JCTC Journal of Chemical Theory and Computation

pubs.acs.org/JCTC Article

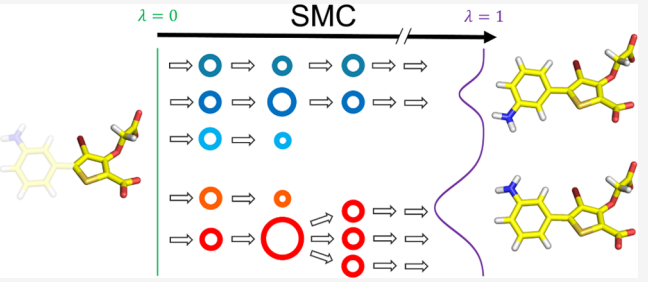
Enhancing Ligand and Protein Sampling Using Sequential Monte Carlo

Miroslav Suruzhon, Michael S. Bodnarchuk, Antonella Ciancetta, Ian D. Wall, and Jonathan W. Essex*

Cite This: J. Chem. Theory Comput. 2022, 18, 3894−3910 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The sampling problem is one of the most widely studied topics in computational chemistry. While various methods exist for sampling along a set of reaction coordinates, many require system-dependent hyperparameters to achieve maximum efficiency. In this work, we present an alchemical variation of adaptive sequential Monte Carlo (SMC), an irreversible importance resampling method that is part of a well-studied class of methods that have been used in various applications but have been underexplored in computational biophysics. Afterward, we apply alchemical SMC on a variety of test cases, including torsional rotations of solvated ligands (butene and a terphenyl derivative), translational and rotational movements of protein-bound ligands, and protein side chain rotation coupled to the ligand degrees of freedom (T4-lysozyme, protein tyrosine phosphatase 1B, and transforming growth factor β). We find that alchemical SMC is an efficient way to explore targeted degrees of freedom and can be applied to a variety of systems using the same hyperparameters to achieve a similar performance. Alchemical SMC is a promising tool for preparatory exploration of systems where long-timescale sampling of the entire system can be traded off against short-timescale sampling of a particular set of degrees of freedom over a population of conformers.



1. INTRODUCTION

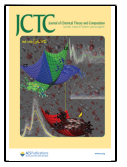
The sampling problem presents one of the biggest challenges in the field of classical computational chemistry, particularly biomolecular simulation.[1,2] Since current computational power is insufficient for studying statistical mechanical problems of systems with more than 10,000 atoms at the relevant millisecond to minute timescales, enhanced sampling methods have been an indispensable tool in the computational chemist's arsenal, trading dynamic detail for the sampling of relevant rare transitions at short timescales.

More generally, the enhanced sampling problem can be referred to as the multimodal global sampling problem, that is, sampling from a probability distribution with multiple relevant modes (i.e., highly populated states), which are highly disconnected and whose locations are generally not known a priori. In a typical physical application, multimodality manifests itself through high kinetic barriers, where the probability of surmounting them decreases exponentially with their heights. As a result, many such transitions are practically impossible at the currently achievable computational timescales. Enhanced sampling research encompasses two key components: the optimal choice of important degrees of freedom (collective variables, CVs), which is system-dependent, and the method used to sample these degrees of freedom. There are many such enhanced sampling methods,[1] some of the most widely known being replica exchange molecular dynamics (REMD),[3−6] metadynamics,[7−9] and umbrella sampling.[10]

A common challenge for enhanced sampling is the need for some prior knowledge of the system under study, which manifests itself beyond the need for a relevant CV. For example, REMD and umbrella sampling benefit greatly from an optimal spacing of the intermediate states, and methods involving nonequilibrium switching[11,12] are most efficient when low-variance pathways are used. An obvious requirement for a robust enhanced sampling method is therefore its ability to adaptively tune itself to the system studied, irrespective of the system complexity.

While there has been a considerable body of work in developing adaptive versions of some of the above methods,[9,13−15] here, we shift our focus to another promising alternative—sequential Monte Carlo (SMC).[16,17] With SMC being one of the oldest enhanced sampling algorithms,[16] it has been rediscovered and further developed in many fields, such as statistics,[18] robotics,[19] meteorology,[20] solid-state

Received: November 30, 2021
Published: May 19, 2022



https://doi.org/10.1021/acs.jctc.1c01198
J. Chem. Theory Comput. 2022, 18, 3894−3910

physics,[21,22] and quantum chemistry,[23] often under different names (particle filtering,[24] weighted-ensemble annealing,[25] population annealing,[21] Rosenbluth sampling,[16] configurational bias Monte Carlo,[26] and diffusion quantum Monte Carlo[23]). While SMC has already been used in classical computational chemistry as a way to improve sampling methods utilizing nonequilibrium switching,[27−29] its usage in biomolecular systems has been mostly restricted to polymer growing and protein folding,[30−32] and its use with more sophisticated force field models has been underexplored.

Most relevant to this work are the recent publications by Christiansen et al.,[33,34] where the authors used an adaptive tempered version of SMC to explore peptide conformations using molecular force field models. In this work, we will extend this methodology to an alchemical setting, where instead of uniformly increasing the temperature of the whole system, a small subset of the molecular interactions will be completely decoupled instead. This approach is particularly suitable for exploring specific molecular degrees of freedom of interest and has been utilized in other methods, such as Hamiltonian replica exchange molecular dynamics (H-REMD)[35,36] and nonequilibrium candidate Monte Carlo,[12,37] and is closely related to alchemical free energy (AFE) methods.[38] We will also apply alchemical SMC on a variety of protein−ligand complexes to measure its suitability for handling high-dimensional systems of practical interest.

In the following, we will first present one of the most popular SMC algorithms, sequential importance resampling (SIR).[17] The original version of SIR is not adaptive and conceptually similar to REMD[3,4] and simulated tempering.[39,40] Afterward, we will discuss several modifications to the original method, some of which have been extensively explored in the field of statistics, while others have been derived from physical considerations and nonequilibrium statistical mechanics. These will allow us to apply SMC to practically relevant scenarios. Finally, we will conclude with a variety of test cases, where we will show examples of enhancing torsional angle sampling and ligand binding mode exploration in systems with increasing complexity.

2. FUNDAMENTALS OF SIR

The fundamental assumption behind SIR is that one starts from a distribution that is trivial to sample from (e.g., a uniform distribution). In most practical examples, where the distributions have many correlated dimensions, this is not possible and the initial distribution is chosen so that transitions between a subset of the modes are more likely than in the distribution of interest. Afterward, a population of samples is propagated over a number of intermediate distributions that connect the initial distribution to the final distribution of interest.

The main focus of this work are Boltzmann-like distributions of the form

$$\pi(\lambda, \vec{x}) = e^{-u(\lambda,\vec{x})+f(\lambda)} \tag{1}$$

where $\vec{x}$ are the system coordinates; λ is an adjustable parameter, such that $0 \le \lambda \le 1$; $u(\lambda,\vec{x})$ is the dimensionless potential energy of the system, which can also contain additional terms, such as a pressure−volume term in the case of an isothermal−isobaric ensemble; and $f(\lambda)$ is the dimensionless free energy, which normalizes the distribution. The coupling parameter λ is defined to be 0 at the initial distribution and 1 at the final distribution of interest.

Each SIR iteration consists of three steps (Figure 1): sampling, reweighting, and resampling. Any valid samplers can

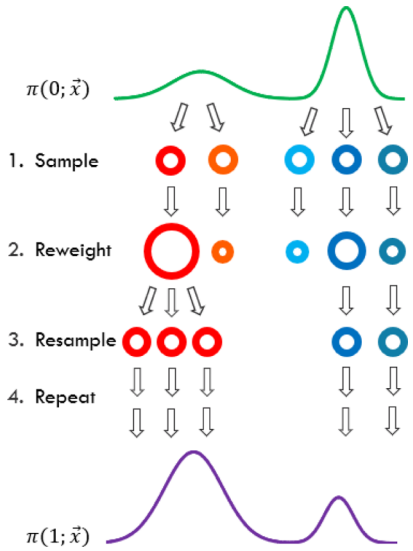


Figure 1. Three stages of each SIR iteration: sampling, reweighting, and resampling. Each unique walker is shown with a different color, and the size of the walker represents its weight. Here, $\pi(0,\vec{x})$ and $\pi(1,\vec{x})$ represent the initial and final distributions, respectively.

be used in the first step, such as Markov chain Monte Carlo (MCMC) or Langevin molecular dynamics (MD), to generate a population of N locally decorrelated samples (walkers). The second step determines the relative transition probability of the j-th walker $p(\lambda_{i+1}|\lambda_i,\vec{x}_j)$ between the current distribution $\pi(\lambda_i,\vec{x})$ and the next distribution in the sequence $\pi(\lambda_{i+1},\vec{x})$, $0 \le \lambda_i < \lambda_{i+1} \le 1$. These relative transition probabilities are normalized and converted into importance sampling weights $w_j(\lambda_{i+1}|\lambda_i)$, which are then assigned to each walker

$$p(\lambda_{i+1}|\lambda_i, \vec{x}_j) \propto \frac{\pi(\lambda_{i+1}, \vec{x}_j)}{\pi(\lambda_i, \vec{x}_j)} = \frac{e^{u(\lambda_i,\vec{x}_j)-u(\lambda_{i+1},\vec{x}_j)}}{\sum_{j=1}^{N} e^{u(\lambda_i,\vec{x}_j)-u(\lambda_{i+1},\vec{x}_j)}}$$
$$\equiv w_j(\lambda_{i+1}|\lambda_i) \tag{2}$$

The final step of an SIR iteration consists of weighted resampling with replacement based on these weights to generate a new set of equally weighted N walkers. This results in the high-weight walkers being copied multiple times and the low-weight walkers being annihilated. This three-step procedure is then repeated for each consecutive distribution until the final distribution has been reached.

One can readily see what sets SIR apart from other enhanced sampling methods: the "survival of the fittest" approach combined with the lack of reversibility and the fact that the method does not satisfy the rather restrictive detailed balance condition, meaning that SIR only explores the best paths and that one can "peek into the future" and adapt the hyperparameters of the method based on this knowledge. This notwithstanding, SIR satisfies a more general stationarity condition, balance,[41] and is known to be completely rigorous in terms of preserving the target distribution $\pi(1,\vec{x})$ in the limit of infinite walkers and infinite sampling at $\pi(0,\vec{x})$.[42] This makes it an asymptotically valid sampling method, similar to all other

https://doi.org/10.1021/acs.jctc.1c01198
J. Chem. Theory Comput. 2022, 18, 3894−3910

sampling methods requiring an infinite amount of samples for convergence (e.g., all methods utilizing MD and/or MCMC).

It can be shown that the expectation value of the unnormalized weights $\tilde{w}_j(\lambda_{i+1}|\lambda_i) \equiv e^{u(\lambda_i,\vec{x}_j)-u(\lambda_{i+1},\vec{x}_j)}$ of the samples generated from $\pi(\lambda_i,\vec{x})$ is an unbiased estimator of the partition function ratio $\frac{Z(\lambda_{i+1})}{Z(\lambda_i)} = e^{f(\lambda_i)-f(\lambda_{i+1})}$ (Zwanzig equation[38]). This means that $Z(1)/Z(0)$ can also be estimated in an unbiased way from the products of the consecutive expectation values of the unnormalized weights. If one is interested in obtaining unbiased expectation values over separate SIR runs, then the final samples from each run need to be reweighted by the total estimated $\widehat{\frac{Z(1)}{Z(0)}}$ for this run,[43] which can be interpreted as the collective relative weight of the final samples. In effect, the samples are weighted by their free energies, as reflected in the partition function ratio. In this case, the unbiased expectation value $\langle O\rangle$ of an observable O over K independent SIR simulations, each having M walkers, is

$$\langle O\rangle = \frac{\sum_{k=1}^{K} \widehat{\frac{Z(1)}{Z(0)}}_k \frac{1}{M}\sum_{i=1}^{M} O_{ik}}{\sum_{k=1}^{K} \widehat{\frac{Z(1)}{Z(0)}}_k} \tag{3}$$

where O_{ik} is the observable evaluated on the i-th walker in the k-th simulation and $\widehat{\frac{Z(1)}{Z(0)}}_k$ is the estimated collective walker weight of the k-th simulation.

It is known that this sample reweighting procedure is not in general unbiased for adaptive SIR, where the strides in the λ space depend on the weights at each step.[44] Although this condition can be circumvented by running adaptive SIR once and using the derived protocol for all consecutive repeats,[45] this approach is not practical for running simulation repeats in parallel, and in this study, we will apply the reweighting procedure during analysis regardless of this and demonstrate its sufficient precision in a wide range of test cases.

3. ADAPTIVE ALCHEMICAL SMC

This section highlights some important considerations about performing SMC on a protein−ligand system, as well as several changes to the base method, most of which have been previously considered in the literature.[33,34,46] Some of these modifications allow us to substitute the system-dependent hyperparameters (e.g., the exact sequence of optimal intermediate distributions) with system-independent hyperparameters (e.g., adaptively choosing the intermediate distributions based on a constant distribution overlap).

3.1. Alchemical Perturbation versus Tempering. Enhancing sampling in the temperature space is valuable when one wants to treat all degrees of freedom equally. However, this approach becomes less feasible for large systems, and enhancing specific degrees of freedom is often more desirable whenever possible. In this work, we consider systems where some degrees of freedom are of greater interest than others. For example, when calculating solvation or protein−ligand binding free energies, the small-molecule rotamers are expected to influence the result more than any other degrees of freedom. Therefore, the molecular torsions together with center of mass (COM) translation and rotation constitute arguably the most important degrees of freedom for most small molecules. These are also the degrees of freedom that have multiple minima, often separated by high-energy barriers.

In these cases, one can use an alchemical approach with a coupling parameter λ, where $\lambda = 0$ denotes all relevant interactions turned off and $\lambda = 1$ represents the target potential energy function of the system (Figure 2). In this regime, one

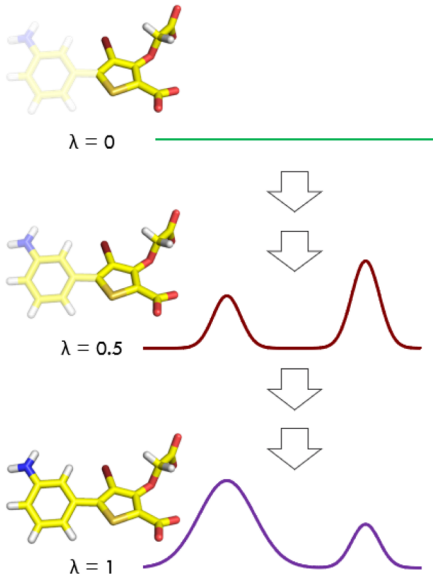


Figure 2. Exploring conformational degrees of freedom with SMC using an alchemical parameter λ. At $\lambda = 0$, all of the nonbonded interactions involving the 3-aminophenyl group are fully decoupled, and the distribution of the torsional angle is uniform. At $\lambda = 0.5$, the 3-aminophenyl group is partially coupled, and at $\lambda = 1$, it is fully interacting, in both cases resulting in two main modes/states.

can readily use any knowledge from the AFE literature. Most notably, an often employed method for deriving the functional form of the intermediate distributions is to introduce a soft-core potential,[47] which disposes of certain singularities in the potential energy function, thereby improving the statistical efficiency of any estimators dependent on the intermediate λ states. This will be invaluable for the systems discussed later, allowing us to make high-energy insertions and rotations without much of a performance penalty.

There are two common ways to turn on the potential energy interactions: the first is to use the soft-core potential only on the Lennard-Jones (LJ) part of the perturbation, followed by a linear coupling of the electrostatics ("split protocol"), and the second method involves concurrent introduction of all relevant potential terms ("unified protocol"), meaning that a soft-core functional form needs to be used for both LJ and electrostatic interactions. It is expected that a unified protocol is generally less desirable due to the presence of soft-core electrostatic terms, meaning that overlapping positive and negative charges are highly energetically favorable and such unphysical structures can dominate the sampling. On the other hand, the split protocol is expected to produce structures biased toward steric favorability since most of the resampling is expected to take place before introducing the electrostatics. In this work, we will explore and evaluate both protocols.

3.2. Adaptively Determining λ_{i+1}. One can use the knowledge obtained from the distribution of the transition probability weights to assess the quality of the configurational space overlap between the current distribution $\pi(\lambda_i,\vec{x})$ and the

https://doi.org/10.1021/acs.jctc.1c01198
J. Chem. Theory Comput. 2022, 18, 3894−3910

next distribution in the sequence $\pi(\lambda_{i+1},\vec{x})$. In general, one can use any measure of the distribution overlap to achieve this. In the SMC literature, an overwhelmingly popular metric is the effective sample size (ESS) estimator R_{ESS}[48]

$$R_{ESS}(\lambda_{i+1}|\lambda_i) = \frac{1}{\sum_{j=1}^{N} w_j(\lambda_{i+1}|\lambda_i)^2} \quad (4)$$

A general problem with most overlap metrics is the difficulty in defining what value range can be considered "good". Although ESS-based measures can be interpreted intuitively more readily than other measures, it has been suggested[49] that R_{ESS} is not necessarily a reliable estimator for the true ESS and should only be seen as a rough heuristic. Instead, one can use a much more conservative measure R_{min}, which acts as a lower bound for the true ESS[49]

$$R_{min} = \frac{1}{\max[w_1(\lambda_{i+1}|\lambda_i),\ \ldots,\ w_N(\lambda_{i+1}|\lambda_i)]} \quad (5)$$

After defining the desired system-independent value of this measure, one can iteratively[34,44,46] determine the next value in the sequence (λ_{i+1}), which results in an overlap metric closest to this threshold using a basic root finding algorithm, such as bisection. Although each iteration of this adaptive algorithm requires energy evaluations of each walker, they are in practice much faster to perform than generating new samples using dynamics, and the speed of this step will likely be limited by the computational implementation.

The utility of adaptively determining the λ protocol in this way is the guaranteed constant overlap between sequential distributions and the independence of the resulting protocol on the nature of the distributions. Furthermore, if one uses the same overlap metric and value, more dissimilar initial and final distributions will automatically result in a higher number of intermediate distributions without any additional system-specific input.

3.3. Adaptively Determining Optimal Sampling Time. An overwhelmingly common way to generate new configurations in biomolecular simulations is MD. This method will be very useful for generating locally decorrelated samples at each λ value. While such decorrelation is not a requirement for the sampling validity of the algorithm, it is in practice desirable to do so since it improves the sample diversity decreased by the resampling procedure. However, the decorrelation time is typically dependent on the system and the nature of the alchemical perturbation. Although it is common practice to choose a value between 1 and 10 ps to achieve local decorrelation, making this step adaptive as well could help maintain the balance between obtaining valid locally decorrelated samples independent of the system and spending as little computational effort as possible.

Since in our typical systems of interest, the equilibrium probability of observing a particular configuration $\vec{x}_j$ at some λ is solely a function of $u(\lambda,\vec{x}_j)$, a natural way to measure sample decorrelation is to measure the Pearson correlation coefficient r_τ between the potential energies of all initial walkers $(u(\lambda,\vec{x}_{j,0});\ j = 1, \ldots, N)$ and the walkers decorrelated for τ timesteps $(u(\lambda,\vec{x}_{j,\tau});\ j = 1, \ldots, N)$. Afterward, the sampling step can only be terminated if r_τ is within some acceptable range,[46] for example, $|r_\tau| \leq 0.1$. In practice, this step also requires an energy evaluation for every walker, and a conceivable implementation could, for instance, involve evaluating these energies every 1 ps, so as to minimize the computational overhead.

3.4. Sampling at $\lambda = 0$. SMC only converges to the correct distribution at $\lambda = 1$ if the initial distribution at $\lambda = 0$ has been sampled exhaustively. In a protein−ligand system, this means running long-timescale protein dynamics—a task that itself often requires other sophisticated enhanced sampling methods to produce satisfactory results. An additional problem is the fact that a very small fraction of the generated structures at $\lambda = 0$ will typically be relevant at $\lambda = 1$ due to the diminishing phase space overlap. In this work, we will not be concerned with long-timescale dynamics, and we will instead explore ligand conformers from a limited set of locally decorrelated equilibrated starting structures. The aim behind this approximation is being able to quickly estimate equilibrium populations biased to the initial structure either as a qualitative tool or as a way to provide information to more expensive methods, such as AFE calculations. Moving beyond this approximation requires a more sophisticated SMC algorithm, which can achieve adequate sampling over time and is thus beyond the scope of this work.

Since this initial stage of SIR is the only checkpoint that generates sample diversity, it is important to take advantage of this. In the test cases we are going to consider, there are three types of sampling moves, for which we know the underlying distribution: torsional rotation, COM rotation, and translation. In all of these cases, we can generate samples typically 1−2 orders of magnitude more than our desired number of walkers due to the fact that all translational and rotational distributions of the noninteracting atoms in these cases are uniform and therefore trivial to sample.

3.4.1. Torsional Rotation. If one removes all nonbonded interactions from at least one side of the torsional bond along with all dihedral terms centered around it, then the initial distribution with respect to the dihedral angle ϕ is uniform, and one can generate configurations by simply drawing random numbers between 0 and 2π. One can use any valid sampling method to achieve this, and in this work, we opt for a low-discrepancy alternative to pseudo-random number generation, which consists of generating equally spaced samples between 0 and 2π with a pseudo-randomly generated offset. In this way, we can be more certain in the representativeness and quality of our samples.

3.4.2. COM Rotation. COM rotation requires all nonbonded interactions between the molecule and the environment to be turned off, and it needs three degrees of freedom to be defined: two spherical coordinate angles on the unit sphere, defining the axis of rotation (θ and ϕ), and the amount of rotation ψ around that axis. To generate uniform rotations on the unit sphere, both ϕ and ψ need to be uniformly distributed between 0 and 2π, while $\theta = \arccos(2X - 1)$ for a uniformly distributed variable $X \in [0, 1)$. As in the previous example, one can use different sampling methods to generate the uniformly distributed variables, and although one can couple the different degrees of freedom to reduce the multidimensional sample discrepancy (i.e., sample "clumping"), in this study, we opt for shuffled one-dimensional grid-based samples with a pseudo-random offset for each degree of freedom. Further research will be needed to test alternative low-discrepancy sampling methods for COM rotation.

3.4.3. COM Translation. Much like COM rotation, COM translation requires the molecule of interest to be decoupled from its environment. The simplest case is COM translation within a cuboidal region, in which case only three uniformly distributed random numbers between −1 and 1 are needed to

https://doi.org/10.1021/acs.jctc.1c01198
J. Chem. Theory Comput. 2022, 18, 3894−3910

define the new reduced coordinates, which can be then scaled to the dimensions of the region of interest. Alternatively, one can uniformly generate points within a sphere with radius R. To achieve this, we can generate the spherical angles θ and ϕ in the same way as in the previous section, while the radius can be expressed as $r = \sqrt[3]{X}$ for a uniformly distributed variable $X \in [0, 1)$. Final scaling by R results in uniform spherical sampling. Similar considerations about low-discrepancy sampling apply here, and we again opt for the same routine for uniform sample generation as in the previous section.

3.4.4. Coupled Moves. Since in all of our examples, we generate random samples for each degree of freedom independently of the others, this procedure is readily extendable to multiple degrees of freedom. However, the presence of more than a few degrees of freedom can quickly lead to a combinatorial explosion, thereby reducing the sampling efficiency, and in this case, one should consider multidimensional low-discrepancy sampling alternatives. However, this approach is beyond the scope of this work, and we will not be utilizing it.

3.5. Using a Conservative Resampling Method. One drawback of SIR is that any loss of walker diversity is irreversible, and in many cases, all of the final samples can be traced to just a few initial samples.[50] It is important, therefore, to minimize unnecessary diversity loss during the resampling step.

The most obvious way to perform weighted resampling is multinomial resampling with replacement. In this case, one draws each new walker independently from the others. This is problematic since there is always a finite, albeit small, probability that the same sample will be resampled in all cases, resulting in sampling that is potentially not representative of the true weights.

More conservative resampling methods have been proposed, the most deterministic and widely used of which is systematic resampling.[51] In this case, it is guaranteed that the number of new samples corresponding to each weight $w_j(\lambda_{i+1}|\lambda_i)$ (derived from eq 2) is between the rounded-down fractional number and the rounded-up fractional number of walkers $Mw_j(\lambda_{i+1}|\lambda_i)$, where M is the number of walkers in the next iteration. For example, if the normalized weight of a particular walker is determined to be 0.27 and the total number of walkers in the next iteration is 10, then the fractional number of copies allotted to this walker is 2.7, meaning that systematic resampling will have a 70% probability of copying this walker three times and 30% probability of copying it twice. Because of this certainty, systematic resampling is highly reliable and will be the algorithm of choice in this study.

3.6. SMC Workflow in Practice. The first step in describing the problem of interest is identifying the relevant degrees of freedom to be explored, which in turn define a set of interactions to be decoupled at $\lambda = 0$. One then supplies an initial structure, the desired number of walkers, and target values for the correlation and decorrelation metrics to the procedure, resulting in an ensemble of structures generated at $\lambda = 1$ (Algorithm 1). While the choice of these hyperparameters is somewhat arbitrary and dependent on the available computational resources, they can be used on a variety of systems, and this is the approach that will be taken in this work.

Algorithm 1 Alchemical SMC

Input
  $\vec{x}_0$ initial system coordinates
  N number of walkers
  $r_{\tau,target}$ target decorrelation metric
  R_{target} target resampling metric
Output
  $(\vec{x}_1,..., \vec{x}_N)$ the walker coordinates at $\lambda = 1$
procedure ALCHSMC($\vec{x}_0, N, r_{\tau,target}, R_{target}$)
  $\lambda = 0$ ▷ decouple relevant interactions
  $(\vec{x}_1,..., \vec{x}_N) \leftarrow$ Equilibrate($\vec{x}_0$) ▷ spawn N walkers
  $(\vec{x}_1,..., \vec{x}_N) \leftarrow$ GenerateConformers($(\vec{x}_1,..., \vec{x}_N)$) ▷ as in Section 4.4
  while $\lambda < 1$ **do**
    $(\vec{x}_{1,0},..., \vec{x}_{N,0}) \leftarrow (\vec{x}_1,..., \vec{x}_N)$ ▷ store initial coordinates
    $r_\tau = 1$ ▷ initial decorrelation metric value
    while $r_\tau > r_{\tau,target}$ **do**
      $(\vec{x}_1,..., \vec{x}_N) \leftarrow$ Sample($(\vec{x}_1,..., \vec{x}_N), \tau$) ▷ τ is the decorrelation time
      $r_\tau \leftarrow$ DecorrelationMetric($(\vec{x}_{1,0},..., \vec{x}_{N,0}), (\vec{x}_1,..., \vec{x}_N)$) ▷ as in Section 4.3
    $R = 0$ ▷ initial resampling metric value
    while $|R - R_{target}| > \epsilon$ **do** ▷ ϵ determines the precision
      $\lambda_{next} \leftarrow$ ProposeLambda() ▷ using bisection, starting from $\lambda = 1$
      $\vec{w} \leftarrow$ Reweight(λ_{next}, λ,$(\vec{x}_1,..., \vec{x}_N)$) ▷ as in Equation (2)
      $R \leftarrow$ ResamplingMetric($\vec{w}$) ▷ as in Section 4.2
    $\lambda \leftarrow \lambda_{next}$
    $(\vec{x}_1,..., \vec{x}_N) \leftarrow$ Resample($(\vec{x}_1,..., \vec{x}_N), \vec{w}$) ▷ as in Section 4.5
  return $(\vec{x}_1,..., \vec{x}_N)$

4. METHODS

4.1. System Setup and Simulation. All of the following SMC simulations have been run using OpenMM 7.4.2,[52] OpenMMTools[53] 0.19.0, and OpenMMSLICER 1.0.0, a plugin for OpenMM developed during the course of this study, available at https://github.com/openmmslicer/openmmslicer. All proteins were protonated using PDB2PQR[54] and subsequently parameterized using the ff14SB[55] protein force field. GAFF2[56] with AM1-BCC charges[57,58] was used for all small molecules. All systems were solvated in cubic boxes of TIP3P[59] water with a length of 3 nm for the solvated ligand systems or 7 nm for the protein−ligand systems. Each system was run independently in six replicates from the same initial coordinates. Each run consisted of an initial minimization, followed by 100 ps of equilibration at $\lambda = 0$ before the SMC run. During this equilibration, all protein backbone atoms were harmonically restrained with force constants of 5 kcal mol^{-1} Å^{-2}. 500 walkers were used for each replicate with 100 initial conformers generated per walker, where all rotatable bonds between alchemical atoms were rotated in addition to the main alchemical moves. An energy decorrelation condition of $|r_{\tau,\text{target}}| \leq 0.1$ alongside a minimum relative configurational space overlap of $\frac{R_{\text{min,target}}}{N_{\text{walkers}}} \geq \frac{1}{5}$ was consistently used throughout the simulations. These values were arbitrarily chosen with the goal of providing a reasonable balance between the computational cost and sampling quality. Systematic resampling was performed in all cases, and all velocities were resampled from the Maxwell−Boltzmann distribution after each iteration.

All short-range nonbonded interactions had a cutoff of 1.2 nm, while long-range electrostatics were calculated using the particle mesh Ewald method.[60] A BAOAB[61] Langevin integrator at 298 K with a 2 fs timestep and a collision rate of 1 ps^{-1} was used, where all water molecules were constrained using the SETTLE[62] algorithm and all other bonds containing hydrogen atoms were constrained using the SHAKE[63] and CCMA[64] algorithms. A Monte Carlo barostat was used for pressure control at 1 atm with rescaling attempts every 50 fs. LJ and electrostatic interactions were switched on either simultaneously (unified protocol) or consecutively from $\lambda = 0$ to $\lambda = 0.8$ and from $\lambda = 0.8$ to $\lambda = 1$, respectively (split protocol). A soft-core potential was used for the LJ interactions

https://doi.org/10.1021/acs.jctc.1c01198
J. Chem. Theory Comput. 2022, 18, 3894−3910

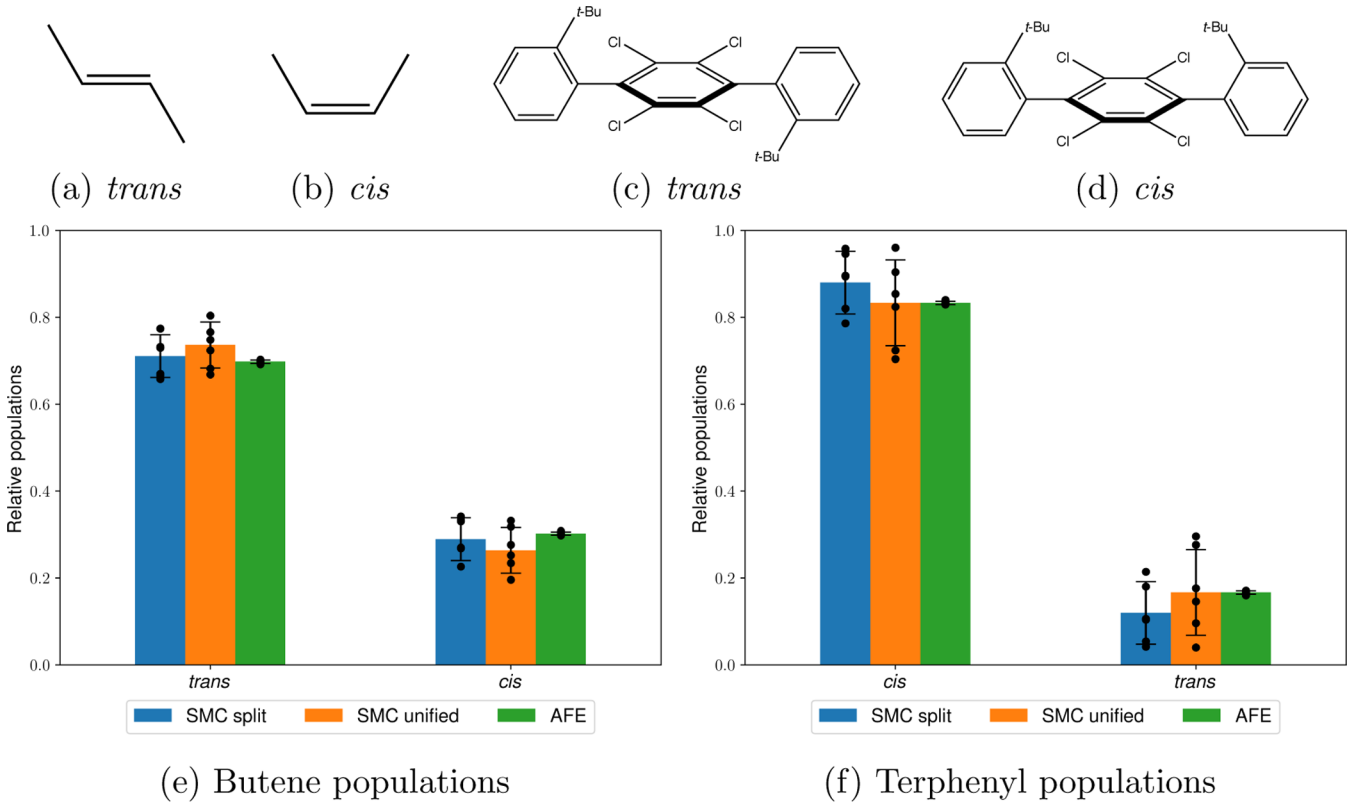


Figure 3. Two butene stereoisomers (a,b) and the two isomers of the terphenyl derivative (c,d) with populations measured via AFE and SMC (e,f) using the split and unified protocols. The heights of the bars represent the mean values weighted using the estimated partition function ratio $\widehat{\frac{Z(1)}{Z(0)}}$, and the error bars represent one weighted standard deviation based on six independent runs (shown as individual data points), as described in Section 4.2.

in both cases and for the electrostatics during the unified protocol with $\alpha = 0.5$ using the following functional form

$$r_{ij,\mathrm{eff}} = (\alpha(1-\lambda)\sigma_{ij}^6 + r_{ij}^6)^{1/6} \qquad \text{(sterics)}$$
$$r_{ij,\mathrm{eff}} = (\alpha(1-\lambda)\sigma_{ij}^2 + r_{ij}^2)^{1/2} \qquad \text{(electrostatics)} \qquad (6)$$

where all interatom distances r_{ij} in the potential energy terms involving alchemically modified atoms are replaced with $r_{ij,\mathrm{eff}}$ in the potential energy function and σ_{ij} is the "particle size" parameter defined by the LJ potential for the ij-th particle pair. In all cases, nonbonded interactions were completely annihilated rather than decoupled from the environment at $\lambda = 0$.

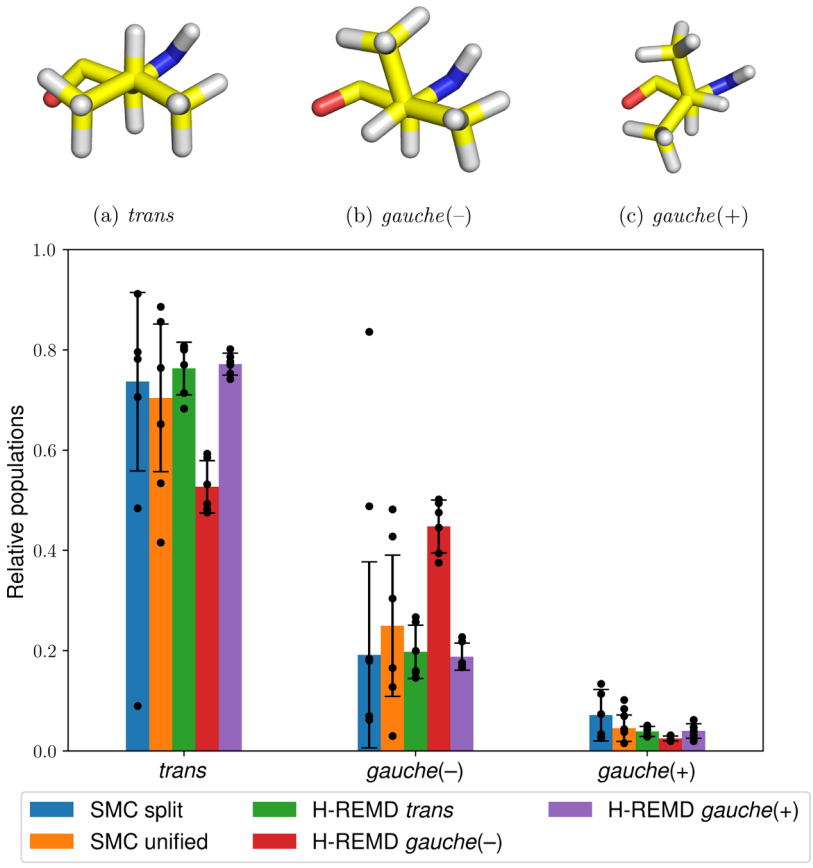
SMC was then validated against established methods in one of two ways. The first approach involved a H-REMD simulation in the λ space between 0 and 1 with multiple intermediates defined similarly to SMC. The resulting conformational populations were afterward obtained from the averaged samples at $\lambda = 1$. The second approach involved AFE calculations, which were only performed when there were only two expected rotamers separated by a high kinetic barrier. In this setting, two separate perturbations were performed in a single-topology fashion from both initial conformations, where the only difference was the rotation of the relevant torsion by 180°, to the nearest common physical intermediate (i.e., to propene in the case of butene and to a phenyl group in place of a substituted phenyl group). The corresponding dihedral terms were not scaled during the AFE calculations so that no unwanted transitions between the rotamers of interest would be observed. The population ratio between both rotamers $\frac{p_{\mathrm{state1}}}{p_{\mathrm{state2}}}$ was then calculated using the formula $kT \ln \frac{p_{\mathrm{state1}}}{p_{\mathrm{state2}}} = \Delta G^{\ominus}_{\mathrm{state1}\to\mathrm{intermediate}} - \Delta G^{\ominus}_{\mathrm{state2}\to\mathrm{intermediate}}$.

Both AFE and H-REMD calculations were performed in sextuplicate in GROMACS[65] 2018.4 patched with PLUMED[66] 2.4.3 using ProtoCaller[67] 1.1 from the same initial structures as those used for the SMC runs (and in the case of AFE, the relevant manually generated rotameric states). In all cases, the alchemically decoupled groups in the H-REMD simulations were the same as those in the SMC simulations. The only exceptions were the T4-lysozyme/3,5-difluoroaniline simulations, where a single ligand carbon atom remained coupled at $\lambda = 0$ to prevent diffusion away from the (closed) binding site without the need of additional restraint potentials. In some cases, several batches of H-REMD simulations were run from different starting conformations to investigate initial structural biasing. These will be indicated later in the text.

The split alchemical protocol was used during both AFE and H-REMD calculations, with 30 initial λ windows used for co-perturbing the soft-core sterics and the bonded interactions and 10 subsequent windows for the electrostatics. All λ values were equally spaced to two significant figures, except for the initial values, which were more closely spaced in an attempt to increase the phase space overlap: 0.001, 0.01, 0.02, 0.03, and 0.05. The Bennett acceptance ratio[68] was used for free energy analysis with snapshots every 5 ps.

The AFE protocol involved an initial 25,000-step steepest descent minimization, followed by a 50 ps *NVT* equilibration

https://doi.org/10.1021/acs.jctc.1c01198
J. Chem. Theory Comput. 2022, 18, 3894−3910



(d) T4-lysozyme Val111 populations

Figure 4. Three Val111 rotamers (a−c) in T4-lysozyme/*p*-xylene and the relative populations of all states using split and unified SMC and H-REMD from the three different initial rotamers (d). The heights of the bars represent the mean values weighted by the estimated partition function ratio $\widehat{\frac{Z(1)}{Z(0)}}$, and the error bars represent one weighted standard deviation based on six independent runs (shown as individual data points), as described in Section 4.2.

and a 50 ps *NPT* equilibration before a 4 ns *NPT* production. The Berendsen barostat[69] was used for equilibration in all cases, while the Parrinello−Rahman barostat was used for the production runs.[70] The LINCS algorithm[71] was used to constrain the non-water hydrogen atoms during both stages, while the rest of the simulation settings matched the ones from the SMC runs. In the H-REMD simulations, the above equilibration schedule was only performed at $\lambda = 1$, and the resulting volume was fixed for all replicas. This was followed by an additional minimization and equilibration only in the *NVT* ensemble and subsequent 4 ns simulations at constant volume. During both H-REMD equilibration and production, adjacent replica swaps were attempted every 1 ps.

4.2. Analysis. All of the measured populations in this study were weighted by the estimated partition function ratio $\widehat{\frac{Z(1)}{Z(0)}}$ for the relevant simulation, as previously described in eq 3. These were used to report weighted averages and weighted sample standard deviations. Since the latter can be low even when there is a high spread of data due to large discrepancies in the replicate weights, all replicate data points will also be added to the plots to visualize the unweighted spread of the resulting values between the runs. On the other hand, the estimated dimensionless free energies and the simulation times have been reported as unweighted averages with unweighted standard deviations in the main text.

To appropriately analyze the relevant kinetically separated states, clustering on the degrees of freedom of interest was performed. In most cases, this was achieved using manually defined cluster boundaries determined from the observed multimodal distributions of the angle of interest. The only exception is the ligand common core clustering analysis performed for transforming growth factor β (TGF-β), where all trajectories at $\lambda = 1$ from the SMC and H-REMD simulations were pooled together and aligned against the protein backbone α-carbon atoms of the initial structure using MDTraj[72] and MDAnalysis.[73,74] Afterward, the three Euler angles providing the best alignment of the common core ligand atoms against their initial coordinates were calculated using the align_vectors routine implemented in SciPy.[75] The sines and cosines of these three Euler angles (six degrees of freedom in total) were used to perform agglomerative clustering with default settings, as implemented in scikit-learn.[76] This analysis resulted in two clusters, whose populations will be reported later in the text alongside two representative structures corresponding to each cluster.

Where applicable, the number of round trips of the H-REMD simulations has been reported. These have been calculated as the total number of round trips of all replicas, where a round trip denotes the transition from $\lambda = 1$ to $\lambda = 0$ and back of a single replica.

https://doi.org/10.1021/acs.jctc.1c01198
J. Chem. Theory Comput. 2022, 18, 3894−3910

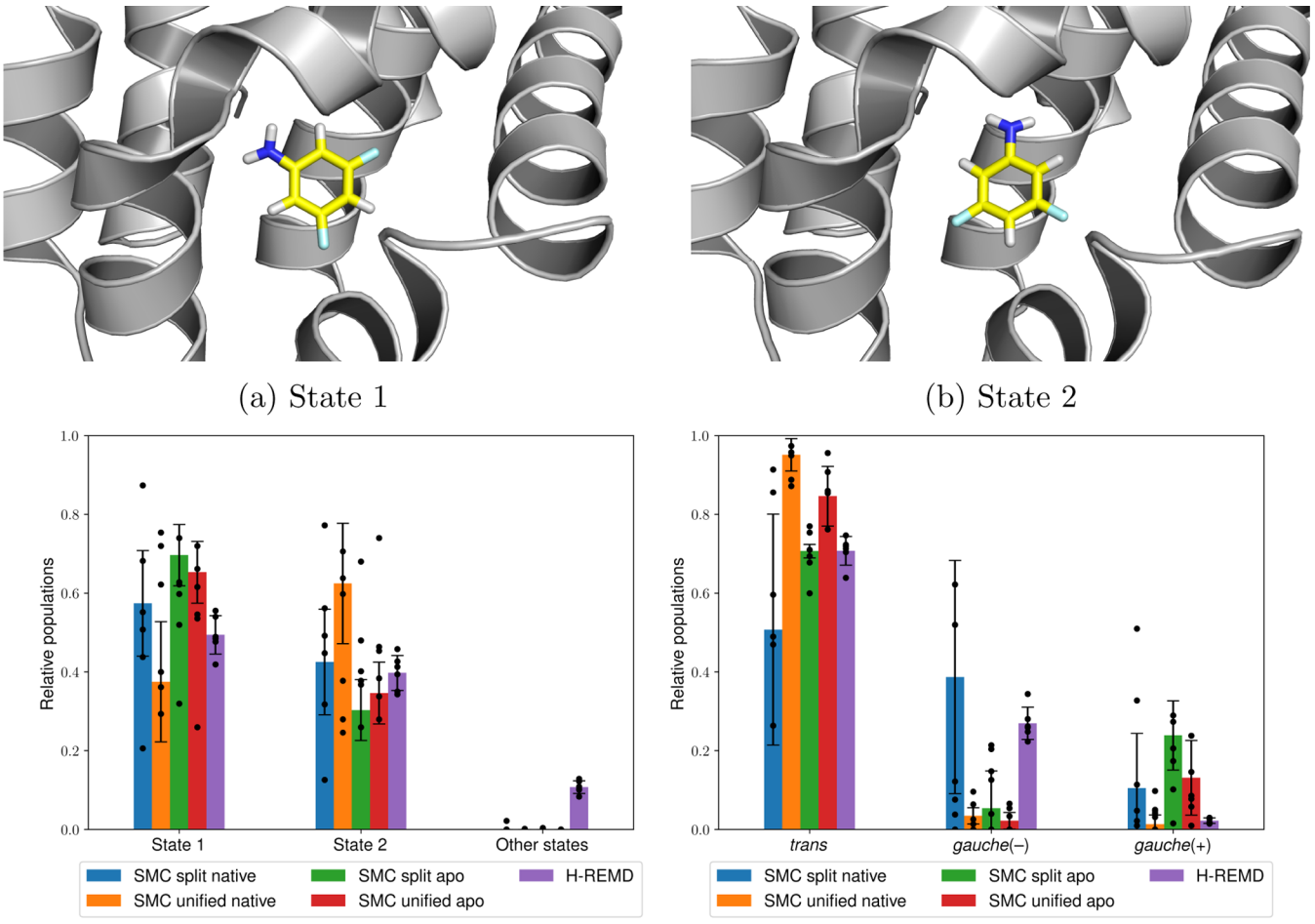


Figure 5. Two 3,5-difluoroaniline binding modes (a,b) bound to T4-lysozyme, the relative populations of both ligand states using the split and unified SMC protocols and H-REMD (c), and the Val111 states from the same simulations (d). The heights of the bars represent the mean values weighted by the estimated partition function ratio $\frac{\widehat{Z(1)}}{Z(0)}$, and the error bars represent one weighted standard deviation based on six independent runs (shown as individual data points), as described in Section 4.2.

In the following results, sampling times have been reported as the aggregate time of all λ windows (AFE and H-REMD) or the total of all walkers (SMC).

5. RESULTS

5.1. Butene in Water. One of the simplest systems involving a high kinetic barrier is the cis−trans isomerization of butene solvated in water (Figure 3). Although not of significant practical interest, this test case is a good demonstration of SMC's capabilities in an ideal setting. To explore this kinetic barrier, all atoms on one side of the double bond, together with all corresponding dihedral terms, were decoupled from their environment at $\lambda = 0$. This enabled us to directly sample this dihedral angle from the uniform distribution at $\lambda = 0$.

The results from SMC using both the unified and split protocols are presented in Figure 3e. Both protocols compare favorably to the converged 160 ns AFE results (70 ± 0% trans), with the split protocol resulting in 71 ± 5% and the unified protocol yielding an average of 74 ± 5%. In addition, both protocols result in similar performances, with 16 ± 1 ns total computational for the adaptive split protocol and 15 ± 1 ns for the adaptive unified protocol. Finally, both protocols result in comparable standard deviations of $-\ln Z_1/Z_0$ (here and henceforth referred to as "dimensionless free energy") with values of 4.79 ± 0.28 and 4.89 ± 0.15 for the split and the unified protocols, respectively, indicating good convergence in both cases.

5.2. Terphenyl in Water. A much more challenging test case with an insurmountable kinetic barrier is the terphenyl derivative shown in Figure 3. It is expected that only alchemical methods can handle such a system since approaching the kinetic barrier with all interactions turned on will result in large repulsive forces and numerical instability. Moreover, alchemically decoupling the *tert*-butylphenyl substituent is also likely to be challenging, making this system a good example of a difficult enhanced sampling problem in solution. Similar to the previous test case, one of the tert-butylphenyl substituents, as well as all dihedral terms corresponding to the rotatable bond, was completely decoupled at $\lambda = 0$ to facilitate sampling.

Figure 3f demonstrates that both the split and unified protocols yield similar results for the main cis conformer: 87 ± 7 and 83 ± 9%, respectively, compared to 83 ± 0% using 160 ns AFE. Moreover, both methods estimate the dimensionless

https://doi.org/10.1021/acs.jctc.1c01198
J. Chem. Theory Comput. 2022, 18, 3894−3910

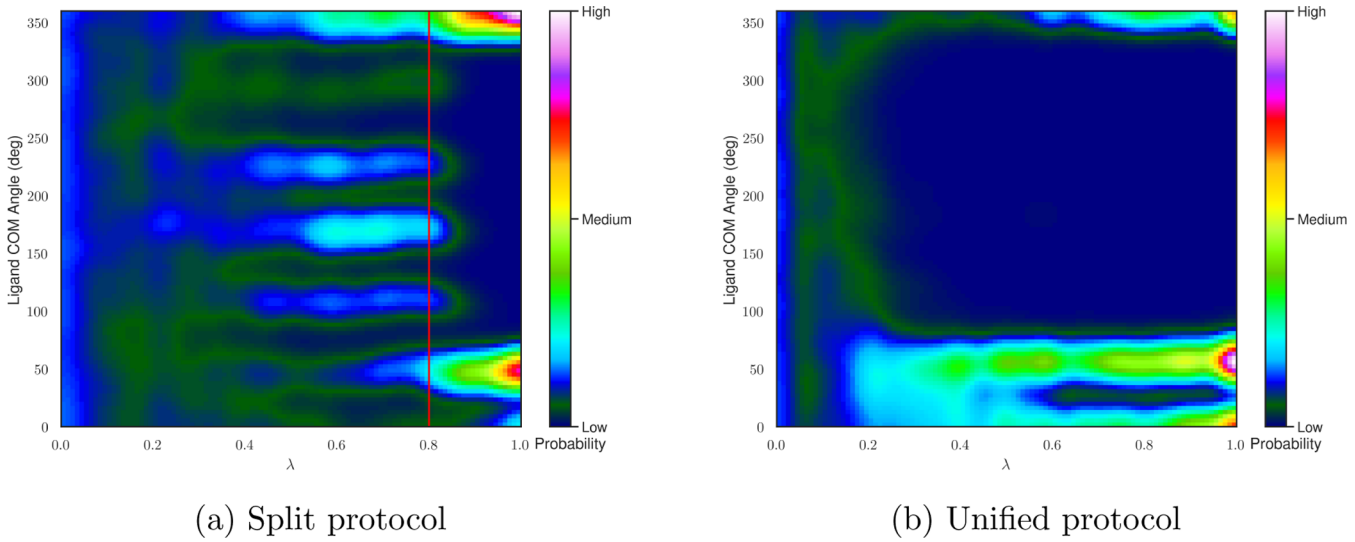


Figure 6. Heat maps of 3,5-difluoroaniline COM angle populations relative to the initial dominant conformer using the split (a) and unified (b) SMC protocols taken from a single representative repeat. The data at discrete λ values have been smoothed in both cases for visual purposes. The solid red line in (a) indicates the alchemical intermediate with fully coupled sterics and fully decoupled electrostatics.

free energy very precisely: 35.60 ± 0.23 for the split protocol and 35.64 ± 0.19 for the unified protocol, indicating good sampling consistency between the SMC alchemical protocols and repeats. Finally, both methods show similar performance, with the split protocol being slightly slower on average (42 ± 2 ns) than the unified protocol (37 ± 3 ns). The longer average simulation times compared to that of the butene perturbation show that the adaptive protocol with the same hyperparameters automatically allocates more computational time to a more difficult problem, as expected.

5.3. T4-Lysozyme/*p*-Xylene. A seemingly simple case that nevertheless showcases the inability of regular MD to provide adequate sampling is the exploration of the active site Val111 rotamers (Figure 4a−c) in model T4-lysozyme L99A with bound *p*-xylene (PDB ID: 187L[77]). It has previously been shown[78] that MD results in highly insufficient rotamer transitions even at 1 μs, suggesting that enhanced sampling is indispensable for this system. We can handle this system similarly to the previous test cases by completely decoupling the Val111 isopropyl group and the corresponding dihedral term to facilitate movement at λ = 0. In this setting, the sampling of *p*-xylene was not enhanced.

The resulting SMC protocols are highly efficient, requiring an average of 25 ± 2 and 21 ± 2 ns per repeat for the split and unified protocols, respectively, while exploring all relevant Val111 rotamers. Although the split protocol results in higher variance than the unified protocol (Figure 4d), both methods result in similar torsional populations and are qualitatively consistent with one another. This is also demonstrated by the relatively high precision of the dimensionless free energy: −43.09 ± 0.85 and −44.32 ± 0.72 for the split and unified protocols, respectively.

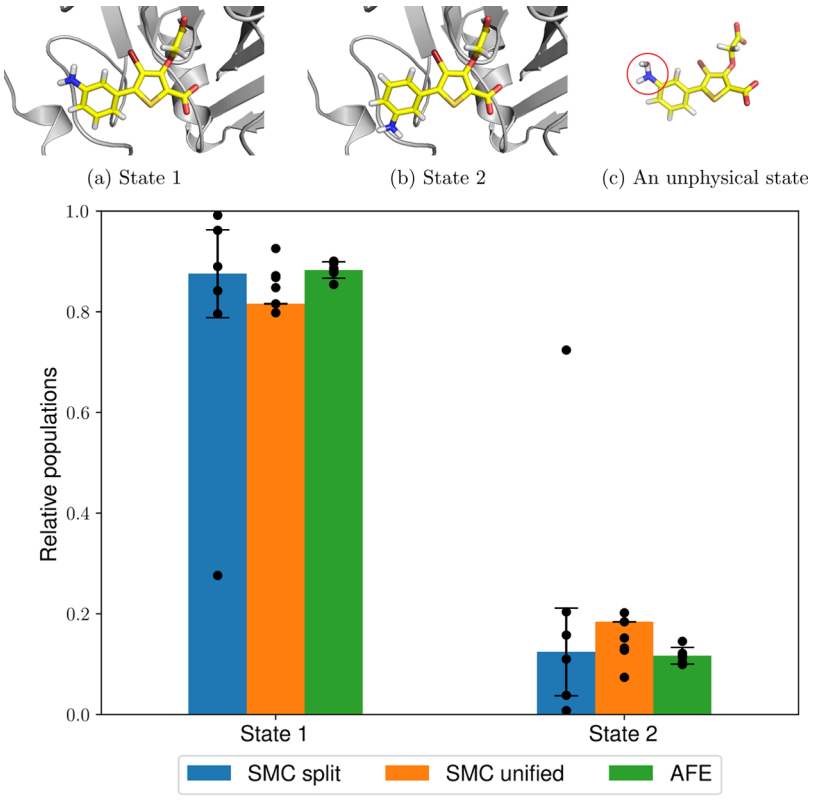
To test the accuracy of the results, they were compared against six H-REMD simulations from each initial Val111 conformer (18 simulations in total) with 160 ns per repeat, or 4 ns per replica. As shown in Figure 4, even after an average of 252 ± 13 round trips per repeat, there is a significant bias in the populations depending on the starting conformation. This discrepancy can be partially attributed to the fact that the H-REMD implementation used does not explicitly draw the decoupled dihedral from the uniform distribution at λ = 0 but instead relies purely on integrator decorrelation to achieve this, meaning that any Val111 state transitions are effectively slowed down even when there are no kinetic barriers. In contrast, the SMC simulations are not biased toward the initial Val111 conformation since all simulations start from a completely decoupled state. Nevertheless, the relative ranking of the populations is consistent between different starting structures, as well as with the SMC simulations using either the split protocol or the unified protocol. Although the predicted dominant rotamer (trans) does not correspond to that in the crystal structure [gauche(−)], the agreement between both enhanced sampling methods suggests that this discrepancy is most likely related to the force field quality and/or long-timescale populations shifts due to, for example, protein rare events, which are beyond the scope of this work.

5.4. T4-Lysozyme/3,5-Difluoroaniline. A more difficult test case is coupling the Val111 motion with translational and rotational movements of the ligand. An example ligand is 3,5-difluoroaniline bound to a L99A/M102Q T4-lysozyme mutant. In this case, the ligand was completely decoupled in addition to the Val111 isopropyl group and uniformly moved at λ = 0 within a sphere with a radius of 0.5 nm centered on its initial COM, suggested by the crystal structure (PDB ID: 1LGX[79]). Since there were two competing ligand binding modes in the electron density, the one with the higher experimentally determined occupancy was chosen for the initial COM evaluation.

The SMC simulations required an average of 48 ± 2 ns simulation time for the split protocol and 40 ± 2 ns for the unified protocol. Both protocols resulted in two main binding modes for the ligand, which are shown in Figure 5. The states are approximately equally probable, with acceptable agreement between the split protocol (57 ± 13: 43 ± 13%), the unified protocol (38 ± 15: 62 ± 15%), 160 ns H-REMD (49 ± 5: 40 ± 4%), and experiment (60:40%). The dimensionless free energies are also consistent, averaging −266.32 ± 3.46 for the split protocol and −266.68 ± 2.27 for the unified protocol.

It is interesting to note the sampling differences between both SMC protocols during the intermediate λ values. As

https://doi.org/10.1021/acs.jctc.1c01198
J. Chem. Theory Comput. 2022, 18, 3894−3910



(d) PTP1B ligand populations

Figure 7. Two thiophene derivative rotamers bound to PTP1B (a,b), unphysical interactions between the amino group and a solvent water molecule commonly observed during the unified protocol (c) (circled in red), and the relative populations of both states using SMC and AFE calculations (d). The heights of the bars represent the mean values weighted by the estimated partition function ratio $\widehat{\frac{Z(1)}{Z(0)}}$, and the error bars represent one weighted standard deviation based on six independent runs (shown as individual data points), as described in Section 4.2.

shown in Figure 6a, the split protocol explores six different binding modes with approximately equal probabilities during the steric coupling step before collapsing into the two main binding modes during the electrostatic coupling step. In contrast, the unified protocol (Figure 6b) collapses almost immediately into the two main binding modes, indicating that in this case, there is higher monotonicity in the population changes over λ.

The same SMC protocols were performed on the same mutant using a different crystal structure (PDB ID: 1LGU[79]), where only mercaptoethanol (part of the crystallization liquor) was bound, making this crystal structure the closest experimentally available structure to an apo form for this mutant. Little difference in the results was observed using both the split protocol (70 ± 8: 30 ± 8%) and the unified protocol (65 ± 8: 35 ± 8%), indicating that the method is not strongly dependent on the initial crystal structure in this case and the results are therefore not biased in an obvious way.

Larger differences were observed for the Val111 rotamers, where there were discrepancies between the populations from both SMC protocols and H-REMD. Since both the native and the apo structures exhibit significant differences between both protocols, it can be concluded that the split and unified protocol results are not consistent with each other in this case. This can be attributed to the different ways in which the different λ schedules affect the time-dependent dynamics of each walker. Since the simulation time for each walker remains very short, the lack of long-timescale sampling can therefore result in biased populations.

5.5. Protein Tyrosine Phosphatase 1B. Another commonly encountered problem is handling dihedral rotations of flexible bound ligands, such as the thiophene derivative bound to protein tyrosine phosphatase 1B (PTP1B) (PDB ID: 2QBS[80]), as shown in Figure 7. In this case, there are two main states of interest (Figure 7a,b), and we can explore this rotation by completely decoupling the 3-aminophenyl group and the relevant dihedral terms at $\lambda = 0$.

Similar to the previous torsional rotation cases, there is a good agreement between the dominant conformer in the split protocol (88 ± 9%), the unified protocol (81 ± 0%), and AFE (88 ± 2%) and the experimental crystal structure. However, in this case, the split protocol results in a much higher unweighted standard deviation (26%), mostly caused by a single outlier. Although the split protocol performs apparently worse than the unified protocol, the latter exhibits extremely poor and variable dimensionless free energy differences: 231.73 ± 20.56 compared to −85.51 ± 2.30 for the former. Since the dimensionless free energies correspond to the negative logarithm of the average relative weight of all walkers sampled from a particular simulation, the unified protocol has a negligible total weight compared to the split protocol due to its strikingly high dimensionless free energy. Therefore, even though the dihedral profiles yielded by the unified protocol appear consistent, the sampling is nevertheless remarkably poor. This can be explained by an energetically favorable

https://doi.org/10.1021/acs.jctc.1c01198
J. Chem. Theory Comput. 2022, 18, 3894−3910

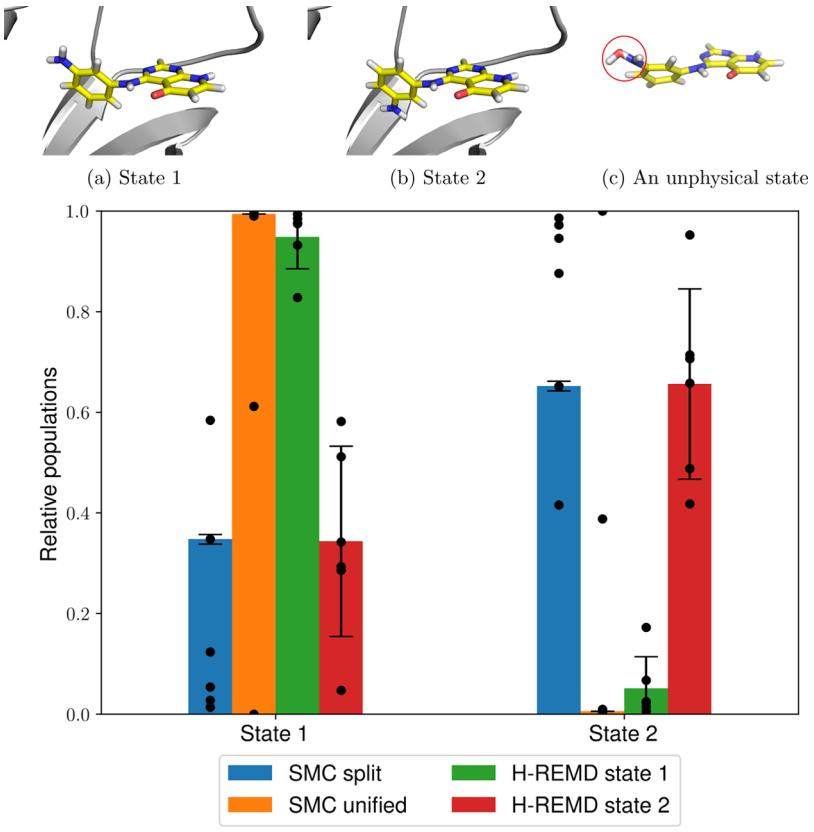


Figure 8. Two TGF-β ligand rotamers (a,b), unphysical interactions between the amino group and a solvent water molecule commonly observed during the unified protocol (c, circled in red), and relative populations of both states using the split and unified protocols and H-REMD starting from either of the states (d). The heights of the bars represent the mean values weighted by the estimated partition function ratio $\frac{\widehat{Z(1)}}{Z(0)}$, and the error bars represent one weighted standard deviation based on six independent runs (shown as individual data points), as described in Section 4.2.

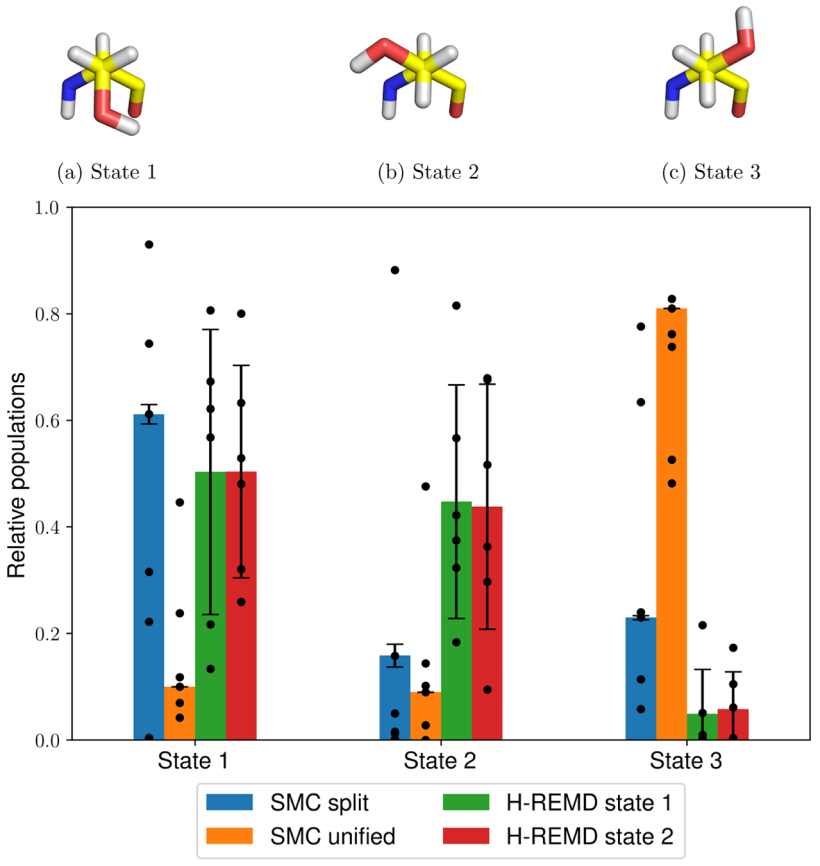
overlap between one of the ligand nitrogen atoms and a water hydrogen atom, coupled with an interaction between the aniline hydrogen and the water oxygen (Figure 7c). These unphysical interactions are not forbidden and quite favorable since introducing a soft-core potential to both sterics and electrostatics removes all potential energy singularities at the atom centers. Although these interactions vanish at $\lambda = 1$, they persist for most of the λ schedule, meaning that in this case, the split protocol is much more preferable. This conclusion is also supported by the average simulation times: 42 ± 2 ns for the split protocol and 55 ± 1 ns for the unified protocol, indicating that these unphysical states hinder the short-timescale dynamics as well.

5.6. Transforming Growth Factor β. The final test case combines a torsional rotation of a flexible ligand bound to TGF-β and a nearby Ser82 side chain rotation. In this case, we have used the initial protein coordinates of TGF-β bound to a ligand containing a related symmetric 4-aminophenyl substituent (PDB ID: 4X2G[81]) combined with the initial binding mode of the 3-aminophenyl-substituted ligand of interest (PDB ID: 4X2J[81]) so that the potential bias toward a particular conformer in the initial PDB file has been minimized. It is known from the PDB file that there are two approximately equally populated alternative conformations of the ligand (Figure 8a,b) and the nearby Ser82 residue (Figure 9a,b). As with the previous examples, this system was handled by decoupling the 3-aminophenyl ligand group concurrently with the Ser82 hydroxymethyl group.

Similarly to PTP1B, the unified protocol has sampling difficulties related to favorable unphysical interactions between an alchemically modified amine group and a water molecule (Figure 8c), resulting in large discrepancies between the dimensionless free energies: −225.51 ± 6.25 for the split protocol, compared to 200.65 ± 49.12 for the unified protocol, showing once again that this type of interaction results in populations with a negligible total weight compared to those obtained from the split protocol. Another point of similarity to the previous test case is the higher average simulation time that is needed by the unified protocol: 90 ± 3 versus 60 ± 6 ns for the split protocol.

In both cases, however, there is a marked increase in the relative weight variance compared to the previous test cases, indicating poor convergence. This is also confirmed by the ligand dihedral profiles (Figure 8d), which show significant quantitative and qualitative differences between the results of both protocols. This observation is reflected by the low efficiency of the 160 ns H-REMD runs, with an average of only 7 ± 4 round trips per repeat. Despite the low number of round trips and the slow convergence, the data from the H-REMD simulations starting from both ligand rotamers suggest that the first conformer (Figure 8a) is likely more stable than the other, implying that the unified protocol is surprisingly qualitatively consistent with H-REMD. The two SMC protocols and H-

https://doi.org/10.1021/acs.jctc.1c01198
J. Chem. Theory Comput. 2022, 18, 3894−3910



(d) TGF-β Ser82 populations

Figure 9. Three TGF-β Ser82 rotamers (a–c) and relative populations of all states using the split and unified protocols and H-REMD starting from either of the states (d). The heights of the bars represent the mean values weighted by the estimated partition function ratio $\frac{\widehat{Z(1)}}{Z(0)}$, and the error bars represent one weighted standard deviation based on six independent runs (shown as individual data points), as described in Section 4.2.

REMD do not agree on the Ser82 populations, however (Figure 9d), meaning that in both cases, there is evidence for insufficient sampling.

Clustering analysis of the ligand common core at $\lambda = 1$ using agglomerative clustering (as described in Section 4) reveals the presence of two distinct, albeit apparently similar, clusters, which correspond to a concerted translational and rotational motion of the ligand common core (Figure 10a,b). It can be seen that the first cluster is overrepresented in the unified protocol structures, as well as the H-REMD simulations starting from the dominant state (Figure 10c). However, the second cluster is the one resulting in the highest total relative weights for both the split and unified SMC runs. Since both clusters are more equally sampled during the comparatively longer H-REMD simulations, this behavior indicates an insufficient level of decorrelation of the SMC results from the initial structure, resulting in significant biasing of the observed ligand dihedral populations. Moreover, these ligand transitions present an orthogonal rare event that is not adequately sampled even at longer timescales and thus results in increased population variance for both SMC and H-REMD.

6. DISCUSSION

The above results show that SMC is extremely efficient at exploring ligand conformers in solution, even for alchemical changes that would be considered difficult to perform in practice. This is not surprising since this is the ideal setting for the method: the ligand degrees of freedom are the only ones that require extensive sampling, while the environment does not need much long-timescale sampling to respond to the ligand motions. Therefore, SMC can be a valuable tool in exploring the degrees of freedom of solvated small molecules and is likely one of the most robust ways to achieve this.

The T4-lysozyme test cases show that a closed binding pocket exhibiting little flexibility also constitutes a favorable application of the method. We have shown that SMC is unaffected by high kinetic barriers and relatively unbiased toward the initial ligand structure, while providing efficient protocols that require no system-specific parameters. These results appear to hold even when exploring coupled motions between a side chain and a ligand.

Similar observations have been made for PTP1B, where the ligand is strongly bound to the protein and the rotatable group of interest faces the solvent. In this case, the efficiency of SMC is similar to that observed in the solvated ligand systems. However, the resulting unweighted population variances from all protein test cases are much higher compared to those from the first two test cases, and this trend carries to the dimensionless free energies. This is expected since protein–ligand systems present a much more challenging sampling problem compared to solvated ligand systems.

TGF-β presents a more challenging system, where rare motions of the unmodified part of the ligand contribute to a

https://doi.org/10.1021/acs.jctc.1c01198
J. Chem. Theory Comput. 2022, 18, 3894–3910

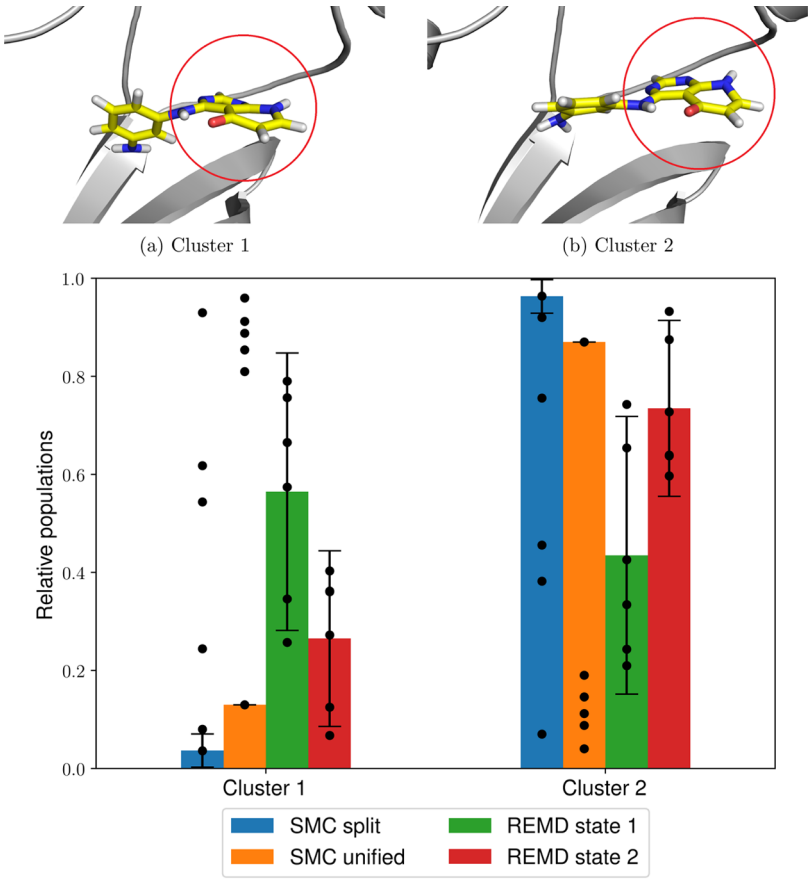


Figure 10. Two TGF-β ligand clusters (a,b, common core circled in red) and their relative populations using the split and unified protocols and H-REMD starting from either of the states (c). The heights of the bars represent the mean values weighted by the estimated partition function ratio $\widehat{\frac{Z(1)}{Z(0)}}$, and the error bars represent one weighted standard deviation based on six independent runs (shown as individual data points), as described in Section 4.2.

much higher observed dihedral population variance than in the previous test systems. This behavior is observed for both SMC and H-REMD, meaning that exploring long-timescale motions for this system is crucial and the short-timescale SMC runs are not sufficient in this case. It is therefore important to be able to identify such potentially problematic systems a priori, and this should be addressed in future work.

The above test cases also show the advantages and disadvantages of the split and unified force field scaling protocols. It has been demonstrated that the unified protocol can result in an unpredictable performance and can suffer from unphysical interactions between atoms with opposite charges, resulting in them collapsing on top of one another. This means that while the unified protocol can in many cases be more efficient than the split protocol, it is also less robust. The split protocol, on the other hand, has been shown to be extremely consistent both in terms of sampling time and free energy estimation but often results in a larger unweighted variance of the sampled populations. It is not yet clear how the above protocols will perform in a system that exhibits a drastic shift in rotamer populations when the electrostatic interactions are switched on, but the above results strongly suggest that the split protocol is by far the safer and more conservative choice for most systems.

All of the above results paint a clear picture of the current advantages and limitations of SMC. SMC excels in cases where one is interested in few degrees of freedom and where the populations of interest remain relatively unchanged over long timescales. In such systems, one can expect high performance with minimal user input, meaning that very different systems can be run with the same hyperparameters without external intervention. Another advantage of SMC is the lack of a need for supplying an initial conformation of the degree of freedom of interest, thereby providing an unbiased estimate of the population over this degree of freedom. In contrast, while H-REMD results in populations with apparently lower variance than SMC, it also exhibits a long-timescale bias toward the initial conformation. Taking this bias into account then results in a similar performance to that of SMC. Moreover, the collective estimated SMC simulation weights $\widehat{\frac{Z(1)}{Z(0)}}$ provide a straightforward way to measure sampling quality, while investigating bias is not as obvious, meaning that SMC is more useful for performing exploratory simulations. On the other hand, AFE calculations result in significantly lower variance than both SMC and H-REMD, but their main disadvantage is that a separate simulation is required for each cluster of interest. These need to be known in advance, and this knowledge is not always available in practice.

https://doi.org/10.1021/acs.jctc.1c01198
J. Chem. Theory Comput. 2022, 18, 3894−3910

SMC is therefore a valuable qualitative exploratory tool, which can quickly provide initial structures that are unbiased over a particular degree of freedom, as well as generate an efficient λ schedule for another more computationally expensive method, such as AFE calculations or H-REMD simulations. The latter methods, on the other hand, can sample over arbitrarily long timescales, thereby being systematically improvable while simultaneously reducing their bias toward the initial protein crystal structure over time. This decorrelation is crucial in, for example, binding free energy calculations, where the initial protein crystal structure can significantly impact the calculated free energies.[82] Therefore, one can use the strengths of both SMC and long-timescale methods to minimize the dependence of the sampled conformations on the choice of initial protein and ligand coordinates.

Owing to the shortness of its simulations, SMC is so far impractical for sampling long-timescale motions and can suffer from initial structure templating, as well as kinetic trapping, which occurs during the alchemical steps. The latter is a significant challenge not only for SMC but also for H-REMD, and it can be triggered by orthogonal rare events, such as slow motions of the unmodified part of the ligand, which makes this behavior difficult to predict. These problems will therefore require key modifications to the SMC method and will be the subject of future work.

7. CONCLUSIONS

We have presented an alchemical version of SMC, a directed irreversible method that can be used for sampling rare events using adaptive importance resampling. Alchemical SMC combines adaptive SMC methods with the knowledge from the AFE literature and is thus ideally suited for protein−ligand and related systems, systems where the requirement for system-specific method parameters would be highly undesirable.

The performance of alchemical SMC was demonstrated in a variety of test cases where regular MD is unable to provide adequate sampling, and we have also measured the relative efficiencies of a split perturbation protocol and a unified scheme, where steric and electrostatic interactions are coupled separately and concurrently, respectively. Our results show that SMC performs best when the results are largely independent of long-timescale motions and other important orthogonal kinetic barriers. In these cases, SMC provides efficient sampling and is unaffected by the exact nature and size of the system. The most consistent and robust results are also observed when the split protocol is used, and we recommend it over the unified protocol for the general case.

We have shown that alchemical SMC is good at generating unbiased conformations over a selection of degrees of freedom. Moreover, it provides a good metric for convergence, the estimated collective weight $\widehat{\frac{Z(1)}{Z(0)}}$, which can be used to assess the sampling quality over different simulation repeats. In this setting, methods such as H-REMD are less useful due to their long-timescale bias, which is often difficult to detect. Similarly, AFE calculations require prior knowledge for the conformers of interest, and their cost scales rapidly with the number of possible states for each degree of freedom. This makes alchemical SMC a good method for performing exploratory simulations with minimal input.

In one of the test cases, SMC exhibits large variance and poor convergence, and this suboptimal performance can be attributed to high dependence on the initial coordinates, meaning that the method needs to be extended to long-timescale sampling. This will be the subject of future research.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.1c01198.

Raw data reported in the publication (XLSX)

Initial structures and Python scripts used to run all simulations using OpenMMSLICER and ProtoCaller (ZIP)

AUTHOR INFORMATION

Corresponding Author

Jonathan W. Essex − *School of Chemistry, University of Southampton, Southampton SO17 1BJ, U.K.;* orcid.org/0000-0003-2639-2746; Email: J.W.Essex@soton.ac.uk

Authors

Miroslav Suruzhon − *School of Chemistry, University of Southampton, Southampton SO17 1BJ, U.K.;* orcid.org/0000-0002-6794-1679

Michael S. Bodnarchuk − *Computational Chemistry, R&D Oncology, AstraZeneca, Cambridge CB4 0WG, U.K.;* orcid.org/0000-0002-9172-1203

Antonella Ciancetta − *Sygnature Discovery, Nottingham NG1 1GR, U.K.; Department of Chemical, Pharmaceutical and Agricultural Sciences—DOCPAS, University of Ferrara, 44121 Ferrara, Italy;* orcid.org/0000-0002-7612-2050

Ian D. Wall − *GSK Medicines Research Centre, Stevenage SG1 2NY, U.K.*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.1c01198

Notes

The authors declare the following competing financial interest(s): JWE's research is part-funded by Sygnature, GSK, AstraZeneca and Syngenta.

ACKNOWLEDGMENTS

The authors would like to thank Russell Viner, Khaled A. Maksoud, and Marley L. Samways for the useful discussions and technical help. This study has been funded by AstraZeneca, GSK, and Syngenta and supported by the EPSRC under EP/V048864/1 and the Centre for Doctoral Training, Theory and Modelling in Chemical Sciences, under Grant EP/L015722/1. The authors acknowledge the University of Southampton high-performance computing cluster Iridis 5, the Hartree Centre high-performance computing cluster JADE, and HECBioSim (Grant EP/R029407/1) for facilitating a part of this study.

REFERENCES

(1) Yang, Y. I.; Shao, Q.; Zhang, J.; Yang, L.; Gao, Y. Q. Enhanced Sampling in Molecular Dynamics. *J. Chem. Phys.* **2019**, *151*, 070902.

(2) Huggins, D. J.; Biggin, P. C.; Dämgen, M. A.; Essex, J. W.; Harris, S. A.; Henchman, R. H.; Khalid, S.; Kuzmanic, A.; Laughton, C. A.; Michel, J.; Mulholland, A. J.; Rosta, E.; Sansom, M. S. P.; van der Kamp, M. W. Biomolecular Simulations: From Dynamics and Mechanisms to Computational Assays of Biological Activity. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2019**, *9*, No. e1393.

(3) Swendsen, R. H.; Wang, J.-S. Replica Monte Carlo Simulation of Spin-Glasses. *Phys. Rev. Lett.* **1986**, *57*, 2607−2609.
(4) Neal, R. M. Sampling from Multimodal Distributions Using Tempered Transitions. *Stat. Comput.* **1996**, *6*, 353−366.
(5) Liu, P.; Kim, B.; Friesner, R. A.; Berne, B. J. Replica Exchange with Solute Tempering: A Method for Sampling Biological Systems in Explicit Water. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13749.
(6) Wang, L.; Friesner, R. A.; Berne, B. J. Replica Exchange with Solute Scaling: A More Efficient Version of Replica Exchange with Solute Tempering (REST2). *J. Phys. Chem. B* **2011**, *115*, 9431−9438.
(7) Laio, A.; Parrinello, M. Escaping Free-Energy Minima. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562−12566.
(8) Barducci, A.; Bussi, G.; Parrinello, M. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Phys. Rev. Lett.* **2008**, *100*, 020603.
(9) Branduardi, D.; Bussi, G.; Parrinello, M. Metadynamics with Adaptive Gaussians. *J. Chem. Theory Comput.* **2012**, *8*, 2247−2254.
(10) Torrie, G. M.; Valleau, J. P. Nonphysical Sampling Distributions in Monte Carlo Free-Energy Estimation: Umbrella Sampling. *J. Comput. Phys.* **1977**, *23*, 187−199.
(11) Isralewitz, B.; Gao, M.; Schulten, K. Steered Molecular Dynamics and Mechanical Functions of Proteins. *Curr. Opin. Struct. Biol.* **2001**, *11*, 224−230.
(12) Nilmeier, J. P.; Crooks, G. E.; Minh, D. D.; Chodera, J. D. Nonequilibrium Candidate Monte Carlo Is an Efficient Tool for Equilibrium Simulation. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, No. E1009.
(13) Tan, Z. Optimally Adjusted Mixture Sampling and Locally Weighted Histogram Analysis. *J. Comput. Graph Stat.* **2017**, *26*, 54−65.
(14) Wojtas-Niziurski, W.; Meng, Y.; Roux, B.; Bernèche, S. Self-Learning Adaptive Umbrella Sampling Method for the Determination of Free Energy Landscapes in Multiple Dimensions. *J. Chem. Theory Comput.* **2013**, *9*, 1885−1895.
(15) Zhang, W.; Chen, J. Efficiency of Adaptive Temperature-Based Replica Exchange for Sampling Large-Scale Protein Conformational Transitions. *J. Chem. Theory Comput.* **2013**, *9*, 2849−2856.
(16) Rosenbluth, M. N.; Rosenbluth, A. W. Monte Carlo Calculation of the Average Extension of Molecular Chains. *J. Chem. Phys.* **1955**, *23*, 356−359.
(17) Liu, J. S.; Chen, R. Sequential Monte Carlo Methods for Dynamic Systems. *J. Am. Stat. Assoc.* **1998**, *93*, 1032−1044.
(18) Doucet, A.; Godsill, S.; Andrieu, C. On Sequential Monte Carlo Sampling Methods for Bayesian Filtering. *Stat. Comput.* **2000**, *10*, 197−208.
(19) Weiss, R.; Glösekötter, P.; Prestes, E.; Kolberg, M. Hybridisation of Sequential Monte Carlo Simulation with Non-Linear Bounded-Error State Estimation Applied to Global Localisation of Mobile Robots. *J. Intell. Rob. Syst.* **2020**, *99*, 335−357.
(20) Leeuwen, P. J.; Künsch, H. R.; Nerger, L.; Potthast, R.; Reich, S. Particle Filters for High-Dimensional Geoscience Applications: A Review. *Q. J. R. Metereol. Soc.* **2019**, *145*, 2335−2365.
(21) Machta, J. Population Annealing with Weighted Averages: A Monte Carlo Method for Rough Free-Energy Landscapes. *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.* **2010**, *82*, 026704.
(22) Wang, W.; Machta, J.; Katzgraber, H. G. Population Annealing: Theory and Application in Spin Glasses. *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.* **2015**, *92*, 063307.
(23) Reynolds, P. J.; Ceperley, D. M.; Alder, B. J.; Lester, W. A. Fixed-node Quantum Monte Carlo for Molecules. *J. Chem. Phys.* **1982**, *77*, 5593−5603.
(24) Del Moral, P. Nonlinear Filtering: Interacting Particle Resolution. *Comptes Rendus Acad. Sci. Ser.* **1997**, *325*, 653−658.
(25) Huber, G. A.; McCammon, J. A. Weighted-Ensemble Simulated Annealing: Faster Optimization on Hierarchical Energy Surfaces. *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.* **1997**, *55*, 4822−4825.
(26) Siepmann, J. I.; Frenkel, D. Configurational Bias Monte Carlo: A New Sampling Scheme for Flexible Chains. *Mol. Phys.* **1992**, *75*, 59−70.
(27) Rousset, M.; Stoltz, G. Equilibrium Sampling from Nonequilibrium Dynamics. *J. Stat. Phys.* **2006**, *123*, 1251−1272.
(28) Lyman, E.; Zuckerman, D. M. Resampling Improves the Efficiency of a "Fast-Switch" Equilibrium Sampling Protocol. *J. Chem. Phys.* **2009**, *130*, 081102.
(29) Rufa, D. A.; Bruce Macdonald, H. E.; Fass, J.; Wieder, M.; Grinaway, P. B.; Roitberg, A. E.; Isayev, O.; Chodera, J. D. Towards Chemical Accuracy for Alchemical Free Energy Calculations with Hybrid Physics-Based Machine Learning/Molecular Mechanics Potentials. bioRxiv, **2020**, 2020.07.29.227959.
(30) Zhang, J.; Chen, R.; Tang, C.; Liang, J. Origin of Scaling Behavior of Protein Packing Density: A Sequential Monte Carlo Study of Compact Long Chain Polymers. *J. Chem. Phys.* **2003**, *118*, 6102−6109.
(31) Tang, K.; Zhang, J.; Liang, J. Fast Protein Loop Sampling and Structure Prediction Using Distance-Guided Sequential Chain-Growth Monte Carlo Method. *PLoS Comput. Biol.* **2014**, *10*, No. e1003539.
(32) Wong, S. W. K.; Liu, J. S.; Kou, S. C. Exploring the Conformational Space for Protein Folding with Sequential Monte Carlo. *Ann. Appl. Stat.* **2018**, *12*, 1628−1654.
(33) Christiansen, H.; Weigel, M.; Janke, W. Accelerating Molecular Dynamics Simulations with Population Annealing. *Phys. Rev. Lett.* **2019**, *122*, 060602.
(34) Christiansen, H.; Weigel, M.; Janke, W. Population Annealing Molecular Dynamics with Adaptive Temperature Steps. *J. Phys.: Conf. Ser.* **2019**, *1163*, 012074.
(35) Sugita, Y.; Okamoto, Y. Replica-Exchange Multicanonical Algorithm and Multicanonical Replica-Exchange Method for Simulating Systems with Rough Energy Landscape. *Chem. Phys. Lett.* **2000**, *329*, 261−270.
(36) 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.
(37) Gill, S. C.; Lim, N. M.; Grinaway, P. B.; Rustenburg, A. S.; Fass, J.; Ross, G. A.; Chodera, J. D.; Mobley, D. L. Binding Modes of Ligands Using Enhanced Sampling (BLUES): Rapid Decorrelation of Ligand Binding Modes via Nonequilibrium Candidate Monte Carlo. *J. Phys. Chem. B* **2018**, *122*, 5579−5598.
(38) Zwanzig, R. W. High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. *J. Chem. Phys.* **1954**, *22*, 1420−1426.
(39) Marinari, E.; Parisi, G. Simulated Tempering: A New Monte Carlo Scheme. *Europhys. Lett.* **1992**, *19*, 451−458.
(40) Lyubartsev, A. P.; Martsinovski, A. A.; Shevkunov, S. V.; Vorontsov-Velyaminov, P. N. New Approach to Monte Carlo Calculation of the Free Energy: Method of Expanded Ensembles. *J. Chem. Phys.* **1992**, *96*, 1776−1783.
(41) Manousiouthakis, V. I.; Deem, M. W. Strict Detailed Balance Is Unnecessary in Monte Carlo Simulation. *J. Chem. Phys.* **1999**, *110*, 2753−2756.
(42) Doucet, A.; Freitas, N.; Gordon, N. An Introduction to Sequential Monte Carlo Methods. *Sequential Monte Carlo Methods in Practice*; Springer, 2001; pp 3−14.
(43) Del Moral, P. *Feynman-Kac Formulae: Genealogical and Interacting Particle Systems with Applications*; Springer: New York; London, 2011.
(44) Zhou, Y.; Johansen, A. M.; Aston, J. A. D. Toward Automatic Model Comparison: An Adaptive Sequential Monte Carlo Approach. *J. Comput. Graph Stat.* **2016**, *25*, 701−726.
(45) Barash, L. Y.; Weigel, M.; Borovský, M.; Janke, W.; Shchur, L. N. GPU Accelerated Population Annealing Algorithm. *Comput. Phys. Commun.* **2017**, *220*, 341−350.
(46) Buchholz, A.; Chopin, N.; Jacob, P. E. Adaptive Tuning of Hamiltonian Monte Carlo within Sequential Monte Carlo. *Bayesian Anal.* **2021**, *16*, 745.

https://doi.org/10.1021/acs.jctc.1c01198
J. Chem. Theory Comput. 2022, 18, 3894−3910

(47) Pham, T. T.; Shirts, M. R. Identifying Low Variance Pathways for Free Energy Calculations of Molecular Transformations in Solution Phase. *J. Chem. Phys.* **2011**, *135*, 034114.
(48) Kong, A. *A Note on Importance Sampling Using Standardized Weights*, Technical Report 348; Department of Statistics, University of Chicago, 1992; Vol. *348*.
(49) Martino, L.; Elvira, V.; Louzada, F. Effective Sample Size for Importance Sampling Based on Discrepancy Measures. *Signal Process.* **2017**, *131*, 386−401.
(50) Naesseth, C. A.; Lindsten, F.; Schön, T. B. Elements of Sequential Monte Carlo. *Found. Trends Mach. Learn.* **2019**, *12*, 307−392.
(51) Doucet, A.; Johansen, A. M. A Tutorial on Particle Filtering and Smoothing: Fifteen Years Later. *Handbook of Nonlinear Filtering*; University of Oxford, 2009; Vol. *12*, p 3.
(52) Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L.-P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D.; Wiewiora, R. P.; Brooks, B. R.; Pande, V. S. OpenMM 7: Rapid Development of High Performance Algorithms for Molecular Dynamics. *PLoS Comput. Biol.* **2017**, *13*, No. e1005659.
(53) Rizzi, A.; Chodera, J.; Naden, L.; Beauchamp, K.; Grinaway, P.; Fass, J.; Wade, A. D.; Rustenburg, B.; Ross, G. A.; Krämer, A.; Macdonald, H. B.; Swenson, D. W.; Simmonett, A.; Rodríguez-Guerra, J.; Rufa, D. A.; Henry, M. F.; Roet, S.; hb0402, *OpenMMTools*, 2019; https://github.com/choderalab/openmmtools (accessed April 09, 2022).
(54) Dolinsky, T. J.; Czodrowski, P.; Li, H.; Nielsen, J. E.; Jensen, J. H.; Klebe, G.; Baker, N. A. PDB2PQR: Expanding and Upgrading Automated Preparation of Biomolecular Structures for Molecular Simulations. *Nucleic Acids Res.* **2007**, *35*, W522−W525.
(55) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696−3713.
(56) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General AMBER Force Field. *J. Comput. Chem.* **2004**, *25*, 1157−1174.
(57) Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: I. Method. *J. Comput. Chem.* **2000**, *21*, 132−146.
(58) Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: II. Parameterization and Validation. *J. Comput. Chem.* **2002**, *23*, 1623−1641.
(59) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926−935.
(60) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089−10092.
(61) Fass, J.; Sivak, D. A.; Crooks, G. E.; Beauchamp, K. A.; Leimkuhler, B.; Chodera, J. D. Quantifying Configuration-Sampling Error in Langevin Simulations of Complex Molecular Systems. *Entropy* **2018**, *20*, 318.
(62) Miyamoto, S.; Kollman, P. A. SETTLE: An Analytical Version of the SHAKE and RATTLE Algorithm for Rigid Water Models. *J. Comput. Chem.* **1992**, *13*, 952−962.
(63) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J. Comput. Phys.* **1977**, *23*, 327−341.
(64) Eastman, P.; Pande, V. S. CCMA: A Robust, Parallelizable Constraint Method for Molecular Simulations. *J. Chem. Theory Comput.* **2010**, *6*, 434−437.
(65) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1*−*2*, 19−25.
(66) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New Feathers for an Old Bird. *Comput. Phys. Commun.* **2014**, *185*, 604−613.
(67) Suruzhon, M.; Senapathi, T.; Bodnarchuk, M. S.; Viner, R.; Wall, I. D.; Barnett, C. B.; Naidoo, K. J.; Essex, J. W. ProtoCaller: Robust Automation of Binding Free Energy Calculations. *J. Chem. Inf. Model.* **2020**, *60*, 1917−1921.
(68) Bennett, C. H. Efficient Estimation of Free Energy Differences from Monte Carlo Data. *J. Comput. Phys.* **1976**, *22*, 245−268.
(69) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81*, 3684−3690.
(70) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, *52*, 7182−7190.
(71) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* **1997**, *18*, 1463−1472.
(72) McGibbon, R. T.; Beauchamp, K. A.; Harrigan, M. P.; Klein, C.; Swails, J. M.; Hernández, C. X.; Schwantes, C. R.; Wang, L.-P.; Lane, T. J.; Pande, V. S. MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophys. J.* **2015**, *109*, 1528−1532.
(73) Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O. MDAnalysis: A Toolkit for the Analysis of Molecular Dynamics Simulations. *J. Comput. Chem.* **2011**, *32*, 2319−2327.
(74) Gowers, R. J.; Linke, M.; Barnoud, J.; Reddy, T. J. E.; Melo, M. N.; Seyler, S. L.; Domański, J.; Dotson, D. L.; Buchoux, S.; Kenney, I. M.; Beckstein, J.; MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. *Proceedings of the 15th Python in Science Conference*, 2016; pp 98−105.
(75) Virtanen, P.; Gommers, R.; Oliphant, T. E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; van der Walt, S. J.; Brett, M.; Wilson, J.; Millman, K. J.; Mayorov, N.; Nelson, A. R. J.; Jones, E.; Kern, R.; Larson, E.; Carey, C. J.; Polat, İ.; Feng, Y.; Moore, E. W.; VanderPlas, J.; Laxalde, D.; Perktold, J.; Cimrman, R.; Henriksen, I.; Quintero, E. A.; Harris, C. R.; Archibald, A. M.; Ribeiro, A. H.; Pedregosa, F.; van Mulbregt, P.; Vijaykumar, A.; Bardelli, A. P.; Rothberg, A.; Hilboll, A.; Kloeckner, A.; Scopatz, A.; Lee, A.; Rokem, A.; Woods, C. N.; Fulton, C.; Masson, C.; Häggström, C.; Fitzgerald, C.; Nicholson, D. A.; Hagen, D. R.; Pasechnik, D. V.; Olivetti, E.; Martin, E.; Wieser, E.; Silva, F.; Lenders, F.; Wilhelm, F.; Young, G.; Price, G. A.; Ingold, G.-L.; Allen, G. E.; Lee, G. R.; Audren, H.; Probst, I.; Dietrich, J. P.; Silterra, J.; Webber, J. T.; Slavič, J.; Nothman, J.; Buchner, J.; Kulick, J.; Schönberger, J. L.; de Miranda Cardoso, J. V.; Reimer, J.; Harrington, J.; Rodríguez, J. L. C.; Nunez-Iglesias, J.; Kuczynski, J.; Tritz, K.; Thoma, M.; Newville, M.; Kümmerer, M.; Bolingbroke, M.; Tartre, M.; Pak, M.; Smith, N. J.; Nowaczyk, N.; Shebanov, N.; Pavlyk, O.; Brodtkorb, P. A.; Lee, P.; McGibbon, R. T.; Feldbauer, R.; Lewis, S.; Tygier, S.; Sievert, S.; Vigna, S.; Peterson, S.; More, S.; Pudlik, T.; Oshima, T.; Pingel, T. J.; Robitaille, T. P.; Spura, T.; Jones, T. R.; Cera, T.; Leslie, T.; Zito, T.; Krauss, T.; Upadhyay, U.; Halchenko, Y. O.; Vázquez-Baeza, Y.; SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat. Methods* **2020**, *17*, 261−272.
(76) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, E. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825−2830.
(77) Morton, A.; Matthews, B. W. Specificity of Ligand Binding in a Buried Nonpolar Cavity of T4 Lysozyme: Linkage of Dynamics and Structural Plasticity. *Biochemistry* **1995**, *34*, 8576−8588.
(78) Burley, K. H.; Gill, S. C.; Lim, N. M.; Mobley, D. L. Enhancing Side Chain Rotamer Sampling Using Nonequilibrium Candidate Monte Carlo. *J. Chem. Theory Comput.* **2019**, *15*, 1848−1862.

https://doi.org/10.1021/acs.jctc.1c01198
J. Chem. Theory Comput. 2022, 18, 3894−3910

(79) Wei, B. Q.; Baase, W. A.; Weaver, L. H.; Matthews, B. W.; Shoichet, B. K. A Model Binding Site for Testing Scoring Functions in Molecular Docking. *J. Mol. Biol.* **2002**, *322*, 339−355.
(80) Wilson, D. P.; Wan, Z.-K.; Xu, W.-X.; Kirincich, S. J.; Follows, B. C.; Joseph-McCarthy, D.; Foreman, K.; Moretto, A.; Wu, J.; Zhu, M.; Binnun, E.; Zhang, Y.-L.; Tam, M.; Erbe, D. V.; Tobin, J.; Xu, X.; Leung, L.; Shilling, A.; Tam, S. Y.; Mansour, T. S.; Lee, J. Structure-Based Optimization of Protein Tyrosine Phosphatase 1B Inhibitors: From the Active Site to the Second Phosphotyrosine Binding Site. *J. Med. Chem.* **2007**, *50*, 4681−4698.
(81) Czodrowski, P.; Hölzemann, G.; Barnickel, G.; Greiner, H.; Musil, D. Selection of Fragments for Kinase Inhibitor Design: Decoration Is Key. *J. Med. Chem.* **2015**, *58*, 457−465.
(82) Suruzhon, M.; Bodnarchuk, M. S.; Ciancetta, A.; Viner, R.; Wall, I. D.; Essex, J. W. Sensitivity of Binding Free Energy Calculations to Initial Protein Crystal Structure. *J. Chem. Theory Comput.* **2021**, *17*, 1806−1821.

https://doi.org/10.1021/acs.jctc.1c01198
J. Chem. Theory Comput. 2022, 18, 3894−3910