JCTC Journal of Chemical Theory and Computation

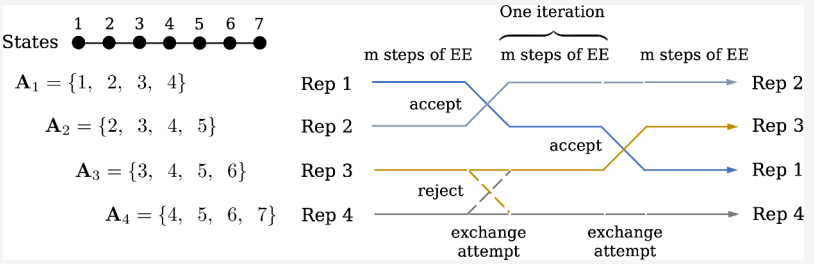
pubs.acs.org/JCTC Article

Replica Exchange of Expanded Ensembles: A Generalized Ensemble Approach with Enhanced Flexibility and Parallelizability

Wei-Tse Hsu and Michael R. Shirts*

Cite This: J. Chem. Theory Comput. 2024, 20, 6062−6081 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information



ABSTRACT: Generalized ensemble methods such as Hamiltonian replica exchange (HREX) and expanded ensemble (EE) have been shown effective in free energy calculations for various contexts, given their ability to circumvent free energy barriers via nonphysical pathways defined by states with different modified Hamiltonians. However, both HREX and EE methods come with drawbacks, such as limited flexibility in parameter specification or the lack of parallelizability for more complicated applications. To address this challenge, we present the method of replica exchange of expanded ensembles (REXEE), which integrates the principles of HREX and EE methods by periodically exchanging coordinates of EE replicas sampling different yet overlapping sets of alchemical states. With the solvation free energy calculation of anthracene and binding free energy calculation of the CB7−10 binding complex, we show that the REXEE method achieves the same level of accuracy in free energy calculations as the HREX and EE methods, while offering enhanced flexibility and parallelizability. Additionally, we examined REXEE simulations with various setups to understand how different exchange frequencies and replica configurations influence the sampling efficiency in the fixed-weight phase and the weight convergence in the weight-updating phase. The REXEE approach can be further extended to support asynchronous parallelization schemes, allowing looser communications between larger numbers of loosely coupled processors such as cloud computing and therefore promising much more scalable and adaptive executions of alchemical free energy calculations. All algorithms for the REXEE method are available in the Python package ensemble_md, which offers an interface for REXEE simulation management without modifying the source code in GROMACS.

INTRODUCTION

Molecular dynamics (MD) has established its significance and versatility in a broad spectrum of scientific disciplines. With sufficiently accurate force fields, it can theoretically serve as a virtual microscope to investigate a plethora of dynamics of interest at an atomistic resolution. However, even with the latest-generation hardware,[1] conventional MD simulations are generally limited to only probing processes with relatively short time scales, leaving real-world challenges such as folding of typical proteins or large-scale conformational transitions still out of reach for direct simulation. As such, the last decades have witnessed the emergence of various advanced/enhanced sampling methods dedicated to addressing this time scale issue.[2]

These methods can be roughly divided into two broad categories based on different strategies for sampling metastable states of interest that are separated by free energy barriers insurmountable by direct sampling with viable computational cost. Methods in the first category work with the phase space purely defined by the configurational degrees of freedom of the system. Frequently, these methods bias the system along a small number of degrees of freedom (the so-called collective variables, or CVs) to encourage diffusive sampling in the configurational space. Representative examples in this category include umbrella sampling,[3] metadynamics,[4] adaptive biasing force,[5] on-the-fly probability enhanced sampling,[6] and their variations.[7−9]

Generalized ensemble methods represent the other category that does not rely on the use of collective variables,[2] but expands the phase space by introducing additional dimensions continuously defined or discretized by intermediate states with

Received: April 12, 2024
Revised: June 19, 2024
Accepted: June 20, 2024
Published: July 15, 2024

https://doi.org/10.1021/acs.jctc.4c00484
J. Chem. Theory Comput. 2024, 20, 6062−6081

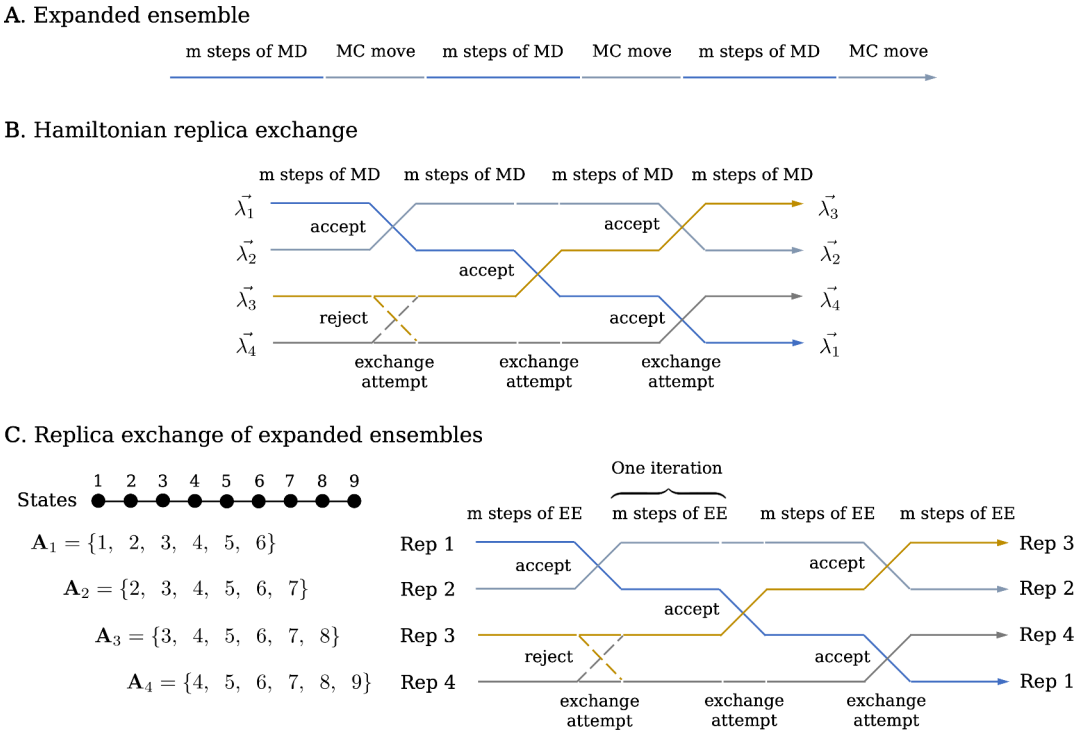


Figure 1. Schematic representations of different generalized ensemble methods. (A) In **expanded ensembles**, MD simulations and MC moves alternate periodically to sample the configurational space and the alchemical space, respectively. (B) In **Hamiltonian replica exchange**, the coordinates of replicas of MD simulations sampling different alchemical intermediate states are periodically exchanged to enhance the sampling in the alchemical space. Each λ vector is a vector of coupling parameters of interest, such as for coupling/decoupling electrostatic interactions, van der Waals interactions, or distance restraints. (C) In the **replica exchange of expanded ensembles** (REXEE) method, the coordinates of replicas of EE simulations are periodically exchanged. $\mathbf{A}_1$, $\mathbf{A}_2$, $\mathbf{A}_3$, and $\mathbf{A}_4$ denote the sets of states different replicas can sample during the simulation.

different temperatures, alchemically modified Hamiltonians, and/or other auxiliary variables. These methods, which can often be expressed as a form of Gibbs sampling,[10,11] alternate the sampling direction between the configurational space and the extended space. The sampling in the configurational space is achieved by molecular dynamics (or occasionally Monte Carlo (MC)), while the sampling between different intermediate states in the extended space is usually done by MC moves (panels A and B in Figure 1). These moves in the state space use approaches such as the Metropolis-Hastings algorithm,[12] the Barker transition algorithm,[13] Gibbs sampling,[11,14] and Metropolized Gibbs sampling.[10,15] Importantly, in the case where the metastability of a system changes along the extended state space, sampling the state space in these additional dimensions allows one to move between different metastable states that exist in the configurational space. For example, by traversing the temperature space either serially with a single simulation (e.g., simulated tempering[16]) or with multiple simulations running in parallel (e.g., temperature replica exchange (TREX), also known as parallel tempering[17]), one can observe the unfolded state of a protein that might be unattainable at a lower temperature.[18,19] Similarly, alchemical free energy methods exploit the fact that configurational free energy barriers might be lower or even absent in states at intermediate values of the alchemical coupling parameter λ. With comprehensive sampling in the alchemical space, either serially (e.g., simulated scaling,[20] expanded ensemble (EE),[21] and λ dynamics[22,23]) or in parallel (e.g., Hamiltonian replica exchange (HREX)[24]), the system can circumvent free energy barriers via nonphysical pathways bridging the coupled and decoupled states, thus allowing the computation of various free energy differences, including solvation free energies,[25−27] binding free energies,[28−30] and mutation free energies.[31,32] Importantly, generalized ensemble methods are not confined to solely augmenting the sampling space with the temperature or alchemical space, but have the capacity to define intermediate states and carry out coordinate exchanges in multidimensional grids defined by varying both temperatures and Hamiltonians,[24,33−35] or either of these with other auxiliary variables.[36]

To aid the sampling in the extended dimension, especially in methods utilizing alchemical intermediate states, varying forms of weights are usually employed. For instance, expanded ensembles can work with methods such as the Wang−Landau algorithm,[37] its $1/t$ variation,[38,39] accelerated weighted histogram (AWH),[40−42] and self-adjusted mixture sampling (SAMS)[43] to iteratively estimate alchemical weights that aim to level out the alchemical free energy profile. In the limiting case where the set of weights gives each alchemical state exactly equal probability, the alchemical weights are equal to the dimensionless free energies. As another example, in the recently proposed alchemical metadynamics[44] where the alchemical variable is treated as a collective variable, Gaussian biasing potentials deposited in the alchemical direction serve as another form of alchemical weights. Hamiltonian replica exchange, while not imposing alchemical weights explicitly, periodically exchanges the coordinates between replicas, which can be regarded as a form of implicit weights enforcing even sampling of each state/replica.

In expanded ensemble simulations, it is common to adopt a two-stage protocol composed of a weight-updating stage followed by a fixed-weight/production stage. Specifically, a

https://doi.org/10.1021/acs.jctc.4c00484
J. Chem. Theory Comput. 2024, 20, 6062−6081

weight-updating algorithm, such as the Wang−Landau algorithm or the similar ones mentioned above, is used in the first stage to iteratively adjust the alchemical weights until the algorithm is converged, e.g., certain criteria on the flatness of visitation rates at each state are reached. Subsequently, the second stage fixes the weights at the values converged by the first stage. Notably, while data generated in the first stage should theoretically reach quasi-equilibrium given sufficient sampling,[43] it is of common practice to only use samples generated in the second stage for free energy estimators that require equilibrium data, such as thermodynamic integration (TI),[45] Bennett acceptance ratio (BAR),[46] and Multistate Bennett acceptance ratio (MBAR).[47] This notion of calculating free energies from equilibrium data also applies to alchemical metadynamics, as discussed in the work[44] by Hsu et al.

However, methods such as Hamiltonian replica exchange and expanded ensemble come with limitations. Each simulation within a generalized ensemble must overlap with at least some other states, or no switches can occur. If one is sampling large changes in the phase space, then large numbers of connecting states will be necessary, potentially leading to diffusion-limited end-to-end traversal time[48] that increases with the square of the number of states in a given direction.

The selection of the number of states can be potentially problematic for HREX in particular, as running the simulation often needs to take into account the available computational resources, such as the number of CPU cores per node or the total number of nodes, as the simulations must be highly parallelized with relatively low latency to function well. To fully leverage the computational power, the number of replicas should ideally be a factor of the number of available cores. This requirement can become restrictive in deciding the number of states in HREX simulations, especially for complex systems that might necessitate a larger number of states to ensure sufficient overlap between adjacent states. For example, given the one-to-one correspondence between replicas and states, an HREX simulation sampling 64 alchemical states would require 64 replicas. In a situation where a 128-core node is available, but 64 states fail to provide the desired overlap, the next logical increment would be 128 states, which could be an unnecessarily large leap. This problem can be exacerbated when the system requires hundreds of alchemical intermediate states to ensure adequate overlap between neighboring states, or when the state space is multidimensional. These scenarios can easily require extensive communication between hundreds of cores that current parallelization schemes do not fully support, thereby hindering the exploration of even more complex systems.

Scenarios that require a large number of states can also pose a challenge to expanded ensemble approaches. Since the sampling along the auxiliary variables is not parallelized, the simulation wall time can therefore be significantly longer than that of the HREX method for equal amounts of sampling at each state. When weights are poorly converged, as can often happen with adaptive methods when there are slow degrees of freedom orthogonal to the generalized coordinate,[38,39,44] the round-trip time in a one-dimensional extended variable space is significantly increased, with potentially some regions of the auxiliary parameter space not visited at all. In replica exchange, one may have a lack of round-trip visits along the generalized variable due to insufficient exchanges, but there is at least sampling at all states. For example, it was reported in refs.[49] and[50] that the Wang−Landau algorithm faced difficulties converging the alchemical weights for several host−guest binding complexes in SAMPL4[51] and SAMPL6[50] SAMPLing challenges.

To alleviate some of the aforementioned issues, we propose the method of **replica exchange of expanded ensembles (REXEE)**, which integrates the core principles of the replica exchange (REX) and expanded ensemble (EE) methods. Specifically, a REXEE simulation runs multiple replicas of EE simulations in parallel and periodically exchanges coordinates between replicas. Each EE replica samples a different but overlapping set of alchemical intermediate states to collectively sample the space between the coupled and decoupled state (see Figure 1C). By design, the REXEE method decouples the number of replicas from the number of states, allowing sampling a large number of intermediate states with significantly fewer replicas than those required in the HREX method and other similar methods. By parallelizing replicas, the REXEE method can also reduce the simulation wall time compared to the EE method, and guarantees that there are always simulations, if not at every replica, at least in all ranges of replicas sets along the auxiliary variables. Importantly, such parallelism also sets the stage for wider applications, such as relative free energy calculations for multi-topology transformations.

Notably, REXEE is not the first method to parallelize enhanced sampling simulations. For example, parallel-tempered metadynamics (PTMetaD)[52] combines parallel tempering[17] with metadynamics,[53,54] allowing efficient sampling in both configurational and temperature spaces. Similarly, replica-exchange on-the-fly parametrization (RE-OTFP)[55] parallelizes temperature-accelerated molecular dynamics (TAMD)[56] simulations, with the additional advantage of a lack of history-dependent bias. The REXEE method, while sharing the goal of enhancing phase space sampling via exchanges between simulations, distinguishes itself by parallelizing the same state space as replica exchange or expanded ensemble across arbitrary numbers of processors.

We also note that there exist replica exchange methods that are well-suited for sampling large numbers of states in a generalized ensemble, with most of them supporting asynchronous parallelization or heterogeneous computational grids. For example, asynchronous replica exchange molecular dynamics[57−60] decentralizes the management of replica simulations by allowing exchanges to be attempted between any two idling replicas. Greedy replica exchange molecular dynamics[61] addressed the synchronization bottleneck in the same spirit, but uses a precomputed schedule of exchange attempts to sample each state equally. Multiplex replica exchange molecular dynamics[62] increases the pool of replicas available for exchange attempts, similarly eliminating the need for replica synchronization. However, we emphasize that these methods are distinctive from the REXEE method as they focus on software implementation design different from the standard replica exchange method. The REXEE method, on the other hand, proposes a new formulation that breaks the one-to-one correspondence between states and replicas, allowing each replica to sample an arbitrary number of states.

Upon the development of the REXEE method, this study sets out to accomplish three objectives. First, we seek to demonstrate that with samples from the fixed-weight stage, the REXEE method can produce free energy estimates on par with those from EE or HREX simulations, while offering greater flexibility in parameter specification and replica configuration.

https://doi.org/10.1021/acs.jctc.4c00484
J. Chem. Theory Comput. 2024, 20, 6062−6081

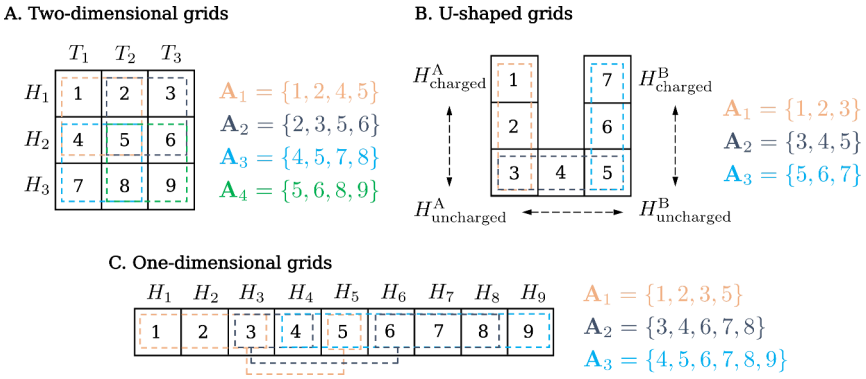


Figure 2. Different possible replica configurations of a REXEE simulation, with each state represented as a grid labeled by the number in its center and characterized by different Hamiltonians and/or temperatures. Different state sets are represented as dashed lines in different colors. Note that the temperature T and Hamiltonian H can be replaced by other physical variables of interest, such as pressure or chemical potential. (A) Two-dimensional grids, or more specifically, a 3-by-3 grid that defines 9 intermediate states with different temperatures and Hamiltonians. For example, state 8 can be represented as (T_2, H_3). (B) Another 2D example where U-shaped grids define 7 intermediate states with different Hamiltonians to discharge portions of molecule A, morph the van der Waals sites of molecule A into those of molecule B, and recharge these sites in molecule B. (C) One-dimensional grids that define 9 intermediate states with different Hamiltonians.

Second, by harnessing the statistics gathered from the overlapping states between replicas, we examine the ability of the REXEE method to converge alchemical weights compared to the weight-updating EE simulations with the Wang−Landau algorithm.[38,39] Importantly, more accurate alchemical weights converged in the weight-updating stage can potentially accelerate the convergence of free energy estimates in the production stage. Lastly, we aim to examine the relationship between REXEE parameters and the performance of the method by comparing various metrics of REXEE simulations with different setups.

In pursuit of these objectives, we applied the REXEE method with different setups to calculate the solvation free energy of anthracene previously studied by Paliwal et al.,[63] as well as the binding free energy of the host−guest binding complex CB7−10 from SAMPL4 SAMPLing challenge.[51] With these test systems, we compared the REXEE method with the weight-updating phase of EE simulations in terms of the quality of the converged weights. Additionally, we compared the REXEE method with the HREX simulations and fixed-weight EE simulations regarding the sampling speed and the accuracy of free energy calculations.

Currently, all necessary algorithms required to enable REXEE have been implemented in the pip-installable Python package ensemble_md (https://github.com/wehs7661/ensemble_md), which wraps around the key GROMACS functionalities and supports GROMACS[64,65] starting from version 2022.5. ensemble_md serves as an interface to automate and manage the initialization and execution of each iteration in a REXEE simulation. It is equipped with several user-friendly command-line interfaces (CLIs) and has been extensively unit-tested, continuously integrated, and thoroughly documented. Our current implementation of the REXEE method supports only synchronous parallelization, with the development of the asynchronous REXEE method left for future work.

THEORY

REXEE Configuration. We consider a REXEE simulation comprised of R parallel replicas of expanded ensembles, each of which is labeled as i = 1, 2, ..., R, respectively. These R replicas are restricted to sampling R different yet overlapping sets of states (termed **state sets**) labeled by m as A_1, A_2, ..., A_R, which collectively traverse N alchemical intermediate states in total, with $N > R$. The label m (m = 1, 2, ..., R) for state sets is a permutation of the label i (i = 1, 2, ..., R) for replicas, and vice versa. The relationship between the replica indices and set indices is

$$\begin{cases} m = f(i) \\ i = f^{-1}(m) \end{cases} \tag{1}$$

where f is a permutation function mapping the replica index i to the state set index m and f^{-1} is its inverse function that performs the reverse mapping. Upon any exchange during the simulation, a new permutation function must be defined. For example, in Figure 1C, the first replica is initialized with sampling the first state set, denoted as $f(1) = 1$. After an accepted exchange, the first replica samples the second state set in the second iteration, which necessitates a new permutation defined by another function f' such that $f'(1) = 2$ (and $f'(2) = 1$). Importantly, such a permutation relationship implies a one-to-one correspondence between i and m, ensuring one and unique state set associated with each replica.

Additionally, we define $s_i \in \{1, 2, ..., N\}$ as the index of the state currently sampled by the i-th replica. For a replica i sampling the state set A_m, s_i is additionally constrained such that $s_i \in A_m$. Importantly, the fact that s_i takes values in {1, 2, ..., N} and $N > R$ implies a many-to-one relationship between the replica index i and the state index s_i, as a certain state may be sampled by multiple replicas. This is in contrast to the one-to-one relationship between the replica index i and the state set index m.

We emphasize that a valid REXEE configuration only requires overlapping state sets and is not restricted to one-dimensional grids, the same number of states for all replicas, nor sequential state indices within the same state sets. For example, Figure 2 shows cases where intermediate states are characterized by more than one thermodynamic variable (panels A and B), where different state sets have different numbers of states (panel C), and where the state indices are not consecutive within the same state sets (panels A and C). While some cases presented in Figure 2 might not necessarily be practical and are just for illustrative purposes, all

https://doi.org/10.1021/acs.jctc.4c00484
J. Chem. Theory Comput. 2024, 20, 6062−6081

configurations in Figure 2 fit into the REXEE formalism, as the overlap between different state sets allows the system to access all states upon exchanges. Currently, the most common case is where the intermediate states are defined in a one-dimensional space, with consecutive state indices within the same state set (e.g., the case in Figure 1C). In a REXEE simulation adopting such a replica configuration, a state shift ϕ between adjacent state sets can be defined to indicate to what extent the set of states has shifted along the auxiliary variable. Additionally, the simulation can be either homogeneous or heterogeneous, depending on whether all replicas have the same number of states and whether or not the state shift is consistent between all adjacent state sets.

In our analysis of the different choices one can make in REXEE and how they affect simulation efficiency, we focus on examples of homogeneous REXEE simulations with consecutive one-dimensional alchemical intermediates defined in each state set, though the approach itself is more general. This decision facilitates simplifying our investigation into the dependence of the simulation performance on REXEE parameters in this exploratory work. Additionally, the majority of existing replica exchange applications adopt only a single auxiliary variable, though we hope the REXEE formalism will make it easier to investigate more complicated setups. We denote the number of states per replica as n_s, so the m-th state set (m = 1, 2, ..., R) at any given time can be expressed as follows:

$$A_m = \{(m-1)\phi + 1, (m-1)\phi + 2, \ldots, (m-1)\phi + n_s\} \quad (2)$$

For such a REXEE simulation, the four configurational parameters N (number of states), R (number of replicas), n_s (number of states per state set (or replica)), and ϕ (shift between adjacent state sets) are related via the following relationship:

$$N = n_s + (R-1)\phi \quad (3)$$

and we additionally define the overlap ratio r as the ratio between the number of overlapping states and the number of states per replica:

$$r = \frac{n_s - \phi}{n_s} \quad (4)$$

For example, the configuration of the REXEE simulation shown in Figure 1C can be expressed as (N, R, n_s, ϕ) = (9, 4, 6, 1) and has an overlap ratio of $r = 5/6 \approx 83\%$, as neighboring sets share 5/6 of their states. As discussed in the Supporting Information (Section 1, "Enumerating the space of configurations of REXEE simulations"), solving eq 3 with a few additional constraints allows efficient enumeration of all possible REXEE configurations aligned along a single auxiliary variable.

Unlike traditional replica exchange methods, the total number of states N does not have to be equal to the number of replicas R in the REXEE method. In fact, it can be shown that for a REXEE simulation sampling with any number of replicas, there exists at least one valid REXEE configuration (see Figure S1B). This allows a much higher degree of flexibility in the parameter specification as compared to traditional replica exchange methods—once the number of replicas is decided, typically as a factor of the number of available cores, the total number of states can be arbitrary.

State Transitions in REXEE Simulations. In a REXEE simulation, regardless of its replica configuration, state transitions occur at both the intrareplica and inter-replica levels. Within each replica of expanded ensemble simulation, transitions between alchemical states within the state set are governed by the selected algorithm in the expanded ensemble simulation, such as the ones mentioned in the section "Introduction". The desired probability distribution of each state and detailed balance condition are determined by whichever of various transition schemes implemented for the expanded ensemble method, as discussed in the study by Chodera et al.[10] Detailed balance for intrareplica exchange ensures the convergence toward an equilibrium distribution for each state set. At the inter-replica level, transitions involve exchanges of configurations between different replicas, which are required to achieve sampling across the entire alchemical space. Detailed balance at this level ensures that the probability influx and outflux are equal for each *set* of states. Notably, these two levels of balance are controlled independently. Both of them need to be obeyed to ensure overall probability distributions and so that the free energy difference across the entire alchemical space can be correctly estimated.

Since the detailed balance at the intrareplica level can be achieved by simply selecting a well-established method used in traditional generalized ensemble methods,[10] we only need to derive the acceptance ratio that ensures the detailed balance at the inter-replica level for the REXEE method. Specifically, we consider replicas i and j that sample the state sets A_m and A_n, respectively. To swap replicas i and j, the state sampled by replica i at the moment, denoted as $s_i \in A_m$, must fall within the state set A_n that is to be swapped, and vice versa. In this case, we call that these replicas i and j are "swappable" and we express the exchange of coordinates x_i and x_j between these two replicas as

$$X = (\ldots, x_m^i, \ldots, x_n^j, \ldots) \rightarrow X' = (\ldots, x_m^j, \ldots, x_n^i, \ldots) \quad (5)$$

with $x_m^i \equiv (x_i, A_m)$ meaning that the i-th replica samples the m-th state set with the coordinates x_i. (See the Supporting Information Section 2, "Derivation of the acceptance ratio for swapping EE replicas in a REXEE simulation" for the strict mathematical definition of the terms involved in eq 5.) Mathematically, the list of swappable pairs $\mathcal{S}$ can be defined as the set of replica pairs as follows:

$$\mathcal{S} = \{(i, j) | s_i \in A_n \text{ and } s_j \in A_m, i \neq j\} \quad (6)$$

As discussed in Section 2 in the Supporting Information, the most straightforward way to derive the acceptance ratio is to assume symmetric proposal probabilities, which can be easily achieved by the design of the used proposal scheme (see the next subsection "Proposal schemes"). Under this assumption, the acceptance ratio of swapping the coordinates (x_i and x_j) between replicas i and j can be expressed as

$$P_{\text{acc}} = \begin{cases} 1 & , \text{if } \Delta \leq 0 \\ \exp(-\Delta) & , \text{if } \Delta > 0 \end{cases} \quad (7)$$

where

$$\Delta = (u_{s_i}(x_j) + u_{s_j}(x_i)) - (u_{s_i}(x_i) + u_{s_j}(x_j)) \quad (8)$$

https://doi.org/10.1021/acs.jctc.4c00484
J. Chem. Theory Comput. 2024, 20, 6062−6081

In eq 8, u_{s_i} and u_{s_j} are the reduced potentials of the states s_i and s_j sampled by replicas i and j, respectively. Notably, the expression of Δ is of the same form as that in the HREX method. This also shows that the REXEE method reduces to the HREX method when each replica is restricted to sampling only one specific state, in which case the state labels s_i and s_j reduce to i and j, respectively (see Section 2 of the Supporting Information for the definition of the reduced potential and the full derivation of the acceptance ratio).

Proposal Schemes. In this section, we discuss a few common proposal schemes for the REXEE method. Given the prevalence of symmetric proposal probabilities in replica exchange methods, we primarily focus on methods that use symmetric proposal probabilities. The only exception is the multiple exchange proposal scheme, which provides an initial exploration of the potential of asymmetric proposal probabilities. Notably, there may easily exist other possible proposal schemes that can achieve the inter-replica level of detailed balance, but we do not further investigate here.

Single Exchange Proposal Scheme. The most straightforward proposal scheme is to randomly draw a pair from the list of swappable pairs $\mathcal{S}$ defined in eq 6, with each pair in the list having an equal probability to be drawn, in which case the proposal probability can be expressed as follows:

$$\alpha(X'|X) = \alpha(x_m^j, x_n^i|x_m^i, x_n^j) = \begin{cases} 1/|\mathcal{S}| & , \text{if } (i,j) \in \mathcal{S} \\ 0 & , \text{if } (i,j) \notin \mathcal{S} \end{cases} \tag{9}$$

Note that this proposal probability is symmetric, i.e., $\alpha(X'|X) = \alpha(X|X')$ for all (i, j) pairs. With this "single exchange proposal scheme", only one exchange is proposed, with the probability defined in eq 9. This proposal scheme has been implemented in the package ensemble_md as a sanity check for the REXEE method.

Neighbor Exchange Proposal Scheme. In traditional replica exchange methods, the neighbor exchange proposal scheme alternates between swapping all replica pairs $2l - 1$ with $2l$ and all pairs $2l$ with $2l + 1$ for each $l = 1, ..., \lfloor R/2 \rfloor$. While this notion of alternating between "odd-even pairs" and "even-odd pairs" is applicable to the REXEE method, one needs to take into account the fact that some pairs may not be swappable, i.e., $(2l - 1, 2l) \notin \mathcal{S}$ or $(2l, 2l + 1) \notin \mathcal{S}$ for specific l values, with $\mathcal{S}$ defined in eq 6.

Another way to perform neighbor exchanges, which has been implemented in the Python package ensemble_md as a sanity check, is to add a constraint to $\mathcal{S}$ defined in eq 6 such that the swappable pairs consist exclusively of neighboring replicas, with each pair having an equal probability to be drawn. Formally, the proposal probability in this case can be expressed as follows:

$$\alpha(X'|X) = \alpha(x_m^j, x_n^i|x_m^i, x_n^j) = \begin{cases} 1/|\mathcal{S}_{\text{neighbor}}| & , \text{if } (i,j) \in \mathcal{S}_{\text{neighbor}} \\ 0 & , \text{if } (i,j) \notin \mathcal{S}_{\text{neighbor}} \end{cases} \tag{10}$$

where

$$\mathcal{S}_{\text{neighbor}} = \{(i,j)|s_i \in A_n \text{ and } s_j \in A_m \text{and} |i - j| = 1\} \tag{11}$$

Similarly, the proposal probability in this case is also symmetric.

Multiple Exchange Proposal Scheme. As opposed to the single exchange or neighbor exchange proposal schemes, one can propose multiple swaps within an exchange interval to further enhance the mixing of replicas. Importantly, whether two replicas are swappable not only depends on the state sets of the two replicas, but also on the states being sampled by the two replicas at the moment. Therefore, it is not always feasible to propose multiple swaps all at once and perform them serially, as a swappable pair might become unswappable after a previous swap is accepted. To address this issue, the "multiple exchange proposal scheme" (see Algorithm 1 in Scheme 1) proposes swaps one at a time, and whenever a proposed swap is accepted, it updates the permutation of the state sets and reidentifies the list of swappable pairs before proposing the next swap. Note, however, since the order of swapping could influence the resulting swapped configurations when there is any replica involved in multiple proposed swaps, this proposal scheme does not have a symmetric proposal probability. Accordingly, an acceptance ratio other than eq 8 has to be carefully designed to deal with this asymmetry so that the detailed balance at the inter-replica level is obeyed. Currently, this proposal scheme has not been implemented in ensemble_md given our focus on proposal schemes with symmetric proposal probabilities.

Scheme 1

Algorithm 1 Multiple exchange proposal scheme

```
Identify the list of swappable pairs 𝒮.
for n = 1 to n_ex, where n_ex is the desired number of swaps do
    if 𝒮 ≠ ∅ then
        Draw (i, j) ∈ 𝒮 with the proposal probability defined in Equation 9.
        Calculate the acceptance ratio P_acc for the drawn pair using Equation 7.
        if the proposed swap (i, j) is accepted then
            Perform the swap (i, j), then update the list of swappable pairs 𝒮.
        end if
    else
        break
    end if
end for
```

Exhaustive Exchange Proposal Scheme. Another approach to carry out multiple swaps serially within an exchange interval is the "exhaustive exchange proposal scheme", as detailed in Algorithm 2 in Scheme 2. In brief, it operates similarly to the single exchange proposal scheme, but exhaustively traverses the list of swappable pairs while updating the list by eliminating any pair involving replicas that appeared in the previously proposed pair. This elimination process circumvents the issue of the result depending on the order of swapping, as no replica will be involved in more than one swap and all the swaps proposed in the same exchange interval are independent of each other. Consequently, this ensures that the proposal probability is symmetric and the detailed balance condition is

Scheme 2

Algorithm 2 Exhaustive exchange proposal scheme

```
Identify the list of swappable pairs 𝒮.
while 𝒮 ≠ ∅ do
    Draw (i, j) ∈ 𝒮 with the proposal probability defined in Equation 9.
    Calculate the acceptance ratio P_acc for the drawn pair using Equation 7.
    if the proposed swap (i, j) is rejected then
        break
    end if
    Perform the swap and update 𝒮 by removing pair(s) involving replicas i and j.
end while
```


https://doi.org/10.1021/acs.jctc.4c00484
J. Chem. Theory Comput. 2024, 20, 6062−6081

obeyed with the use of eq 8. This method has also been implemented in the Python package ensemble_md.

Weight Combination Schemes for Weight-Updating REXEE. One pronounced difference between the REXEE method and other generalized ensemble approaches like EE or HREX lies in the existence of "overlapping states" in REXEE. These states fall within the intersection of at least two state sets (see Figure 1C) and are therefore accessible by multiple simulation replicas. To leverage the samples of these overlapping states collected from multiple replicas in REXEE simulations composed of weight-updating EE simulations, termed weight-updating REXEE simulations, we could combine alchemical weights on the fly for these states across replicas before initializing a subsequent iteration. The hypothesis is that such on-the-fly modifications to the weights could potentially further accelerate the convergence of the alchemical weights during the weight-updating phase, providing a better starting point for the subsequent production phase.

While there are various possible ways to combine weights across replicas, some of them suffer from reference-dependent results, or the issue of interdependence in the weight difference between adjacent states. For example, one intuitive way to combine weights for a state s across replicas is to calculate the negative logarithm of the probability averaged over all replicas accessible to s. Denoting these replicas as $k \in Q_s$, the combined/averaged weight for state s, $\overline{g_s}$ (with its corresponding probability $\overline{p_s}$), can be expressed as

$$\overline{g_s} = -\ln \overline{p_s} = -\ln\left(\frac{1}{|Q_s|}\sum_{k\in Q_s} p_s^k\right) = -\ln\left(\frac{1}{|Q_s|}\sum_{k\in Q_s} e^{-g_s^k}\right) \tag{12}$$

where g_s^k is the alchemical weight of state s in the state set A_k, and p_s^k is its corresponding probability. However, as free energy differences are only defined up to a constant, it is standard to set the weight of some reference state to 0. Different state sets include different states and thus must generally define different reference states. As such, different choices of references could lead to different resulting combined weights and there is no clear justification about which reference should be favored.

An alternative approach is to take advantage of the weight differences between adjacent states. Given the weight differences between neighboring states for each replica, this method then calculates the average for the weight differences accessible by multiple replicas. This average can be either a simple average or the average weighted by the inverse statistical variance of the alchemical weights combined. In the latter case, measurements having lower variability are assigned with higher contributions, so the resulting weighted mean is less sensitive to outliers. In this study, we write the weight difference between the states s and $s+1$ in replica i as $\Delta g_{(s,s+1)}^i = g_{s+1}^i - g_s^i$, and the set of replicas that can access both s and $s+1$ as $Q_{(s,s+1)}$. Then, for the case where the inverse-variance weighting is used, we have the averaged weight difference between s and $s+1$ as

$$\overline{\Delta g_{(s,s+1)}} = \frac{\sum_{k\in Q_{(s,s+1)}} (\Delta g_{(s,s+1)}^k/(\sigma_{(s,s+1)}^k)^2)}{\sum_{k\in Q_{(s,s+1)}} 1/(\sigma_{(s,s+1)}^k)^2} \tag{13}$$

with its propagated error as

$$\delta_{(s,s+1)} = \sqrt{\left(\sum_{k\in Q_{(s,s+1)}} (\sigma_{(s,s+1)}^k)^{-2}\right)^{-1}} \tag{14}$$

where $\sigma_{(s,s+1)}^k$ is the standard deviation calculated from the time series of $\Delta g_{(s,s+1)}^k$ since the last update of the Wang−Landau incrementor in the EE simulation sampling the k-th state set. From the averaged weight differences between all adjacent states, we can obtain a profile of alchemical weights that can be used to initialize the next iteration.

Free Energy Calculations. We term REXEE simulations composed of fixed-weight EE replicas as **fixed-weight REXEE** simulations, and those composed of weight-updating EE replicas as **weight-updating REXEE** simulations. For free energy calculations, the protocol used for the fixed-weight REXEE simulations is similar to those for HREX and fixed-weight EE simulations, albeit with extra consideration for overlapping states. For each state set, one should concatenate the trajectories from all replicas, truncate the nonequilibrium region[66] and then decorrelate the concatenated data. Then, for each replica in the fixed-weight REXEE simulation, one can use free energy estimators such as TI,[45] BAR,[46] and MBAR[47] to calculate the alchemical free energies for different state sets. For the overlapping states, one can use eqs 13 and 14 to calculate the mean of the associated free energy differences $\overline{\Delta G_{(s,s+1)}}$ and the accompanying propagated error $\delta_{(s,s+1)}$, with $\Delta g_{(s,s+1)}^k$ replaced by $\Delta G_{(s,s+1)}^k$, the free energy difference computed by the chosen free energy estimator. In this context, $\sigma_{(s,s+1)}^k$ used in eqs 13 and 14 should be the uncertainty associated with $\Delta G_{(s,s+1)}^k$ calculated by the estimator. Importantly, free energy differences involving overlapping states are likely to have smaller uncertainties because more uncorrelated samples can be collected given a larger pool of samples gathered from multiple replicas.

Notably, in eqs 13 and 14, we consider only the path of $0 \to 1 \to 2 \to \cdots \to s \to s+1 \to \cdots \to N$ for transitioning from the coupled state (state 0) to the decoupled state (state N) in the contexts of both weight combinations and free energy calculations. Other pathways, such as $0 \to 2 \to 4 \to \cdots \to s \to s+2 \to \cdots \to N$ (assuming N is an even number), as well as more complex or irregular paths, may also be viable paths for weight combinations and free energy calculations. It should be noted, however, that the resulting weight/free energy difference between states 0 and N, along with its associated uncertainty, will vary based on the chosen path. Although all paths connecting the end states yield valid results, we advocate for the approach that considers all adjacent states, as used in eqs 13 and 14. The primary reason for this preference is that the path composed of only one-state moves offers the greatest flexibility, simplicity and can accommodate any degree of overlap between state sets. Moreover, in instances where the uncertainties $\sigma_{(s,s+1)}^k$ from different replicas k are close, path selection will have minimal impact on the final values $\Delta g_{(1,N)}$ or $\Delta G_{(1,N)}$ and their respective uncertainties.

An alternative way to estimate the uncertainty of $\Delta G_{(s,s+1)}$ in eq 13 is to perform bootstrapping, which is path-independent. In theory, it is also more accurate and rigorous than the propagated error in eq 14, but more computationally expensive. Specifically, in one bootstrap iteration, one can replicate the input data set that is already concatenated, truncated, and decorrelated by drawing N_{data} samples with replacement, where N_{data} is the size of the data set. From the

https://doi.org/10.1021/acs.jctc.4c00484
J. Chem. Theory Comput. 2024, 20, 6062−6081

sampled data set, one can repeat the protocol above to get a free energy estimate. Then, the uncertainty can be estimated by taking the standard deviation of the free energy estimates obtained from a predetermined number of bootstrap iterations.

METHODS

The ensemble_md Python Package. ensemble_md is a pip-installable Python package that houses all algorithms required for implementing the REXEE method. It is compatible with GROMACS starting from version 2022.5. As an adaptable wrapper around GROMACS functionalities, ensemble_md provides a layer of high-level abstraction over the complexity of performing a REXEE simulation. Specifically, it offers several user-friendly CLIs, such as explore_REXEE, run_REXEE and analyze_REXEE. The CLI explore_REXEE solves eq 3 with additional constraints to efficiently enumerate all possible REXEE configurations, providing the user a quick overview of the parameter space for configuring a REXEE simulation. The CLI run_REXEE streamlines the workflow of conducting a REXEE simulation by orchestrating the iterative cycle of GROMACS simulations. This includes preparing the simulation inputs for each iteration, launching simulations, and performing brief, on-the-fly data analysis for parameter adjustment between iterations. Lastly, the CLI analyze_REXEE integrates an assortment of data analysis methods tailored for REXEE simulations. Tasks automated by analyze_REXEE include trajectory stitching, replica- and state-based transition analysis, time series analysis, Markov State Model (MSM)[67−69] analysis, data visualization, and free energy calculations.

At its core, ensemble_md implements basic functionalities that are central to the REXEE method. This includes methods for swapping input configurations between replicas, calculating the acceptance ratio, and managing the behavior of the next EE iteration by tweaking the input simulation parameters. It also provides different proposal weights and weight combination methods introduced in the section "Theory". By launching subprocess calls of the GROMACS executable, ensemble_md eliminates the need to alter the source code of GROMACS, while ensuring negligible overhead in Python execution. Owing to the modularity of its core functionalities, ensemble_md is also easily extensible. It provides building blocks for easy extension of the REXEE approach (e.g., by enabling customized swapping schemes), or formulating novel simulation approaches (e.g., by combining the replica exchange method with other enhanced sampling methods). In addition, the majority of ensemble_md's functions are agnostic of the MD simulation engine, paving the way for integration with other MD engines by merely adding a few engine-specific functions to handle different file extensions and data types. Lastly, ensemble_md ensures quality through extensive unit testing, continuous integration, and comprehensive documentation.

Simulations of Anthracene. To showcase the usage of the REXEE approach, we applied the method to the calculation of the solvation free energy of anthracene, a selection made based on several compelling factors. Primarily, anthracene presents a balanced level of challenge for free energy methodologies—it is cheap enough to allow efficient exploration of diverse simulation parameter setups while still providing a sufficient challenge because of the large number of atomic sites introduced to rigorously test the efficacy of free energy methods. Furthermore, the absence of configurational metastable states in the anthracene system eliminates the possibility of having configurational free energy barriers orthogonal to the alchemical variable, simplifying the comparison of alchemical sampling among generalized ensemble methods, including EE, HREX, and the REXEE methods. Lastly, anthracene has been extensively studied in the work by Paliwal et al.,[63] which offers a comparative reference. With this system, we compared the free energy estimates from fixed-weight REXEE simulations with different setups to the benchmark values obtained from HREX and fixed-weight EE simulations. We also compared REXEE and EE methods in their ability to converge alchemical weights in the weight-updating stage. All simulations described in the following subsections were performed with GROMACS 2022.5, with the aid of the package ensemble_md for the REXEE simulations. To enable a straightforward comparison of computational costs, all simulations were carried out on identical computational architectures. Example simulation inputs, including initial configurations, topologies, and simulation parameters are available at https://github.com/wehs7661/ensemble_md/tree/master/ensemble_md/data.

System Preparation. For the anthracene system, we selected one of the configurations from the study by Paliwal et al.[63] as the initial configuration for the subsequent downstream investigation. We equilibrated the system in the NVT ensemble and then the NPT ensemble. Both equilibration processes were performed for 200 ps, with the velocity rescaling method[70] used in both to maintain the reference temperature at 300 K, and a Berendsen barostat[71] used in the latter to fix the pressure at 1 bar. Afterward, we carried out a 5 ns NPT MD simulation with a Parrinello−Rahman barostat[72,73] keeping the pressure at 1 bar. A switching function was used for the calculation of van der Waals interaction, where the switch started at 0.8 nm and the cutoff distance was set at 0.9 nm. For efficient calculations of long-range electrostatic interactions, the PME (particle mesh Ewald) method[74] was used with a cutoff distance of 0.9 nm and a grid spacing of 0.1 nm. Bonds involving hydrogen bonds were constrained by the LINCS algorithm[75] with 2 iterative corrections. The highest order in the expansion of the constraint coupling matrix was specified as 12. Upon completion of the MD simulation, we extracted the configuration with the box volume closet to the average volume of the MD trajectory. This configuration then served as the input configuration for subsequent EE, HREX, and REXEE simulations elaborated in the following sections, as all these simulations were conducted in the NVT ensemble to avoid potential issues with λ dependence of pressure.

Benchmark Simulations. To establish benchmark values of solvation free energy for comparison with the results from REXEE simulations, we performed EE and HREX simulations for the anthracene system, both in the NVT ensemble. Both benchmark simulations utilized 8 alchemical intermediate states in total, which is a convenient number to parallelize the replicas in the HREX simulation. These 8 alchemical states were used to only decouple the van der Waals interactions, as the anthracene model had zero charge on all atoms. Their coupling parameters, available at https://github.com/wehs7661/ensemble_md/tree/master/ensemble_md/data, were chosen to ensure sufficient overlap between adjacent states. To avoid singularities in standard Lennard-Jones potentials, soft-core interpolation[76] between the end states was applied. The parameters α and p in the soft-core potentials were specified as 0.5 and 1, respectively.

https://doi.org/10.1021/acs.jctc.4c00484
J. Chem. Theory Comput. 2024, 20, 6062−6081

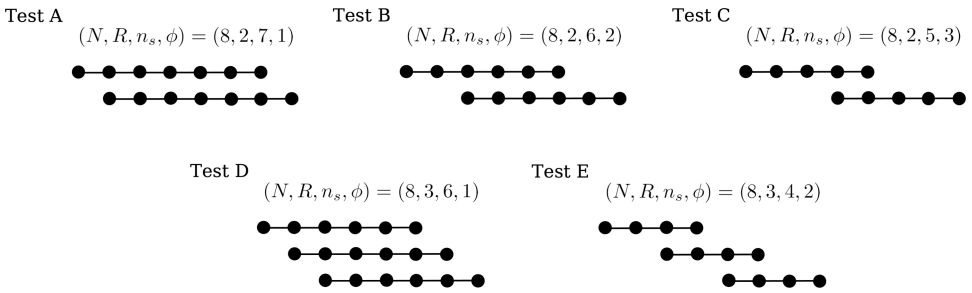


Figure 3. Schematic representations of REXEE configurations of Tests A to E for anthracene solvation free energy calculation. The black dots represent alchemical states and different rows of dots represent different replicas.

For the EE benchmark, we adopted the prevalent two-stage protocol described in the section "Introduction". In the weight-updating phase, we used the $1/t$ variant of the Wang−Landau algorithm[38,39] to converge the alchemical weights for the 8 alchemical states, with the initial Wang−Landau incrementor set to 0.5 k_BT. Monte Carlo moves in the alchemical space were proposed every 100 integration steps with Metropolized Gibbs sampling. We adopted the default value of 0.8 for the flatness ratio R_{flat}, a ratio between the histogram counts of a state and the histogram counts averaged over all states. A cutoff of 0.8 means the histogram is considered flat only if all states have an R_{flat} value smaller than 0.8, and its reciprocal $1/R_{flat}$ larger than 0.8. Whenever the histogram is considered flat, all histogram counts are reset to 0 with the Wang−Landau incrementor scaled by a scaling factor, which was set to 0.8 in our case. This weight-updating procedure ceased when the Wang−Landau incrementor fell below 0.001 k_BT, reaching what we term as the **Wang−Landau (WL) convergence**. The weights converged by the Wang−Landau algorithm were then fixed for the EE simulation in the production phase, which was performed for 200 ns. The fixed-weight EE simulation used exactly the same set of parameters as those used in the weight-updating stage, except that no weight-updating settings were specified. To determine the number of uncorrelated samples, we truncated the nonequilibrium regime and decorrelated the time series of the Hamiltonian difference between adjacent states[66] generated by the fixed-weight EE simulation. Then, we applied MBAR[47] to the uncorrelated samples to compute the free energy difference between the coupled and decoupled states, which is the solvation free energy of anthracene.

The HREX benchmark simulation used one replica for each of the 8 intermediate states. All 8 replicas were seeded with the same initial configuration and each of them was performed for 25 ns, summing up to the same total simulation time of 200 ns as the production phase of the EE benchmark simulation. The parameters used in the HREX benchmark simulation are identical to those in the fixed-weight EE simulation, except that no alchemical weights were assigned and no expanded ensemble settings were used. For free energy calculations, we truncated and decorrelated the time series of Hamiltonian differences for each replica individually, then concatenated data from different replicas for MBAR[47] calculations.

Fixed-Weight REXEE Simulations. To make the comparison between REXEE simulations and the benchmark simulations straightforward, we used the same 8 alchemical intermediate states ($N = 8$) for all REXEE simulations. All EE replicas of all REXEE simulations have the same set of parameters as the EE benchmark simulation, except that the different EE replicas in REXEE were restricted to different state sets.

With the anthracene system, we tested different setups of fixed-weight REXEE simulations and classified them into two groups. Group 1 includes Tests 1 to 3, which test the effect of different frequencies for exchanging replicas, with the corresponding simulation lengths per iteration (i.e., exchange period) of 4, 10, and 100 ps, respectively. All three tests were configured with $(N, R, n_s, \phi) = (8, 4, 5, 1)$. In parallel, Group 2 includes Tests A to E, which explore 5 different (N, R, n_s, ϕ) combinations varying the number of replicas and the level of overlap. All 5 tests were performed at a fixed exchange frequency of swapping replicas every 4 ps, a frequency shown to lead to faster mixing in the sampling space and decent free energy estimates in Group 1 (see the section "Results and Discussion"). The 5 considered (N, R, n_s, ϕ) combinations include (8, 2, 7, 1), (8, 2, 6, 2), (8, 2, 5, 3), (8, 3, 6, 1), and (8, 3, 4, 2) (see Figure 3). All REXEE simulations in Groups 1 and 2 have the same effective simulation lengths as those of benchmarks, which are 200 ns. All simulations adopted the exhaustive exchange proposal scheme and were all initiated with the same set of alchemical weights obtained in the weight-updating phase of the EE benchmark simulation.

To assess the performance of the REXEE method, we compared the free energy estimates of all 8 REXEE simulations with the benchmark simulations. Specifically, for each state set, we stitched all Hamiltonian time series, then truncated and decorrelated the combined time series[66] for MBAR[47] calculations. The free energy profiles obtained from different state sets are then combined as described in the "Theory" section. Furthermore, from each demuxed trajectory, we quantified the sampling speed of REXEE simulations in both the replica space and the state space. If Hamiltonian replica exchange is being used as a configurational sampling technique, then faster sampling in state space is a necessary (though not sufficient) requirement for improved configurational sampling. Specifically, we calculated the replica-space relaxation time (τ_r) given by below:

$$\tau_r = \frac{\tau}{1 - \lambda_2} \tag{15}$$

where τ is the exchange period and λ_2 is the second-largest eigenvalue calculated from the replica-space transition matrix. As a measure of the sampling speed in the replica space, τ_r provides an estimate of the simulation time required for the autocorrelation function of the replica index to decay to $1/e$ of its initial value, and a shorter replica-space relaxation time is indicative of faster mixing in the replica space. As for the state-space sampling speed, we adopted two straightforward metrics, including the correlation time of the state index (τ_λ) and the number of round trips (n_r) in the alchemical space. Notably,

https://doi.org/10.1021/acs.jctc.4c00484
J. Chem. Theory Comput. 2024, 20, 6062−6081

for the state index correlation time τ_λ, we report averages over R available trajectories with uncertainties being the standard deviation. For the number of round trips n_r, we report the sum over all R trajectories, so that all values are based on the same simulation length. Its uncertainty was estimated as $1/\sqrt{R}$ times the standard deviation calculated from the R trajectories, as the uncertainty should scale with $1/\sqrt{R}$ as the sample size scales with R. Lastly, we estimated the uncertainty of the replica-space relaxation time τ_λ by applying bootstrapping to synthesized replica-space transition matrices that mock the empirical transition matrix. For metrics applicable to the benchmark simulations (i.e., τ_λ and n_r), we also applied them to assess the sampling speed of the benchmark simulations. Notably, the comparison of sampling speed and free energy calculations were not only carried out between the REXEE simulations and the benchmark simulations, but also between the REXEE simulations with different setups.

Weight-Updating REXEE Simulations. In addition to fixed-weight REXEE simulations, we also performed 6 weight-updating REXEE simulations, which are the 6 combinations of three exchange periods (4 ps, 10 and 100 ps) and whether the weight combination scheme based on simple averages was used or not. All 6 simulations were configured by $(R, n_s, \phi) = (4, 5, 1)$, a setup that has a reasonable overlap between replicas and resulted in a decent performance of the simulation given a sufficiently high exchange frequency (see the section "Results and Discussion"). Each of the 4 replica simulations in all 6 tests was scheduled for 50 ns, leading to a total simulation length of 200 ns for each test. All tests utilized the exhaustive exchange proposal scheme for swapping the coordinates between replicas.

To assess REXEE's ability to converge alchemical weights, we adopted two metrics. First, we measured the WL convergence time for each test, which refers to the time it takes for all replicas to converge the weights for their respective state set according to the Wang−Landau algorithm's criteria. Second, for each REXEE simulation, we calculated the root-mean-squared error (RMSE) between the free energy profile averaged over the period since the last update of the Wang−Landau incrementor and the reference profile calculated from the EE benchmark simulation. Due to the dynamic nature of the weight-updating phase, we performed all 6 simulations in 3 replicates and report the average to lower the influence of noise, with the uncertainty being the standard deviation. The metrics of each test were then compared with those averaged over 3 replicates of weight-updating EE simulation that used the $1/t$ variant of the Wang−Landau algorithm to converge the alchemical weights for all states.

Simulations of the CB7−10 Host−Guest Binding Complex. To further understand the REXEE approach, we applied the method to the binding free energy calculation of the host−guest binding complex CB7−10. CB7−10, composed of a ring molecule cucurbit[7]uril (CB7) as the host and 1-(2-aminoethyl)piperazine as the guest ligand, is one of the binding complexes in SAMPL4 SAMPLing challenge.[51]

In contrast to anthracene, CB7−10 has at least three predominant metastable states, including the unbound state and a binding site on each side of the symmetrical host ring. This presents a more complex scenario than solvation free energy calculations, but not to the extent that comprehensive sampling in configurational space becomes overly challenging, given that its slow degrees of freedom are not strictly orthogonal to the alchemical direction. Furthermore, CB7−10 offers the advantage of testing the accuracy of not only solute-water but also solute−solute interactions, making it an ideal candidate for evaluating free energy methods in practical contexts. With this system, we assessed REXEE simulations with various setups, evaluating the sampling speed in both the alchemical space and the configurational space, as well as the accuracy of free energy calculations and the convergence of alchemical weights in the weight-updating stage. All REXEE simulations were compared with each other and also with the EE benchmark simulation. Free energy estimates from these simulations were compared with the values reported in the study by Monroe et al.,[49] which also employed the EE method. All simulations were carried out in the NVT ensemble using GROMACS 2022.5 on identical computational architecture, with the package ensemble_md employed specifically for REXEE simulations. Example simulation inputs, including initial configurations, topologies, and simulation parameters are available at https://github.com/wehs7661/ensemble_md/tree/master/ensemble_md/data.

System Preparation. Starting from the coordinate file provided in SAMPL4 SAMPLing challenge repository[77] and the topology file adopted in the study by Monroe et al.,[49] we solvated the system in a cubic box with 1.5 nm between the solute and box edges. After charge neutralization with three chloride ions, the system was then proceeded with energy minimization followed by NVT equilibration, NPT equilibration, and finally, a 5 ns NPT MD simulation. The structure whose box volume was closest to the average volume of the MD trajectory was extracted to serve as the input for REXEE and EE benchmark simulations, which were all performed in an NVT ensemble. All parameters used in these steps are the same as those used for the preparation of the anthracene system, except that switching functions were used in this case for both the calculations of van der Waals and electrostatic interactions in the last step, with the range for switching being 0.85 to 0.9 nm and 0.89 to 0.9 nm, respectively.

Decomposition of the Binding Free Energy Calculation. We used the double decoupling method[78] with the thermodynamic cycle shown in Figure 4 to decompose the

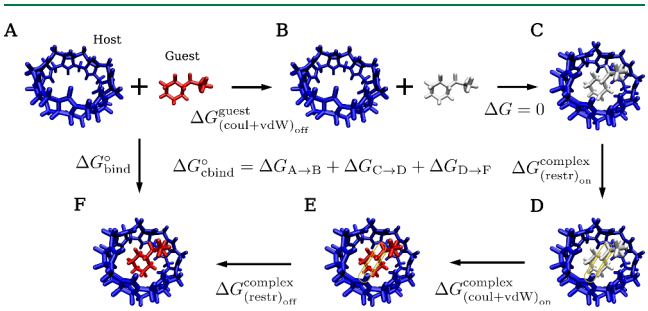


Figure 4. Double decoupling thermodynamic cycle for the binding free energy calculation of the CB7−10 host−guest binding complex. The host molecule is colored in blue, while the guest ligand is colored in either white (fully decoupled from the host) or red (fully coupled with the host). States A and F refer to the bound and unbound states of the binding complex, respectively. In State B, the ligand is fully decoupled from the host and later moved into the binding site of the host in State C. In States D and E, the paper clips (yellow) represent the distance restraint applied between the host and guest molecules, without and with nonbonded interactions between the two entities, respectively.

https://doi.org/10.1021/acs.jctc.4c00484
J. Chem. Theory Comput. 2024, 20, 6062−6081

binding free energy calculation of CB7−10 binding complex. Starting from state A, we split the entire cycle into two alchemical processes (from states A to B and from states D to F) bridged by the transfer of the ligand into the binding cavity of the host molecule (from states B to C) and the application of a distance restraint (from states C to D). The associated free energy difference of each alchemical process was calculated from a single fixed-weight EE or REXEE simulation. The simulation for the transition from states A to B, which we term the "solvent simulation", gradually decouples the ligand from its surroundings. After the ligand is fully decoupled, we can freely move it into the binding cavity of the host molecule without any energy penalty, so the associated free energy difference $\Delta G_{B\to C}$ is 0. Then, it is of common practice to apply a distance restraint between the host and guest molecules to prevent the guest ligand from drifting away, which effectively shortens the decorrelation time. The change in the free energy associated with this process can be calculated analytically:

$$\Delta G^{\text{complex}}_{(\text{restr})_{\text{on}}} = -k_B T \ln\left[\frac{1}{V_0}\left(\left(\frac{2\pi k_B T}{K}\right)^{3/2} + \frac{8\pi r_0 k_B T}{K} + 2\pi r_0^2\left(\frac{2\pi k_B T}{K}\right)^{1/2}\right)\right] \tag{16}$$

where r_0 is the reference distance, K is the force constant and V_0 is the molecular volume (1.6605 nm^3) corresponding to the 1 mol/L reference concentration (for the derivation of eq 16, please refer to Section 3 of the Supporting Information). The thermodynamic cycle can then be closed by the so-called "complex simulation" going from states D to F, which turns back on the nonbonded interactions and switches off the distance restraint. Finally, the binding free energy $\Delta G^{\circ}_{\text{bind}}$ can be calculated as the sum of the free energy differences associated with the process going from states A through B, C, D, E to F. To simplify the comparison between the EE and REXEE methods, we only compared the methods in the complex simulation for the calculation of $\Delta G_{D\to F}$, which is typically the more challenging part of binding free energy calculations. We therefore only used the EE method to calculate $\Delta G_{A\to B}$ from the solvent simulation for all comparisons.

Solvent Simulation. As described above, the solvent simulation, which samples the transition from states A to B in Figure 4, was performed using only the EE method with the common two-stage protocol. We adopted the simulation parameters used in the study by Monroe et al. wherever possible for a more straightforward comparison of the free energy estimates. This includes the specification of the alchemical path and the settings for Monte Carlo moves in the state space, alongside the parameters for both the Wang−Landau algorithm and the soft-core potential. Specifically, 40 alchemical states in total were used, 21 of which were used to decouple the Coulombic interactions and the others for the van der Waals interactions. In the weight-updating stage, the $1/t$ variant of the Wang−Landau algorithm was used to adaptively estimate the weight for each alchemical state, with an initial Wang−Landau incrementor of 10 $k_B T$, a flatness ratio R_{flat} of 0.8, a scaling factor of 0.7, and a cutoff of 0.001 $k_B T$ for halting the weight-updating process. Monte Carlo moves in the state space were attempted every 100 integration steps using the Metropolized Gibbs sampling method. The parameters α and p for the soft-core potential[76] were set to 0.5 and 1, respectively. The converged weights were used to seed the 200 ns production run, from which the data was extracted and fed to the MBAR estimator for free energy calculations after proper truncation and decorrelation. Both simulations were conducted in the NVT ensemble to avoid any potential issues with λ dependence of pressure. This is as opposed to the EE simulations performed in Monroe et al.'s study, which were performed in the NPT ensemble, as changes in GROMACS since that time made it complicated to run the expanded ensemble simulations in the NPT ensemble. However, in an aqueous solution at 1 atm, the difference between Helmholtz and Gibbs free energies is within statistical error if run at properly corresponding pressure/volumes.

Benchmark Simulations. We performed EE simulations with the two-stage protocol to calculate a benchmark value for the free energy difference $\Delta G_{D\to F}$, to which the free energy estimates from the REXEE simulations were compared. In the weight-updating stage, we set the initial Wang−Landau incrementor to 10 $k_B T$ to allow more rapid equilibration given the large free energy difference in desolvation of a charged complex. A λ-dependent distance restraint was applied between the ligand and the host molecules to reduce the movement of the ligand in the unbound state, which could otherwise explore the entire simulation box. Specifically, for each state, the equation for the force constant was $k(\lambda) = \lambda K$, where K was set as 1000 kJ/mol/nm^2 (all λ values are available at https://github.com/wehs7661/ensemble_md/tree/master/ensemble_md/data). The initial distance between the centers of mass of the two entities, r_0 = 0.428 nm, was used as the reference distance, at which structures correspond to the bound state. All the other simulation parameters used in the weight-updating EE simulation are the same as those used in the solvent simulation in the weight-updating stage. The fixed-weight EE simulation for production used the same set of parameters as the weight-updating simulation except that no weight-updating parameters were specified and the weights were initialized at the converged values obtained in the weight-updating stage. Similarly, after truncation and decorrelation, we applied MBAR to compute the free energy difference $\Delta G_{D\to F}$.

To assess the binding and unbinding dynamics of the host−guest binding complex, we performed a clustering analysis for the fixed-weight EE simulation and used it as a benchmark for later comparison with REXEE simulations. We first removed periodicity and jumps of molecules across the simulation box from the trajectory, centered the binding complex, and performed clustering using the single linkage algorithm for the fully coupled configurations, with an RMSD cutoff of 0.13 nm, a value also adopted in the study by Monroe et al.[49] Given the trajectory with each frame assigned to a cluster, we then calculated the number of flips between the two most dominant clusters, which were expected to be the guest molecule bound to the two different portals of the host molecule, to quantify the sampling efficiency in the configurational space.

Fixed-Weight REXEE Simulations. We performed 8 fixed-weight REXEE simulations for the CB7−10 host−guest binding complex to sample the same 40 alchemical states as those defined in the EE benchmark simulation. These simulations, denoted as Tests 1 to 8, explore 8 different REXEE configurations, with the (N, R, n_s, ϕ) combination including (40, 4, 13, 9), (40, 4, 19, 7), (40, 4, 37, 1), (40, 6, 10, 6), (40, 6, 15, 5), (40, 6, 35, 1), (40, 8, 12, 4), and (40, 8, 33, 1). These configurations were chosen such that for each R value (number of replicas) of 4, 6, 8, the configurations with the highest and lowest overlap between adjacent state sets were included, with additions of an intermediate overlap for R = 4

https://doi.org/10.1021/acs.jctc.4c00484
J. Chem. Theory Comput. 2024, 20, 6062−6081

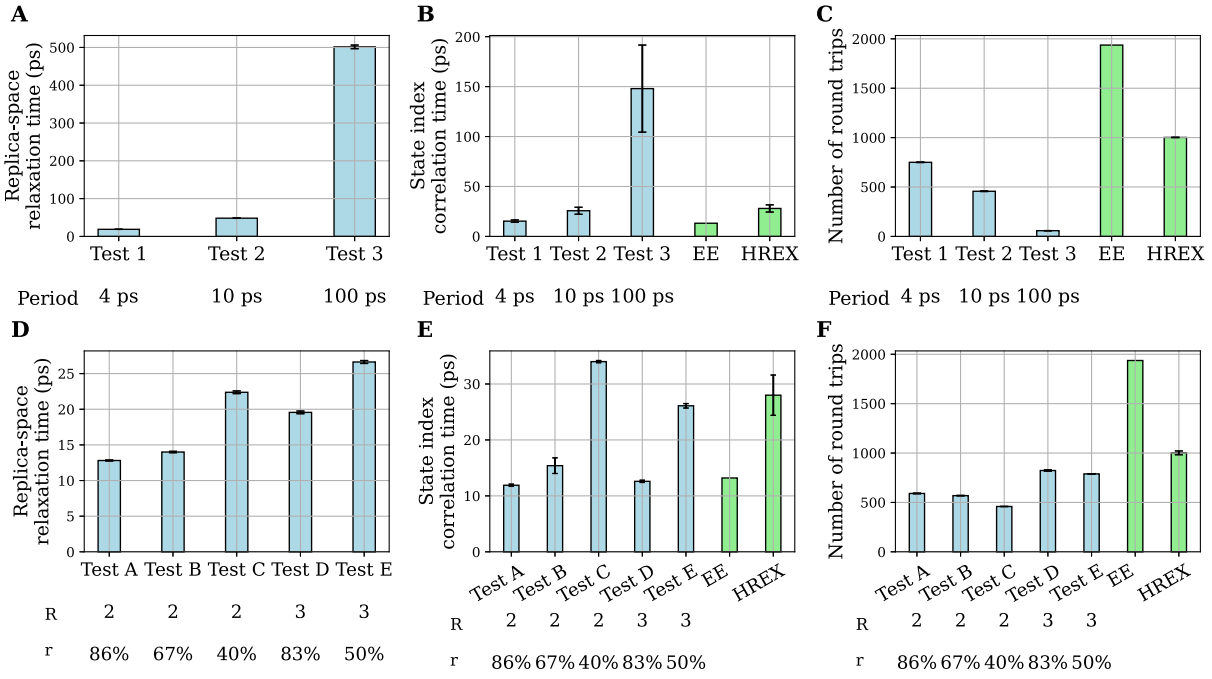


Figure 5. Different metrics for assessing the sampling speed for all fixed-weight REXEE simulations of anthracene, with parameters differing between tests annotated below the *x*-axis for easier comparisons. In panels D to F, *R* and *r* represent the number of replicas and the overlap ratio, respectively. Results from the benchmark simulations are in light green. The uncertainties of each metric were calculated using the protocol described in section "Methods", with some error bars being too small to be visible. Overall, the figure shows that faster sampling in the replica space and the state space can be achieved by faster exchanges (shorter exchange periods) or higher state overlaps between adjacent state sets. (A) Replica-space relaxation time from tests in Group 1. (B) State index correlation time from tests in Group 1. (C) Total number of round trips from tests in Group 1. (D) Replica-space relaxation time from tests in Group 2. (E) State index correlation time from tests in Group 2. (F) Total number of round trips from tests in Group 2.

and 6 to capture a broader insight into the performance spectrum of these setups. All 8 tests use the exhaustive exchange proposal scheme to propose exchanges every 4 ps, a frequency that is highest among those tested in the anthracene simulations and led to the highest sampling efficiency in the alchemical space (see the section "Results and Discussion"). All tests have the same aggregate simulation length of 200 ns and used the same set of parameters as utilized in the EE benchmark simulation. They were all initiated with the same weights obtained from the weight-updating phase of the EE benchmark simulation. We utilized the same data analysis protocol used for the anthracene fixed-weight simulations to analyze the CB7−10 simulations, with an exception in the time series decorrelation protocol for free energy calculations. Specifically, we found that the data decorrelation method[66] occasionally underestimated the statistical inefficiency of CB7−10 trajectories, and we therefore applied the geometric mean of the statistical inefficiency over all trajectories to average out the uncertainty intrinsic to the method.

To examine the sampling efficiency in the alchemical space, we adopted the same protocol as the one used for analyzing fixed-weight REXEE simulations of anthracene. To assess the configurational sampling of the CB7−10 REXEE simulations, we first stitched and recovered a continuous trajectory for each starting configuration, and applied the same protocol for clustering analysis described in the previous section to get the flipping rate for each simulation. Given that the aggregate length of each REXEE simulation is the same as the length of the benchmark EE simulation, for each REXEE simulation, we summed up the number of flips across all trajectories and compared the sum with the flip count calculated from the EE benchmark simulation.

Weight-Updating REXEE Simulations. In contrast to the weight-updating REXEE simulations performed for the anthracene system, which focused on one single REXEE configuration and investigated different exchange frequencies and the use of weight combinations, the 8 weight-updating REXEE simulations performed for the CB7−10 system explored different REXEE configurations using the same exchange frequency. Referred to as Tests 1 to 8, these tests examine the same 8 REXEE configurations as those explored in the 8 tests of fixed-weight REXEE simulations described in the previous section. They did not use any weight combination given that the results from the anthracene system showed that weight combinations generally impeded the convergence of weights (see the section "Results and Discussion"). Additionally, given that the exchange frequency did not have a noticeable influence on the weight convergence in the anthracene system, we adopted an intermediate swapping period of 10 ps in the exhaustive exchange proposal scheme (see the section "Results and Discussion"). Then, we again utilized the WL convergence time and RMSE with respect to the EE benchmark simulation to assess the weight convergence of REXEE simulations. All REXEE simulations were done in 3 replicates and averages across replicates are reported with standard deviations as uncertainties. For comparison, we also performed 3 replicates of weight-updating EE simulations and assessed them with the same metrics.

https://doi.org/10.1021/acs.jctc.4c00484
J. Chem. Theory Comput. 2024, 20, 6062−6081

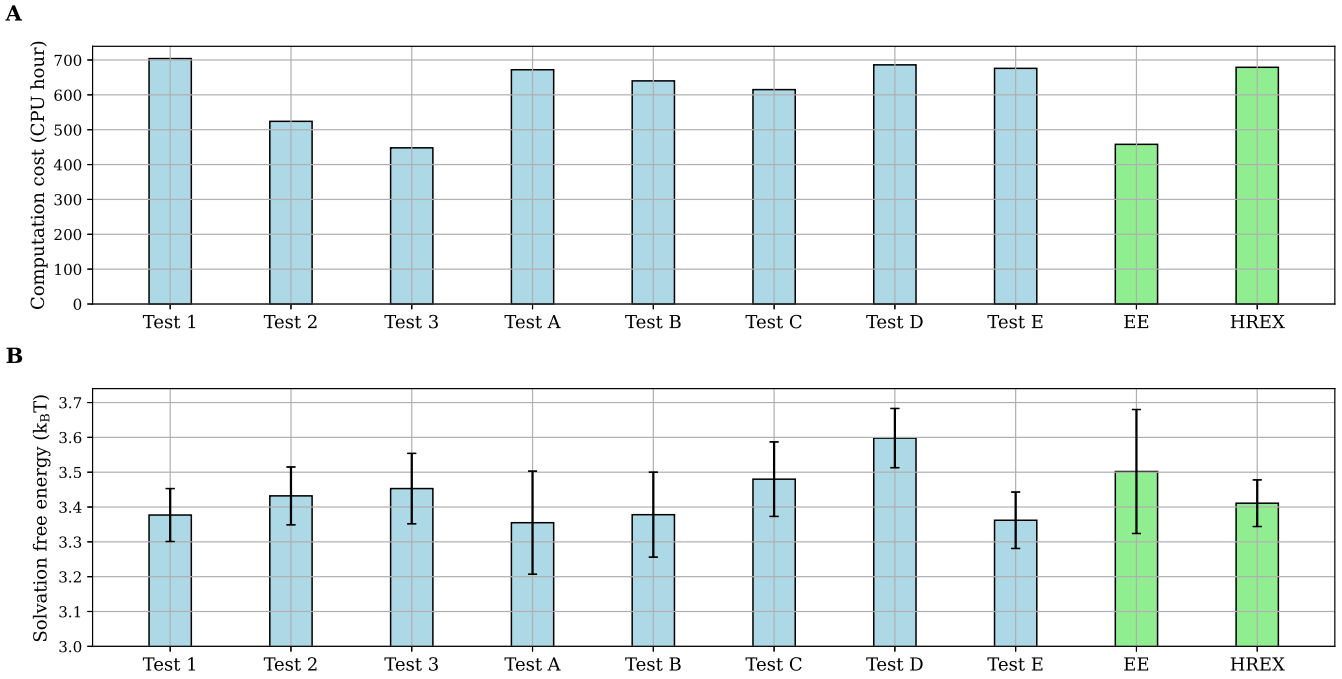


Figure 6. (A) The computation costs (in CPU hours) of fixed-weight REXEE, fixed-weight EE, and HREX simulations. Tests 1 to 3 from Group 1, which respectively adopted an exchange period of 4, 10, and 100 ps, exhibit decreasing computational costs due to less frequent simulation initialization. Additionally, Tests A to E in Group 2, which adopt the same exchange period (4 ps) but different REXEE configurations, incurred similar computational costs, with some stochastic variability. (B) The estimates of the solvation free energy calculated from fixed-weight REXEE, fixed-weight EE, and HREX simulations. All REXEE tests, regardless of the exchange frequency and REXEE configuration, produced free energy estimates consistent with both EE and HREX benchmarks. The error bars of the free energy estimates were statistical errors calculated by the MBAR estimator.

RESULTS AND DISCUSSION

Simulations of Anthracene. *Benchmark Simulations.* With the weights fixed at the values converged by the Wang–Landau algorithm, the solvation free energy of anthracene estimated by the benchmark EE simulation was 3.502 ± 0.178 k_BT, which is statistically consistent with the estimate of 3.411 ± 0.067 k_BT derived from the HREX benchmark simulation with the same effective simulation length. While these two values do not agree within uncertainty to the value reported in the work by Paliwal et al.,[63] this disagreement does not affect our demonstration of the REXEE method, since internal consistency is achieved between the EE and HREX benchmarks and between results from the REXEE simulations and the benchmarks, as we show in the next subsection. We have included a discussion about the potential reasons for the discrepancy between our free energy estimates and the result reported in the work by Paliwal et al. in Section 4.1 in the Supporting Information.

We observed that the EE benchmark simulation showed a shorter state index correlation time and more round trips than the HREX benchmark simulation (see Figure 5). This can be attributed to the fact that EE simulations have provably higher exchange acceptance rates than HREX simulations given the same alchemical path.[79]

Fixed-Weight REXEE Simulations. As hypothesized, faster sampling in either replica or state space can be achieved by both higher exchange frequency and larger overlap ratio, as assessed by metrics including the replica-space relaxation time, state index correlation time, and the total number of round trips. Specifically, panels A, B, and C in Figure 5 compare these three metrics between REXEE simulations with different exchange frequencies. We find that higher exchange frequency correlates with faster replica-space and state-space sampling, as shown by a shorter replica-space relaxation time, shorter state index correlation time, and more round trips in the state space. Notably, the observation of improved sampling efficiency in the state space with faster exchanges is consistent with earlier studies[80,81] that demonstrated enhanced mixing in pure replica exchange simulations with higher exchange frequencies. Interestingly, Test 1, distinguished by the highest exchange frequency, exhibited a shorter state index correlation time than the HREX benchmark simulation and was on par with the EE benchmark. Using the same high exchange frequency, tests in Group 2 also generally show a shorter state index correlation time than the HREX benchmark (see the bars corresponding to Tests A, B, and D in Figure 5E). This shows that REXEE can preserve the advantage of faster mixing as seen in the EE benchmark, which is unsurprising given that the state-space sampling in a REXEE simulation is conducted by the EE replicas. However, the use of these higher exchange frequencies demanded more computational resources, as can be seen from the comparison between Tests 1 to 3 in Group 1 in Figure 6A. This is a natural outcome of requiring more iterations to reach the simulation length equivalent to the other tests, which necessitates frequent simulation initialization and results in a longer total GROMACS start time. Still, these additional overheads are within a reasonable range as they did not make REXEE simulations more expensive than the HREX benchmark simulation.

As mentioned, tests in Group 2 propose an exchange every 4 ps, a frequency shown in Group 1 that allowed for rapid alchemical sampling without introducing an excessive computational cost relative to the HREX benchmark

https://doi.org/10.1021/acs.jctc.4c00484
J. Chem. Theory Comput. 2024, 20, 6062−6081

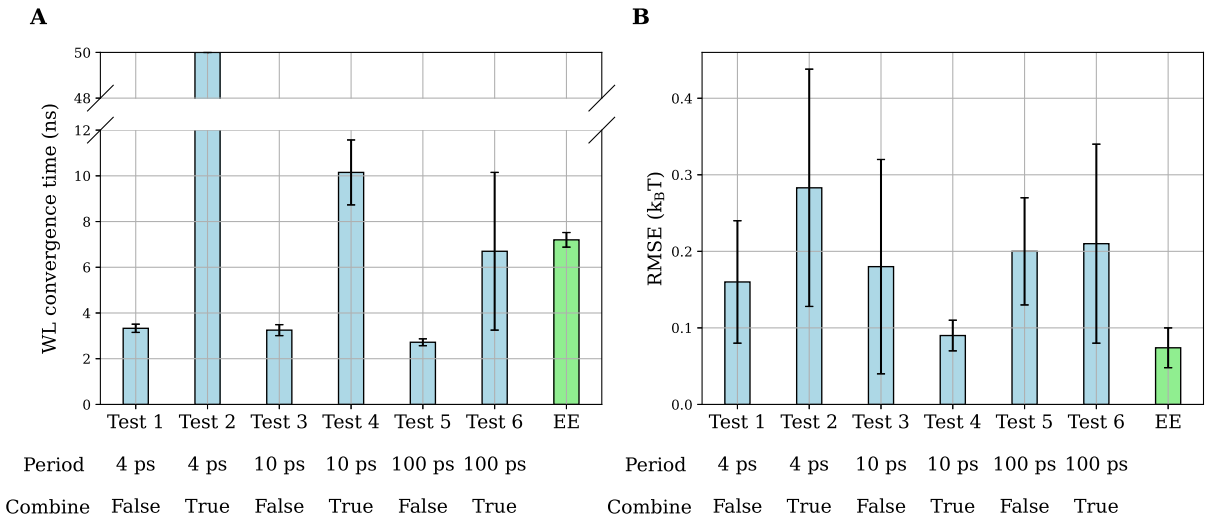


Figure 7. (A) The Wang−Landau (WL) convergence time for all weight-updating REXEE simulations and the weight-updating EE simulation. For Test 2, 50 ns is reported because the simulation did not converge weights for all replicas within the scheduled length of 50 ns per replica. (B) The RMSE relative to the reference free energy profile for all weight-updating REXEE simulations and the weight-updating EE simulation. For both panels, parameters differing between tests are annotated below the *x*-axis for easier comparisons, including the exchange period and whether the weight combination was used. The reported values are averages over 3 replicas, with the uncertainty being the standard deviation. The figure shows that using weight combination schemes did not shorten the WL convergence time nor effectively lower the RMSE value. Additionally, it shows that the exchange frequency has a limited effect on both the WL convergence time and RMSE value.

simulation. Panel D in Figure 5 reveals a clear trend: given the same number of replicas, increasing the overlap between adjacent state sets accelerates mixing in the replica space, as readily observable in comparisons between Tests A, B, and C and between Tests D and E in panel D. This enhanced replica-space mixing, which we attribute to the more swappable pairs at exchanges given higher state overlap, in turn contributes to the acceleration of the state-space sampling as well (see Figure 5E). Notably, the tests in Group 2, though varied in their REXEE configurations, shared a similar computational cost given the same exchange frequency (see Figure 6A). While there is a noticeable trend that tests with higher overlap incurred higher computational costs, the difference is generally marginal.

We infer that the boundary of each state set can create an intrinsic barrier for the system to reach states outside the current state set it is sampling, as observed that all REXEE simulations exhibit fewer round trips than the EE and HREX benchmark simulations (see panels C and F in Figure 5). While a higher exchange frequency helps diminish the barrier by increasing the flux across the boundaries (see Figure 5C for Tests 1 to 3 in Group 1), different extents of overlap have a limited effect (see Figure 5F). The importance of increasing the number of round trips usually lies in the pursuit of enhanced configurational sampling, especially when the fully coupled and decoupled states favor different metastable states in the configurational space, such as the bound and unbound state of a binding complex. Since the anthracene system does not have long-lasting configurational metastable states, the number of round trips does not directly influence the accuracy of free energy calculations. In fact, Figure 6B confirms the robustness of the REXEE method, as all REXEE simulations provided estimates of the solvation free energy of anthracene statistically consistent with both the EE and HREX benchmarks. Interestingly, the trend exhibited by Tests 1, 2, and 3 suggests that a faster exchange frequency results in lower uncertainty in free energy calculations. However, from Tests A to E there is no discernible trend of how the REXEE configuration (i.e., amount of overlap or number of replicas with fixed total simulation time) influences the accuracy of the solvation free energy estimate.

Weight-Updating REXEE Simulations. To assess the ability of the REXEE method in weight convergence, we compare different weight-updating REXEE simulations with weight-updating EE simulations in terms of the WL convergence time and the RMSE relative to the reference free energy profile.

Surprisingly, we find that simulations that did not apply weight combination had a WL convergence time of 2 to 3 ns, while their counterparts that combined weights across replicas had significantly longer weight-converging processes, as shown in Figure 7A, with Test 2 notably failing to reach the criteria of the Wang−Landau algorithm for weight convergence within the scheduled simulation length of 50 ns. Taking this with the observation from Figure 7B that weight combination did not reduce the RMSE values of the free energies, we conclude that weight combinations based on simple averages did not introduce any advantage in weight convergence. We additionally tested more complicated weight combination schemes with different exchange frequencies, including weight combinations based on inverse-variance weighted means (using eqs 13 and 14) and weight combinations with histogram corrections or weight corrections, but none of them outperformed simulations that did not use any weight combination scheme in terms of both WL convergence time and RMSE values.

We reason that this surprising result that combining weights does not improve convergence is because the exchanges of coordinates between replicas have already caused each replica to visit all of the configurations that the different replicas were initialized with, and thus have "seen" the different configurations and incorporated them into the accumulated weights. Therefore, additionally combining weights across replicas may not provide any additional advantage. This is distinct from multiple walkers metadynamics,[82] where each "walker" only

https://doi.org/10.1021/acs.jctc.4c00484
J. Chem. Theory Comput. 2024, 20, 6062−6081

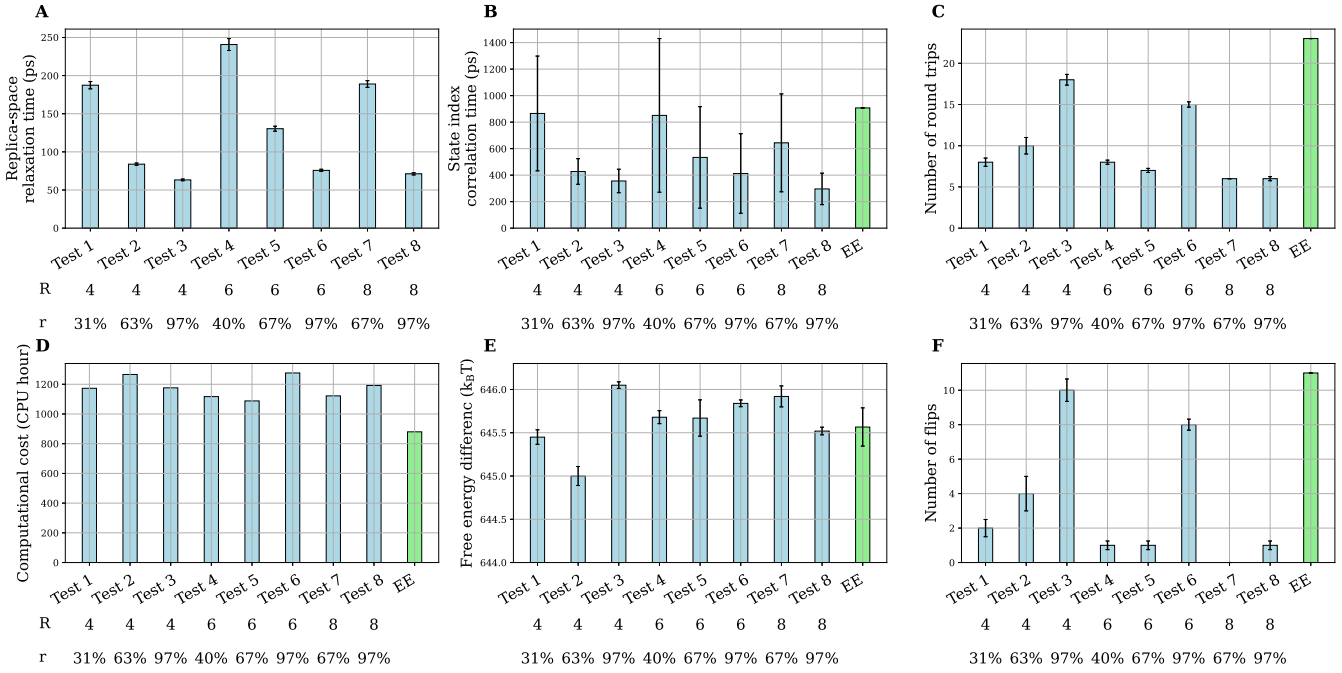


Figure 8. Results from fixed-weight REXEE simulations of the CB7−10 binding complex, with the number of replicas (R) and the overlap ratio (r) annotated. Metrics include the (A) replica-space relaxation time, (B) state index correlation time, (C) total number of round trips in the alchemical space, (D) free energy difference $\Delta G_{D\to F}$, (E) computational cost, and (F) the number of flips between the two largest clusters of the binding complex. Results from the benchmark simulations are colored in light green. For each metric, the uncertainty was estimated using the same method used for the anthracene simulations, except that the uncertainty of the free energy benchmark from the EE simulation was calculated as the standard deviation over 3 replicates, instead of the statistical error calculated by the MBAR estimator that can occasionally be an underestimate. Overall, the trends shown in each metric of sampling are consistent with those observed in the anthracene simulations, e.g., there exists a positive correlation between the overlap ratio and state-space/replica-space sampling.

samples a single configurational ensemble, and thus can benefit from weight information from alternate configurations.

One possible reason for the poor results using weight combinations is that small changes in weights can drastically affect sampling, as state probabilities are exponential in the free energy differences between states. If one of the weights being combined is particularly bad, it will disrupt sampling for the other weights, and will therefore lower the convergence rate. This may also explain the trend in WL convergence time revealed in the comparison between Tests 2, 4, and 6 in Figure 7A, where a higher exchange frequency led to more frequent "contamination" with any poorly converged weights and thus harmed the overall weight convergence. This is in contrast to the negligible impact of the exchange frequency on the WL convergence time as revealed by the comparison between Tests 1, 3, and 5, where weight combination was not used, over all the different exchange frequencies employed. Notably, although these three REXEE simulations achieved weight convergence in shorter wall times compared to weight-updating EE simulations, they incurred slightly higher computational costs due to running four replicas, as opposed to the single replica performed in the EE simulations. Lastly, Figure 7B shows that the weights of all tests converged to values statistically consistent with those obtained from weight-updating EE simulations. Overall, for anthracene, it appears that the performance of weight convergence is relatively independent of REXEE parameters as long as no weight combination scheme is applied.

Simulations of the CB7−10 Host−Guest Binding Complex. *Benchmark Simulations.* The benchmark EE simulation, which is 200 ns in length and had alchemical weights fixed at the values obtained from a weight-updating simulation, estimated the free energy difference $\Delta G_{D\to F} = \Delta G^{complex}_{(restr)_{off}} + \Delta G^{complex}_{(coul+vdW)_{off}}$ as -645.773 ± 0.050 k_BT. Additionally, the free energy difference $\Delta G_{A\to B} = \Delta G^{guest}_{(coul+vdW)_{off}}$ was estimated as 622.180 ± 0.050 k_BT by the solvent simulation. With the correction term $\Delta G_{C\to D} = \Delta G^{complex}_{(restr)_{on}}$ calculated as 2.260 k_BT using eq 16, this leads to an estimate of the binding free energy ΔG°_{bind} as -21.33 ± 0.07 k_BT. This value is statistically inconsistent with the binding free estimate reported in the work by Monroe et al.[49] (-13.99 ± 0.54 k_BT), mainly due to differences between GROMACS versions and that an incorrect correction term $\Delta G_{C\to D}$ was used in the reference work. A more detailed discussion about the disagreement between our estimate and the value reported in the reference work is included in Section 4.2 in the Supporting Information. Again, we emphasize that this discrepancy does not affect the demonstration of the REXEE method since internal consistency is shown between estimates from the REXEE simulations and the benchmark carried out with EE with the same code versions, as discussed in the next subsection.

Fixed-Weight REXEE Simulations. In panels A, B, and C in Figure 8, we present the results of the three metrics assessing the sampling efficiency in the alchemical space, including the replica-space relaxation time, state index correlation time, and the total number of round trips. Overall, the exhibited trend is consistent with the observation from the anthracene REXEE simulations that a higher overlap ratio would lead to a shorter replica-space relaxation time (i.e., faster replica-space sam-

https://doi.org/10.1021/acs.jctc.4c00484
J. Chem. Theory Comput. 2024, 20, 6062−6081

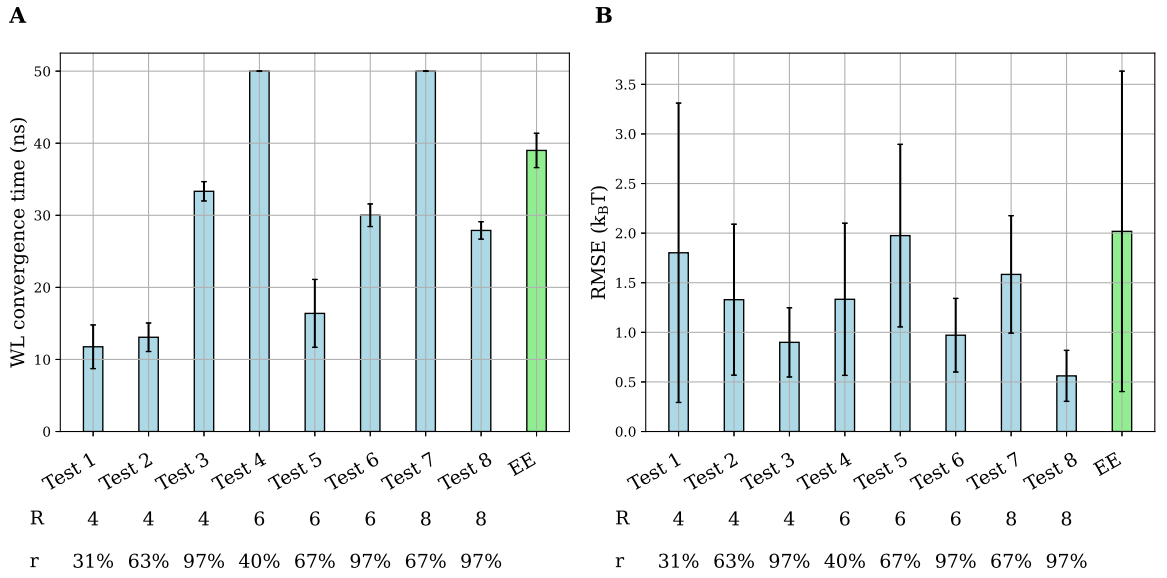


Figure 9. (A) The Wang−Landau (WL) convergence time for all weight-updating REXEE and EE simulations, with the number of replicas (R) and the overlap ratio (r) annotated. For Tests 4 and 7, 50 ns is reported because the both simulations did not converge the weights for all their replicas when reaching the scheduled length of 50 ns per replica. (B) The RMSE relative to the reference free energy profile for all weight-updating REXEE simulations and the weight-updating EE simulation. For both panels, the reported values are averages over 3 replicates, with the uncertainty being the standard deviation. The figure shows that while simulations with different overlap ratios may vary in the WL convergence time, they achieve similar RMSE values, suggesting little impact of the state set overlap on the quality of the converged weights.

pling) and a shorter state index correlation time (i.e., faster state-space sampling) and that the REXEE simulations have fewer round trips than the EE benchmark simulation. Interestingly, some tests (Tests 2, 3, 6, and 8) even show statistically shorter state index correlation time than the EE benchmark simulation (see Figure 8B), which might be attributable to the fact that the exchange period (4 ps) is much shorter than the intrinsic correlation time in the configuration space. Another trend in the CB7−10 simulations consistent with that for the anthracene system is that all REXEE simulations, which utilized the same exchange frequency but different REXEE configurations, incurred the same level of computational cost, with a reasonable overhead introduced by frequent exchanges of coordinates as compared to the benchmark EE simulation (see Figure 8D).

For free energy calculations, there was some indication that the uncertainty of the free energy estimate from the EE benchmark simulation might be underestimated. This conclusion is based on observations from two additional replicates of the benchmark simulation. Specifically, we found that the standard deviation of the three replicates, 0.27 k_BT, was much higher than the statistical error of 0.05 k_BT computed by the MBAR estimator for each replicate. This discrepancy likely lies in the fact that the data decorrelation method[66] occasionally underestimated the statistical inefficiency/overestimated the number of effective samples in CB7−10 trajectories. To provide a more accurate baseline for comparison with the REXEE simulations, we report the average and standard deviation over the 3 replicates of the EE benchmark simulation in Figure 8E, which has a value of −645.57 ± 0.27 k_BT. This refined benchmark leads to a binding free energy estimate of −21.13 ± 0.27 k_BT, which is consistent with the estimate considering only the first replicate as reported in the previous section, i.e., − 21.33 ± 0.07 k_BT. Upon comparison with the refined benchmark, we find that free energy estimates $\Delta G_{D\to F}$ from all REXEE simulations are statistically consistent with the benchmark, regardless of the replica configuration. However, we note that since the issue of underestimating the statistical inefficiency in data decorrelation also occurred in REXEE trajectories, the lower uncertainties observed in the REXEE simulations compared to the EE benchmark do not necessarily provide sufficient evidence of the superior accuracy of the REXEE method in free energy calculations.

Finally, Figure 8F shows that REXEE simulations exhibited fewer transitions between the two dominant clusters corresponding to configurations of the ligand bound to different (symmetrical) sides of the host ring molecule as compared to the EE benchmark. The reason for this difference was further explored in Figure S2 in the Supporting Information, where an almost complete correlation was observed between the number of flips and the number of round trips in the alchemical space, with a Kendall's tau correlation coefficient of 0.93. This correlation comes from the fact that the guest molecule can unbind from and rebind to the host molecule more easily in the fully decoupled state, so more switches between the two alchemical end states bring more opportunities for more configurational flips, indicating the correlation between the alchemical and configurational degrees of freedom of the binding complex. Importantly, this correlation suggests that the previously identified inherent barrier to sampling across state-set boundaries, which accounts for the reduced number of round trips in the alchemical space in REXEE simulations, is likely also responsible for the decreased number of transitions between the two configurational clusters. Thus, REXEE configurations that have increased motion in state space, such as higher exchange rates and greater overlap, should generally be preferred. However, the variability in the number of flips between different REXEE configurations, and the significant uncertainty observed in the number of flips in additional replicates for

https://doi.org/10.1021/acs.jctc.4c00484
J. Chem. Theory Comput. 2024, 20, 6062−6081

some tests require further investigation, which is beyond the scope of the paper.

Weight-Updating REXEE Simulations. For challenging systems like binding complexes, the quality of the alchemical weights that a weight-updating simulation converges to is of importance, as the weights determine the flatness of the biased free energy surface that seeds a subsequent fixed-weight simulation. In Figure 9, we plotted the WL convergence times and the RMSE values relative to the reference free energy profile for all weight-updating REXEE and EE simulations of CB7−10. As a result, most tests achieved weight convergence for all replicas within 35 ns, with Tests 4 and 7 as notable exceptions, which failed to meet convergence criteria within the scheduled 50 ns simulation time frame. Notably, those who converged within a shorter time than the weight-updating EE simulation still consumed higher computational costs, as synchronous REXEE simulations require all replicas to continue running even if some of them have converged the weights for their respective state sets. Despite these higher computational costs, Figure 9B indicates that some REXEE tests could reach lower RMSE values than weight-updating EE simulations. Interestingly, while Tests 1 and 2 have statistically shorter WL convergence time than Test 3, they ended up converging to weights that have a larger deviation from the reference free energy profile. This indicates that the known issue of the Wang−Landau algorithm of getting "burned in" to inaccurate free energy estimates can still persist in weight-updating REXEE simulations, and a prolonged WL convergence time may not necessarily represent a negative outcome, especially considering the pursuit of lower RMSE values for a better starting point in subsequent fixed-weight simulations. This observation also resonates with the fact that Tests 4 and 7 reached an RMSE value on par with others in spite of their inability to converge as fast as other tests. Lastly, the comparable RMSE values across all tests regardless of REXEE configurations suggest that for complex systems like CB7−10, selecting the number of replicas based on available computational resources, with a preference for an intermediate to high overlap ratio, could be a viable strategy.

CONCLUSIONS

In this study, we proposed the method of replica exchange of expanded ensembles (REXEE), a novel generalized ensemble method that integrates the working principles of the methods of Hamiltonian replica exchange and expanded ensemble. The REXEE method decouples the number of states from the number of replicas in its algorithmic design, escaping the one-to-one correspondence inherent to the HREX method and therefore offering much greater flexibility in parameter specification—REXEE allows the number of states needed for simulation to be chosen independently of the number of simulations possible on available computational resources. In addition, the parallelizability of the REXEE method opens the door to wider applications compared to the EE method, such as one-shot free energy calculations in multitopology contexts, e.g., serial mutations and scaffold-hopping transformations. All necessary algorithms for running and analyzing REXEE simulations have been encapsulated in the Python package ensemble_md, which is a flexible and easily extensible wrapper that not only eliminates the need to modify the source code of GROMACS, but also provides convenient automation and management of REXEE simulations.

With this implementation, we evaluated the effectiveness of the REXEE method with various setups, focusing on the calculation of the solvation free energy of anthracene and the binding free energy of the CB7−10 complex. As a result, we confirmed that the REXEE method did not compromise its ability to accurately compute free energy differences while offering enhanced flexibility and parallelizability. Accordingly, REXEE may be particularly useful in scenarios where sampling a large number of states is desired, e.g., in high-dimensional grids defined in an extended ensemble, as higher flexibility in replica configuration and shorter simulation wall times are generally more favorable in these scenarios. In addition, we observed that increasing the exchange frequency or overlap ratio in fixed-weight REXEE simulations enhances sampling/mixing in both the replica and state spaces without introducing exorbitant overhead as compared to EE and HREX simulations.

In weight-updating simulations, we demonstrated that the exchange frequency and REXEE configuration have little impact on the WL convergence time and the RMSE value relative to the reference free energy profile. Moreover, we concluded that weight combination schemes could actually introduce additional errors in the alchemical weights, leading to longer WL convergence times without necessarily improving the accuracy of the weights. Lastly, while weight-updating REXEE simulations may incur slightly higher computational costs than weight-updating EE simulations due to the overhead of exchanges, we showed that they generally exhibited lower RMSE values, which indicated more accurate converged weights. This improved convergence may be facilitated by spawning the different weight-updating EE replica simulations with different initial configurations to better consider configurational heterogeneity, promising a better starting point to seed subsequent fixed-weight simulations.

Importantly, the current REXEE implementation could be further enhanced by incorporating asynchronous parallelization schemes, which permit less constrained communications between a larger count of loosely coupled processors, making it more adaptable to environments like cloud computing, compared to the currently implemented synchronous REXEE method. In the weight-updating phase, asynchronous REXEE can save computational power by dropping replicas that have converged the corresponding alchemical weights. This contrasts with synchronous parallelization schemes, where the termination of a single replica halts the entire simulation ensemble. Moreover, asynchronous REXEE provides an elevated level of flexibility in parameter specification compared to synchronous REXEE. It not only accommodates heterogeneous parameter configurations, such as different numbers of states per replica or state shifts for different replicas, but also allows for adaptive changes to the parameters in response to data collected, such as changing the number of states per replica or even the number of replicas to optimize the simulation's performance. Lastly, the development of the package ensemble_md and the foundational principles of the REXEE method can be integrated with our recent endeavors in alchemical metadynamics by having multiple walkers sample the joint space of alchemical variable and configurational collective variables, with each walker exploring different alchemical state sets. This can potentially result in the development of new simulation approaches like the ensemble of alchemical metadynamics, promising enhanced flexibility,

https://doi.org/10.1021/acs.jctc.4c00484
J. Chem. Theory Comput. 2024, 20, 6062−6081

parallelizability, and configurational sampling for a wider range of systems.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.4c00484.

Theory for enumerating the space of configurations of REXEE simulations, accompanied by additional figures demonstrating the independence of the number of replicas and the number of states in REXEE. Derivation of the acceptance ratio for swapping EE replicas in a REXEE simulation. Derivation of the analytical correction term used in our double decoupling thermodynamic cycle for the binding free energy calculation of CB7−10 binding complex. Detailed comparisons between our free energy estimates and references reported in previous studies, including the solvation free energy of anthracene and binding free energy of CB7−10 binding complex. A supplementary figure showing the strong correlation between the number of flips between the two dominant configurational clusters of CB7−10 binding complex and the number of found trips in the alchemical space (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Michael R. Shirts − *Department of Chemical and Biological Engineering, University of Colorado Boulder, Boulder, Colorado 80305, United States;* orcid.org/0000-0003-3249-1097; Email: michael.shirts@colorado.edu

Author

Wei-Tse Hsu − *Department of Chemical and Biological Engineering, University of Colorado Boulder, Boulder, Colorado 80305, United States;* orcid.org/0000-0001-6167-5480

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.4c00484

Author Contributions

W.-T.H. and M.R.S. primarily conceptualized the project and designed the methodology. W.-T.H. implemented the sampling method in the Python package ensemble_md. Experiments were performed and analyzed by W.-T.H., with contributions from M.R.S. for validation. W.-T.H wrote the original manuscript draft; editing and review of the manuscript were done by M.R.S. M.R.S. supervised the project and obtained resources.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This study was supported by the grant OAC-1835720 from the National Science Foundation and grant R01GM123296 from the National Institutes of General Medical Sciences. The computational work done in this publication used resources provided from the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1548562. Specifically, it used the Bridges-2 system, which is supported by NSF ACI-1928147, located at the Pittsburgh Supercomputing Center (PSC). We thank Anika Friedman for proofreading and comments on the manuscript.

■ REFERENCES

(1) Padua, D. *Encyclopedia of parallel computing*; Springer Science & Business Media, 2011.

(2) Hénin, J.; Lelièvre, T.; Shirts, M. R.; Valsson, O.; Delemotte, L. Enhanced sampling methods for molecular dynamics simulations. *Living J. Comp. Mol. Sci.* **2022**, *4*, 1583.

(3) Torrie, G. M.; Valleau, J. P. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.* **1977**, *23*, 187−199.

(4) Laio, A.; Parrinello, M. Escaping free-energy minima. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12562−12566.

(5) Darve, E.; Rodríguez-Gómez, D.; Pohorille, A. Adaptive biasing force method for scalar and vector free energy calculations. *J. Chem. Phys.* **2008**, *128*, 144120.

(6) Invernizzi, M. OPES: On-the-fly probability enhanced sampling method. *arXiv preprint* arXiv:2101.06991 2021.

(7) Piana, S.; Laio, A. A bias-exchange approach to protein folding. *J. Phys. Chem. B* **2007**, *111*, 4553−4559.

(8) Invernizzi, M.; Parrinello, M. Exploration vs convergence speed in adaptive-bias enhanced sampling. *J. Chem. Theory Comput.* **2022**, *18*, 3988−3996.

(9) Dama, J. F.; Rotskoff, G.; Parrinello, M.; Voth, G. A. Transition-tempered metadynamics: robust, convergent metadynamics via on-the-fly transition barrier estimation. *J. Chem. Theory Comput.* **2014**, *10*, 3626−3633.

(10) 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.

(11) Geman, S.; Geman, D. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell.* **1984**, *PAMI-6*, 721−741.

(12) Hastings, W. K. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **1970**, *57*, 97.

(13) Barker, A. A. Monte carlo calculations of the radial distribution functions for a proton-electron plasma. *Aust. J. Phys.* **1965**, *18*, 119−134.

(14) Liu, J. S.; Liu, J. S. *Monte Carlo strategies in scientific computing*; Springer, 2001; Vol. *75*.

(15) Liu, J. S. Peskun's theorem and a modified discrete-state Gibbs sampler. *Biometrika* **1996**, *83*, 681.

(16) Marinari, E.; Parisi, G. Simulated tempering: a new Monte Carlo scheme. *EPL* **1992**, *19*, 451.

(17) Sugita, Y.; Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **1999**, *314*, 141−151.

(18) Paschek, D.; García, A. E. Reversible temperature and pressure denaturation of a protein fragment: a replica exchange molecular dynamics simulation study. *Phys. Rev. Lett.* **2004**, *93*, 238105.

(19) Zhou, R. Replica exchange molecular dynamics method for protein folding simulation. *Prot. Fold. Protoc.* **2006**, *350*, 205−223.

(20) 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.

(21) Lyubartsev, A.; Martsinovski, A.; Shevkunov, S.; Vorontsov-Velyaminov, P. New approach to Monte Carlo calculation of the free energy: Method of expanded ensembles. *J. Chem. Phys.* **1992**, *96*, 1776−1783.

(22) Knight, J. L.; Brooks III, C. L. λ-Dynamics free energy simulation methods. *J. Comput. Chem.* **2009**, *30*, 1692−1700.

(23) Knight, J. L.; Brooks III, C. L. Multisite λ dynamics for simulated structure−activity relationship studies. *J. Chem. Theory Comput.* **2011**, *7*, 2728−2739.

(24) Sugita, Y.; Kitao, A.; Okamoto, Y. Multidimensional replica-exchange method for free-energy calculations. *J. Chem. Phys.* **2000**, *113*, 6042−6051.

(25) Mobley, D. L.; Guthrie, J. P. FreeSolv: a database of experimental and calculated hydration free energies, with input files. *J. Comput. Aided Mol. Des.* **2014**, *28*, 711−720.
(26) Mobley, D. L.; Liu, S.; Cerutti, D. S.; Swope, W. C.; Rice, J. E. Alchemical prediction of hydration free energies for SAMPL. *J. Comput. Aided Mol. Des.* **2012**, *26*, 551−562.
(27) Scheen, J.; Wu, W.; Mey, A. S.; Tosco, P.; Mackey, M.; Michel, J. Hybrid alchemical free energy/machine-learning methodology for the computation of hydration free energies. *J. Chem. Inf. Model.* **2020**, *60*, 5331−5339.
(28) Khalak, Y.; Tresadern, G.; Aldeghi, M.; Baumann, H. M.; Mobley, D. L.; de Groot, B. L.; Gapsys, V. Alchemical absolute protein−ligand binding free energies for drug design. *Chem. Sci.* **2021**, *12*, 13958−13971.
(29) Lee, T.-S.; Allen, B. K.; Giese, T. J.; Guo, Z.; Li, P.; Lin, C.; McGee Jr, T. D.; Pearlman, D. A.; Radak, B. K.; Tao, Y.; et al. Alchemical binding free energy calculations in AMBER20: Advances and best practices for drug discovery. *J. Chem. Inf. Model.* **2020**, *60*, 5595−5623.
(30) Abel, R.; Wang, L.; Mobley, D. L.; Friesner, R. A. A critical review of validation, blind testing, and real-world use of alchemical protein-ligand binding free energy calculations. *Curr. Top. Med. Chem.* **2017**, *17*, 2577−2585.
(31) Piomponi, V.; Fröhlking, T.; Bernetti, M.; Bussi, G. Molecular simulations matching denaturation experiments for N6-Methyladenosine. *ACS Cent. Sci.* **2022**, *8*, 1218−1228.
(32) Hayes, R. L.; Brooks III, C. L. A strategy for proline and glycine mutations to proteins with alchemical free energy calculations. *J. Comput. Chem.* **2021**, *42*, 1088−1094.
(33) Bergonzo, C.; Henriksen, N. M.; Roe, D. R.; Swails, J. M.; Roitberg, A. E.; Cheatham, T. E., III Multidimensional replica exchange molecular dynamics yields a converged ensemble of an RNA tetranucleotide. *J. Chem. Theory Comput.* **2014**, *10*, 492−499.
(34) 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.
(35) Ebrahimi, P.; Kaur, S.; Baronti, L.; Petzold, K.; Chen, A. A. A two-dimensional replica-exchange molecular dynamics method for simulating RNA folding using sparse experimental restraints. *Methods* **2019**, *162*, 96−107.
(36) Cruzeiro, V. W. D.; Roitberg, A. E. Multidimensional replica exchange simulations for efficient constant pH and redox potential molecular dynamics. *J. Chem. Theory Comput.* **2019**, *15*, 871−881.
(37) Wang, F.; Landau, D. P. Efficient, multiple-range random walk algorithm to calculate the density of states. *Phys. Rev. Lett.* **2001**, *86*, 2050.
(38) Belardinelli, R.; Pereyra, V. Fast algorithm to calculate density of states. *Phys. Rev. E* **2007**, *75*, 046701.
(39) Belardinelli, R.; Pereyra, V. Wang-Landau algorithm: A theoretical analysis of the saturation of the error. *J. Chem. Phys.* **2007**, *127*, 184105.
(40) Lidmar, J. Improving the efficiency of extended ensemble simulations: The accelerated weight histogram method. *Phys. Rev. E* **2012**, *85*, 056708.
(41) Lindahl, V.; Lidmar, J.; Hess, B. Accelerated weight histogram method for exploring free energy landscapes. *J. Chem. Phys.* **2014**, *141*, 044110.
(42) Lundborg, M.; Lidmar, J.; Hess, B. The accelerated weight histogram method for alchemical free energy calculations. *J. Chem. Phys.* **2021**, *154*, 204103.
(43) Tan, Z. Optimally adjusted mixture sampling and locally weighted histogram analysis. *J. Comput. Graph. Stat.* **2017**, *26*, 54−65.
(44) Hsu, W.-T.; Piomponi, V.; Merz, P. T.; Bussi, G.; Shirts, M. R. Alchemical metadynamics: Adding alchemical variables to metadynamics to enhance sampling in free energy calculations. *J. Chem. Theory Comput.* **2023**, *19*, 1805−1817.
(45) Kirkwood, J. G. Statistical mechanics of fluid mixtures. *J. Chem. Phys.* **1935**, *3*, 300−313.
(46) Bennett, C. H. Efficient estimation of free energy differences from Monte Carlo data. *J. Comput. Phys.* **1976**, *22*, 245−268.
(47) Shirts, M. R.; Chodera, J. D. Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.* **2008**, *129*, 124105.
(48) Escobedo, F. A.; Martinez-Veracoechea, F. J. Optimization of expanded ensemble methods. *J. Chem. Phys.* **2008**, *129*, 154107.
(49) Monroe, J. I.; Shirts, M. R. Converging free energies of binding in cucurbit [7] uril and octa-acid host−guest systems from SAMPL4 using expanded ensemble simulations. *J. Comput. Aided Mol. Des.* **2014**, *28*, 401−415.
(50) Rizzi, A.; Jensen, T.; Slochower, D. R.; Aldeghi, M.; Gapsys, V.; Ntekoumes, D.; Bosisio, S.; Papadourakis, M.; Henriksen, N. M.; De Groot, B. L.; et al. The SAMPL6 SAMPLing challenge: assessing the reliability and efficiency of binding free energy calculations. *J. Comput. Aided Mol. Des.* **2020**, *34*, 601−633.
(51) Muddana, H. S.; Fenley, A. T.; Mobley, D. L.; Gilson, M. K. The SAMPL4 host−guest blind prediction challenge: an overview. *J. Comput. Aided Mol. Des.* **2014**, *28*, 305−317.
(52) Bussi, G.; Gervasio, F. L.; Laio, A.; Parrinello, M. Free-energy landscape for β hairpin folding from combined parallel tempering and metadynamics. *J. Am. Chem. Soc.* **2006**, *128*, 13435−13441.
(53) Laio, A.; Parrinello, M. Escaping free-energy minima. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12562−12566.
(54) Barducci, A.; Bonomi, M.; Parrinello, M. Metadynamics. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2011**, *1*, 826−843.
(55) Paz, S. A.; Vanden-Eijnden, E.; Abrams, C. F. Polymorphism at 129 dictates metastable conformations of the human prion protein N-terminal β-sheet. *Chem. Sci.* **2017**, *8*, 1225−1232.
(56) Abrams, C. F.; Vanden-Eijnden, E. Large-scale conformational sampling of proteins using temperature-accelerated molecular dynamics. *Biophys. J.* **2010**, *98*, 26a.
(57) Gallicchio, E.; Levy, R. M.; Parashar, M. Asynchronous replica exchange for molecular simulations. *J. Comput. Chem.* **2008**, *29*, 788−794.
(58) Gallicchio, E.; Xia, J.; Flynn, W. F.; Zhang, B.; Samlalsingh, S.; Mentes, A.; Levy, R. M. Asynchronous replica exchange software for grid and heterogeneous computing. *Comput. Phys. Commun.* **2015**, *196*, 236−246.
(59) Xia, J.; Flynn, W. F.; Gallicchio, E.; Zhang, B. W.; He, P.; Tan, Z.; Levy, R. M. Large-scale asynchronous and distributed multi-dimensional replica exchange molecular simulations and efficiency analysis. *J. Comput. Chem.* **2015**, *36*, 1772−1785.
(60) Radak, B. K.; Romanus, M.; Gallicchio, E.; Lee, T.-S.; Weidner, O.; Deng, N.-J.; He, P.; Dai, W.; York, D. M.; Levy, R. M., et al. A framework for flexible and scalable replica-exchange on production distributed CI. *Proceedings of the Conference on Extreme Science and Engineering Discovery Environment: Gateway to Discovery*. 2013; pp 1−8.
(61) Lockhart, C.; O'Connor, J.; Armentrout, S.; Klimov, D. K. Greedy replica exchange algorithm for heterogeneous computing grids. *J. Mol. Model.* **2015**, *21*, 243.
(62) Rhee, Y. M.; Pande, V. S. Multiplexed-replica exchange molecular dynamics method for protein folding simulation. *Biophys. J.* **2003**, *84*, 775−786.
(63) Paliwal, H.; Shirts, M. R. A benchmark test set for alchemical free energy transformations and its use to quantify error in common free energy methods. *J. Chem. Theory Comput.* **2011**, *7*, 4115−4134.
(64) Hess, B.; Kutzner, C.; Van Der Spoel, D.; Lindahl, E. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435−447.
(65) Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; Van Der Spoel, D.; et al. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **2013**, *29*, 845−854.

https://doi.org/10.1021/acs.jctc.4c00484
J. Chem. Theory Comput. 2024, 20, 6062−6081

(66) Chodera, J. D. A simple method for automated equilibration detection in molecular simulations. *J. Chem. Theory Comput.* **2016**, *12*, 1799−1805.
(67) Husic, B. E.; Pande, V. S. Markov state models: From an art to a science. *J. Am. Chem. Soc.* **2018**, *140*, 2386−2396.
(68) Bowman, G. R.; Pande, V. S.; Noé, F. *An introduction to Markov state models and their application to long timescale molecular simulation*; Springer Science & Business Media, 2013; Vol. *797*.
(69) Scherer, M. K.; Trendelkamp-Schroer, B.; Paul, F.; Pérez-Hernández, G.; Hoffmann, M.; Plattner, N.; Wehmeyer, C.; Prinz, J.-H.; Noé, F. PyEMMA 2: A software package for estimation, validation, and analysis of Markov models. *J. Chem. Theory Comput.* **2015**, *11*, 5525−5542.
(70) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.
(71) Berendsen, H. J.; Postma, J. v.; Van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684−3690.
(72) Parrinello, M.; Rahman, A. Crystal structure and pair potentials: A molecular-dynamics study. *Phys. Rev. Lett.* **1980**, *45*, 1196.
(73) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182−7190.
(74) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577−8593.
(75) Hess, B.; Bekker, H.; Berendsen, H. J.; Fraaije, J. G. LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997**, *18*, 1463−1472.
(76) Beutler, T. C.; Mark, A. E.; van Schaik, R. C.; Gerber, P. R.; Van Gunsteren, W. F. Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations. *Chem. Phys. Lett.* **1994**, *222*, 529−539.
(77) Iorga, B. I.; Mobley, D. L. *samplchallenges/SAMPL4: Version 1.0: Historical data*. 2021; DOI: 10.5281/zenodo.5508284, Version 1.0: Historical data.
(78) Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. The statistical-thermodynamic basis for computation of binding affinities: a critical review. *Biophys. J.* **1997**, *72*, 1047−1069.
(79) Park, S. Comparison of the serial and parallel algorithms of generalized ensemble simulations: An analytical approach. *Phys. Rev. E* **2008**, *77*, 016709.
(80) Sindhikara, D.; Meng, Y.; Roitberg, A. E. Exchange frequency in replica exchange molecular dynamics. *J. Chem. Phys.* **2008**, *128*, 024103.
(81) 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.
(82) Raiteri, P.; Laio, A.; Gervasio, F. L.; Micheletti, C.; Parrinello, M. Efficient reconstruction of complex free energy landscapes by multiple walkers metadynamics. *J. Phys. Chem. B* **2006**, *110*, 3533−3539.

https://doi.org/10.1021/acs.jctc.4c00484
J. Chem. Theory Comput. 2024, 20, 6062−6081