OneOPES, a combined enhanced sampling method to rule them all

Valerio Rizzi[▽,†,‡,¶] Simone Aureli[▽,†,‡,¶] Narjes Ansari[▽,§] and Francesco Luigi Gervasio[*,†,‡,¶,‖]

†*School of Pharmaceutical Sciences, University of Geneva, Rue Michel Servet 1, 1206, Genève, Switzerland*
‡*Institute of Pharmaceutical Sciences of Western Switzerland (ISPSO), University of Geneva, 1206, Genève, Switzerland*
¶*Swiss Institute of Bioinformatics, University of Geneva, 1206, Genève, Switzerland*
§*Atomistic Simulations, Italian Institute of Technology, Via Enrico Melen 83, 16152, Genova, Italy*
‖*Department of Chemistry, University College London, London, WC1E 6BT, United Kingdom*

E-mail: francesco.gervasio@unige.ch

[▽]V.R., S.A. and N.A. contributed equally to the manuscript.

Abstract

Enhanced sampling techniques have revolutionised molecular dynamics (MD) simulations, enabling the study of rare events and the calculation of free energy differences in complex systems. One of the main families of enhanced sampling techniques uses physical degrees of freedom called collective variables (CVs) to accelerate a system's dynamics and recover the original system's statistics. However, encoding all the relevant degrees of freedom in a limited number of CVs is challenging, particularly in large biophysical systems. Another category of techniques, such as parallel tempering, simulates multiple replicas of the system in parallel, without requiring CVs. However, these methods may explore less relevant high-energy portions of the phase space and become computationally expensive for large systems. To overcome the limitations of both approaches, we propose a replica exchange method called OneOPES that combines the power of multi-replica simulations and CV-based enhanced sampling. This method efficiently accelerates the crossing of both enthalpic and entropic barriers without the need for optimal CVs definition, parameters fine tuning nor the use of a large number of replicas, as demonstrated by its successful applications to protein-ligand binding and protein folding benchmark systems. Our approach shows promise as a new direction in the development of enhanced sampling techniques for molecular dynamics simulations, providing an efficient and robust framework for the study of complex and unexplored problems.

Introduction

The development of enhanced sampling techniques has allowed molecular dynamics (MD) simulations to explore considerably longer time-scales, unlocking the possibility of routinely studying rare events and calculating free energy differences in a number of complex problems[1–4].

One of the main families of enhanced sampling techniques tackles the problem by accelerating a system's dynamics along selected physical degrees of freedom (DOFs) called collective variables (CVs) and by recovering the original system's statistics by factoring out the deposited bias. In time, many CV-based sampling methods have been developed that are actively used by the community[4]. Typically, the maximum number of CVs that can be simultaneously biased is limited, due to the exponentially increasing cost of exhaustively exploring multi-dimensional free energy spaces. The precursor of such methods is Umbrella Sampling[5], while Metadynamics[6] is one of the most popular. In Well-Tempered Metadynamics (WT-MetaD)[7], one iteratively builds a bias potential that, for a long enough time, is proven to converge and provide exact statistical properties[8]. However, the definition of "long enough time" is determined by the CVs' capability to capture all the slow DOFs of a system that are relevant for the process studied.

The quality of the CVs crucially determines the simulation timescale needed to reach convergence[9], as the slowest ignored DOFs dictate the speed at which transitions between metastable states occur, making the standard MD timescale problem reappear and translate into a search for better CVs. Encoding all the relevant DOFs in a limited number of CVs is no easy task and it is especially unfeasible in the large and complex systems that characterise the biophysical world. Common CVs such as distances and angles are quite straightforward to figure out and do not require extensive knowledge of the problem at hand, but they are hopeless to capture complex many-body transitions. More sophisticated techniques to build optimal CVs by combining simpler ones, or by finding an optimal path between initial and final states, or by using machine learning on large datasets have been successful, but they still often require significant human effort or large amounts of data[10–29].

Another category of techniques addresses the timescale problem through parallelism by simulating multiple realisations of the system, or replicas, at the same time[30–32]. Typically, these methods do not require the introduction of CVs to be biased. Among them, replica exchange methods simulate parallel replicas differing from each other through the variation of internal system's parameters[30,33,34] and periodically attempt exchanging their coordinates, given a physical acceptance criterion. One prominent example of such methods is Parallel tempering[33,35] and its evolutions[36,37], where a progressive temperature increase in the replicas leads to all kinetic barriers being lowered and enthalpy-driven processes being accelerated. Parallel tempering methods have the advantage of being able to sample large phase space regions without the necessity to reach very long timescales and without being limited by missing slow DOFs in the CVs definition. However, high temperatures and no CV-defined direction can lead the simulations to explore

less relevant high-energy portions of the phase space, reducing overall efficiency and not improving their capability to overcome entropic barriers[38,39]. Furthermore, in large systems, a considerable number of replicas must often be employed to ensure effective exchanges, making Parallel Tempering methods expensive[40]. While solutions to some of these limitations are arising[41,42], a possible way forward is to combine the power of multi-replica simulations and CV-based enhanced sampling.

A number of successful methods have attempted, to varying degrees, to combine the power of both approaches, including Multiple-Walkers Metadynamics (MW-MetaD)[43], Parallel-Tempering Metadynamics (PT-MetaD)[44], Bias-Exchange Metadynamics (BE-MetaD)[45] and further evolutions[46–50]. These methods showed significant promise as the combination of CV and replica-based algorithms is able to efficiently accelerate the crossing of both enthalpic and entropic barriers without the necessity of an optimal CV development, as shown by their successful application to complex biological systems[51,52]. However, they also inherited some of the intrinsic limitations of their predecessors, namely a reduced but still present dependence of the free energy convergence on the quality of the CVs, a problematic setup of optimal parameters, and a significant computational cost. The rise of novel enhanced sampling techniques such as On-the-fly Probability Enhanced Sampling (OPES)[53] and its variants[54,55] has prompted us to formulate an OPES-based replica exchange method. Its aim is to provide a framework that produces converged results at a reasonable computational cost while being less reliant on the setup parameters and the CV quality. The overall strategy exploits the qualities of a combination of existing OPES variants in a parallel strategy that we call *OneOPES*.

In standard OPES[53], one estimates the unbiased probability distribution by depositing weighted Gaussian kernels along chosen CVs. In a thought-provoking paper[55], it was shown that, when CVs are excellent, the rapidly converging bias potential leads to a high transition frequency and very accurate results in a short computational time. On the other hand, when CVs are suboptimal, the OPES bias potential determines a slow phase space exploration that in turn forces simulations to extend for long times before reaching convergence. OPES explore[55] addresses this point and builds a more rapidly varying bias potential which leads to a faster phase space exploration, at the price of a slower and noisier convergence. Meanwhile, another conceptually different OPES variant called OPES Multithermal[54] has been developed, where a system simultaneously visits a range of temperature distributions without changing the thermostat or having to run multiple replicas.

In OneOPES, inspired by the approach of Ref. 47, we set up a mixture of replicas including OPES variants of different exploratory intensities and combine them with explicit replica exchange. All replicas include a standard OPES that sets a common baseline by building a bias potential that is sufficient to trigger transitions. The first replica is the most convergence-focused replica and only includes the standard OPES bias potential. As such, it will be used to calculate equilibrium properties through reweighting. The higher-order replicas are more and more exploratory, as they include OPES Explore with an increasing exploratory power regulated by the bias factor. The most exploratory

replicas also include OPES Multithermal so that the effect of suboptimal CVs is mitigated and all kinetic barriers are further lowered. As done in Ref[45,47], different CVs may be employed in different explorative replicas. In a nutshell, exchanges between exploratory and convergence-dedicated replicas ensure that the former simulations bring variety into the latter ones, as convergence-dedicated ones moderate the exuberance of the exploratory ones.

We test OneOPES on a set of case studies of increasing order of complexity, namely Alanine dipeptide, protein-ligand binding in Trypsin-Benzamidine, and the folding of Chignolin. In alanine dipeptide, we compare the use of ideal and bad quality CVs. In Trypsin-Benzamidine and chignolin, we perform OneOPES in combination with standard but suboptimal CVs that would undermine the convergence of the reconstructed free energy when used in combination with standard OPES and other CV-based approaches. Furthermore, we show that our new algorithm is able to provide at no additional cost significant features of the process such as entropy, enthalpy, and the melting temperature of the protein. Our results compare very favourably with existing state-of-the-art simulations[23,56,57]. The examples provide a scenario for the intended use that we envision for OneOPES: to efficiently exploit the available resources in the study of real-world applications, striking a balance between the human effort needed to design optimal CVs and the computational effort to run long simulations.

Methods

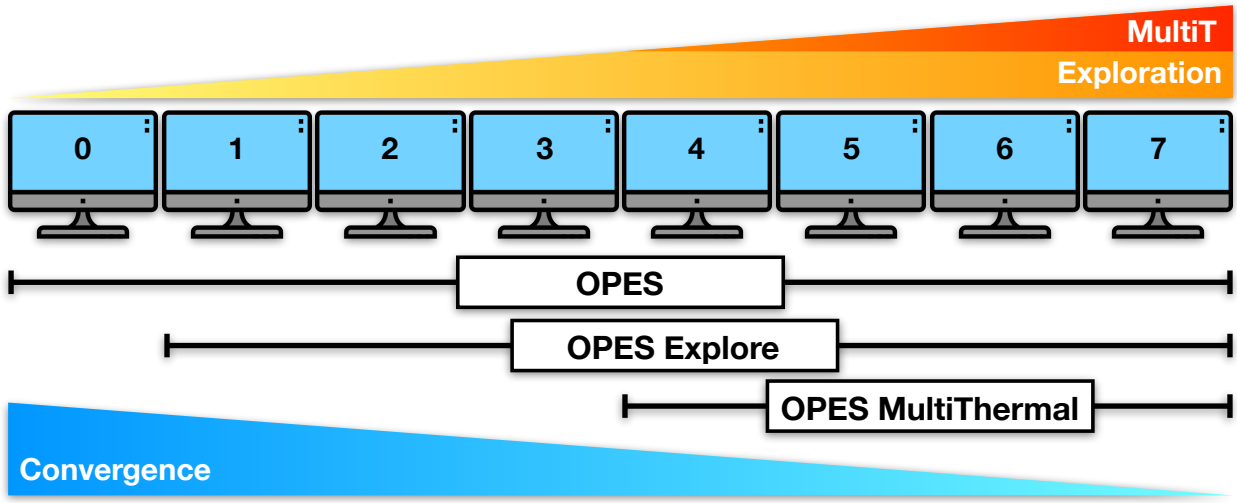

Figure 1: Schematic representation of the OneOPES replica exchange method. Replica **0** only includes OPES and is the most convergence-focused replica, while replica **7** is the most exploration-focused one as it includes all three OPES variants.

OneOPES is an explicit replica-exchange technique, whose framework is designed

as a progressive stratification of three different methodologies (see Fig. 1), which are gradually layered in a sequence of replicas. Here, we present the optimal strategy that we have devised to study the examples that we propose. In this strategy, only two parameters are optimised depending on the system. We include a total of 8 replicas per system, but the method can be trivially modified and tuned to include a different number of replicas. OneOPES incorporates three enhanced sampling techniques, i.e. OPES, OPES Explore, and OPES Multithermal, and is entirely implemented in the popular open-source plug-in PLUMED2[58]. Below, we give an overview of the main features of each of the three techniques and then further discuss their combined use in OneOPES.

The first layer of OneOPES is represented by OPES[53] that is the main simulation bias to drive transitions and reweight trajectories. The same OPES input parameters are used in all the replicas but, at variance with implementations such as MW[43], each replica builds its own local bias potential. OPES is a recent evolution of MetaD aimed at making the system sample a broadened target probability distribution $p_{\mathrm{O}}^{\mathrm{tg}}(\mathbf{s}) \propto [P_{\mathrm{O}}(\mathbf{s})]^{\frac{1}{\gamma_{\mathrm{O}}}}$ called the Well-tempered distribution, where $P_{\mathrm{O}}(\mathbf{s})$ is the unbiased marginal distribution, $\mathbf{s}$ are the chosen CVs and γ_{O} is the bias factor that controls the broadening. To achieve this, Gaussian Kernels are used to reconstruct $P_{\mathrm{O}}(\mathbf{s})$ which in turn determines the bias potential $V_{\mathrm{O}}(\mathbf{s})$ through a recursive strategy that at step n reads

$$\begin{aligned} V_{\mathrm{O}}(\mathbf{s}) &= \left(1 - \frac{1}{\gamma_{\mathrm{O}}}\right) k_{\mathrm{B}} T \log\left(\frac{P_{\mathrm{O}}(\mathbf{s})}{Z_{\mathrm{O}}} + \epsilon_{\mathrm{O}}\right) \\ P_{\mathrm{O}}(\mathbf{s}) &= \frac{\Sigma_k^n w_k G_k(\mathbf{s}, \mathbf{s}_k)}{\Sigma_k^n w_k} \end{aligned} \tag{1}$$

where k_{B} is the Boltzmann constant, T the temperature set by the thermostat, Z_{O} a normalisation factor, and $G_k(\mathbf{s}, \mathbf{s}_k)$ is the Gaussian kernel deposited at step k with a weight of $w_k = \mathrm{e}^{V_{k-1}(\mathbf{s}_k))/k_{\mathrm{B}}T}$. The regularisation term ϵ_{O} and γ_{O} conjure up the crucial `BARRIER` parameter ΔE_{O} through the relation $\epsilon_{\mathrm{O}} = \mathrm{e}^{-\Delta E_{\mathrm{O}}/(k_{\mathrm{B}}T(1-1/\gamma_{\mathrm{O}}))}$. ΔE_{O} sets a limit on the maximum amount of bias energy that OPES can inject in the system and is the first parameter to be tuned in OneOPES. Ideally, it should be comparable to or slightly larger than the expected maximum kinetic barrier of the process under investigation so that OPES is able to drive transitions away from the initial basin while not visiting irrelevant high energy states. Further details are provided in the original paper where the OPES method was first introduced[53].

The second layer of OneOPES is represented by OPES Explore[55] whose role is to help promote transitions and increase phase space exploration in the dynamics. It is embedded in replicas **1-7** as an auxiliary bias to assist the main one provided by OPES.

OPES Explore is a recent evolution of OPES whose bias potential at step n is as follows

$$
\begin{aligned}
V_{\mathrm{E}}(\mathbf{s}) &= (\gamma_{\mathrm{E}} - 1) k_{\mathrm{B}} T \log \left(\frac{p_{\mathrm{E}}^{\mathrm{tg}}(\mathbf{s})}{Z_{\mathrm{E}}} + \epsilon_{\mathrm{E}} \right) \\
p_{\mathrm{E}}^{\mathrm{tg}}(\mathbf{s}) &= \frac{1}{n} \sum_{k}^{n} G_k(\mathbf{s}, \mathbf{s}_k).
\end{aligned}
\tag{2}
$$

An unweighted average of Gaussian kernels is used to estimate on-the-fly the well-tempered distribution $p_{\mathrm{E}}^{\mathrm{tg}}(\mathbf{s})$, whose explorative power is regulated by γ_{E}. The resulting bias potential $V_{\mathrm{E}}(\mathbf{s})$ is more coarse than its corresponding OPES one and does not converge as rapidly to a quasi-static one, allowing an augmented exploration[26,55,59]. As we will see below, a gradual introduction of OPES Explore in the replicas is beneficial to the exchange rate between replicas, therefore we progressively increase the intensity of the OPES explore bias through the replicas, making sure that it always remains a secondary one and does not overcome the maximum value of the dominant bias $V_{\mathrm{O}}(\mathbf{s})$. To achieve this, in the examples we set $\Delta E_{\mathrm{E}} = k_{\mathrm{B}} T$ and introduce a gradual increase of the bias factor parameter in analogy with Ref. 47: $\gamma_{\mathrm{E}} = 1.5$ for replica **1**, $\gamma_{\mathrm{E}} = 3$ for replica **2** and $\gamma_{\mathrm{E}} = 5$ for replicas **4-7**.

The third layer is represented by OPES Multithermal[54] that is aimed at improving the convergence capabilities of the strategy in the presence of suboptimal CVs. In the examples, it is included in replicas **4-7**. By enhancing the fluctuations of the potential energy U, OPES Multithermal allows the system to sample the multicanonical ensemble corresponding to temperatures T_j with $j = 1, \ldots, N_{\mathrm{T}}$ in the range $[T_{\mathrm{min}}, T_{\mathrm{max}}]$. The free energy difference between each temperature $\Delta F(T_j)$ is iteratively updated while the bias potential is built. At step n, the bias potential is as follows

$$
\begin{aligned}
V_n^{\mathrm{MT}}(U) &= -\log \left(\frac{1}{N_{\mathrm{T}}} \Sigma_j^{N_{\mathrm{T}}} \; e^{-\frac{1}{k_{\mathrm{B}}} \left(\frac{1}{T_j - T} \right) U + \Delta F_n(T_j)} \right) \\
\Delta F_n(T_j) &= -\log \left(\frac{\Sigma_k^n e^{-\Delta \frac{1}{k_{\mathrm{B}}} \left(\frac{1}{T_j - T} \right) U + V_{k-1}^{\mathrm{MT}}}}{\Sigma_k^n e^{V_{k-1}^{\mathrm{MT}}}} \right).
\end{aligned}
\tag{3}
$$

By effectively heating and cooling the system, OPES Multithermal helps to overcome free energy barriers in a similar fashion to Parallel-Tempering techniques. It is particularly useful to accelerate the sampling along unknown DOFs that are not taken into consideration by the CVs $\mathbf{s}$. The temperature range is the second key parameter to tune in the OneOPES strategy. Intuitively, its optimal value must strike a balance between being broad enough to significantly help to sample and not too broad to trigger transitions to high energy states.

All in all, at any given step, each replica i presents a potential energy $U_i(x_i)$ and a

total bias potential $V_i^{\mathrm{TOT}}(x_i)$ given by the sum of the OPES biases (V_{O}, V_{E} and V^{MT}) that are applied to it. If we define $\overline{U}(x_i)) = U(x_i) + V_i^{\mathrm{TOT}}(x_i)$, swaps of coordinates between neighbouring replicas i and j are attempted and regulated by the Metropolis–Hastings algorithm with an acceptance of

$$\alpha = \min\left(1,\ \mathrm{e}^{\frac{1}{k_{\mathrm{B}}T}\left(-\overline{U}_i(x_j)+\overline{U}_i(x_i)-\overline{U}_j(x_i)+\overline{U}_j(x_j)\right)}\right). \tag{4}$$

In the replica exchange literature, it is known that the frequency at which exchanges are attempted is an important parameter to optimise. Here, for simplicity, we set it to be always equal to 20 times the OPES deposition `PACE` parameter, which in turn is set to 500 integration steps. As a rule of thumb, exchange acceptances about or above 30% are often considered ideal. Large temperature intervals $[T_{\mathrm{min}}, T_{\mathrm{max}}]$ applied directly on replica **4** can hamper exchanges with replica **3** and act as an exchange bottleneck. To alleviate this problem, one can use a gradual increase in the temperature range between replicas as we will show in some of the examples.

The combination of the three layers eases the creation of a double gradient along our replicas framework, i.e. an "exploration gradient" (from **0** to **7**) and a "convergence gradient" (from **7** to **0**), see Fig. 1. The most explorative replicas act as a generator of transitions between different states, that are distilled towards the convergent replica **0**. If relevant slow DOFs are not included in the CVs, these transitions occur out-of-equilibrium thanks to the exploratory power of Opes Explore or the energy fluctuations of Opes Multithermal. These out-of-equilibrium transitions allow to visit a system's free energy minima possibly bypassing the correct transition state region. Therefore, the resulting free energy surface (FES) from these calculations would reproduce well the minima, but would not be able to reliably describe the transition state region. As we will see in the examples, the better the CV used, the better reconstruction of the FES in all regions, including the transition state.

Computational details

All calculations are run using the GROMACS 2022.4 engine[60] patched with the PLUMED 2.8 plugin[58] with the exchange algorithm implemented in Ref.[46]. All the simulation details are provided in the Supporting Information (SI). We apply OneOPES to study three standard biophysical examples: conformational changes in Alanine dipeptide, protein-ligand binding in Trypsin-Benzamidine, and protein folding in the Chignolin miniprotein. In each case, we perform 5 independent simulations and calculate the corresponding free energy difference as a function of simulation time. We compare the average and standard deviation of the independent simulations free energy difference and FES with exact or highly converged results. For replicating a general non optimised usage, we bias standard non-optimised CVs and limit the total simulation time to a relatively short one. In Alanine dipeptide we bias first the ideal ϕ angle and then we perform simulations biasing the suboptimal ψ angle alone. In Trypsin-Benzamidine we bias the funnel axis

z and radius r[61], in Chignolin we bias a harmonic linear discriminant analysis (HLDA) CV based on six interatomic contacts within the protein[18,62]. To highlight the impact of OPES Multithermal, we compare OneOPES with a less powerful version that we call *OneOPES light* that does not include OPES Multithermal.

Results

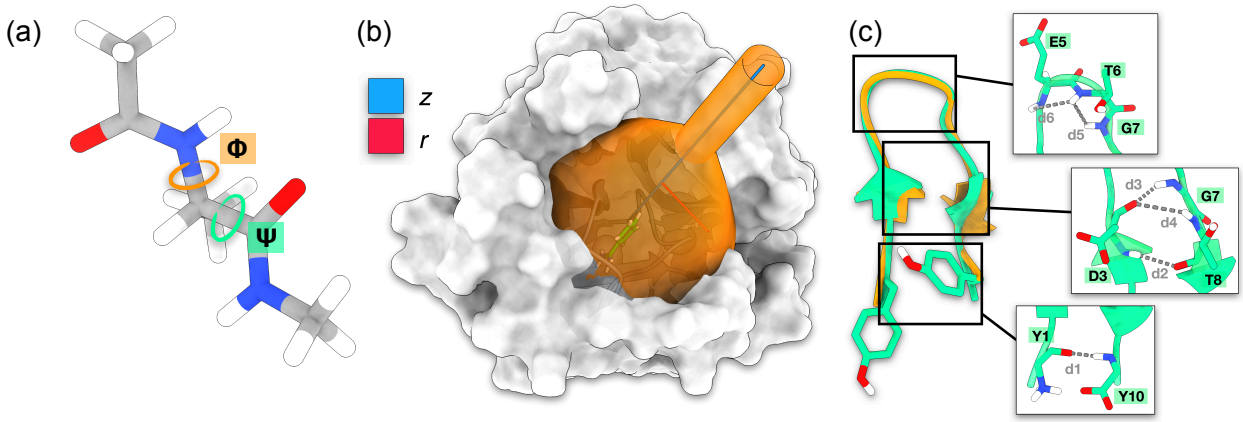

Figure 2: Graphical depiction of the systems that we investigate with OneOPES. In (a), we show Alanine dipeptide with the ϕ and ψ dihedrals coloured in orange and green, respectively. In (b), we present the Trypsin-Benzamidine complex, with the height z and the radius r of the funnel that we employ as CVs coloured in blue and red, respectively. In (c), we show the Chignolin miniprotein. We superimpose the Wild-Type structure in orange with the double mutant CLN025 that we simulate in green. The residues and the intraprotein contacts included in the HLDA CV are displayed in the panel insets and highlighted through grey dashed lines.

Alanine Dipeptide

Alanine dipeptide in vacuum is a prototypical system, routinely used in the early phase of method development[1]. The system presents two conformational states that depend upon dihedral angles ϕ and ψ. In this regard, ϕ is a nearly-ideal CV to capture the conformational change as it distinguishes well the two states and the corresponding transition state. Conversely, ψ is far from ideal as it barely distinguishes the states and is almost orthogonal to the transition state.

At first, we apply OneOPES on ϕ to verify the strategy's convergence in combination with optimal CVs. The simulations converge to the exact results in a timescale of a few nanoseconds (see Fig. S1). In particular, we notice that OneOPES light gives equivalent results with respect to the OneOPES ones and that the FES is well described in all regions,

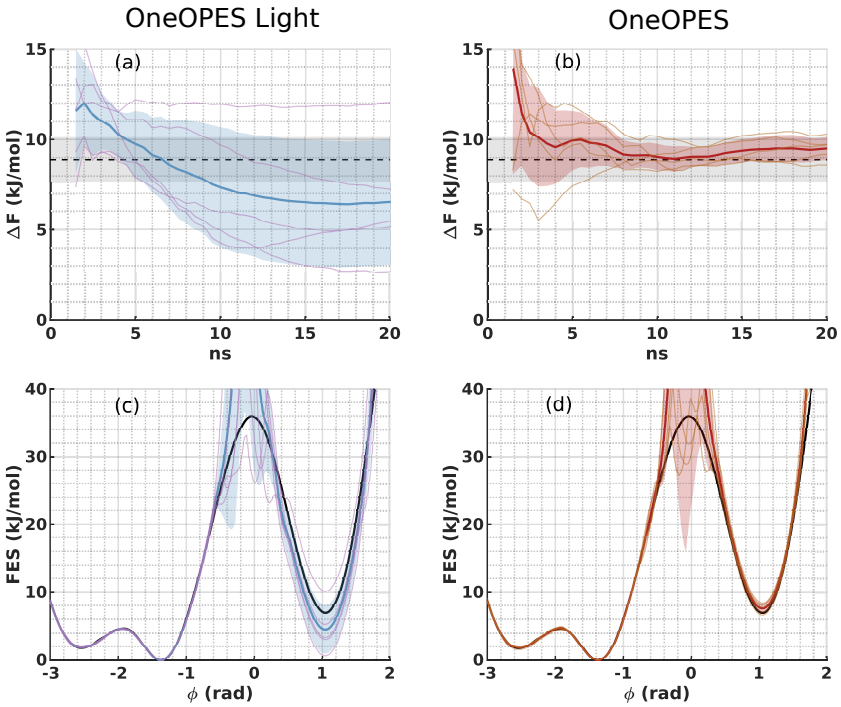

Figure 3: Alanine Dipeptide set of 5 independent simulations where we bias the suboptimal CV ψ with OneOPES light (panels (a) and (c)) and OneOPES (panels (b) and (d)). In (a) and (b), we show the average ΔF in time in a dark blue and a dark red solid line and their standard deviation in semitransparent regions in light blue and light red, respectively. ΔF values corresponding to individual OneOPES simulations are shown in solid purple and orange lines, respectively. The expected ΔF is indicated by a dashed black line with an error of 0.5 $k_B T$ in shaded grey. In (c) and (d), we show the one-dimensional FES reweighted over ϕ after 20 ns for OneOPES light and OneOPES simulations, respectively. The same colour scheme applies as in panels (a) and (b).

including the transition state. This is not surprising and confirms that using optimal CVs in enhanced sampling simulations, OneOPES included, is the best route to obtain well converged results in short simulation times.

Regrettably, optimal CVs are hard to acquire in realistic systems. To replicate the effect of using bad quality CVs, we perform a more demanding test on Alanine Dipeptide by biasing only the suboptimal CV ψ. In Fig. 3 we show the resulting free energy difference and the FES of five independent OneOPES simulations. In panels (a) and (c) we show the OneOPES light results where, while the average ΔF and FES roughly agree with the expected outcome, their standard deviation is large and has no tendency of diminishing for longer simulation time.

In panels (b) and (d) we use OneOPES and observe a much improved match between mean values and exact results, with a variance between simulations that shrinks in time. In presence of very ineffective CVs, it is clear that the role of OPES Multithermal is

crucial to bring the simulations to convergence. As expected, we notice that, while the free energy in the basins closely matches the exact one, the transition region does not. This occurs because the transition region is not effectively sampled and it is skipped over by the high-energy transitions triggered by OPES Multithermal.

In the Figs. S4 and S5, we compare the ϕ dynamics and the 2-dimensional FES of replica **0** and **7** respectively in OneOPES light and OneOPES simulations. As expected, we notice that, when biasing the suboptimal CV ψ, only the most explorative replica **7** is able to sample the transition state region.

Trypsin-Benzamidine

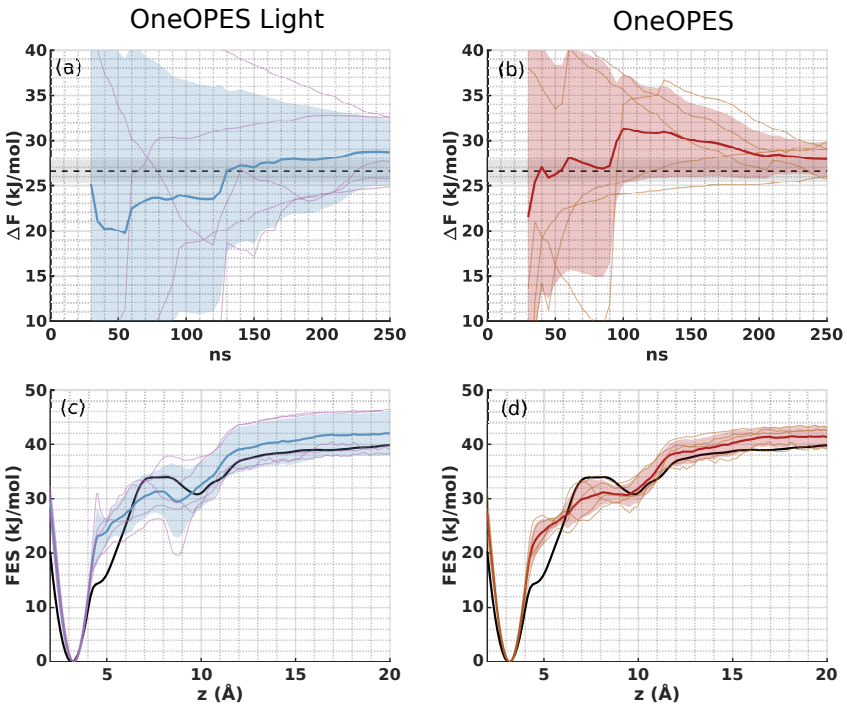

Figure 4: Set of 5 independent Trypsin-Benzamidine simulations where we bias the funnel coordinates z and r[61] with OneOPES light (panels (a) and (c)) and OneOPES (panels (b) and (d)). In (a) and (b), we show the average ΔF in time in a dark blue and a dark red solid line and their standard deviation in semitransparent regions in light blue and light red, respectively. ΔF corresponding to individual simulations are shown in solid purple and orange lines, respectively. The expected ΔF is taken from Ref.[56] and is shown in a dashed black line with an error of 0.5 $k_B T$ in shaded grey. In (c) and (d), we show the one-dimensional FES reweighted over z after 250 ns for OneOPES light and OneOPES simulations, respectively. The same colour scheme applies as in panels (a) and (b).

A more arduous test is the ligand-binding benchmark Trypsin-Benzamidine. While

this system has been routinely used for years as a benchmark for ligand-binding methods[61,63–73], it is far from trivial and still offers a significant challenge. High-resolution crystallographic experiments have recently demonstrated that individual water molecules play a crucial role in the system's binding/unbinding process[74], so that the introduction of specialised water-focused CVs proved decisive in bringing simulations to convergence in our recent work[56]. While the information provided by such CVs is invaluable, it is nevertheless an unfeasible task to replicate its development and optimisation in a high-throughput context.

In this paper, we pursue a different approach and we test OneOPES in combination with standard CVs that only capture the motion of the ligand with respect to the binding pocket and neglect the water dynamics (see SI). These CVs are clearly not optimal for the problem at hand but represent a more suitable option for future applications where one wants to extract binding free energies of sufficient quality from a number of systems, without focusing on a case-by-case CV optimisation.

In Fig. 4 we compare the free energy difference and the FES of OneOPES light (panels (a) and (c)) and OneOPES (panels (b) and (d)) with the converged results from Ref.[56]. The mean binding free energy value of 5 OneOPES light simulations reaches the expected one in about 125 ns and its standard deviation slowly decreases in time, so that at 250 ns it is about 4 kJ/mol. The slow convergence of different simulations to the expected result is determined by ignored DOFs in the CVs that here we know include long-lived water molecules directly affecting the binding free energy.

In the OneOPES simulations, we observe an agreement in mean binding free energy value that is accompanied by an improved consistency among different simulations. The binding free energy standard deviation at 250 ns reaches a value of about 2 kJ/mol. The faster convergence of independent simulation is evidently connected with OPES Multithermal speeding up the dynamics of long-lived water molecules. Further analysis of this effect can be found in Figs. S6-S7 where we present the CVs dynamics, including the water-focused Deep-LDA CV[56], and the 2-dimensional FES. Once again, in both cases, we notice that the transition state region of the FES is not well reproduced which indicates that the CVs in use are suboptimal and trigger high-energy transitions.

Chignolin

Protein folding examples are difficult to study with CV-based enhanced sampling methods due to the difficulty for CVs to capture a process that is characterised by a number of intermediate metastable states[76]. A further source of problems is the fact that a protein's folded and unfolded state are intrinsically different, as the former is enthalpy dominated while the latter is entropy dominated. Developing CVs that are optimal in capturing the whole folding/unfolding transition is a very complex task that today can be accomplished only in the simplest cases[23].

We focus on such a case and use OneOPES to study the folding of CLN025, a double mutant (G1Y, G10Y) of the Chignolin miniprotein, a system largely employed in the last decade as a benchmark to study fast-folding proteins through both long un-

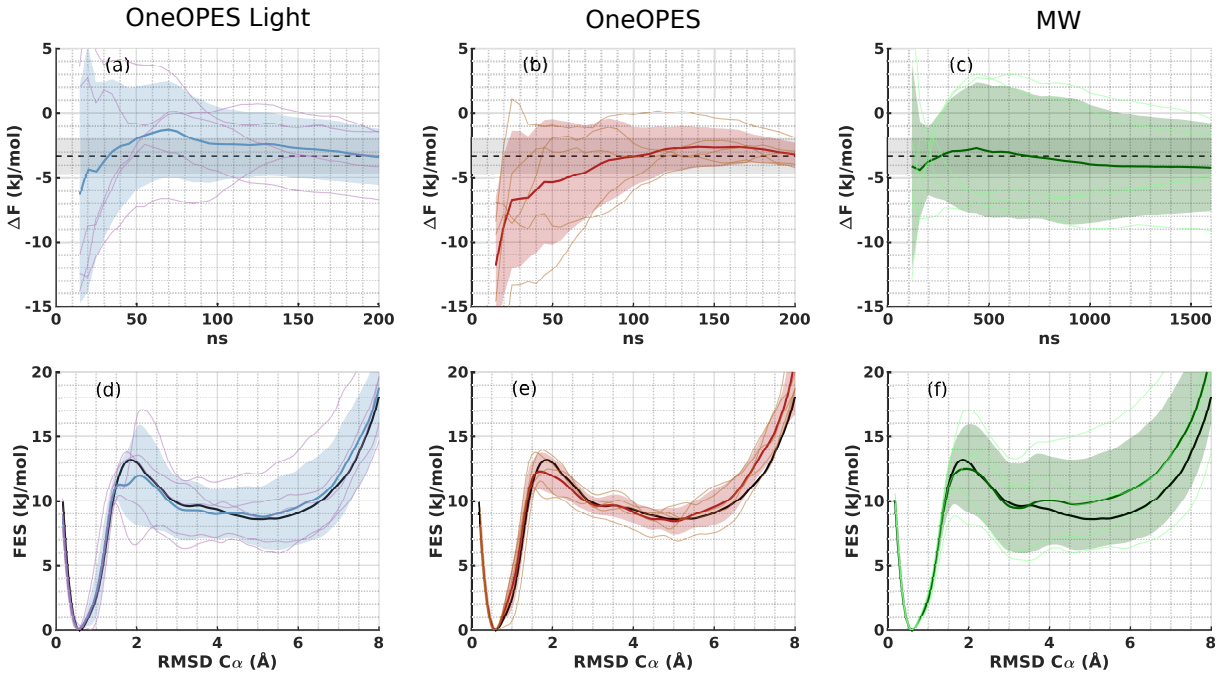

Figure 5: Set of 5 independent Chignolin simulations where we bias the HLDA CV[18,62] with OneOPES light (panels (a) and (d)), OneOPES (panels (b) and (e)) and MW (panels (c) and (f)). In (a),(b) and (c), we show the average folding free energy ΔF in time in a dark blue, a dark red and a dark green solid line and their standard deviation in semitransparent regions in light blue, light red and light green, respectively. ΔF corresponding to individual simulations are shown in solid purple, orange and aquamarine lines, respectively. The expected ΔF is taken from Ref.[75] ans is indicated by a dashed black line with an error of 0.5 $k_B T$ in shaded grey. In (d), (e) and (f), we show the one-dimensional FES reweighted over the RMSD Cα at the end of the simulations for OneOPES light, OneOPES and MW, respectively. The same colour scheme applies as in panels (a), (b) and (c).

biased molecular dynamics simulations[75] and more recently enhanced sampling simulations[23,54,62,77–80]. We will employ a relatively simple CV that is based on the linear combination of six interprotein contacts whose weights are obtained through harmonic linear discriminant analysis (HLDA)[18,62] (see SI for additional details). Following the scheme used in the previous examples, we perform 5 OneOPES Light and 5 OneOPES simulations, and compare their results with Ref.[75]. As a title of comparison with a commonly used technique, we also perform 5 MW simulations with an analogous setup as the OneOPES (see SI).

As visible in Fig. 5 (a), OneOPES Light average folding free energy reaches the expected value of $\Delta F = -3.3$ kJ/mol at around the 100 ns with a spread in the estimates of 2 kJ/mol, which remains about constant in time. In Fig. 5 (b) we show the corresponding OneOPES results and see that, while average ΔF is still achieved in $\sim$100 ns,

the spread between estimates decreases up to $\sim$1 kJ/mol. Once again, the use of OPES Multithermal helps in promoting a consistent convergence among different simulations. The MW results in Fig. 5 (c), show a good agreement in the average ΔF but also a very large variation between the simulations, with little tendency for the error to shrink for longer time.

Fig. 5 (d), (e) and (f) show the average and individual 1D FES reweighted along the RMSD in OneOPES Light, OneOPES and MW respectively. While all methods capture well the folded state minimum, only OneOPES shows a consistent agreement with the expected result in the unfolded state between different simulations. Remarkably, the FES is reproduced well also in the barrier region which signals the CV's ability to drive transitions close to the transition state.

OPES Multithermal allows the evaluation of physical properties in a range of temperatures away from the temperature set by the thermostat $T = 340K$ through a reweighting procedure[54], so we exploit this feature and use the simulations' replica **4** to estimate the system's folding ΔF between 330K and 350K (see Fig. 11 SI). Through this we estimate the melting temperature of Chignolin to be $388K \pm 14K$, which is in good agreement with the value of $381K$ with a 68% confidence interval of $361 - 393K$ from Ref. 75. Moreover, by performing a linear fit of the Van 't Hoff equation $\Delta F = \Delta H - T\Delta S$ we can also estimate the enthalpy ΔH and entropy $-T\Delta S$ of folding, which are -29.7 ± 4.1 kJ/mol and 26.1 ± 4.1 kJ/mol respectively (see Tab. S1 for more fitting data).

Conclusions

Collective variable-based enhanced sampling MD simulations rely on optimal CVs that approximate the reaction coordinate and encapsulate all the relevant slow DOFs so that they can accelerate their sampling. A serious drawback of these approaches is the effort required to define optimal CVs that capture complex processes such as folding, the flexibility of a receptor[81,82] or the role of water at a ligand/protein's interface[83–93]. Methods based on machine learning are increasingly successful in providing optimal CVs, but they need significant amount of data that is not always available[10,12,14–29].

As a result, the use of suboptimal CVs is often necessary. To address this problem, we propose a novel replica-exchange framework named OneOPES, based on the combination of the recently developed OPES method and its variants OPES Explore and OPES Multithermal. OneOPES is able to compensate the CVs shortcomings by setting up a hierarchy of replicas in convergence and explorative power, making out-of-equilibrium barrier crossings occur mostly on explorative replicas and letting configurations exchange towards convergence-focused replicas. We have shown that OneOPES can consistently recover the free energy of the increasingly complex examples that we benchmarked (i.e. Alanine dipeptide, Trypsin-Benzamidine, and Chignolin) within an error of 0.5 $k_B T$ at a reasonable computational cost, even in combination with suboptimal CVs. At the same time, it unlocks the possibility to infer thermodynamical properties of the system under investigation such as the enthalpy, the entropy, and the melting

temperature.

We emphasise that although OneOPES is very effective and represents a significant advance over other OPES and Metadynamics variants, we do not expect it to be miraculous in combination with very poor quality CVs. If the CVs are not relevant or inadequate for the problem at hand, OneOPES would still fail to converge. Still, in combination with reasonable CVs that distinguish the important states of the system but are not fine-tuned (since crucial but hard-to-capture DOFs are missing from the chosen set of CVs) the correct free energy landscape can be still recovered with a reasonable amount of sampling. Such an approach would fit well with the study of complex biophysical phenomena, ranging from conformational changes to ligand binding. In particular, we believe that our results endorse OneOPES as a valuable tool that can be safely used beyond benchmark systems in the study of complex and so far unexplored systems whose optimal CVs are still unknown. In venturing in this direction, OneOPES can be easily combined with state-of-the-art machine-learning CV design techniques, stretching to the limit the boundary of the systems that can be studied by modern computational techniques.

Data Availability

OneOPES is implemented in PLUMED[58], from version 2.8 onward, in combination with GROMACS[46,60]. The input files to replicate all the simulations and the analysis scripts will be soon uploaded on the PLUMED NEST repository[94] https://www.plumed-nest.org.

Supplementary Information Available

Acknowledgements

We acknowledge PRACE and the Swiss National Supercomputing Centre (CSCS) for large supercomputer time allocations on Piz Daint, project IDs: pr126, s1107, s1169. We acknowledge the Swiss National Science Foundation and Bridge for financial support (projects number: 200021_204795 and 40B2-0_203628). We thank Michele Parrinello, Umberto Raucci, Dhiman Ray, Michele Invernizzi and Ioannis Galdadas for helpful discussions.

References

(1) Laio, A.; Gervasio, F. L. Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. *Reports on Progress in Physics* **2008**, *71*, 126601.

(2) Valsson, O.; Tiwary, P.; Parrinello, M. Enhancing Important Fluctuations: Rare Events and Metadynamics from a Conceptual Viewpoint. *Annual Review of Physical Chemistry* **2016**, *67*, 159–184.

(3) Yang, Y. I.; Shao, Q.; Zhang, J.; Yang, L.; Gao, Y. Q. Enhanced sampling in molecular dynamics. *The Journal of Chemical Physics* **2019**, *151*, 070902.

(4) Henin, J.; Lelievre, T.; Shirts, M. R.; Valsson, O.; Delemotte, L. Enhanced Sampling Methods for Molecular Dynamics Simulations [Article v1.0]. *Living Journal of Computational Molecular Science* **2022**, *4*, 1–60.

(5) Torrie, G.; Valleau, J. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics* **1977**, *23*, 187–199.

(6) Laio, A.; Parrinello, M. Escaping free-energy minima. *Proceedings of the National Academy of Sciences* **2002**, *99*, 12562–12566.

(7) Barducci, A.; Bussi, G.; Parrinello, M. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Physical Review Letters* **2008**, *100*, 020603.

(8) Dama, J. F.; Parrinello, M.; Voth, G. A. Well-Tempered Metadynamics Converges Asymptotically. *Physical Review Letters* **2014**, *112*, 240602.

(9) Pietrucci, F. Strategies for the exploration of free energy landscapes: Unity in diversity and challenges ahead. *Reviews in Physics* **2017**, *2*, 32–45.

(10) Ma, A.; Dinner, A. R. Automatic Method for Identifying Reaction Coordinates in Complex Systems. *The Journal of Physical Chemistry B* **2005**, *109*, 6769–6779.

(11) Branduardi, D.; Gervasio, F. L.; Parrinello, M. From A to B in free energy space. *Journal of Chemical Physics* **2007**, *126*.

(12) Pietrucci, F.; Andreoni, W. Graph Theory Meets Ab Initio Molecular Dynamics: Atomic Structures and Transformations at the Nanoscale. *Physical Review Letters* **2011**, *107*, 085504.

(13) Díaz Leines, G.; Ensing, B. Path Finding on High-Dimensional Free Energy Landscapes. *Physical Review Letters* **2012**, *109*, 020601.

(14) Pérez-Hernández, G.; Paul, F.; Giorgino, T.; De Fabritiis, G.; Noé, F. Identification of slow molecular order parameters for Markov model construction. *The Journal of Chemical Physics* **2013**, *139*, 015102.

(15) Tiwary, P.; Berne, B. J. Spectral gap optimization of order parameters for sampling complex molecular systems. *Proceedings of the National Academy of Sciences* **2016**, *113*, 2839–2844.

(16) M. Sultan, M.; Pande, V. S. tICA-Metadynamics: Accelerating Metadynamics by Using Kinetically Selected Collective Variables. *Journal of Chemical Theory and Computation* **2017**, *13*, 2440–2447.

(17) McCarty, J.; Parrinello, M. A variational conformational dynamics approach to the selection of collective variables in metadynamics. *The Journal of Chemical Physics* **2017**, *147*, 204109.

(18) Mendels, D.; Piccini, G.; Parrinello, M. Collective Variables from Local Fluctuations. *The Journal of Physical Chemistry Letters* **2018**, *9*, 2776–2781.

(19) Chen, W.; Ferguson, A. L. Molecular enhanced sampling with autoencoders: On-the-fly collective variable discovery and accelerated free energy landscape exploration. *Journal of Computational Chemistry* **2018**, *39*, 2079–2102.

(20) Hovan, L.; Comitani, F.; Gervasio, F. L. Defining an Optimal Metric for the Path Collective Variables. *Journal of Chemical Theory and Computation* **2019**, *15*, 25–32.

(21) Gkeka, P. et al. Machine Learning Force Fields and Coarse-Grained Variables in Molecular Dynamics: Application to Materials and Biological Systems. *Journal of Chemical Theory and Computation* **2020**, *16*, 4757–4775.

(22) Bonati, L.; Rizzi, V.; Parrinello, M. Data-Driven Collective Variables for Enhanced Sampling. *The Journal of Physical Chemistry Letters* **2020**, *11*, 2998–3004.

(23) Bonati, L.; Piccini, G.; Parrinello, M. Deep learning the slow modes for rare events sampling. *Proceedings of the National Academy of Sciences* **2021**, *118*.

(24) Hooft, F.; Pérez de Alba Ortíz, A.; Ensing, B. Discovering Collective Variables of Molecular Transitions via Genetic Algorithms and Neural Networks. *Journal of Chemical Theory and Computation* **2021**, *17*, 2294–2306.

(25) Trizio, E.; Parrinello, M. From Enhanced Sampling to Reaction Profiles. *The Journal of Physical Chemistry Letters* **2021**, *12*, 8621–8626.

(26) Raucci, U.; Rizzi, V.; Parrinello, M. Discover, Sample, and Refine: Exploring Chemistry with Enhanced Sampling Techniques. *The Journal of Physical Chemistry Letters* **2022**, 1424–1430.

(27) Sun, L.; Vandermause, J.; Batzner, S.; Xie, Y.; Clark, D.; Chen, W.; Kozinsky, B. Multitask Machine Learning of Collective Variables for Enhanced Sampling of Rare Events. *Journal of Chemical Theory and Computation* **2022**, *18*, 2341–2353.

(28) Mouaffac, L.; Palacio-Rodriguez, K.; Pietrucci, F. Optimal reaction coordinates and kinetic rates from the projected dynamics of transition paths. **2023**, 1–19.

(29) Ray, D.; Trizio, E.; Parrinello, M. Deep Learning Collective Variables from Transition Path Ensemble. **2023**, 1–22.

(30) Swendsen, R. H.; Wang, J.-S. Replica Monte Carlo Simulation of Spin-Glasses. *Physical Review Letters* **1986**, *57*, 2607–2609.

(31) Zuckerman, D. M.; Chong, L. T. Weighted Ensemble Simulation: Review of Methodology, Applications, and Software. *Annual Review of Biophysics* **2017**, *46*, 43–57.

(32) Husic, B. E.; Pande, V. S. Markov State Models: From an Art to a Science. *Journal of the American Chemical Society* **2018**, *140*, 2386–2396.

(33) Sugita, Y.; Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chemical Physics Letters* **1999**, *314*, 141–151.

(34) Mitsutake, A.; Okamoto, Y. Replica-exchange extensions of simulated tempering method. *The Journal of Chemical Physics* **2004**, *121*, 2491.

(35) Hansmann, U. H. Parallel tempering algorithm for conformational studies of biological molecules. *Chemical Physics Letters* **1997**, *281*, 140–150.

(36) Liu, P.; Kim, B.; Friesner, R. A.; Berne, B. J. Replica exchange with solute tempering: A method for sampling biological systems in explicit water. *Proceedings of the National Academy of Sciences* **2005**, *102*, 13749–13754.

(37) Wang, L.; Friesner, R. A.; Berne, B. J. Replica Exchange with Solute Scaling: A More Efficient Version of Replica Exchange with Solute Tempering (REST2). *The Journal of Physical Chemistry B* **2011**, *115*, 9431–9438.

(38) Nymeyer, H. How Efficient Is Replica Exchange Molecular Dynamics? An Analytic Approach. *Journal of Chemical Theory and Computation* **2008**, *4*, 626–636.

(39) Rosta, E.; Hummer, G. Error and efficiency of replica exchange molecular dynamics simulations. *The Journal of Chemical Physics* **2009**, *131*, 165102.

(40) Di Lorenzo, G. et al. Phosphorylation of FAM134C by CK2 controls starvation-induced ER-phagy. *Science Advances* **2022**, *8*, 1–18.

(41) Bonomi, M.; Parrinello, M. Enhanced Sampling in the Well-Tempered Ensemble. *Physical Review Letters* **2010**, *104*, 190601.

(42) Invernizzi, M.; Krämer, A.; Clementi, C.; Noé, F. Skipping the Replica Exchange Ladder with Normalizing Flows. *The Journal of Physical Chemistry Letters* **2022**, *13*, 11643–11649.

(43) Raiteri, P.; Laio, A.; Gervasio, F. L.; Micheletti, C.; Parrinello, M. Efficient Reconstruction of Complex Free Energy Landscapes by Multiple Walkers Metadynamics. *The Journal of Physical Chemistry B* **2006**, *110*, 3533–3539.

(44) Bussi, G.; Gervasio, F. L.; Laio, A.; Parrinello, M. Free-Energy Landscape for β Hairpin Folding from Combined Parallel Tempering and Metadynamics. *Journal of the American Chemical Society* **2006**, *128*, 13435–13441.

(45) Piana, S.; Laio, A. A bias-exchange approach to protein folding. *Journal of Physical Chemistry B* **2007**, *111*, 4553–4559.

(46) Bussi, G. Hamiltonian replica exchange in GROMACS: a flexible implementation. *Molecular Physics* **2014**, *112*, 379–384.

(47) Gil-Ley, A.; Bussi, G. Enhanced Conformational Sampling Using Replica Exchange with Collective-Variable Tempering. *Journal of Chemical Theory and Computation* **2015**, *11*, 1077–1085.

(48) Bonomi, M.; Camilloni, C.; Cavalli, A.; Vendruscolo, M. Metainference: A Bayesian inference method for heterogeneous systems. *Science Advances* **2016**, *2*, 1–9.

(49) Amirkulova, D. B.; White, A. D. Combining enhanced sampling with experiment-directed simulation of the GYG peptide. *Journal of Theoretical and Computational Chemistry* **2018**, *17*, 1840007.

(50) 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. *Journal of Chemical Theory and Computation* **2023**,

(51) Sutto, L.; Gervasio, F. L. Effects of oncogenic mutations on the conformational free-energy landscape of EGFR kinase. *Proceedings of the National Academy of Sciences* **2013**, *110*, 10616–10621.

(52) Mattedi, G.; Acosta-Gutiérrez, S.; Clark, T.; Gervasio, F. L. A combined activation mechanism for the glucagon receptor. *Proceedings of the National Academy of Sciences* **2020**, *117*, 15414–15422.

(53) Invernizzi, M.; Parrinello, M. Rethinking Metadynamics: From Bias Potentials to Probability Distributions. *The Journal of Physical Chemistry Letters* **2020**, *11*, 2731–2736.

(54) Invernizzi, M.; Piaggi, P. M.; Parrinello, M. Unified Approach to Enhanced Sampling. *Physical Review X* **2020**, *10*, 041034.

(55) Invernizzi, M.; Parrinello, M. Exploration vs Convergence Speed in Adaptive-Bias Enhanced Sampling. *Journal of Chemical Theory and Computation* **2022**, *18*, 3988–3996.

(56) Ansari, N.; Rizzi, V.; Parrinello, M. Water regulates the residence time of Benzamidine in Trypsin. *Nature Communications* **2022**, *13*, 5438.

(57) Piana, S.; Robustelli, P.; Tan, D.; Chen, S.; Shaw, D. E. Development of a Force Field for the Simulation of Single-Chain Proteins and Protein–Protein Complexes. *Journal of Chemical Theory and Computation* **2020**, *16*, 2494–2507.

(58) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New feathers for an old bird. *Computer Physics Communications* **2014**, *185*, 604–613.

(59) Raucci, U.; Sanchez, D. M.; Martínez, T. J.; Parrinello, M. Enhanced Sampling Aided Design of Molecular Photoswitches. *Journal of the American Chemical Society* **2022**, 1–15.

(60) 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.

(61) Limongelli, V.; Bonomi, M.; Parrinello, M. Funnel metadynamics as accurate binding free-energy method. *Proceedings of the National Academy of Sciences* **2013**, *110*, 6358–6363.

(62) Mendels, D.; Piccini, G.; Brotzakis, Z. F.; Yang, Y. I.; Parrinello, M. Folding a small protein using harmonic linear discriminant analysis. *The Journal of Chemical Physics* **2018**, *149*, 194113.

(63) Gervasio, F. L.; Laio, A.; Parrinello, M. Flexible Docking in Solution Using Metadynamics. *Journal of the American Chemical Society* **2005**, *127*, 2600–2607.

(64) Buch, I.; Giorgino, T.; De Fabritiis, G. Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations. *Proceedings of the National Academy of Sciences* **2011**, *108*, 10184–10189.

(65) Tiwary, P.; Limongelli, V.; Salvalaglio, M.; Parrinello, M. Kinetics of protein-ligand unbinding: Predicting pathways, rates, and rate-limiting steps. *Proceedings of the National Academy of Sciences of the United States of America* **2015**, *112*, E386–E391.

(66) Plattner, N.; Noé, F. Protein conformational plasticity and complex ligand-binding kinetics explored by atomistic simulations and Markov models. *Nature Communications* **2015**, *6*, 7653.

(67) Dickson, A.; Lotz, S. D. Multiple Ligand Unbinding Pathways and Ligand-Induced Destabilization Revealed by WExplore. *Biophysical Journal* **2017**, *112*, 620–629.

(68) Brotzakis, Z. F.; Limongelli, V.; Parrinello, M. Accelerating the Calculation of Protein–Ligand Binding Free Energy and Residence Times Using Dynamically Optimized Collective Variables. *Journal of Chemical Theory and Computation* **2019**, *15*, 743–750.

(69) Wolf, S.; Lickert, B.; Bray, S.; Stock, G. Multisecond ligand dissociation dynamics from atomistic simulations. *Nature Communications* **2020**, *11*, 2918.

(70) Miao, Y.; Bhattarai, A.; Wang, J. Ligand Gaussian Accelerated Molecular Dynamics (LiGaMD): Characterization of Ligand Binding Thermodynamics and Kinetics. *Journal of Chemical Theory and Computation* **2020**, *16*, 5526–5547.

(71) Hüfner-Wulsdorf, T.; Klebe, G. Role of Water Molecules in Protein–Ligand Dissociation and Selectivity Discrimination: Analysis of the Mechanisms and Kinetics of Biomolecular Solvation Using Molecular Dynamics. *Journal of Chemical Information and Modeling* **2020**, *60*, 1818–1832.

(72) Votapka, L. W.; Stokely, A. M.; Ojha, A. A.; Amaro, R. E. SEEKR2: Versatile Multiscale Milestoning Utilizing the OpenMM Molecular Dynamics Engine. *Journal of Chemical Information and Modeling* **2022**, *62*, 3253–3262.

(73) Ray, D.; Stone, S. E.; Andricioaei, I. Markovian Weighted Ensemble Milestoning (M-WEM): Long-Time Kinetics from Short Trajectories. *Journal of Chemical Theory and Computation* **2022**, *18*, 79–95.

(74) Schiebel, J.; Gaspari, R.; Wulsdorf, T.; Ngo, K.; Sohn, C.; Schrader, T. E.; Cavalli, A.; Ostermann, A.; Heine, A.; Klebe, G. Intriguing role of water in protein-ligand binding studied by neutron crystallography on trypsin complexes. *Nature Communications* **2018**, *9*, 3559.

(75) Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, D. E. How Fast-Folding Proteins Fold. *Science* **2011**, *334*, 517–520.

(76) Noé, F.; De Fabritiis, G.; Clementi, C. Machine learning for protein folding and dynamics. *Current Opinion in Structural Biology* **2020**, *60*, 77–84.

(77) Shaffer, P.; Valsson, O.; Parrinello, M. Enhanced, targeted sampling of high-dimensional free-energy landscapes using variationally enhanced sampling, with an application to chignolin. *Proceedings of the National Academy of Sciences* **2016**, *113*, 1150–1155.

(78) Sultan, M. M.; Pande, V. S. Automated design of collective variables using supervised machine learning. *The Journal of Chemical Physics* **2018**, *149*, 094106.

(79) Belkacemi, Z.; Gkeka, P.; Lelièvre, T.; Stoltz, G. Chasing Collective Variables Using Autoencoders and Biased Trajectories. *Journal of Chemical Theory and Computation* **2022**, *18*, 59–78.

(80) Paissoni, C.; Camilloni, C. How to Determine Accurate Conformational Ensembles by Metadynamics Metainference: A Chignolin Study Case. *Frontiers in Molecular Biosciences* **2021**, *8*, 1–10.

(81) D'Annessa, I.; Raniolo, S.; Limongelli, V.; Di Marino, D.; Colombo, G. Ligand Binding, Unbinding, and Allosteric Effects: Deciphering Small-Molecule Modulation of HSP90. *Journal of Chemical Theory and Computation* **2019**, *15*, 6368–6381.

(82) Lukauskis, D.; Samways, M. L.; Aureli, S.; Cossins, B. P.; Taylor, R. D.; Gervasio, F. L. Open Binding Pose Metadynamics: An Effective Approach for the Ranking of Protein–Ligand Binding Poses. *Journal of Chemical Information and Modeling* **2022**,

(83) Ladbury, J. E. Just add water! The effect of water on the specificity of protein-ligand binding sites and its potential application to drug design. *Chemistry & Biology* **1996**, *3*, 973–980.

(84) Rhee, Y. M.; Sorin, E. J.; Jayachandran, G.; Lindahl, E.; Pande, V. S. Simulations of the role of water in the protein-folding mechanism. *Proceedings of the National Academy of Sciences* **2004**, *101*, 6456–6461.

(85) Homans, S. Water, water everywhere — except where it matters? *Drug Discovery Today* **2007**, *12*, 534–539.

(86) Juraszek, J.; Bolhuis, P. G. Rate constant and reaction coordinate of Trp-cage folding in explicit water. *Biophysical Journal* **2008**, *95*, 4246–4257.

(87) Pietrucci, F.; Marinelli, F.; Carloni, P.; Laio, A. Substrate Binding Mechanism of HIV-1 Protease from Explicit-Solvent Atomistic Simulations. *Journal of the American Chemical Society* **2009**, *131*, 11811–11818.

(88) Hummer, G. Under water's influence. *Nature Chemistry* **2010**, *2*, 906–907.

(89) Limongelli, V.; Marinelli, L.; Cosconati, S.; La Motta, C.; Sartini, S.; Mugnaini, L.; Da Settimo, F.; Novellino, E.; Parrinello, M. Sampling protein motion and solvent effect during ligand binding. *Proceedings of the National Academy of Sciences of the United States of America* **2012**, *109*, 1467–1472.

(90) Mahmoud, A. H.; Masters, M. R.; Yang, Y.; Lill, M. A. Elucidating the multiple roles of hydration for accurate protein-ligand binding prediction via deep learning. *Communications Chemistry* **2020**, *3*, 19.

(91) Rizzi, V.; Bonati, L.; Ansari, N.; Parrinello, M. The role of water in host-guest interaction. *Nature Communications* **2021**, *12*, 93.

(92) Ansari, N.; Rizzi, V.; Carloni, P.; Parrinello, M. Water-Triggered, Irreversible Conformational Change of SARS-CoV-2 Main Protease on Passing from the Solid State to Aqueous Solution. *Journal of the American Chemical Society* **2021**, *143*, 12930–12934.

(93) Samways, M. L.; Bruce Macdonald, H. E.; Taylor, R. D.; Essex, J. W. Water Networks in Complexes between Proteins and FDA-Approved Drugs. *Journal of Chemical Information and Modeling* **2023**, *63*, 387–396.

(94) Bonomi, M.; Bussi, G.; Camilloni, C.; Tribello, G. A. Promoting transparency and reproducibility in enhanced molecular simulations. *Nature Methods* **2019**, *16*, 670–673.
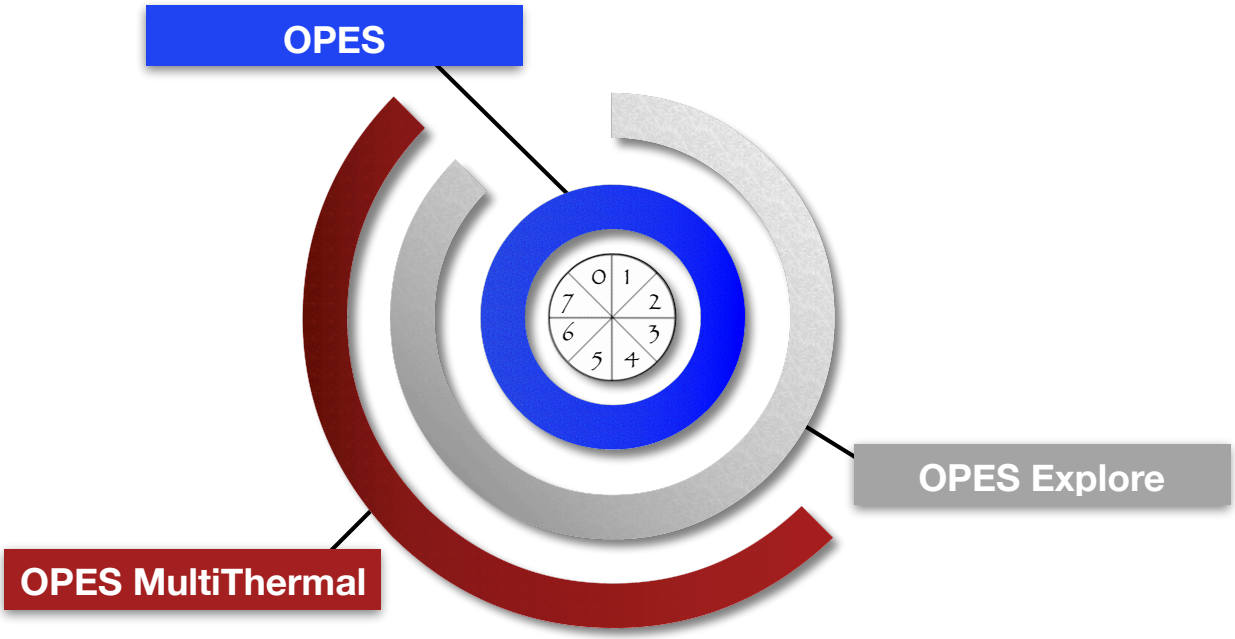
Table of Content Graphic