysis[43] by extracting the total e ective relaxation time of the binding energy at = 1 0 ($_b$ in Eq.1 is shortened as for better notation here after), which includes all relaxation e ects both from MD simulations and replica exchange mixing. To include all of the relaxation times at di erent values and the replica exchange mixing time, we borrow the idea of the di usion coe cient in coordinate space[84] and de ne the di usion coe cient of the replica exchange system in the space as

$$D = \frac{< (_i \quad _j)^2 >}{2 < _{ii} >} \tag{4}$$

where the average is (< >) calculated through all states, $_{ii}$ is life time of state $_i$ measuring the total simulation time of a replica stays in the $_i$ state before jumping to any di erent state $_j$, $_i$ $_j$ is the jump di erence of a replica from the $_i$ state to

any di erent state $_j$.

System preparation and computational details

BCD toy model

The toy model complex is b-cyclodextrin-heptanoate as depicted in Fig. 2 was used to benchmark the ported IMPACT MD engine and the WCG implementation of AsyncRE framework. It is one of host-guest systems well investigated in our previous algorithm developments and benchmarking work.[51,52] Besides the standard tests for the IMPACT MD engine to reproduce single point energies, we performed standard 1D BEDAM simulations at 16 values (0.0, 0.001, 0.002, 0.004, 0.01, 0.04, 0.07, 0.1, 0.2, 0.4, 0.6, 0.7, 0.8, 0.9, 0.95, and 1.0) using the OPLS-AA force eld,[85,86] the AGBNP2 implicit solvent model,[69] and three di erent implementations of RE framework: a) SyncRE using MPI targeting homogeneous HPC clusters; b) AsyncRE framework implemented by Python interface for median-size campus computing grids such as Temple Grid; c) AsyncRE framework implemented by C++ for massive-scale distributed grids such as the IBM WCG grid. The MD period for the SyncRE simulations is 100 steps (0.2 ps) and 10,000 steps (20ps) for both types of AsyncRE simulations. From Fig.2b we can see that all simulations (after 10ns) converge to the consistent binding energy distributions at $= 1\ 0$ which supports the correctness of the di erent implementation of RE frameworks (SyncRE on NSF XSEDE and other HPC clusters, AsyncRE on heterogeneous campus grids. and massive AsyncRE on IBM WCG).

HIV-1 integrase

The 301 ligands (53 experimental binders and 248 nonbinders, 451 protein-ligand complexes considering the protonation states and other tautormeric extensions of ligands) of HIV-1 IN complexes were prepared previously[53] for the HIV-1 integrase virtual screening SAMPL4 challenge[54,55] with the aim of predicting likely binders to the LEDGF site of integrase[87] using the BEDAM method for binding free energy calculations[8,50 52].

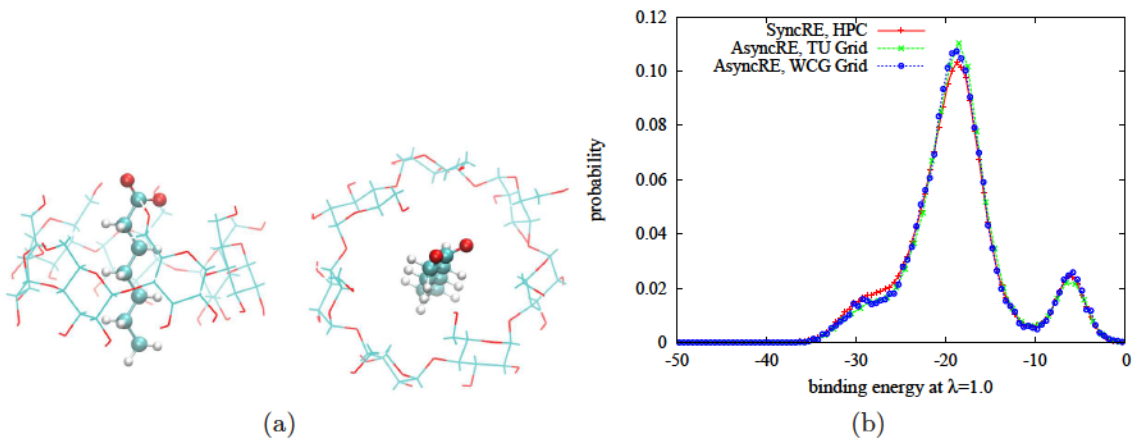

Figure 2: (a) Side and top view of b-cyclodextrin-heptonoate complex, (b) binding energy distributions of λ=1.0 at the temperature of 300K calculated from the BEDAM simulations using three different RE implementations.

Docked structures for all ligands were obtained by the flexible docking procedure using AutoDock Vina for the SAMPL4 challenge[54] and initial reference crystal complex with the PDB ID 3NF8 as shown in Fig. 3. Only the best docking conformation of each complex was selected to perform the binding free energy refinement for the original simulations using the SyncRE framework on HPC resources and submitted to the SAMPL4 challenge.[53] The tremendous computing resources of IBM WCG grid allow us to include the remaining top 8 predicted poses from AutoDock Vina in addition to the best one for each of 451 protein-ligand complexes and extend the number of replicas per complex from the original 20 to 100 as listed in the Supporting Information. This corresponds to 900 replicas per complex (100 replicas/pose $\times$ 9 poses). Furthermore all simulation parameters, force field parameters, and restraint methods were kept the same as the previous work[53] as described in more detail in the Supporting Information. We performed different types of AsyncRE BEDAM simulations for the SAMPL4 ligands as listed in Table 1 using the WCG resources. The first set of WCG simulations for the SAMPL4 library were carried out using the independence sampling scheme with the same 20 λ values as the original SAMPL4 submission but without exchange feature and the simulations were extended to all 9 predicted poses from AutoDock. A large MD period of 100,000 steps was selected in order to minimize the number of resubmissions

(hence the overhead) due to the fact that an individual job on the client side is limited to a few hours of the CPU running time (after removing any pauses and interrupts) and a maximum return time of 7 days (namely a job is considered as failed if it can not be returned in 7 days) based on the previous setting for other projects on WCG. The remaining two sets of WCG simulations of the whole SAMPL4 library use the AsyncRE framework to perform exchanges among 100 replicas (100 states) but with di erent MD periods of 10,000 (20 ps) and 50,000 (100 ps) steps per exchange cycle respectively. For a comparison, we also include attening AsyncRE simulations that were carried out previously[76] on the Temple University Grid.

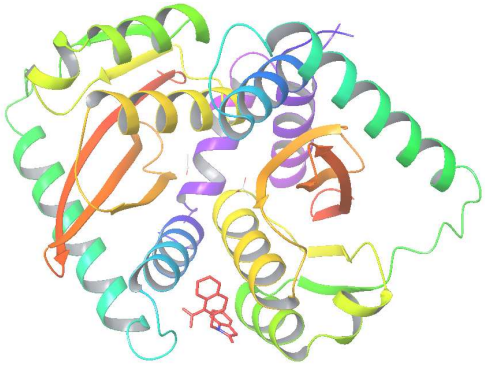

Figure 3: Graphic representation of HIV-1 IN complex from PDB ID 3NF8. The red sticks on the bottom show the corresponding ligand bound to the LEDGF site.

Results and Discussion

Optimization of simulation parameters

In the conventional parallel synchronous replica exchange (SyncRE) framework, each replica is assigned a di erent thermodynamic state (a value in this report) and periodically, after all of the replicas complete a given number of steps (MD period), replicas attempt to exchange their current state assignments with typically one of their nearest neighbors at adjacent thermodynamic states. Exchange attempts are accepted based on well-established microscopic reversibility requirements.[18 22,32,33] To improve the ef- ciency and convergence, conventional REMD has been extended mostly through the

Table 1: Summary table for di erent types of BEDAM simulations performed for all SAMPL4 HIV-1 IN complexes.

abbreviation	WCG-Indp[a]	WCG-10K[b]	WCG-50K[c]	Flat-All[d]
starting pose[e]	docking	docking	docking	docking
simulation time per replica	3.0ns	1.2ns	3.0ns	3.0ns
exchange scheme	No	AsyncRE	AsyncRE	AsyncRE
number of replicas[f]	20	100	100	20
used docking poses	9	9	9	1
total of replicas per complex	180	900	900	20
MD period	200ps	20ps	100ps	2ps
attening type	no	no	no	torsional+ nonbonded
binders	53	53	53	53
nonbinders	248	248	248	248
total MD simulations (s)	243.5	487.1	1217.7	27.1
computer resource	WCG	WCG	WCG	Temple Grid

[a]WCG simulations without replica exchange (independence sampling);
[b]WCG AsyncRE simulations with a MD period of 10K steps;
[c]WCG AsyncRE simulations with a MD period of 50K steps;
[d]AsyncRE simulations using the local Temple Grid with attening potentials applied to ligands;
[e]All docking poses are from the AutoDock Vina predictions for the original SAMPL4 submission;
[f]See all lambda values listed in Supporting Information for more details.

modi cation of the Hamiltonian [24,25,27 29,36,88 95] as for example our recent addition of biasing potentials for BEDAM simulations.[76] On the other hand, some studies focused on expanding the exchange dimensions[19,21,36,96] and optimizing the setting of simulation parameters not limited to the number and distribution of replicas (such as values[97] and temperatures[18,98 100]), the number of exchange attempts,[74,101] and exchange frequency (MD period).[82,83,102,103] Although the e ciency and convergence analysis of SyncRE MD in comparison with conventional MD without exchange has been performed in many previous studies.[78 80,104 108] there still exist debates on how to best select simulation parameters such as the length of individual MD simulations (MD or exchange period) and the number of exchanges attempted after a MD period. Early results showed that the e ciency of REMD could be signi cantly reduced when the MD period is smaller than a certain size,[104 106] while recent studies[82,83] found that the e ciency increases monotonically as the MD period becomes smaller. Previous ndings[74,101,106,109] also illustrated that the number of exchange attempts should be chosen as frequently as feasible so as to reach a so-called in nite swap limit.[75,102,103,110]

The characteristics of the WCG distributed computing grid are unsuitable for SyncRE, we implemented an asynchronous RE protocol which di ers from SyncRE in various respects[43,75]: 1) Only a portion of all simulated replicas are available locally on the server side to participate in the exchange process and the rest (half of the total number of replicas for our WCG simulations) reside on the client side for individual MD simulations; 2) Because the randomness of available replicas makes the nearest-neighboring exchange scheme of the SyncRE not feasible, we developed a random pairwise exchange scheme whereby many pairs of replicas (not limited to the nearest neighboring replicas) in the waiting pool are randomly selected for exchange attempts so as to reach the in nite swap limit;[43,74,75] 3) AsyncRE simulations generally require relatively large MD periods (> 1 picoseconds = 500 MD steps) to minimize overheads (such as preparation and transportation of input and output les between the coordination server and the computing clients, random pauses of the computing client, and the job queuing time on the server) relatively to the running time of a MD period. Due to the characteristics of AsyncRE

and the dynamic and heterogeneous nature of computing grids, it is di cult to perform a direct comparison on e ciency and convergence between the AsyncRE simulations on the WCG and the SyncRE simulations. To optimize and guide our AsyncRE simulations on WCG, we consider two simulation parameters for our benchmarking tests using the AVX38789 ligand from the SAMPL4 library: a) the total number of replicas varying from 20 to 1000 (Note that Table 1 only lists the simulations using 20 and 100 replicas for the whole SAMPL4 library.) to test the e ect of larger numbers of replicas which can result in more overlaps among the binding energy distributions of replicas and increase the accepted ratio of exchanges; and b) the MD period varying from 10 thousand to 50 thousand steps to reduce the communication overhead at the expense of frequency of exchanges but this might lead to slower equilibration.

Table 2 shows the accepted ratios of exchanges (the total number of successful exchanges divided by the total number of attempted exchanges) calculated from di erent benchmarking tests of WCG AsyncRE simulations for the AVX38789 ligand in complex with HIV-1 IN. While its value is not a ected by the MD period, the accepted ratio almost doubles as the number of states is increased from 20 to 1000 albeit not in a linear way. For example, the accepted ratio of accepted changes to exchange attempts increases only slightly (0.2 to 0.22) when the number of states is increased from 100 to 1000. The increase of accepted ratio is based on the fact that increasing the number of states increases the overlap of binding energy distributions of two exchanged replicas. The corresponding accepted ratios calculated from the SyncRE simulations for the original SAMPL4 submission are around 0.5 where the exchanges are attempted only between nearest neighboring states (with more overlaps in binding energy distributions) and results in more successful exchanges when the total number of attempted exchanges is xed. Instead, AsyncRE simulations require many more attempted exchanges to reach a similar swap limit.[43,74,75] Due to the insensitivity to the MD period, the accepted ratio is not a good quantity to optimize the simulation parameters. Similar trends can be found on the mean square jump of values ($< {}_{ij}{}^2 >$) as shown in Fig. 4a. On the contrary, the mean life time of replicas ($< {}_{ii} >$ in the unit of picoseconds) decreases as the MD period is

decreased and its value approaches to a limit (the MD period) as the number of replicas is increased to 1000 as exhibited in Fig.4b. Figure 4c displays the di usion coe cients (D) in the space (de ned in Eq. 4) which include both e ects of the mean square jump of values and the mean life time of replicas. It is pronounced that the di usion coe cient decreases as the MD period is increased when the number of replica is xed. The di usion coe cient also exhibits a large increase (almost doubled) when the number of replicas is increased from 20 to 100. However, there are no signi cant changes (only 10%) to the di usion coe cient as the total number of replicas is increased from 100 to 1000 with the MD period is xed. Based on this observation we selected 100 replicas for each AsyncRE WCG simulation of the whole SAMPL4 library. Because there are more than 4 fold increase of the di usion coe cient, in principle we should select 10 thousand steps for the MD period instead of 50 thousand if there exists no overhead (or the same percentage of overhead) for exchanges and queuing on the server between two adjacent MD periods. However, reducing the MD period will increase the number of cycles required of replica resubmission (when the total number of simulated MD steps is xed for each replica) and increase the percentage of the overhead that dynamically depends on the available WCG resources and the settings of clients and server. We will discuss the e ects of the MD period in more detail in the following section where we analyze two sets of WCG AsyncRE simulations for the whole SAMPL4 library using MD periods of 10 and 50 thousand steps respectively.

Table 2: Accepted ratios of exchanges calculated from di erent benchmarking tests of WCG AsyncRE simulations for the AVX38789 ligand in complex with HIV-1 IN.

	20 s	100 s	1000 s
10K steps	0.134	0.198	0.217
30K steps	0.138	0.198	0.217
50K steps	0.139	0.193	0.218

Overhead analysis due to the heterogeneity of WCG simulations

We conducted two sets of AsyncRE simulations on WCG (with MD periods of 10K and 50K

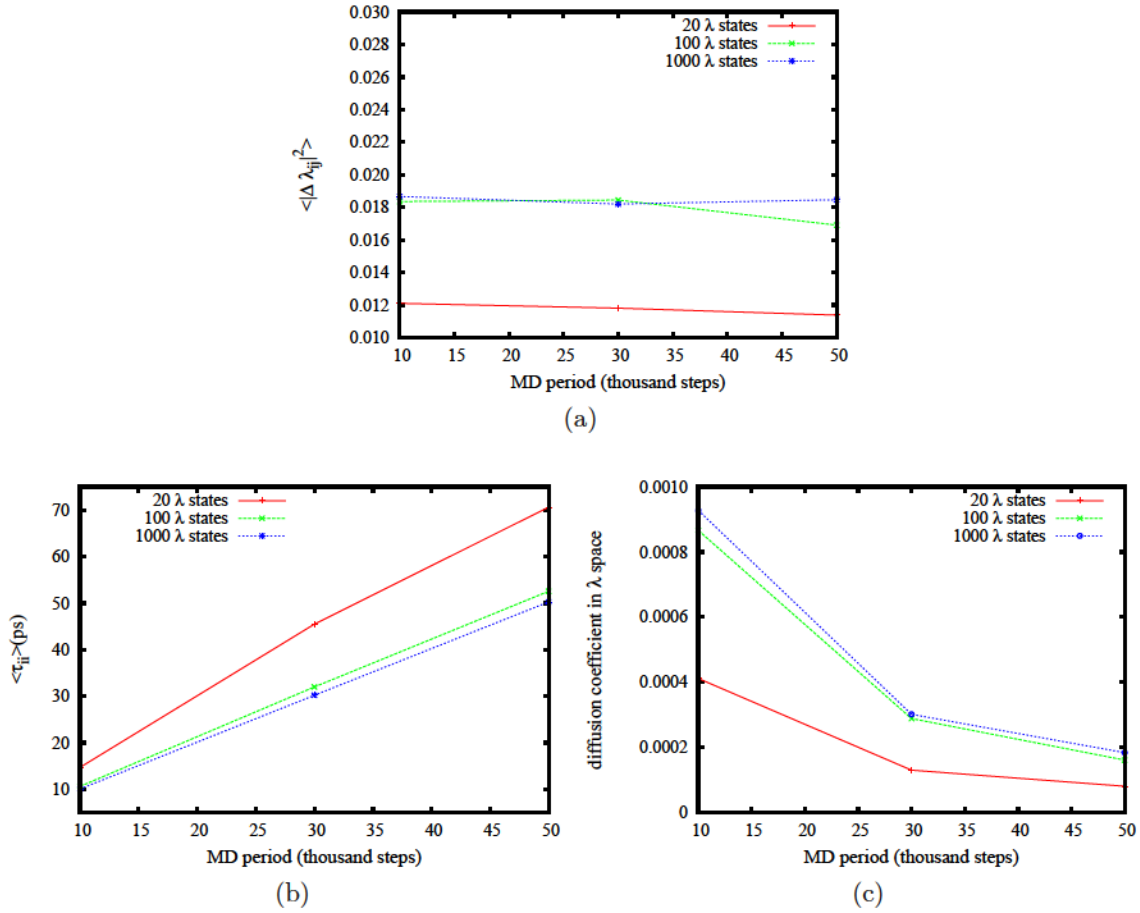

Figure 4: Optimization measures. (a) mean square of lambda jumps, (b) mean lifetime of lambda state, and (c) the diffusion coefficient calculated from BEDAM simulations with different combinations of the number of lambda states (20, 100 and 1000) and the MD period from WCG benckmarking tests on the AVX38789 ligand in complex with HIV-1 IN.

respectively). Each set includes 451 complexes and 9 predicted poses from AutoDock for each complex. There are 100 replicas for each BEDAM simulation started from each pose. In total, there are 2 451 9 100 = 811 800 replicas and with half of them (405,900) submitted for MD simulations on the WCG client side and half of them checkpointed to disk and participating in exchanges on the WCG master server side. The total available computing units for the FightAIDS@Home Phase 2 project uctuates around 100,000. Therefore at any one time roughly three quarters of submitted MD jobs are waiting for execution on the WCG BOINC server and only one quarter are running on client CPUs. To perform the overhead analysis we de ne four quantities that can be found or calculated from the returned info les of WCG simulations: 1) the MD CPU time which represents the wall clock time for a computing unit to complete a period of MD simulation of 10K or 50K steps without any pauses (temporally idling due to the occupation of the volunteer CPU by other tasks) on the client side; 2) the replica client time which denotes the total wall clock time that a replica resides on a client node including restarting if any from a crashed job and the delay in returning a completed MD job (A MD job is considered as failed if it can not be returned within one week); 3) the replica server time which accounts for the total wall clock time that a replica stays on the WCG server side including the time spent on the exchange process of AsyncRE and the time spent on the MD execution queue of BOINC server when the replica is selected for the next cycle of MD period after exchanges; 4) The replica cycle time (=replica client time + replica server time) which is calculated from the wall clock time di erence between two sequentially returned MD jobs for the same replica (Its mean value multiplied by the total number of RE cycles is equal to the mean value of wall clock time to nish the total simulation per replica). There is no direct way to obtain the replica server time (from the returned info les of WCG simulations) but we can get the value by subtracting the replica client time from the corresponding replica cycle time after performing a one to one match.

The probability distributions of MD CPU times calculated from two sets of WCG AsyncRE simulations are displayed in Figure 5a. The most probable value (mode) for a client to complete a MD period of 50K steps is 5 hours, almost 5 times longer than

that for 10K steps (1 hour) which is also consistent with the mean values as listed in Table 3. Figure 5b shows the corresponding distributions of replica client times, with the most probable values (1.1 and 9.1 hours for 10K and 50K respectively) not more than twice larger in comparison with the corresponding values of MD CPU time as listed in Table 3. The mean values of the replica client times behave in a di erent way for these two sets of WCG AsyncRE simulations: The mean value from 10K simulations (6.9 hours) is almost 5 times larger than the mean value (1.3 hours) of MD CPU time but there is only twice larger for 50K simulations. Namely the 10K simulations have much larger overhead (5 .vs. 2) on the client side due to the fact that although the returning time of a MD job can be distributed in a wide distribution due to the heterogeneity of CPU speeds and operation settings of client computers, the maximum return time for a MD job is the same (7 days) for both sets of simulations as exhibited by the unsymmetrical long tail distributions in Figs.5b. Similar distributions of replica cycle times are shown in Fig. 5c. The mode values (see Table 3) from the simulations of 10K and 50K MD periods are around 60 and 70 hours respectively and there is an exponential tail at the longer time side due to the random selection of exchanged replicas to be submitted to a MD simulation on the client side, with the mean values of 74.6 and 113.2 hours (for 10K and 50K respectively). The reason that those two modes are so close is that the replica server time (calculated by subtracting the replica client time from the replica cycle time) is more than 60 hours for those two sets of simulations, and their mean values (67.7 and 99.6 hours for 10K and 50K respectively) take the major part (91% and 88% for 10K and 50K respectively) of the mean replica cycle time (74.6 and 113.2 hours for 10K and 50K respectively) as listed in Table 3. For an ideal case when there is no overhead from the server, theoretically the minimum value of mean replica sever time should be the same as the mean value of replica client time since only half number of total simulated replicas are assigned to run MD jobs on the client side and half of them stay on the server side. Namely, on average a replica has to wait for an additional period of replica client time to be resubmitted for the next cycle of MD simulation. Hence, the minimum value of the mean replica cycle time is twice (13.9

and 27.2 hours for 10K and 50K respectively) as the mean value of the replica client time. Assuming that we can reduce the mean value of replica server time by a factor of 4 if we do not over submit the WCG grid and the mean value of replica client time is the same as before, then the mean values of replica cycle times are 23.8 and 38.5 hours for 10K and 50K respectively (1.7 and 1.4 times of the minimum values of mean replica cycle time). Hence, when considering the overhead from both the client and the server sides together, the nal overall overheads are 8.5 (5 on the client side 1.7 on the server side) and 2.8 (2 on the client side 1.4 on the server side) times for 10K and 50K simulations respectively when the WCG grid is not over submitted. Because of the heterogeneity of WCG simulations and the di erent percentages of overhead as illustrated in Fig.5 and Table 3, the average value of the total wall clock time to nish a total of 600,000 MD steps (1.2ns) for each replica using 10K MD period is around 3 times larger than that when using a 50K MD period.

Although the percentage of overhead is considerable large (8.5 and 2.8 times for 10K and 50K respectively without over-submission) , the advantage of WCG simulations becomes more obvious when we consider the total number of available computing units and the total number of simulations we can submit. Table 4 lists the estimations of wall clock time for di erent types of BEDAM simulations performed for all 451 SAMPL4 HIV-1 IN complexes. The total MD length per replica has been normalized to 1.2ns for a consistent comparison. Our original SAMPL4 submission required 2 millions service unit (SU, 1 core per hour) in total on XSEDE HPC clusters and 48 days to nish all jobs when using 1000 cores at the same time. One set of WCG AsyncRE simulations for the SAMPL4 library needs 90M SUs (around 9 years allocations for a big computing project) and would take 2160 days to nish on XSEDE if the same 1000 cores are used to running all jobs without considering the queuing time. Using the current available resources (100,000 cores) for the FightAIDS@Home Phase 2 project the required time to nish 1.2ns per replica for one set of AsyncRE simulations is only 56 days for the 50K MD period and 180 days for 10K MD period respectively. We should emphasize that the number of required days can be reduced greatly (19 and 59 days respectively)

if the WCG grid is not over submitted (with 400,000 CPUs available) to reduce the average queuing time on the BOINC server (and the replica server time), by 14 and 36 days respectively for the ideal case without any overhead from the server. Furthermore, a signi cant reduction can also be achieved on the client side by using a much smaller value of the maximum return time of MD jobs and asking the clients return the completed jobs as soon as possible.

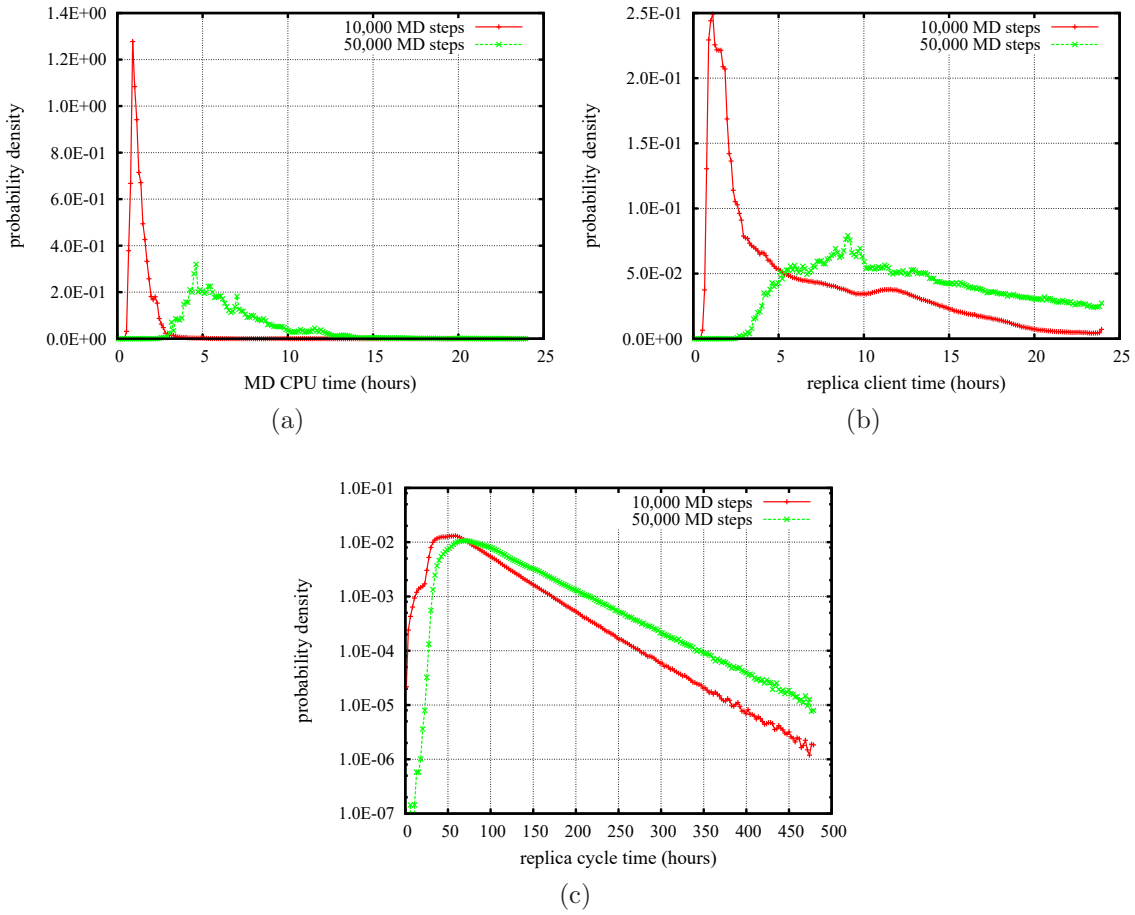

Figure 5: Statistical distributions calculated from 405,900 replicas for each SAMPL4 library and 600,000 MD steps in total for each replica: (a) MD CPU time (wall clock time to complete a MD period after removing the idling time due to random pauses of MD simulation at client side); (b) replica client time (wall clock time for a replica stays on the client side), (b) replica cycle time (wall clock time for a replica stays on the server side).

Table 3: Execution time statistics calculated from two sets of AsyncRE WCG simulations for the 451 SAMPL4 HIV-1 IN complexes (in the unit of hours).

MD period	di erent time	mode	mean	standard deviation
10K Steps	replica cycle time	58.8	74.60	49.67
	replica server time	-	67.67	-
	replica client time	1.14	6.93	5.95
	MD CPU time	0.9	1.34	0.64
50K Steps	replica cycle time	70.8	113.15	64.34
	replica server time	-	99.57	-
	replica client time	9.06	13.58	7.81
	MD CPU time	4.62	6.46	2.70

Table 4: Wall clock time estimates for di erent types of BEDAM simulations of the 451 SAMPL4 HIV-1 IN complexes. In this analysis, the total MD length per replica is normalized to 1.2ns for consistency.

abbreviation	WCG-Indp	WCG-10K	WCG-50K	Flat-All
Exchange scheme	No	AsyncRE	AsyncRE	AsyncRE
number of replicas	20	100	100	20
docked poses	9	9	9	1
total MD (s)	97.4	487.1	487.1	10.8
total CPU hours required on XSEDE	18M	90M	90M	2M
total wall clock time on XSEDE (days) base on 1000 cores available	432	2160	2160	48
total wall clock time on WCG (days) based on 100,000 cores available	25	180	56	
total wall clock time on WCG (days) based on 400,000 cores available		59	19	
total wall clock time on WCG(days) without overhead from the server		36	14	

Binding free energy predictions of whole SAMPL4 library

To perform quantitative comparisons for the retrospective predictions of the whole library from di erent types of BEDAM simulations, we calculated the receiver operating characteristic (ROC) curves as shown in Fig. 6 using the experimental crystallographic data available.[55] Note that there are 53 binders based on the available crystal structure data for the SAMPL4 library and they are expected to be weak binders since most of them have a nities more than 200 M via surface plasmon resonance (SPR) and only 8 ligands were provided binding a nities by the SAMPL4 Challenge organizers.[55] To aggregate the WCG simulations which include 451 complexes (301 unique ligands plus protonation and tautomeric states) and 9 poses for each complex, we have applied the following strategy: a) select the minimum value of binding free energies calculated from 9 BEDAM simulations started from 9 di erent poses for each complex; b) then select the minimum value from the simulations with di erent chemical extensions (such as the di erent protonation and tautormeric states) but for the same ligand as done for previous work.[53,76] The ROC curves calculated from the binding free energy (G_b^o) predictions are shown in Fig. 6a and corresponding values of the area under the curves (AUC) are listed in Table 5. It can be seen that the free energy scores from all BEDAM simulations (obtained with or without exchange) increase the AUC value in comparison with that from the AutoDock predictions. The AUC value from the BEDAM simulations with Independence sampling (represented by WCG-Indp) have the smallest increase (9%). The AUC values calculated from the other two sets of WCG simulations with the exchange feature have larger improvements of 20% (50,000 MD steps, denoted as WCG-50K) and 22% (10,000 MD steps, denoted as WCG-10K). The most signi cant improvement (50%) is from the BEDAM simulations with both the selected intramolecular torsional and nonbonded interactions attened (represented by Flat-All). We should emphasize that it is well-known that docking methods are better suitable for libraries with very diverse ligands instead of the focused libraries as is the SAMPL4 library in this report. Furthermore, the AutoDock predictions are the raw output and there are post processing procedures such as pharmacophore model ltering and

visual checking that can improve the docking predictions.[54 53,54] Our BEDAM binding free energy prediction provides an alternative approach to the re nement of docking results. In real applications, our BEDAM calculations should start from the best docking predictions after post analysis of ltering and prioritization using other fast models. In comparison with the AUC value from the original SAMPL4 BEDAM simulations using SyncRE framework on XSEDE,[53] both WCG AsyncRE simulations with exchange show only slight improvements (1% for WCG-50K and 2% for WCG-10K respectively) but have already achieved the same rank (2nd out of 25 submissions) in SAMPL4 Challenge. This result is mainly due to two facts: 1) The WCG AsyncRE simulations have an exchange frequency more than 100 times slower than the SyncRE simulations on XSEDE; 2) Although the AsyncRE simulations on the WCG have 45 times larger aggregated simulation time (9 poses and 5 times more lambda values), simulations from pose 6 to 9 (as we show below) have worse predictions than that of pose 1 to 5 and reduce the e ective aggregated simulation time to only 25 times larger. The corresponding ROC curves using the average binding energy ($G^o_{II} =< u >_{RL}$) predictions are illustrated in Fig. 6b. It is obvious that all AUC values using average binding energies from all BEDAM simulations are greatly reduced in comparison with that from the binding free energy predictions. Especially at the cuto of top 20% of inactives the distinguishing binders from nonbinders is close to a random selection. The inferior prediction from the average binding energies implies that besides the binding energy component the reorganization free energy component is also crucial and they are correlated in a way to enhance the discrimination between binders and nonbinders.

The AUC values focus on the performance of predictions for the whole library. Instead the enrichment factor measure as de ned in the Supporting Information is better suitable for evaluating the prediction power of early recognition from a library, which tells us how many more actives (binders in our analysis) can be found within an early recognition fraction (such as 20%) of the ordered library based on a given score relative to a random distribution. Table 5 also shows calculated enrichment factors at 20% (EF20) and 40% (EF40) cuto . Similar observations as the AUC values can be

found: All EFs using binding free energies (G_b^o) calculated from BEDAM simulations have more than 20% improvements in comparison with the raw AutoDock predictions; The WCG simulations with independence sampling have the smallest increases (21% for EF20 and 26% for EF40); The EF values for the two sets of WCG AsyncRE simulations show improvements of 20 30% for EF20 and 40 50% for EF40; Obviously the attening simulations have the largest increases (46% for EF20 and 53% for EF40). From Table 5 it is also clear that the EF values using average binding energies are worse than that from binding free energies indicating the importance of reorganization free energies.

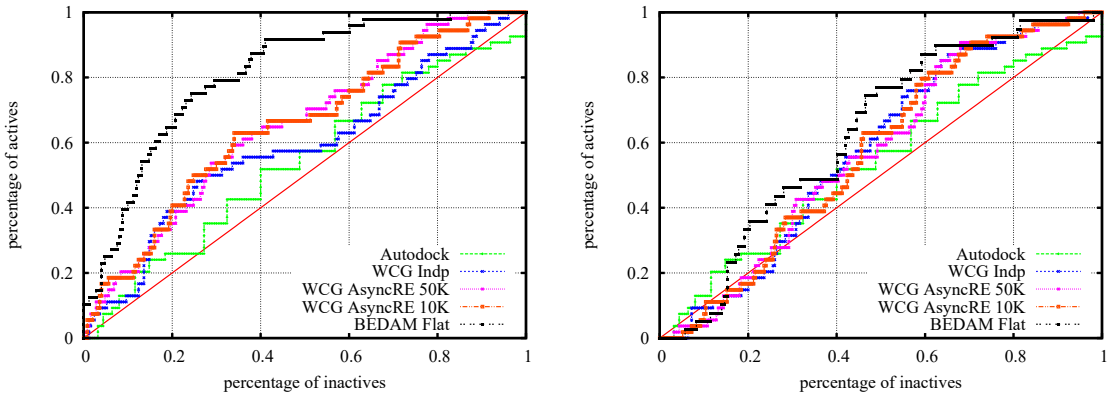

Figure 6: (a) ROC curves of di erent types of BEDAM-based screening for the SAMPL4 library using binding free energy (G_b^o) scores; (b) Same as in (a) but using average binding energy ($G_{II}^o =< u >_{RL}$) scores.

Binding free energy predictions from di erent docking poses

As with most free energy calculation methods, the BEDAM method relies on docked poses or crystal structures to start a simulation. The tremendous WCG resources allow us to re ne the docking predictions not only based on the best predicted pose but also alternative poses according to the docking scores. The WCG results shown in Fig. 6 and Table 5 were obtained by selecting the minimum (most favorable) binding free energy scores among the 9 scores calculated from individual BEDAM simulations started from each of 9 di erent poses of the same complex. In Fig. S2 of Supporting Information we display corresponding ROC curves of the binding free energies calculated from all trajectories of 9

Table 5: Area under curve (AUC) and enrichment factor (EF) values calculated from binding free energy (G_b^o) and average binding energy ($G_{II}^o =< u >_{RL}$) scores from di erent BEDAM simulations (1.2ns per replica) of SAMPL4 library.

		AutoDock	WCG-Indp	WCG-50K	WCG-10K	FLat-All
Pose min[c]	AUC(G_b^o)	0.541	0.593	0.650	0.658	0.812
	AUC($< u >_{RL}$)		0.577	0.572	0.573	0.529
	EF20[a] (G_b^o)	1.314	1.591	1.595	1.689	2.745
	EF40[b](G_b^o)	1.070	1.349	1.489	1.582	1.989
	EF20[a]($< u >_{RL}$)	1.314	0.843	0.938	0.845	1.417
	EF40[b]($< u >_{RL}$)	1.070	1.210	1.210	1.117	1.224
Pose avg[d]	AUC(G_b^o)	0.535	0.566	0.583	0.598	
	AUC($< u >_{RL}$)		0.622	0.581	0.598	

[a]EF20 = enrichment factors at 20% cuto ; [b]EF40= enrichment factors at 40% cuto ;
[c]aggregated to the minimum value from 9 di erent poses; [d]average value from 9 poses.

poses by simple averaging (denoted as pos avg). It is clear that such equal treatment for all 9 poses results in a worse prediction of AUC values as listed in Table 5, implying that many simulations are still trapped in a local equilibrium state instead of reaching a global one. Such local equilibrium properties become more obvious when we check the Spearman ranking order correlations between the results calculated from 1.2ns trajectories and that from 3.0ns as shown in Fig. S3 of Supporting Information. Both the WCG AsyncRE simulations using a MD period of 50K steps and the attening BEDAM simulations have rank order correlations larger than 0.95 (highly correlated) at two di erent simulation times. To optimize the computing resources for future simulations we are interested in the prediction performance from individual docking poses in addition to that aggregated from all 9 poses for the same complex. Fig. S4 exhibits the ROC curves of binding free energies calculated from the WCG AsyncRE BEDAM simulations using the MD period of 10K steps and started from each one of 9 di erent poses predicted by AutoDock (See Fig. S5 in Support Information for corresponding simulations using 50K MD period). The corresponding AUC values are plotted in Fig. 7 (see Table S1 for the values in detail). Generally the AUC values calculated from individual poses are smaller than that obtained by aggregating to the minimum value of binding free energies from all 9

poses of the same complex implying that aggregation by the minimum value does help to improve the accuracy of overall prediction. They are also correlated among the di erent WCG simulations (See Fig. S3 for the comparisons of individual binding free energies) which is consistent with the fact that di erent types of simulations were started from the same individual poses. Using a smaller MD period increases the AUC values, which is also in agreement with the di usion coe cient analysis shown in 4c. A more careful comparison also shows that the AUC values have large uctuations from pose 1 to 5 implying the best pose (top 1) predicted by AutoDock might not be the best candidate for BEDAM simulations (for example, WCG BEDAM simulations from the top 2 and top 5 poses have larger AUC values). We can also nd that the AUC values decrease to lower values from pose 6 to 9 which means that it is not necessary to include all 9 poses into the WCG AsyncRE simulations and top 1 to 5 poses from the AutoDock predictions are good enough for this SAMPL4 library. We should point out that this is only an observation based on Fig. 7 and for a more rigorous comparison we should perform the accumulative analysis of AUC values by aggregating the minimum values of binding free energies from top 1 to top 9 poses. The conclusion might di er depending on which ligand and also the parameter selection for the AutoDock predictions. In fact for more advanced BEDAM simulations using attening potentials, we can only use one pose but achieve a high AUC value as shown in Fig.6 and Table 5.

Conclusion

We implemented and optimized the AsyncRE framework of BEDAM method on the IBM WCG volunteer grid for massive-scale binding free energy calculations using the OPLS-AA force eld for describing receptor-ligand complexes and the AGBNP model for implicit solvent e ects. We discussed the general procedure on how to select the simulation parameters of BEDAM and re ne the AutoDock prediction. We also performed the rst massive-scale binding free energy calculations using distributed computing gird and AsyncRE framework with 2 451 9 = 8 118 complexes and 811 800 replicas submitted

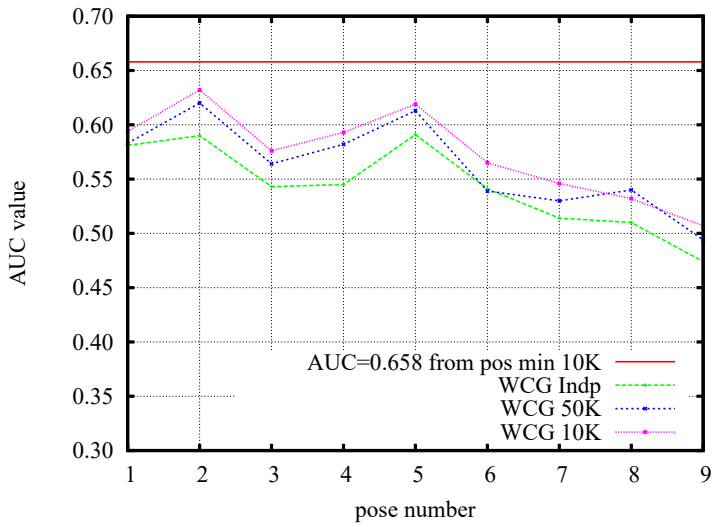

Figure 7: AUC values of ROC curves using binding free energy (G_b^o) scores calculated from di erent types of BEDAM simulations started from di erent individual poses (1 to 9) predicted by AutoDock for the SAMPL4 library . The red line corresponds the AUC value obtained by aggregating to the minimum (most favorable) binding free energy from all BEDAM simulations (WCG AsyncRE 10K) started from 9 di erent poses of the same complex. All other three lines shows the AUC values of binding free energies using individual poses only (1 to 9) without aggregation from di erent poses.

to the WCG volunteer gird at same time. Namely for the WCG AsyncRE simulations we extended the number of lambda states to 100 and included 9 poses from the AutoDock predictions for each HIV-1 IN complex to remediate the slower di usion in lambda space due to the use of much longer MD periods than the HPC SyncRE simulations to reduce the overhead. The free energy scores obtained from WCG AsyncRE simulations are comparable with those from the SyncRE simulations and show signi cant improvements over the initial AutoDock predictions although they are worse than those from more advanced REMD that includes the attening potentials to selected degrees of freedom of ligands. Running asynchronous REMD on WCG is a tradeo : We sacri ce rapid simultaneous RE exchanges (in depth) with little overhead in order to gain many more simulations (in breadth) on heterogeneous hardware but more overhead per replica.

The binding free energies from the attening AsyncRE BEDAM simulations show signi cant improvement on discriminating binders and nonbinders in comparison with the initial AutoDock predictions and BEDAM simulations without the attening feature. Using attening simulations can also reduce the number of poses for each simulated

complex and the requirements on computing resources since the attening potentials can help to accelerate sampling of internal degrees of freedom of the ligand/receptor and adjust the pose during a simulation. However there is also a tradeo between the selected number of degrees of freedom for attening and the simulation time to sample all of the relevant space similar to the REST methods in FEP calculations.[27,88] Namely improvement from the convergence of simulations in the relevant coordinate space might be weakened by including unnecessary internal coordinates into attening.[27,88] Hence, using attening BEDAM simulation is the best strategy to re ne the AutoDock results when we have prior knowledge about ligand libraries and/or particular receptors. For example we can perform cluster analysis of the binding pocket to identify related degrees of freedom of receptor exibility using crystallographic data, MD simulations, or induced- t docking studies. On the other hand we can consider the degrees of freedom from ligand exibility using our chemical knowledge and other studies such as exible docking that can specify the rotatable dihedral angles to explore the conformational exibility of ligands. The attening BEDAM simulations reported here were performed in a AsyncRE framework using a local campus grid on Temple University including 5000 CPUs. The implementation of attening BEDAM on the IBM WCG grid is underway.

The use of AsyncRE simulations without attening on the WCG Grid is a good choice when we do not have su cient prior knowledge concerning which degrees of freedom to apply attening potentials to or when the number of complexes is very large and an automatic work ow to select related internal degrees of freedom is not available. Considering the smaller overhead and insigni cant reduction to AUC values, choosing an MD period of 50K and 100 states to run on the WCG is a better choice than that of 10K. We also found that the AUC values from the BEDAM simulations started from individual poses decrease signi cantly at large number of pose (after 5). This means that we could perform simulations for the 5 top poses predicted by AutoDock which will reduce the requirement of computing resources close to 45%. We note that a factor of 4 attributable to overhead is from the waiting (queuing) times on the IBM BOINC server after jobs are submitted (this is included in the replica server time). The reason for this

is that the number of jobs submitted to the grid is more than three times the number of CPUs available (0.1 million) to our FightAIDS@Home Phase 2 project (though there are 3.0 million CPUs in the entire WCG grid). This part of the overhead can be removed when the available CPUs are increased to a desired number or the number of jobs is reduced using fewer poses or ligands. The real overhead (8.5 and 2.8 times for 10K and 50K respectively without over-submission) can also be reduced further by using a much smaller value of the maximum return time of MD jobs on the client side or/and increasing the frequency of checking returned MD jobs and performing exchanges on the server side. In fact, we were able to tweak the settings of client machines and the server in such a way that the overall is reduced to less than a factor of 2 on our local Temple University Grid. We should point out that although the largest benchmark test on the SAMPL4 library is 1000 lambda values per complex, the AsyncRE framework and computing resources support running hundreds of thousands of replicas per complex if it is required such as in multi-dimensional REMD simulations. For the WCG simulations of the whole SAMPL4 library, we submitted jobs including 8,118 complexes and 811,800 replicas at the same time. In principal, we can use all 811,800 replicas for the same complex. We did not run the simulation in this way because the gain from increasing the total number of replicas becomes less signi cant after its value is increased to 100 based on the di usion coe cient analysis.

In summary, combining the BEDAM method with the distributed computing network enables the massive-scale re nement directly on docking predictions and generates good candidates for further lead optimization using more high resolution FEP methods in explicit solvent. We expect that such protocols will become more routine in the near future as massive-scale distributed computing resources such as WCG gird, Amazon Cloud, and Internet of Things, become more widely available.

Supporting Information

More information about the system preparations, AsyncRE attening BEDAM, reweighting analysis, and other related topics are included. This material is available free of charge via the Internet pubs.acs.org.

Acknowledgments

This work has been supported by grants from the National Institutes of Health (GM30580 and U54 GM103368) and from the National Science Foundation (NSF 1665032). REMD simulations were carried out mainly on the IBM WCG grid with some performed on various resources: BOINC distributed networks at Temple University, SuperMIC, Stampede and Comet clusters of XSEDE resources (supported by TG-MCB100145). This research was also supported in part by local computing resources, the cb2rr HPC cluster (funded by the NIH instrumentation grant, S10-OD020095-01) and the owlsnest high performance cluster at Temple University (funded by the NSF research instrumentation grant, 1625061). We gratefully thank the IBM WCG team for porting our AsyncRE framework and IMPACT simulation package onto WCG and providing IT support for running the project. Also thank the WCG volunteers for their generous contribution of computing resources. Without the support from IBM and the WCG volunteers, this project would not have been possible. The authors also gratefully acknowledge support from the IT and HPC teams at Temple University.

References

(1) Gallicchio, E.; Levy, R. M. Recent Theoretical and Computational Advances for Modeling Protein-Ligand Binding A nities. *Adv. Prot. Chem. Struct. Biol.* **2011**, *85*, 27 80.

(2) Chodera, J. D.; Mobley, D. L.; Shirts, M. R.; Dixon, R. W.; Branson, K.;

Pande, V. S. Alchemical Free Energy Methods for Drug Discovery: Progress and Challenges. *Curr. Opin. Struct. Biol.* **2011**, *21*, 150 160.

(3) Mobley, D. L.; Dill, K. A. Binding of Small-Molecule Ligands to Proteins: What You See is not always What You Get . *Structure* **2009**, *17*, 489 498.

(4) Jorgensen, W. L. E cient Drug Lead Discovery and Optimization. *Acc. Chem. Res.* **2009**, *42*, 724 733.

(5) Gilson, M. K.; Zhou, H.-X. Calculation of Protein-Ligand Binding A nities. *Annu. Rev. Biophys. Biomol. Struct.* **2007**, *36*, 21 42.

(6) Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. The Statistical-Thermodynamic Basis for Computation of Binding A nities: A Critical Review. *Biophys. J.* **1997**, *72*, 1047 1069.

(7) Che, J.; Dzubiella, J.; Li, B.; McCammon, J. A. Electrostatic Free Energy and its Variations in Implicit Solvent Models. *J Phys Chem B* **2008**, *112*, 3058 3069.

(8) Gallicchio, E.; Lapelosa, M.; Levy, R. M. The Binding Energy Distribution Analysis Method (BEDAM) for the Estimation of Protein-Ligand Binding A nities. *J. Chem. Theory Comput.* **2010**, *6*, 2961 2977.

(9) Boehr, D. D.; Nussinov, R.; Wright, P. E. The Role of Dynamic Conformational Ensembles in Biomolecular Recognition. *Nat. Chem. Biol.* **2009**, *5*, 789 796.

(10) Zwier, M. C.; Chong, L. T. Reaching Biological Timescales with All-Atom Molecular Dynamics Simulations. *Curr. Opin. Pharmacol.* **2010**, *10*, 745 752.

(11) Shaw, D. E.; Maragakis, P.; Lindor -Larsen, K.; Piana, S.; Dror, R. O.; Eastwood, M. P.; Bank, J. A.; Jumper, J. M.; Salmon, J. K.; Shan, Y.; Wriggers, W. Atomic-Level Characterization of the Structural Dynamics of Proteins. *Science* **2010**, *330*, 341 346.

(12) Lane, T. J.; Shukla, D.; Beauchamp, K. A.; Pande, V. S. To Milliseconds and Beyond: Challenges in the Simulation of Protein Folding. *Curr. Opin. Struct. Biol.* **2013**, *23*, 58 65.

(13) Bowers, K. J.; Dror, R. O.; Shaw, D. E. Zonal Methods for the Parallel Execution of Range-Limited-N-Body Simulations. *J. Comput. Phys.* **2007**, *221*, 303 329.

(14) Bowers, K.; Chow, E.; Xu, H.; Dror, R.; Eastwood, M.; Gregersen, B.; Klepeis, J.; Kolossvary, I.; Moraes, M.; Sacerdoti, F.; Salmon, J.; Shan, Y.; Shaw, D. Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters. Proceedings of the ACM/IEEE Conference on Supercomputing (SC06). Tampa, Florida, 2006.

(15) Goetz, A. W.; Williamson, M. J.; Xu, D.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. *J. Chem. Theory. Comput.* **2012**, *8*, 1542 1555.

(16) Swendsen, R.; Wang, J.-S. Replica Monte Carlo Simulation of Spin-Glasses. *Phys. Rev. Lett.* **1986**, *57*, 2607 2609.

(17) Hansmann, U. H. Parallel Tempering Algorithm for Conformational Studies of Biological Molecules. *Chem. Phys. Lett.* **1997**, *281*, 140 150.

(18) Sugita, Y.; Okamoto, Y. Replica-Exchange Molecular Dynamics Method for Protein Folding. *Chem. Phys. Lett.* **1999**, *314*, 141 151.

(19) Sugita, Y.; Kitao, A.; Okamoto, Y. Multidimensional Replica-Exchange Method for Free-Energy Calculations. *J. Chem. Phys.* **2000**, *113*, 6042 6051.

(20) Kokubo, H.; Tanaka, T.; Okamoto, Y. Ab initio Prediction of Protein-Ligand Binding Structures by Replica-Exchange Umbrella Sampling Simulations. *J. Comput. Chem.* **2011**, *32*, 2810 2821.

(21) Kokubo, H.; Tanaka, T.; Okamoto, Y. Two-Dimensional Replica-Exchange Method for Predicting Protein-Ligand Binding Structures. *J. Comput. Chem.* **2013**, *34*, 2601 2614.

(22) Mitsutake, A.; Okamoto, Y. Replica-Exchange Extensions of Simulated Tempering Method. *J. Chem. Phys.* **2004**, *121*, 2491 2504.

(23) Laio, A.; Parrinello, M. Escaping Free-Energy Minima. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12562 12566.

(24) Fajer, M.; Hamelberg, D.; McCammon, J. A. Replica-Exchange Accelerated Molecular Dynamics (REXAMD) Applied to Thermodynamic Integration. *J Chem Theory Comput* **2008**, *4*, 1565 1569.

(25) Arrar, M.; de Oliveira, C. A. F.; Fajer, M.; Sinko, W.; McCammon, J. A. w-REXAMD: A Hamiltonian Replica Exchange Approach to Improve Free Energy Calculations for Systems with Kinetically Trapped Conformations. *J. Chem. Theory Comput.* **2013**, *9*, 18 23.

(26) Liu, P.; Kim, B.; Friesner, R. A.; Berne, B. J. Replica Exchange with Solute Tempering: A Method for Sampling Biological Systems in Explicit Water. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 13749 13754.

(27) Wang, L.; Friesner, R. A.; Berne, B. J. Replica Exchange with Solute Scaling: A More E cient Version of Replica Exchange with Solute Tempering (REST2). *J. Phys. Chem. B* **2011**, *115*, 9431 9438.

(28) A entranger, R.; Tavernelli, I.; Iorio, E. A Novel Hamiltonian Replica Exchange MD Protocol to Enhance Protein Conformational Space Sampling. *J. Chem. Theory Comput.* **2006**, *2*, 217 228.

(29) Liu, P.; Voth, G. A. Smart Resolution Replica Exchange: An E cient Algorithm for Exploring Complex Energy Landscapes. *J. Chem. Phys.* **2007**, *126*, 045106 045111.

(30) Zheng, L.; Chen, M.; Yang, W. Random Walk in Orthogonal Space to Achieve E cient Free-Energy Simulation of Complex Systems. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 20227 20232.

(31) Zheng, L. Q.; Chen, M. G.; Yang, W. Simultaneous Escaping of Explicit and Hidden Free Energy Barriers: Application of the Orthogonal Space Random Walk Strategy in Generalized Ensemble Based Conformational Sampling. *J. Chem. Phys.* **2009**, *130*, 234105 23414.

(32) Sugita, Y.; Okamoto, Y. Replica-Exchange Multicanonical Algorithm and Multicanonical Replica-Exchange Method for Simulating Systems with Rough Energy Landscape. *Chem. Phys. Lett.* **2000**, *329*, 261 270.

(33) Mitsutake, A.; Sugita, Y.; Okamoto, Y. Generalized-Ensemble Algorithms for Molecular Simulations of Biopolymers. *Biopolymers* **2001**, *60*, 96 123.

(34) 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.

(35) 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*, 2559 2565.

(36) Jiang, W.; Luo, Y.; Maragliano, L.; Roux, B. Calculation of Free Energy Landscape in Multi-Dimensions with Hamiltonian-Exchange Umbrella Sampling on Petascale Supercomputer. *J. Chem. Theory Comput.* **2012**, *8*, 4672 4680.

(37) Marinari, E.; Parisi, G. Simulated Tempering: A New Monte Carlo Scheme. *Europhys. Lett.* **1992**, *19*, 451 458.

(38) Geyer, C. J.; Thompson, E. A. Annealing Markov Chain Monte Carlo with Applications to Ancestral Inference. *J. Am. Stat. Assoc.* **1995**, *90*, 909 920.

(39) Li, H.; Fajer, M.; Yang, W. 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.* **2007**, *126*, 024106.

(40) Hagen, M.; Kim, B.; Liu, P.; Friesner, R. A.; Berne, B. J. Serial Replica Exchange. *J. Phys. Chem. B* **2007**, *111*, 1416 1423.

(41) Rauscher, S.; Neale, C.; Pomes, R. Simulated Tempering Distributed Replica Sampling, Virtual Replica Exchange, and Other Generalized-Ensemble Methods for Comformational Sampling. *J. Chem. Theory Comput.* **2009**, *10*, 2640 2662.

(42) Rodinger, T.; Howell, P. L.; Pomes, R. Distributed Replica Sampling. *J. Chem. Theory Comput.* **2006**, *2*, 725 731.

(43) Xia, J.; Flynn, W. F.; Gallicchio, E.; Zhang, B. W.; He, P.; Tan, Z.; Levy, R. M. Large-Scale Asynchronous and Distributed Multidimensional Replica Exchange Molecular Simulations and E ciency Analysis. *J. Comput. Chem.* **2015**, *36*, 1772 1785.

(44) Gallicchio, E.; Xia, J.; Flynn, W. F.; Zhang, B. W.; Samlalsingh, S.; Mentes, A.; Levy, R. M. Asynchronous Replica Exchange Software for Grid and Heterogeneous Computing. *Comput. Phys. Commun.* **2015**, *196*, 236 246.

(45) Zhou, H.-X.; Gilson, M. K. Theory of Free Energy and Entropy in Noncovalent Binding. *Chem. Rev.* **2009**, *109*, 4092 4107.

(46) Woo, H.-J.; Roux, B. Calculation of Absolute Protein-Ligand Binding Free Energy from Computer Simulations. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 6825 6830.

(47) Mobley, D. L.; Chodera, J. D.; Dill, K. A. On the Use of Orientational Restraints and Symmetry Corrections in Alchemical Free Energy Calculations. *J. Chem. Phys.* **2006**, *125*, 084902 084917.

(48) Banks, J. L.; Beard, J. S.; Cao, Y.; Cho, A. E.; Damm, W.; Farid, R.; Felts, A. K.; and Halgren, T. A.; Mainz, D. T.; Maple, J. R.; Murphy, R.; Philipp, D. M.; Repasky, M. P.; Zhang, L. Y.; Berne, B. J.; Friesner, R. A.; Gallicchio, E.; Levy, R. M. Integrated Modeling Program, Applied Chemical Theory (IMPACT). *J. Comp. Chem.* **2005**, *26*, 1752 1780.

(49) Fukunishi, H.; Watanabe, O.; Takada, S. On the Hamiltonian Replica Exchange Method for E cient Sampling of Biomolecular Systems: Application to Protein Structure Prediction. *J. Chem. Phys.* **2002**, *116*, 9058 9067.

(50) Lapelosa, M.; Gallicchio, E.; Levy, R. M. Conformational Transitions and Convergence of Absolute Binding Free Energy Calculations. *J. Chem. Theory Comput.* **2012**, *8*, 47 60.

(51) Gallicchio, E.; Levy, R. M. Prediction of SAMPL3 Host-Guest A nities with the Binding Energy Distribution Analysis Method (BEDAM). *J. Comput. Aided Mol. Des.* **2012**, *26*, 505 516.

(52) Wickstrom, L.; He, P.; Gallicchio, E.; Levy, R. M. Large Scale A nity Calculations of Cyclodextrin Host-Guest Complexes: Understanding the Role of Reorganization in the Molecular Recognition Process. *J. Chem. Theory Comput.* **2013**, *9*, 3136 3150.

(53) Gallicchio, E.; Deng, N.; He, P.; Perryman, A. L.; Santiago, D. N.; Forli, S.; Olson, A. J.; Levy, R. M. Virtual Screening of Integrase Inhibitors by Large Scale Binding Free Energy Calculations: the SAMPL4 Challenge. *J. Comp. Aided Mol. Des.* **2014**, *28*, 475 490.

(54) Perryman, A. L.; Santiago, D. N.; Forli, S.; Santos-Martins, D.; Olson, A. J. Virtual Screening with AutoDock Vina and the Common Pharmacophore Engine of a Low Diversity Library of Fragments and Hits against the Three Allosteric Sites of HIV Integrase: Participation in the SAMPL4 Protein Ligand Binding Challenge. *J. Comp. Aided Mol. Des.* **2014**, *28*, 1 13.

(55) Mobley, D. L.; Liu, S.; Lim, N. M.; Wymer, K. L.; Perryman, A. L.; Forli, S.; Deng, N.; Su, J.; Branson, K.; Olson, A. J. Blind Prediction of HIV Integrase Binding from the SAMPL4 Challenge. *J. Comp. Aided Mol. Des.* **2014**, *28*, 327 345.

(56) Lawrenz, M.; Shukla, D.; Pande, V. S. Cloud Computing Approaches for Prediction of Ligand Binding Poses and Pathways. *Scientific Reports* **2015**, *5*, 7918 7922.

(57) Kohlho , K. J.; Shukla, D.; Lawrenz, M.; Bowman, G. R.; Konerding, D. E.; Belov, D.; Altman, R. B.; Pande, V. S. Cloud-Based Simulations on Google Exacycle Reveal Ligand Modulation of GPCR Activation Pathways. *Nat. Chem.* **2014**, *6*, 15 21.

(58) Rodinger, T.; Howell, P. L.; Pomes, R. Calculation of Absolute Protein-Ligand Binding Free Energy Using Distributed Replica Sampling. *J. Chem. Phys.* **2008**, *129*, 155102 155113.

(59) Jayachandran, G.; Shirts, M. R.; Park, S.; Pande, V. S. Parallelized-Over-Parts Computation of Absolute Binding Free Energy with Docking and Molecular Dynamics. *J. Chem. Phys.* **2006**, *125*, 084901.

(60) Huang, X.; Bowman, G. R.; Pande, V. S. Convergence of Folding Free Energy Landscapes via Application of Enhanced Sampling Methods in a Distributed Computing Environment. *J. Chem. Phys.* **2008**, *128*, 205106.

(61) Rhee, Y. M.; Pande, V. S. Multiplex-Replica Exchange Molecular Dynamics Method for Protein Folding Simulation. *Biophys. J.* **2003**, *84*, 775 786.

(62) Shen, H.; Czaplewski, C.; Liwo, A.; Scheraga, H. A. Implementation of a Serial Replica Exchange Method in a Physics-Based United-Residue (UNRES) Force Field. *J. Chem. Theory Comput.* **2008**, *4*, 1386 1400.

(63) Lockhart, C.; O Connor, J.; Armentrout, S.; Klimov, D. K. Greedy Replica Exchange Algorithm for Heterogeneous Computing Grids. *J. Mol. Model* **2015**, *21*, 1 12.

(64) Forli, S.; J., O. A. Computational Challenges of Structure-Based Approaches Applied to HIV. *Curr. Top. Microbiol. Immunol.* **2015**, *389*, 31 51.

(65) Engelman, A.; Kessl, J. J.; Kvaratskhelia, M. Allosteric Inhibition of HIV-1 Integrase Activity. *Curr. Op. Chem. Biol.* **2013**, *17*, 339 345.

(66) Widom, B. Potential-Distribution Theory and the Statistical Mechanics of Fluids. *J. Phys. Chem.* **1982**, *86*, 869 872.

(67) Gallicchio, E.; Levy, R. AGBNP: An Analytic Implicit Solvent Model Suitable for Molecular Dynamics Simulations and High-Resolution Modeling. *J. Comput. Chem.* **2004**, *25*, 479 499.

(68) Ravindranathan, K. P.; Gallicchio, E.; Levy, R. M. Conformational Equilibria and Free Energy Pro les for the Allosteric Transition of the Ribose-binding Protein. *J. of Mol. Biol.* **2005**, *353*, 196 210.

(69) Gallicchio, E.; Paris, K.; Levy, R. M. The AGBNP2 Implicit Solvation Model. *J. Chem. Theory Comput.* **2009**, *5*, 2544 2564.

(70) Tan, Z.; Gallicchio, E.; Lapelosa, M.; Levy, R. M. Theory of Binless Multi-State Free Energy Estimation with Applications to Protein-Ligand Binding. *J. Chem. Phys.* **2012**, *136*, 144102.

(71) Shirts, M. R.; Chodera, J. D. Statistically Optimal Analysis of Samples from Multiple Equilibrium States. *J. Chem. Phys.* **2008**, *129*, 124105.

(72) Zhang, B. W.; Xia, J.; Tan, Z.; Levy, R. M. A Stochastic Solution to the Unbinned WHAM Equations. *J. Phys. Chem. Lett.* **2015**, *6*, 3834 3840.

(73) Tan, Z.; Xia, J.; Zhang, B. W.; Levy, R. M. Locally Weighted Histogram Analysis and Stochastic Solution for Large-Scale Multi-State Free Energy Estimation. *J. Chem. Phys.* **2016**, *144*, 034107 034123.

(74) Chodera, J. D.; Shirts, M. R. Replica Exchange and Expanded Ensemble Simulations as Gibbs Sampling: Simple Improvements for Enhanced Mixing. *J. Chem. Phys.* **2011**, *135*, 194110.

(75) Zhang, B. W.; Dai, W.; Gallicchio, E.; He, P.; Xia, J.; Tan, Z.; Levy, R. M. Simulating Replica Exchange: Markov State Models, Proposal Schemes, and the In nite Swapping Limit Replica Exchange Simulations of Binding Free Energies: Markov State Models, Proposal Schemes, and Reweighting Techniques. *J. Phys. Chem. B* **2016**, *120*, 8289 8301.

(76) Xia, J.; Flynn, W.; Levy, R. M. Improving Prediction Accuracy of Binding Free Energies and Poses of HIV Integrase Complexes Using the Binding Energy Distribution Analysis Method with Flattening Potentials. *J. Chem. Info. Model.* **2018**, *58*, 1356 1371.

(77) Mentes, A.; Deng, N. J.; Vijayan, R. S. K.; Xia, J.; Gallicchio, E.; Levy, R. M. Binding Energy Distribution Analysis Method: Hamiltonian Replica Exchange with Torsional Flattening for Binding Mode Prediction and Binding Free Energy Estimation. *J. Chem. Theory. Comput.* **2016**, *5*, 2459 2470.

(78) Rosta, E.; Hummer, G. Error and E ciency of Replica Exchange Molecular Dynamics Simulations. *J. Chem. Phys.* **2009**, *131*, 165102.

(79) Rosta, E.; Hummer, G. Error and E ciency of Simulated Tempering Simulations. *J. Chem. Phys.* **2010**, *132*, 034102.

(80) Nymeyer, H. How E cient Is Replica Exchange Molecular Dynamics? An Analytic Approach. *J. Chem. Theory Comput.* **2008**, *4*, 626 636.

(81) Zhang, X.; Bhatt, D.; Zuckerman, D. M. Automated Sampling Assessment for Molecular Simulations Using the E ective Sample Size. *J. Chem. Theory Comput.* **2010**, *6*, 3048 3057.

(82) Sindhikara, D.; Meng, Y.; Roitberg, A. E. Exchange Frequency in Replica Exchange Molecular Dynamics. *J. Chem. Phys.* **2008**, *128*, 024103.

(83) Sindhikara, D. J.; Emerson, D. J.; Roitberg, A. E. Exchange Often and Properly in Replica Exchange Molecular Dynamics. *J. Chem. Theory Comput.* **2010**, *6*, 2804 2808.

(84) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Oxford University Press: New York, 1993.

(85) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Developement and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225 11236.

(86) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. Evaluation and Reparameterization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides. *J. Phys. Chem. B* **2001**, *105*, 6474 6487.

(87) Jurado, K. A.; Wang, H.; Slaughter, A.; Feng, L.; Kessl, J. J.; Koh, Y.; Wang, W.; Ballandras-Colas, A.; Patel, P. A.; Fuchs, J. R.; Kvaratskhelia, M.; Engelman, A. Allosteric Integrase Inhibitor Potency is Determined through the Inhibition of HIV-1 Particle Maturation. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 8690 8695.

(88) Wang, L.; Berne, B. J.; Friesner, R. A. On Achieving High Accuracy and Reliability in the Calculation of Relative Protein-Ligand Binding A nities. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 1937 1942.

(89) Huang, X.; Hagen, M.; Kim, B.; Friesner, R. A.; Zhou, R.; Berne, B. J. Replica

Exchange with Solute Tempering: E ciency in Large scale systems. *J. Phys. Chem. B* **2007**, *111*, 5405 5410.

(90) Lyman, E.; Ytreberg, F. M.; Zuckerman, D. M. Resolution Exchange Simulation. *Phys. Rev. Lett.* **2006**, *96*, 4.

(91) Bussi, G.; Gervasio, F. L.; Laio, A.; Parrinello, M. Free-Energy Landscape for Hairpin Folding from Combined Parallel Tempering and Metadynamics. *J. Am. Chem. Soc.* **2006**, *128*, 13435 13441.

(92) Kannan, S.; Zacharias, 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.

(93) Roitberg, A. E.; Okur, A.; Simmerling, C. Coupling of Replica Exchange Simulations to a Non-Boltzmann Structure Reservoir. *J Phys Chem B* **2007**, *111*, 2415 2418.

(94) Okur, A.; Roe, D. R.; Cui, G. L.; Hornak, V.; Simmerling, C. Improving Convergence of Replica-Exchange Simulations through Coupling to a High-Temperature Structure Reservoir. *J. Chem. Theory Comput.* **2007**, *3*, 557 568.

(95) Li, H. Z.; Li, G. H.; Berg, B. A.; Yang, W. Finite Reservoir Replica Exchange to Enhance Canonical Sampling in Rugged Energy Surfaces. *J. Chem. Phys.* **2006**, *125*, 5.

(96) Min, D.; Chen, M.; Zheng, L.; Jin, Y.; Schwartz, M. A.; Sang, Q.-X. A.; Yang, W. Enhancing QM/MM Molecular Dynamics Sampling in Explicit Environments via an Orthogonal-Space-Random-Walk-Based Strategy. *J. Phys. Chem. B* **2011**, *115*, 3924 3935.

(97) Hritz, J.; Oostenbrink, C. Optimization of Replica Exchange Molecular Dynamics by Fast Mimicking. *J. Chem. Phys.* **2007**, *127*, 204104.

(98) Trebst, S.; Troyer, M.; Hansmann, U. H. E. Optimized Parallel Tempering Simulations of Proteins. *J. Chem. Phys.* **2006**, *124*, 174903.

(99) Li, X.; O Brien, C. P.; Collier, G.; Vellore, N. A.; Wang, F.; Latour, R. A.; Bruce, D. A.; Stuart, S. J. An Improved Replica-Exchange Sampling Method: Temperature Intervals with Global Energy Reassignment. *J. Chem. Phys.* **2007**, *127*, 164116.

(100) Zhang, W.; Chen, J. E ciency of Adaptive Temperature-Based Replica Exchange for Sampling Large-Scale Protein Conformational Transitions. *J. Chem. Theory Comput.* **2013**, *9*, 2849 2856.

(101) Zheng, W.; Andrec, M.; Gallicchio, E.; Levy, R. M. Simulating Replica Exchange Simulations of Protein Folding with a Kinetic Network Model. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 15340 15345.

(102) Plattner, N.; Doll, J. D.; Dupuis, P.; Wang, H.; Liu, Y.; Gubernatis, J. E. An In nite Swapping Approach to the Rare-Event Sampling Problem. *J. Chem. Phys.* **2011**, *135*, 134111.

(103) Lu, J.; Vanden-Eijnden, E. In nite Swapping Replica Exchange Molecular Dynamics Leads to a Simple Simulation Patch Using Mixture Potentials. *J. Chem. Phys.* **2013**, *138*, 084105.

(104) Zhang, W.; Wu, C.; Duan, Y. Convergence of Replica Exchange Molecular Dynamics. *J. Chem. Phys.* **2005**, *123*, 154105.

(105) Periole, X.; Mark, A. E. Convergence and Sampling E ciency in Replica Exchange Simulations of Peptide Folding in Explicit Solvent. *J. Chem. Phys.* **2007**, *126*, 014903.

(106) Abraham, M. J.; Gready, J. E. Ensuring Mixing E ciency of Replica-Exchange Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2008**, *4*, 1119 1128.

(107) Denschlag, R.; Lingenheil, M.; Tavan, P. E ciency Reduction and Pseudo-Convergence in Replica Exchange Sampling of Peptide Folding-Unfolding Equilibria. *Chem. Phys. Lett.* **2008**, *458*, 244 248.

(108) Zuckerman, D. M.; Lyman, E. A Second Look at Canonical Sampling of Biomolecules Using Replica Exchange Simulation. *J. Chem. Theory Comput.* **2006**, *2*, 1200 1202.

(109) Zheng, W.; Andrec, M.; Gallicchio, E.; Levy, R. M. Simple Continuous and Discrete Models for Simulating Replica Exchange Simulations of Protein Folding. *J. Phys. Chem. B* **2008**, *112*, 6083 6093.

(110) Plattner, N.; Doll, J. D.; Meuwly, M. Overcoming the Rare-Event Sampling Problem in Biological Systems with In nite Swapping. *J. Chem. Theory Comput.* **2013**, *9*, 4215 4224.

for Table of Contents use only