

## Review Article

# Computational methods for exploring protein conformations

 Jane R. Allison<sup>1,2</sup>

<sup>1</sup>Digital Life Institute and Maurice Wilkins Centre for Molecular Biodiscovery, School of Biological Sciences, University of Auckland, Auckland, New Zealand; <sup>2</sup>Biomolecular Interaction Centre, University of Canterbury, Christchurch, New Zealand

**Correspondence:** Jane R. Allison ([j.allison@auckland.ac.nz](mailto:j.allison@auckland.ac.nz))



Proteins are dynamic molecules that can transition between a potentially wide range of structures comprising their conformational ensemble. The nature of these conformations and their relative probabilities are described by a high-dimensional free energy landscape. While computer simulation techniques such as molecular dynamics simulations allow characterisation of the metastable conformational states and the transitions between them, and thus free energy landscapes, to be characterised, the barriers between states can be high, precluding efficient sampling without substantial computational resources. Over the past decades, a dizzying array of methods have emerged for enhancing conformational sampling, and for projecting the free energy landscape onto a reduced set of dimensions that allow conformational states to be distinguished, known as collective variables (CVs), along which sampling may be directed. Here, a brief description of what biomolecular simulation entails is followed by a more detailed exposition of the nature of CVs and methods for determining these, and, lastly, an overview of the myriad different approaches for enhancing conformational sampling, most of which rely upon CVs, including new advances in both CV determination and conformational sampling due to machine learning.

## Introduction

It is now generally accepted that in their biological environment, proteins exist not in a single, rigid structure, but as a dynamic ensemble of conformations distributed across a free energy landscape according to their Boltzmann-weighted probability of occurrence. The structure of this landscape is typically rugged, comprising a large number of conformations of similar energy and possible transition pathways between these, due to proteins having numerous degrees of freedom. The nature of the conformational ensemble and associated energy landscape depend on the protein and its environment, and may change in response to events such as interaction with cellular binding partners. Characterisation of the free energy landscape and thus the accessible conformational ensemble is essential in order to understand the function and malfunction of proteins.

Experiments that report on protein structure, such as X-ray diffraction [1], nuclear magnetic resonance [2] and Förster resonance energy transfer [3], generally provide either the average values of a large number of structural properties or distributions of a small number of structural properties. Neither is sufficient to fully characterise the conformational ensemble, which requires knowing not only the distributions of a large number of property values, but also which combinations of property values occur simultaneously, and the likelihood of each of these combinations. Molecular simulation can, at least in principle, provide this type of information. There are also myriad ways to bias simulations towards conformations that fit the experimental data [4–6].

The physico-chemical properties and molecular connectivity of a molecule or group of molecules are described by a ‘force field’, a mathematical relationship for the potential energy of the system in terms of the coordinates of its substituent particles (often atoms). The coordinates define the

Received: 2 June 2020  
 Revised: 7 July 2020  
 Accepted: 9 July 2020

Version of Record published:  
 5 August 2020

molecular conformation, and the potential energy relates to its probability of occurrence. In Markov chain Monte Carlo (MCMC), conformational space is stochastically sampled, whereas in molecular dynamics (MD) simulations, the main focus here, conformational space is deterministically sampled by propagating Newton's equations of motion. In its simplest formulation, this generates time-dependent conformational dynamics, although many of the methods described here trade time continuity for enhanced sampling.

The quality of the force field determines whether the conformations that are sampled in a simulation are realistic. The usefulness of a simulation, however, depends primarily on three factors:

- 1) The molecules included in/excluded from the simulation.
- 2) The level of detail required to examine the biological process or properties of interest.
- 3) The degree of sampling (in terms of temporal and spatial scale).

These three factors are inextricably linked: inclusion of more molecules, representation at a higher level of detail, and more extensive conformational sampling all increase the computational cost. The optimal choice, therefore, is always a trade-off, and depends on the available computational resources.

With regard to the first factor, historically, MD simulations of proteins typically comprised a single protein molecule in vacuum [7] and, later, in aqueous solution [8]. Nowadays, increases in computing power mean that large protein complexes comprising multiple subunits and/or ligands and protein-membrane systems can be simulated [9]. There is also increasing interest in simulating proteins in crowded conditions reminiscent of the interior of a cell [10,11]. While these developments are exciting, the trade-off between system size, the temporal and spatial scale of the motions of interest, and the computational effort required remains.

The choice of the most appropriate level of detail at which to run a simulation depends on the system and process of interest. Force fields are constantly being improved [12–14], but must always balance accuracy with computational efficiency. Standard atomic-level protein force fields, such as AMBER [15], CHARMM [16], and OPLS [17], explicitly model each atom or, in the case of the GROMOS [18] force fields, all heavy and non-aliphatic hydrogen atoms. The interactions between bonded and non-bonded atoms are described using simple mathematical functions, the parameters for which are fitted to structural or thermodynamic quantum-mechanical (QM) or experimental data [19].

To study the reaction mechanism of an enzyme, a mixed QM/MM representation is required. The reaction site is modelled using a QM description, allowing bond breaking and formation, and the remainder of the system is modelled at an atomic (molecular mechanics, MM) level. The QM part of the simulation is extremely computationally expensive, however, limiting the applicability of QM/MM approaches to small systems or short time-scales [20].

To model larger-scale processes with a high dependence on electronic polarisation, such as computing the binding free energy of a ligand to a protein, a polarisable model may be most appropriate [19,21]. Both approaches are too computationally expensive to be a good choice when these aspects are not of interest, however.

At the other end of the scale, coarse-grained representations [22,23], in which multiple atoms are subsumed into beads, allow much larger systems to be simulated for longer time-scales and thus sample larger-scale conformational motions. They can be limited, however, by the common need for additional terms to maintain native structure [24] (e.g. an elastic network model (ENM) [25]), which can limit conformational sampling. The recent implementation of a Gō model as an alternative structural restraint for the Martini coarse-grained model appears to hold much promise in this light [26]. As with QM/MM, mixed resolution or multi-scale simulations provide a compromise between the limitations and benefits of atomic-level and coarse-grained representations [27]. The different resolutions may be deployed sequentially, or concurrently [28]; in the latter, the resolution of each particle can be fixed, adaptable, or decoupled via virtual sites. These various strategies and their different implementations were recently reviewed by Machado et al. [29].

The main focus of this review is the third factor, the degree of sampling. An MD simulation can be considered ergodic if it samples all conformations accessible under the conditions in which it is run (e.g. temperature, pressure) at the correct probability of occurrence. Ergodicity is required if the goal is to determine the underlying free energy landscape. If the aim is to sample only part of the free energy landscape, for example, to follow a particular process, then the extent of sampling of the relevant regions of conformational space becomes important. Convergence of the sampling of conformational space towards the correct Boltzmann-weighted conformational ensemble is unfortunately difficult to measure, although a number of potential methods have been proposed [30]. Convergence may also refer to the approach of the estimated value of a quantity (e.g. the time-average of a structural property calculated from the simulation) to its true value

(e.g. the ensemble-average of the same structural property determined experimentally). Often, however, a simulation is being run because the true value is not known, in which case, the overlap between independent estimates (e.g. derived from multiple independent simulations or statistically independent segments of a single simulation) and their associated confidence intervals can be used as a proxy for convergence [30]. A simulation is then said to be converged when the estimated value of a property or properties no longer depends substantially on the length of the simulation or on its initial conditions.

Convergence is in principle possible using MD or MC, but in practice, even just sampling of biologically relevant time-scales (microseconds–milliseconds) is seldom achieved. In general, this is due to the rugged nature of the conformational free energy landscape, on which different regions, representing pools of accessible conformations, may be separated by high free energy barriers that are unlikely to be traversed on the simulation time-scale achievable using available computing resources. For example, an MD simulation with an atomistic force field requires an integration time steps on the order of femtoseconds ( $10^{-15}$  s), whereas biologically interesting events such as protein folding take milliseconds ( $10^{-3}$  s) or longer, and so require  $>10^{12}$  integration time steps. At each step, the interactions between tens or hundreds of thousands of atoms must be evaluated, such that it typically requires weeks or months of simulation on a high-performance computing system to obtain microsecond–millisecond length simulations. Even this may not be enough to assess the probability of all possible events or to estimate the relative population of all possible conformations, that is, to determine the free energy landscape. The sampling problem is exacerbated if the initial coordinates are of poor quality or structurally distant from the region(s) of the free energy landscape that are of interest, and is particularly fraught for intrinsically disordered proteins and protein regions [31,32].

One solution to the sampling problem has been to increase the speed of MD software packages through parallelisation and use of GPUs [33–37]. Another is to build dedicated hardware specifically designed for MD simulation, such as Anton [38] and MDGRAPE-4 [39]. Anton provided the first millisecond all-atom MD simulation of a protein, and its successor can perform multi-microsecond simulation of even larger systems in a single day [40].

MD simulations themselves can also be run in parallel. An early proponent of this, and of citizen science, is the *folding@home* project, in which a huge number of MD simulations are run on computers volunteered by private citizens [41]. Methods and software aimed at allowing scientists to easily run multiple simulations in parallel are also emerging [42–44]. Each simulation is typically much shorter than the time-scales of interest and covers just a tiny fraction of the complete conformational ensemble, and so must be combined and reweighted. An increasingly popular way of doing so is to construct Markov state models (MSMs), memoryless transition networks describing the populations and kinetics of interconversion between metastable conformational states [45,46]. MSMs have been used to study slow dynamical processes that would otherwise only be accessible using specialised computing infrastructure, including protein folding and conformational transitions. There are also path-based methods that use many short simulations to study slow processes and rare events, such as transition path sampling [47] and *milestoning* [48], although for the former, the simulations often need to be run sequentially.

Another approach, which forms the basis of this review, is the development of algorithmic methods for enhancing conformational sampling during an MD simulation. It is impossible to discuss modern methods for enhancing sampling without discussing collective variables (CVs), however. CVs are useful for both interpreting the huge amount of detailed data produced during an MD simulation, and directing conformational sampling to efficiently cover the underlying free energy landscape. This review therefore begins by discussing the nature of collective variables and methods for determining these, before describing different approaches to enhancing conformational sampling either to follow a particular process or to sample the free energy landscape, including the flurry of new methods that leverage machine learning techniques. The goal is to provide a general overview of the field rather than delve too deep into technical details.

## Collective variables

Conformational ensembles are inherently high dimensional, which makes them difficult to visualise and analyse. Each conformation generated during an MD simulation, for instance, is represented by the coordinates of the  $N$  atoms comprising the system, meaning that the conformational ensemble that is produced exists on a  $3N$ -dimensional free energy landscape. Only some of these dimensions are likely to be informative, however. Dimensionality reduction methods are increasingly being used to organise conformational ensembles, and in doing so, to determine which properties are important for organising the conformations or to a biological

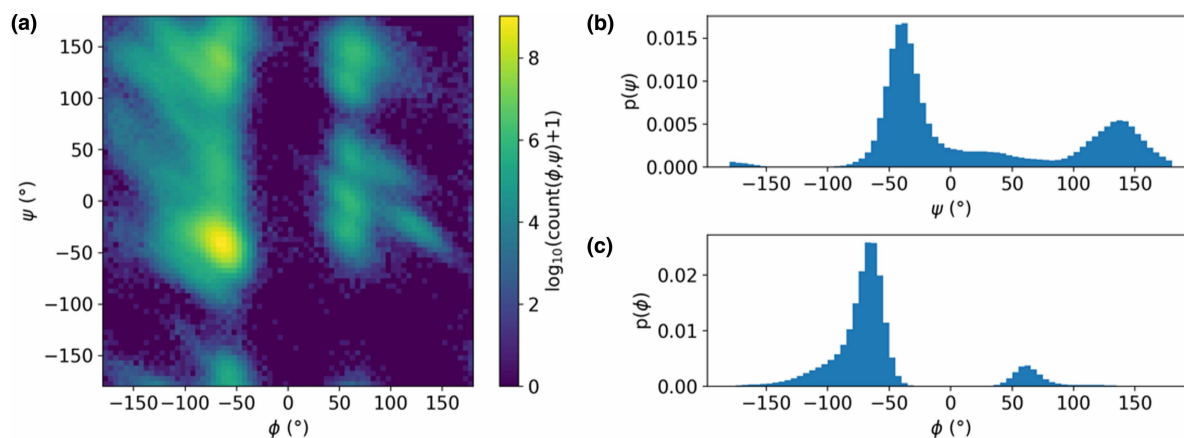
process of interest, and which are simply noise [49,50]. The generalised coordinates produced by dimensionality reduction are often called collective variables (CVs) (Figure 1), but are also referred to as reaction coordinates, order parameters, and features.

A CV is useful if conformational states of interest can be distinguished when the conformational ensemble is projected onto it. For instance, where the goal is to exhaustively sample a conformational ensemble, the conformational states of interest are the metastable states between which transitions are rare. For a process that involves transition between an initial and final conformation, a change in the CV represents progression along the path connecting the initial and final states. Sometimes more than one CV may be required to completely distinguish conformational states.

In addition to aiding the analysis and interpretation of conformational ensembles and MD simulation data, sampling can be directed along one or more CVs to enhance conformational sampling. To effectively enhance sampling, a CV should satisfy the following properties [51–53]:

- Clearly demarcate (meta)stable states of the system that are of interest;
- Account for the highest-variance or slowest conformational transitions;
- Limited in number (to allow the conformations along each CV to be exhaustively sampled);
- Be calculable as an explicit function of the system coordinates.

It is very difficult to intuit good CVs that will usefully enhance sampling, and accelerating irrelevant CVs may not improve the sampling over standard MD. Because of this, there is an enormous literature on CV design, including several recent reviews [52–56]. It is also important to consider interpretability of CVs—this is not necessary to enhance sampling, but is desirable for understanding mechanisms. It may be possible to, however, enhance sampling on one CV and then reweight the trajectory according to another more easily interpretable CV for analysis. Conceptually simple CVs include inter-atomic distances, angles or dihedrals, or the radius of gyration, of a subset of atoms in the system. Even in combination, these may not be adequate to describe the often complex conformational changes that take place during a biologically important event or across the entirety of a free energy landscape, however. In some cases, prior knowledge of the system or process can be used to determine appropriate CVs. It is also possible to use experimental observables directly as CVs [57,58]. In other cases, analysis of a preliminary simulation, or simulations of the end states of a transition, can provide clues.



**Figure 1. Illustration of projection of a free energy landscape onto commonly used CVs.**

(a) Ramachandran maps project the conformational free energy landscape onto the backbone  $\phi$  and  $\psi$  dihedral angle values. The example shown here is for a 100 ns MD simulation of hen egg white lysozyme (PDB ID: 1aki). (b, c) Projections of the conformational free energy landscape onto a single CV: (b)  $\psi$  and (c)  $\phi$ . All angle values are in degrees. Projection of the free energy landscape onto the combination of both backbone dihedral angles is useful because it clearly separates the two major regions of secondary structure, namely (right-handed)  $\alpha$ -helices and  $\beta$ -strands, although it is less effective at providing a more detailed degree of separation, such as between parallel and antiparallel  $\beta$ -strands — for this, additional CVs are required.  $\psi$  alone (b) could be a useful CV, as it preserves this separation, whereas projection onto  $\phi$  (c) conflates  $\alpha$ -helical and  $\beta$ -strand structure.

CV choice is delicate, however, and a poor choice can add user bias and reduce the reliability of the CV-enhanced sampling. This concern has stimulated the development of methods for using machine learning, with differing degrees of user input, to determine CVs, although in some cases, this may come at the expense of interpretability [55].

An early but still popular machine learning approach to determining high-variance CVs is to run a preliminary MD simulation and analyse it using principal component analysis (PCA). This identifies linear projections along which the conformational variance is maximal. The first few eigenvectors of the correlation matrix, which represent the leading high-variance modes, can be used as CVs [59]. More recently, harmonic linear discriminant analysis (HLDA) has been used to obtain CVs that are linear combinations of a small set of user-specified descriptors thought to be capable of discriminating between metastable states, based on short unbiased simulations of each state [60]. Combination of HLDA with neural networks allows compression of a larger number of descriptors into a lower-dimensional space [61].

Methods for learning CVs that describe slow modes from preliminary simulations have largely arisen from the MSM field, where they are critical to MSM construction. A key development was that of the variational approach to conformational dynamics (VAC) [62,63] and the more general variational approach of Markov processes (VAMP), which allow iterative assessment of many input functions i.e. CVs or linear combinations thereof, describing possible state decompositions. Input functions with the highest eigenvalue correspond to the indicator function whose eigenfunctions best approximate those of the continuous transfer operator describing transitions between metastable states [62–64].

A popular special case of VAC [65] is time structure-based (or time-lagged) independent component analysis (tICA) [65–68]. tICA allows creation of slow CVs suitable for enhanced sampling methods such as metadynamics [69] by linearly combining potentially simple CVs such as dihedral angles or pairwise contact distances such that their decorrelation time is maximised [45,70]. Importantly, it uses the tICs, which are the time equivalent of the principal components, to explicitly encode kinetic correspondence rather than using structural similarity as a proxy. Kernel tICA [71] and landmark kernel tICA [72] broaden the range of applicable input functions by removing the need for linearity.

Neural networks [73], including nonlinear [50,74,75] and time-lagged variational [76] autoencoders, and a Bayesian framework that operates according to similar principles [77], have also been developed to find the optimal slow CVs. In some cases (e.g. VAMPnets [73]), these methods map directly from molecular coordinates to Markov states, and so are not useful for identifying CVs for other purposes. Additionally, the slowest modes are not always the modes of interest [45,78], although a solution to this problem was recently provided by the deflated variational approach to Markov processes (dVAMP) [79].

Other recent approaches to learning CVs for enhanced sampling include spectral gap optimisation of order parameters (SGOOP) [80], which estimates the best combination of low-dimensional candidate CVs according to the maximum path entropy estimate of the spectral gap for dynamics viewed as a function of those CVs; EncoderMap, based on a neural network autoencoder [81]; and identification of the essential internal coordinates by using supervised machine learning to assign molecular structures to metastable states [82].

Machine learning and related methods are increasingly being used to provide a less heuristic approach to CV discovery. The success of machine learning, however, relies upon abundant and suitable training data; a chicken and egg situation in the case of the rare events for which enhanced sampling is required. A solution is to iterate between sampling and machine learning; such adaptive methods are discussed towards the end of this review.

## Enhanced sampling algorithms

Enhanced sampling algorithms speed up sampling of rare events, typically by adjusting the simulation temperature or Hamiltonian, with the latter including tuning the existing terms or adding bias potentials. While use of an enhanced sampling algorithm does not guarantee convergence, done correctly, it will improve conformational sampling and thus increase the likelihood of convergence. A huge number of enhanced sampling algorithms are available, and various methods for categorising these have been suggested [83,84]. Here, the different methods are divided into four categories, depending on whether sampling is enhanced along a CV, how the CV is determined, and whether or not the biasing and/or the CV adapt during the simulation:

1. Sampling enhanced by scaling the temperature
2. Sampling enhanced along one or more CVs

3. Sampling adaptively enhanced along one or more CVs
4. Sampling adaptively enhanced along one or more CVs learnt on-the-fly

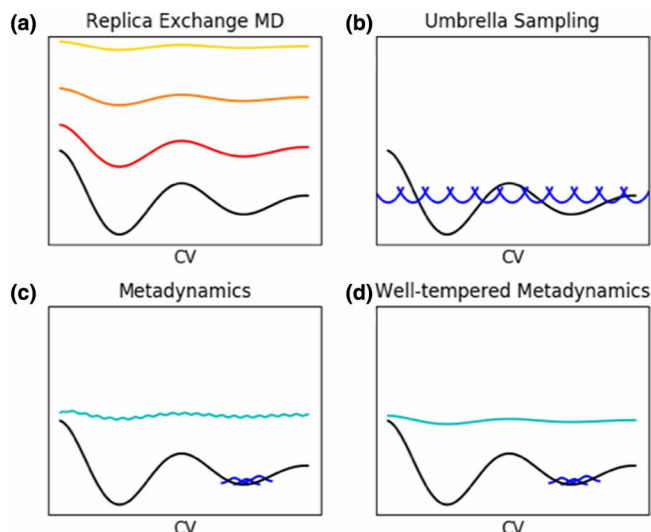
### Sampling enhanced by scaling the temperature

Notwithstanding the space dedicated to determining CVs, the simplest way to enhance sampling is to increase the temperature and thus the kinetic energy of the system, effectively lowering the heights of barriers between conformations (Table 1). Such approaches will only improve convergence if the major barriers to conformational sampling are not temperature-dependent [85].

Of these methods, Temperature replica exchange MD (T-REMD, parallel tempering), which combines MD and MC [86], deserves explanation, as it forms the basis for a wide range of methods. Multiple independent MD simulations are run in parallel across a ladder of different temperatures (Figure 2a). Exchange of coordinates between two temperatures is attempted at regular intervals and accepted or rejected according to the Metropolis criterion. Only the lowest temperature simulation, run at the temperature of interest, is kept and analysed. The exchanges mean that sampling is discontinuous, so special analysis methods are required to extract kinetic information and follow processes [87–89]. T-REMD was one of the earliest enhanced sampling methods and remains widely used. For small systems, the increase in computational resources required to run multiple simulations in parallel is compensated for by the increase in sampling. Unfortunately, however, efficient exchange between neighbouring replicas requires sufficient overlap of their energy distributions, and, as a consequence, the number of replicas required to cover a given temperature range grows according to the square root of the number of particles in the system [90].

**Table 1 Category 1: Sampling enhanced by scaling the temperature**

Category 1. No/general CV		
Name	Description	Citations
Simulated annealing	System is heated and then gradually cooled. May involve multiple iterations to sample different minima on the free energy landscape. One of the oldest techniques, but recently shown to increase sampling by at least an order of magnitude. Does not sample from a Boltzmann distribution.	[107,108]
Simulated tempering	Like simulated annealing, but samples from a Boltzmann distribution.	[109,110]
T-REMD: Temperature replica exchange MD	Multiple independent replicas in parallel, with coordinates exchanged at regular intervals. Sensitive to the choice of control parameters; substantial literature regarding their optimisation.	[86,91,111–117]
R-REMD: Reservoir REMD	T-REMD with the highest temperature replica replaced with a pre-generated reservoir of structures. Dependent on reservoir adequately covering conformational space.	[118–121]
M-REMD: Multiplexed REMD	T-REMD with several independent simulations at each temperature. Exchanges can occur between these and between temperatures. Takes advantage of highly parallel computing.	[122]
TAMD: Temperature-accelerated MD	Explores free-energy landscape of a large set of CVs at the physical temperature using an artificially high fictitious temperature.	[123]
REST and REST2: Replica exchange with solute tempering	Only the temperature of the solute differs between replicas. Increases the probability of exchanges by reducing the effective system size compared at each exchange attempt.	[91,124]
SGLD: Self-guided Langevin dynamics	SGLD increases the temperature of low-frequency motions only, with the SGLD temperatures scaled across replicas. The implementation of Wu et al. uses the SGLD partition function to remove the problems caused by the <i>ad hoc</i> force term of Lee and Olson [125].	[125,126]



**Figure 2. Schematic illustration of three key enhanced sampling methods.**

In all cases, the black line represents a free energy landscape projected onto a single CV, for simplicity. (a) Replica exchange MD, in which multiple independent replicas are run under different conditions, such as at increasingly high temperatures (red to yellow lines), which smooth the free energy landscape; (b) Umbrella sampling, where the blue harmonic potentials represent the ‘umbrellas’ that restraint conformational sampling along the CV; (c) metadynamics, where the potential energy surface is smoothed along one or more CVs by adding Gaussian functions (blue) to regions of the conformational space that have already been visited until ultimately (cyan) the entire surface is filled; (d) well-tempered metadynamics, where the rate and size of the Gaussian functions (blue) are reduced as sampling progresses, resulting in a smooth free energy surface (or a pre-specified distribution function, cyan) and avoiding over-filling.

## Sampling enhanced along one or more CVs

Perhaps the conceptually simplest approaches to CV-based enhanced sampling are those in which sampling is directed or restrained along a CV that describes a particular process (Table 2 and Figure 2b). These can be useful when the start and end point are known, and if a CV that describes the transition between these can be defined. They are less useful, however, for global exploration of conformational space.

Hamiltonian REMD (H-REMD) refers to REMD with each replica simulated under a different Hamiltonian (Figure 2a). T-REMD is a special case of H-REMD in which the scaling of the temperature is equivalent to scaling the whole Hamiltonian simultaneously [90]. The benefit of H-REMD is that only the part(s) of the Hamiltonian that limit conformational sampling can be scaled. In this way, H-REMD, like replica exchange with solute tempering (REST) [91], solves the T-REMD problem of the rapidly increasing numbers of replicas required as system size increases [90]. A wide variety of different modifications of the Hamiltonian have been suggested (Table 2). H-REMD is useful when there is little prior knowledge of the nature of the conformational ensemble or free energy landscape and, in particular, what the barriers to conformational sampling might be. Substantial simulation effort can be expanded on sampling modified Hamiltonians that are not of interest, however [55]. Additionally, the choice of how to modify the Hamiltonian can rely on and thus be biased by prior knowledge.

## Sampling adaptively enhanced along one or more CVs

Adaptive approaches to enhanced sampling (Table 3) learn about the underlying free energy surface during a simulation and use that knowledge to drive the simulation away from conformations (in CV-space) that have already been visited. In some cases, the size or nature of the added biasing potential may change during the simulation. These approaches not only enhance sampling, but also allow reconstruction of the free energy surface as a function of the chosen CVs [51,53]. Early approaches include accelerated MD [92], adaptive variations of umbrella sampling (US), and methods such as local elevation [93] and conformational flooding [94] that add a history-dependent potential to one or a few system properties, thus constructing the bias potential

**Table 2 Category 2: Sampling enhanced along one or more CVs**

Name	Description	Citations
SMD: <u>s</u> teered MD	An external force is applied to induce rare transitions along a CV to occur at a faster rate. Computational analogy to atomic force microscopy. Added force may induce physically unrealistic conformational transitions, and in general, does not sample from a Boltzmann distribution.	[127–129]
US: <u>u</u> mbrella <u>s</u> ampling	Uses a harmonic biasing potential to restrain the simulation to a series of windows along a pre-defined CV. If reweighted, can be used to determine the free energy surface and thus the change in free energy along the CV.	[130]
H-REMD: <u>H</u> amiltonian REMD	Like T-REMD, but each replica is simulated under a different Hamiltonian. Classic versions involve scaling the protein backbone and side chain dihedral angle potentials or the non-bonded interactions.	[90,131–137]
Resolution H-REMD	Each replica is simulated at a different level of resolution, e.g. atomic-level to coarse-grained.	[119,138,139]
Partial- and local-H-REMD	Only terms of the Hamiltonian involving the part of the system for which sampling is slow are exchanged.	[140]
2D-REMD	Two-dimensional H-REMD with scaling of temperature and inter-molecular interactions. Also used coarse-grained representation to calculate $K_d$ for an IDP allosteric regulator.	[141]
REAMD: <u>R</u> eplica <u>e</u> xchange of aMD	Combination of aMD with REMD; each replica has a different level of acceleration. Avoids the statistical reweighting problem of aMD.	[85,142]
ENM-H-REMD: <u>E</u> lastic <u>n</u> etwork <u>m</u> odel H-REMD.	Each replica is simulated with a different degree of a distance-dependent biasing potential that drives the structure away from its initial conformation in directions compatible with an ENM. Primarily enhances sampling around the initial structure.	[143]
HS-H-REMD: hydrogen bond <u>s</u> witching H-REMD	Exchanges take place between three replicas; two with either an attractive or repulsive hydrogen bonding potential added to the Hamiltonian. Similar performance to T-REMD with fewer replicas.	[144]

on-the-fly. Local elevation and conformational flooding are similar to what has become the primary method for adaptive enhanced sampling along CVs: metadynamics [95] (Figure 2c). Its main novelties were adaptation of the free energy rather than potential energy surface, and generalisation of the bias potential to act upon any CV or multidimensional set of CVs. These CVs may be as simple as the backbone dihedral angles, or can be almost arbitrarily complex. There are a huge number of variations of metadynamics, many of which involve combining it with other enhanced sampling techniques. Others, such as well-tempered metadynamics (Figure 2d), adjust the weight of the biasing potential as the biased free energy landscape approaches some target distribution (most simply, a uniform distribution), with the most recent versions doing so via machine learning. The underlying free energy landscape can then be reconstructed as the mirror image of the ultimate bias potential. Table 3 focuses on approaches aimed at increasing conformational sampling rather than calculation of kinetic properties. Interested readers are directed to recent comprehensive reviews dedicated to metadynamics for further details [51–53].

### Sampling adaptively enhanced along one or more CVs learnt on-the-fly

While traditionally, CVs are defined prior to running an enhanced sampling simulation, more recently developed adaptive sampling methods (compared here [96]) iterate between or combine these two phases, making them more applicable to systems where there is little prior knowledge from which to estimate appropriate CVs (Table 4). The learning phases utilise MSMs, tICA, and a variety of machine learning techniques.

### Emerging approaches to conformational sampling

In addition to being used to learn CVs, machine learning has been used to do away with the need to run MD or MC simulations or determine CVs almost entirely.



**Table 3 Category 3: Sampling adaptively enhanced along one or more CVs**

Part 1 of 2

Name	Description	Citations
aMD: <u>a</u> ccelerated MD	'Boost' potential applied when potential energy drops below a user-specified cut-off to increase rate of escape from minima. Reweighting of the resulting conformational ensemble to account for the applied bias is not always straightforward.	<a href="#">[92,145,146]</a>
aUS: <u>a</u> daptive US	Iterates between sampling along a CV according to an umbrella potential and updating the umbrella potential according to an estimate of the probability distribution along the CV to improve sampling of under-sampled regions.	<a href="#">[147,148]</a>
SH-US: <u>s</u> elf- <u>h</u> ealing US	Automatically updates the umbrella potential on-the-fly until the umbrella potentials cancel out the free energy profile.	<a href="#">[149]</a>
Multidimensional aUS	Like aUS, but with the umbrella potentials applied across more than one CV.	<a href="#">[150]</a>
Local elevation	Generates a history-dependent bias potential by adding Gaussians centred on the currently occupied value of one or more system properties to persuade the system to visit new areas of conformational space.	<a href="#">[93]</a>
Conformational flooding	Like local elevation but formulated more generally to act on coarse-grained conformational coordinates.	<a href="#">[94]</a>
LEUS: <u>L</u> ocal <u>e</u> levation <u>u</u> mbrella <u>s</u> ampling	A short LE build-up phase is used to construct an optimized biasing potential along conformationally relevant degrees of freedom that is then used in a (comparatively longer) US sampling phase.	<a href="#">[83]</a>
Metadynamics	Like local elevation, but the biases are added to the free energy rather than potential energy surface, and the bias potential is generalised to act upon any CV or multidimensional set of CVs.	<a href="#">[95]</a>
Multiple walkers (altruistic) metadynamics	Many metadynamics runs are performed in parallel, all of which contribute to filling in the free energy landscape.	<a href="#">[151]</a>
WTE metadynamics: <u>w</u> ell- <u>t</u> empered <u>e</u> nsemble metadynamics	The energy is used as collective variable to sample the well-tempered ensemble. Note that this is different to well-tempered metadynamics.	<a href="#">[152]</a>
Bias-exchange metadynamics	A number of independent metadynamics simulations are run in parallel, each biasing a different CV, with exchange of coordinates between biases. The REMD and metadynamics act synergistically to overcome barriers.	<a href="#">[153]</a>
Parallel-bias metadynamics	Single-replica variant of bias-exchange metadynamics in which the CV that is biased is switched during the simulation according to the Metropolis criterion, avoiding the need to have as many replicas as CVs.	<a href="#">[154]</a>
T-REMD (parallel tempering) metadynamics	Multiple metadynamics simulations are performed in parallel at different temperatures, all of which contribute to filling in the free energy landscape. Improves the exploration of low probability regions and sampling of degrees of freedom not included in the CV, but requires a large number of replicas for all but very small systems.	<a href="#">[155]</a>
REST metadynamics	Like T-REMD metadynamics, but only the solute experiences different temperatures.	<a href="#">[156]</a>
WTE-metadynamics REMD	Combines WTE-metadynamics with T-REMD by running WTE-metadynamics at each temperature. Overlap and thus exchange between replicas is increased, and canonical averages of properties of interest can be obtained with reweighting.	<a href="#">[157]</a>

Continued

Downloaded from <http://portlandpress.com/biochemsoctrans/article-pdf/48/4/1707/1991950/bst-2020-0193c.pdf> by guest on 30 September 2022

**Table 3 Category 3: Sampling adaptively enhanced along one or more CVs**

Part 2 of 2

Name	Description	Citations
<b>Metadynamics with on-the-fly adjustment of the biasing frequency or weight</b>		
WT-metadynamics: <u>w</u> ell- <u>t</u> empered metadynamics	The height of the Gaussian functions and the rate at which they are deposited decreases during the simulation and inversely to the time spent at a given value of the CV(s) to prevent over-filling.	[158]
TT metadynamics: <u>t</u> ransition- <u>t</u> empered metadynamics	Like WT-metadynamics, but decreases the height of the Gaussians according to the number of round trips between basins in the free energy landscape. Useful for calculating the free energy surface along a few well-chosen collective variables (CVs) at a time, but requires a priori estimation of the basin positions.	[159]
$\mu$ -tempered metadynamics	Like WT-metadynamics, but allows use of wide Gaussians and a high filling rate without slowing convergence.	[160]
WT-metadynamics-REMD	Multiple WT-metadynamics simulations are run in parallel, each biasing multiple CVs simultaneously. The degree of bias increases across the ladder of replicas.	[161]
Metabasin metadynamics	The energy level to which the metadynamics can fill the free energy landscape is restricted, to either a pre-defined level or relative to unknown barrier energies, with both these and the Gaussian shape estimated on-the-fly. Reduces need to carefully choose CVs to avoid sampling irrelevant high-energy regions.	[162]
<b>Metadynamics with on-the-fly adjustment of the biasing frequency or weight to achieve a target probability distribution function</b>		
OPES: <u>o</u> n- <u>t</u> he- <u>f</u> ly <u>p</u> robability- <u>e</u> nhanced <u>s</u> ampling	A recent reconsideration of metadynamics that begins with a coarse-grained estimate of the free energy landscape and converges towards a more detailed representation using a weighted kernel density estimation and on-the-fly compression algorithm.	[163]
VES: <u>v</u> ariationally <u>e</u> nhanced <u>s</u> ampling; deep-VES	Use an artificial neural network to determine a smoothly differentiable bias potential as a function of a pre-selected small number of CVs that drives the system towards a user-defined target probability distribution in which free energy barriers are lowered.	[164] [165]
TALOS: <u>t</u> argeted <u>a</u> dversarial <u>l</u> earning <u>o</u> ptimized <u>s</u> ampling	Uses a generative adversarial network competing game between a sampling engine and a virtual discriminator to construct the bias potential.	[166]

Deep generative MSMs (DeepGenMSMs) [97] use a generative neural network to learn a model that maps the high-dimensional conformational space to a low-dimensional latent space, describes transitions between metastable states in this latent space, and maps the latent space back to conformations. This model can then predict the future evolution of a system, including previously unseen conformations.

Boltzmann generators [98] use a deep generative neural network to construct invertible transformations between the conformational landscape and a simple Gaussian coordinate system. Sampling of the equilibrium conformational probability distribution, including states not included in the training, can then take place in the simple coordinate system, and the Gaussian coordinates transformed back into conformations. Training the neural network requires furnishing of high- and low-probability states, and thus requires some preliminary, possibly enhanced, conformational sampling [99]. Additionally, this method will require further development to be applicable to the high dimensional space of, e.g. a protein in explicit water [99]. Regardless, this is a very promising approach.

Another new method capable of predicting conformations not incorporated into the training data are dynamic graphical models (DGMs) [100]. The global molecular conformation is partitioned into local

**Table 4 Category 4: Sampling enhanced along one or more CVs learnt on-the-fly**

Name	Description	Citations
On-the-fly HTMD: on-the-fly <u>high-throughput MD</u>	Iterates between multiple short MD simulations (HTMD) and use of an MSM to learn a simplified model of the system to decide from where to respawn the next batch of simulations.	[167]
Extended DM-d-MD: extended <u>diffusion-map-directed MD</u> ; iMapD: <u>intrinsic map dynamics</u>	Uses diffusion maps, a non-linear manifold machine learning technique for dimensionality reduction to select regions of conformational space from an initial unbiased MD simulation from which to launch new rounds of MD simulations. Unbiased simulations are used because CVs based on diffusion maps do not explicitly map to atomic coordinates, and so cannot be used in US or metadynamics, which require calculation of the gradient of the CV with respect to the atomic coordinates [55]	[168,169]
VAC-metadynamics	Uses tICA to analysis an initial WT-metadynamics simulation to obtain more effective CVs that are used in a second WT-metadynamics simulation. Not strictly iterative.	[70]
RAVE: <u>reweighted autoencoded variational Bayes for enhanced sampling</u>	Iterates between enhanced sampling simulations and deep learning using variational autoencoders to learn an optimum but still physically interpretable reaction coordinate, as well as the probability distribution along this coordinate, which are then used to bias the enhanced sampling simulations.	[170]
REAP: <u>reinforcement learning based adaptive sampling</u>	Uses reinforcement learning to estimate the importance of CVs on-the-fly while exploring the conformational landscape. Requires an initial unbiased MD simulation from which to generate a dictionary of CVs and their trial weights.	[171]
MESA: <u>molecular enhanced sampling with autoencoders</u>	Iterates between umbrella sampling along trial CVs and using an auto-associative artificial neural network with a nonlinear encoder and decoder to learn CVs.	[172]

substructures, changes to which depend only on themselves and their neighbours. DGMs seem particularly well suited to IDPs and other systems with a very large number of metastable states.

In contrast with the general trend towards removing, as much as possible, human influence on the choice of CVs and thus the directions in which conformational sampling is enhanced, one intriguing method for enhancing sampling that takes the opposite approach is interactive MD, in which the user can add force directly by interacting with the simulation while it is running. While this functionality is not new [101,102], it has recently been given a new lease of life by taking advantage of virtual reality (VR) has been used to allow the user to interact directly with the molecule [103]. A recent update allows interactive ensemble MD simulations, and provides output suitable for MSM workflows [104].

## Perspectives

- Converged sampling of the entire free energy landscape, and thus conformational ensemble, is crucial to properly estimate equilibrium properties of molecular systems and to assess the likelihood of states and transition pathways between them.
- Current efforts increasingly use machine learning to determine optimal CVs along which to direct sampling and project the free energy landscape, and iterate between learning CVs and adaptive enhanced sampling techniques, thus reducing the impact of poor CV choice. The latest techniques do away with the need to learn CVs or carry out enhanced sampling altogether, other than to provide training data.

- As has long been the case in the method development field, there remains an urgent need to prove the potential of new methods on the types of complex biological systems that users are interested in, not just simple model systems. There is an increasing focus on improving the reproducibility and reliability of simulations [105]; ideally, use of multiple different force fields and simulation methods should become the norm, aided by tools for improving interoperability between different MD codes [106], and convergence should be monitored, for which robust and widely applicable tools are required. Together, these will greatly improve the quality of investigations of protein conformational dynamics.

### Competing Interests

The author declares that there are no competing interests associated with this manuscript.

### Open Access

Open access for this article was enabled by the participation of University of Auckland in an all-inclusive *Read & Publish* pilot with Portland Press and the Biochemical Society under a transformative agreement with CAUL.

### Funding

J.R.A. is supported by a Rutherford Discovery Fellowship (15-MAU-001/UOA-1508).

### Author Contribution

J.R.A. designed, drafted and revised this work and approved it for publication.

### Acknowledgements

J.R.A. acknowledges the contribution towards organising these ideas made by discussions with and attempts to explain simulation techniques to colleagues and students.

### Abbreviations

aMD, accelerated MD; aUS, adaptive US; CV, collective variables; DeepGenMSMs, Deep generative MSMs; DGM, dynamic graphical model; DM-d-MD, diffusion-map-directed MD; dVAMP, deflated variational approach to Markov processes; ENM, elastic network model; GPU, graphical processing unit; HLDA, harmonic linear discriminant analysis; H-REMD, Hamiltonian REMD; HS-H-REMD, hydrogen bond switching H-REMD; HTMD, high-throughput MD; iMapD, intrinsic map dynamics; LEUS, Local elevation US; MCMC, Markov chain Monte Carlo (MCMC); MD, molecular dynamics; MESA, molecular enhanced sampling with autoencoders; MM, molecular mechanics; M-REMD, Multiplexed REMD; MSM, Markov state models; OPES, on-the-fly probability-enhanced sampling; PCA, principal component analysis; QM, quantum-mechanical; RAVE, reweighted autoencoded variational Bayes for enhanced sampling; REAMD, Replica exchange of aMD; REAP, reinforcement learning based adaptive sampling; REMD, replica exchange MD; REST, Replica exchange with solute tempering; R-REMD, Reservoir REMD; SGLD-H-REMD, Self-guided Langevin dynamics H-REMD; SGOOP, spectral gap optimisation of order parameters; SH-US, self-healing US; SMD, steered MD; TALOS, targeted adversarial learning optimised sampling; TAMD, Temperature-accelerated MD; tICA, time structure-based (or time-lagged) independent component analysis; T-REMD, Temperature REMD; TT, transition-tempered; US, umbrella sampling; VAC, variational approach to conformational dynamics; VAMP, variational approach of Markov processes; VES, variationally enhanced sampling; VR, virtual reality; WT, well-tempered; WTE, well-tempered ensemble.

### References

- 1 Yonath, A. (2011) X-ray crystallography at the heart of life science. *Curr. Opin. Struct. Biol.* **21**, 622–626 <https://doi.org/10.1016/j.sbi.2011.07.005>
- 2 Bax, A. and Clore, G.M. (2019) Protein NMR: boundless opportunities. *J. Magn. Reson.* **306**, 187–191 <https://doi.org/10.1016/j.jmr.2019.07.037>
- 3 Lerner, E., Cordes, T., Ingargiola, A., Alhadid, Y., Chung, S., Michalet, X. et al. (2018) Toward dynamic structural biology: two decades of single-molecule Förster resonance energy transfer. *Science* **359**, eaan1133 <https://doi.org/10.1126/science.aan1133>

- 4 Allison, J.R. (2017) Using simulation to interpret experimental data in terms of protein conformational ensembles. *Curr. Opin. Struct. Biol.* **43**, 79–87 <https://doi.org/10.1016/j.sbi.2016.11.018>
- 5 Bottaro, S. and Lindorff-Larsen, K. (2018) Biophysical experiments and biomolecular simulations: a perfect match? *Science* **361**, 355 <https://doi.org/10.1126/science.aat4010>
- 6 Gaalswyk, K., Muniyat, M.I. and MacCallum, J.L. (2018) The emerging role of physical modeling in the future of structure determination. *Curr. Opin. Struct. Biol.* **49**, 145–153 <https://doi.org/10.1016/j.sbi.2018.03.005>
- 7 McCammon, J.A., Gelin, B.R. and Karplus, M. (1977) Dynamics of folded proteins. *Nature* **267**, 585–590 <https://doi.org/10.1038/267585a0>
- 8 Levitt, M. and Sharon, R. (1988) Accurate simulation of protein dynamics in solution. *Proc. Natl Acad. Sci. U.S.A.* **85**, 7557 <https://doi.org/10.1073/pnas.85.20.7557>
- 9 Patel, D.S., Qi, Y. and Im, W. (2017) Modeling and simulation of bacterial outer membranes and interactions with membrane proteins. *Curr. Opin. Struct. Biol.* **43**, 131–140 <https://doi.org/10.1016/j.sbi.2017.01.003>
- 10 Ostrowska, N., Feig, M. and Trylska, J. (2019) Modeling crowded environment in molecular simulations. *Front. Mol. Biosci.* **6**, 86 <https://doi.org/10.3389/fmolb.2019.00086>
- 11 Sugita, Y. and Feig, M. (2020) Chapter 14 All-atom Molecular Dynamics Simulation of Proteins in Crowded Environments. In *Cell NMR Spectroscopy: From Molecular Sciences to Cell Biology: The Royal Society of Chemistry*, pp. 228–248
- 12 Lindorff-Larsen, K., Maragakis, P., Piana, S., Eastwood, M.P., Dror, R.O. and Shaw, D.E. (2012) Systematic validation of protein force fields against experimental data. *PLoS ONE* **7**, e32131 <https://doi.org/10.1371/journal.pone.0032131>
- 13 Robustelli, P., Piana, S. and Shaw, D.E. (2018) Developing a molecular dynamics force field for both folded and disordered protein states. *Proc. Natl Acad. Sci. U.S.A.* **115**, E4758 <https://doi.org/10.1073/pnas.1800690115>
- 14 Beauchamp, K.A., Lin, Y.S., Das, R. and Pande, V.S. (2012) Are protein force fields getting better? A systematic benchmark on 524 diverse NMR measurements. *J. Chem. Theory Comput.* **8**, 1409–1414 <https://doi.org/10.1021/ct2007814>
- 15 Maier, J.A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K.E. and Simmerling, C. (2015) ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* **11**, 3696–3713 <https://doi.org/10.1021/acs.jctc.5b00255>
- 16 Huang, J., Rauscher, S., Nawrocki, G., Ran, T., Feig, M., de Groot, B.L. et al. (2017) CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **14**, 71–73 <https://doi.org/10.1038/nmeth.4067>
- 17 Harder, E., Damm, W., Maple, J., Wu, C., Reboul, M., Xiang, J.Y. et al. (2016) OPLS3: a force field providing broad coverage of drug-like small molecules and proteins. *J. Chem. Theory Comput.* **12**, 281–296 <https://doi.org/10.1021/acs.jctc.5b00864>
- 18 Diem, M. and Oostenbrink, C. (2020) Hamiltonian reweighing To refine protein backbone dihedral angle parameters in the GROMOS force field. *J. Chem. Inf. Model.* **60**, 279–288 <https://doi.org/10.1021/acs.jcim.9b01034>
- 19 Lemkul, J.A. (2020) Chapter One: pairwise-additive and polarizable atomistic force fields for molecular dynamics simulations of proteins. *Prog. Mol. Biol. Transl. Sci.* **170**, 1–71 <https://doi.org/10.1016/bs.pmbts.2019.12.009>
- 20 Amaro, R.E. and Mulholland, A.J. (2018) Multiscale methods in drug design bridge chemical and biological complexity in the search for cures. *Nat. Rev. Chem.* **2**, 0148 <https://doi.org/10.1038/s41570-018-0148>
- 21 Bradshaw, R.T., Dziedzic, J., Skylaris, C.-K. and Essex, J.W. (2020) The role of electrostatics in enzymes: do biomolecular force fields reflect protein electric fields? *J. Chem. Inf. Model.* **60**, 3131–3144 <https://doi.org/10.1021/acs.jcim.0c00217>
- 22 Dama, J.F., Siniatskiy, A.V., McCullagh, M., Weare, J., Roux, B., Dinner, A.R. et al. (2013) The theory of ultra-coarse-graining. 1. General principles. *J. Chem. Theory Comput.* **9**, 2466–2480 <https://doi.org/10.1021/ct4000444>
- 23 Marrink, S.J. and Tieleman, D.P. (2013) Perspective on the martini model. *Chem. Soc. Rev.* **42**, 6801–6822 <https://doi.org/10.1039/c3cs60093a>
- 24 Bond, P.J. and Sansom, M.S.P. (2006) Insertion and assembly of membrane proteins via simulation. *J. Am. Chem. Soc.* **128**, 2697–2704 <https://doi.org/10.1021/ja0569104>
- 25 Periolo, X., Cavalli, M., Marrink, S.-J. and Ceruso, M.A. (2009) Combining an elastic network with a coarse-grained molecular force field: structure, dynamics, and intermolecular recognition. *J. Chem. Theory Comput.* **5**, 2531–2543 <https://doi.org/10.1021/ct9002114>
- 26 Poma, A.B., Cieplak, M. and Theodorakis, P.E. (2017) Combining the MARTINI and structure-Based coarse-Grained approaches for the molecular dynamics studies of conformational transitions in proteins. *J. Chem. Theory Comput.* **13**, 1366–1374 <https://doi.org/10.1021/acs.jctc.6b00986>
- 27 Hills RD, J., Lu, L. and Voth, G.A. (2010) Multiscale coarse-Graining of the protein energy landscape. *PLoS Comput. Biol.* **6**, e1000827 <https://doi.org/10.1371/journal.pcbi.1000827>
- 28 Ayton, G.S., Noid, W.G. and Voth, G.A. (2007) Multiscale modeling of biomolecular systems: in serial and in parallel. *Curr. Opin. Struct. Biol.* **17**, 192–198 <https://doi.org/10.1016/j.sbi.2007.03.004>
- 29 Machado, M.R., Zeida, A., Darré, L. and Pantano, S. (2019) From quantum to subcellular scales: multi-scale simulation approaches and the SIRAH force field. *Interface Focus* **9**, 20180085 <https://doi.org/10.1098/rsfs.2018.0085>
- 30 Grossfield, A., Patrone, P.N., Roe, D.R., Schultz, A.J., Siderius, D.W. and Zuckerman, D.M. (2019) Best practices for quantification of uncertainty and sampling quality in molecular simulations [Article v1.0]. *Living J. Comp. Mol. Sci.* **1**, 5067 <https://doi.org/10.33011/livecoms.1.1.5067>
- 31 Levine, Z.A. and Shea, J.-E. (2017) Simulations of disordered proteins and systems with conformational heterogeneity. *Curr. Opin. Struct. Biol.* **43**, 95–103 <https://doi.org/10.1016/j.sbi.2016.11.006>
- 32 Best, R.B. (2017) Computational and theoretical advances in studies of intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* **42**, 147–154 <https://doi.org/10.1016/j.sbi.2017.01.006>
- 33 Abraham, M.J., Murtola, T., Schulz, R., Páll, S., Smith, J.C., Hess, B. et al. (2015) GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25 <https://doi.org/10.1016/j.softx.2015.06.001>
- 34 Phillips, J.C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E. et al. (2005) Scalable molecular dynamics with NAMM. *J. Comput. Chem.* **26**, 1781–1802 <https://doi.org/10.1002/jcc.20289>
- 35 Bowers, K.J., Chow, E., Xu, H., Dror, R.O., Eastwood, M.P., Gregersen, B.A. et al. (2006) Scalable algorithms for molecular dynamics simulations on commodity clusters. *Proceedings of the 2006 ACM/IEEE conference on Supercomputing: Association for Computing Machinery*, Tampa, Florida, pp. 84
- 36 Harvey, M.J., Giupponi, G. and Fabritiis, G.D. (2009) ACEMD: accelerating biomolecular dynamics in the microsecond time scale. *J. Chem. Theory Comput.* **5**, 1632–1639 <https://doi.org/10.1021/ct9000685>

- 37 Case, D.A., Cheatham Iii, T.E., Darden, T., Gohlke, H., Luo, R., Merz Jr, K.M. et al. (2005) The amber biomolecular simulation programs. *J. Comput. Chem.* **26**, 1668–1688 <https://doi.org/10.1002/jcc.20290>
- 38 Shaw, D.E., Dror, R.O., Salmon, J.K., Grossman, J.P., Mackenzie, K.M., Bank, J.A. et al. (2009) *Millisecond-scale molecular dynamics simulations on Anton. Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis*, Association for Computing Machinery, Oregon, Portland, p. Article 39
- 39 Ohmura, I., Morimoto, G., Ohno, Y., Hasegawa, A. and Tajiri, M. (2014) MDGRAPE-4: a special-purpose computer system for molecular dynamics simulations. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **372**, 20130387 <https://doi.org/10.1098/rsta.2013.0387>
- 40 Shaw, D.E., Grossman, J.P., Bank, J.A., Batson, B., Butts, J.A., Chao, J.C. et al. (2014) *Anton 2: raising the bar for performance and programmability in a special-purpose molecular dynamics supercomputer. Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, IEEE Press, New Orleans, Louisiana, pp. 41–53
- 41 Shirts, M. (2000) Pande VS. screen savers of the world unite! *Science* **290**, 1903–1904 <https://doi.org/10.1126/science.290.5498.1903>
- 42 Harada, R., Takano, Y., Baba, T. and Shigeta, Y. (2015) Simple, yet powerful methodologies for conformational sampling of proteins. *Phys. Chem. Chem. Phys.* **17**, 6155–6173 <https://doi.org/10.1039/C4CP05262E>
- 43 Pronk, S., Pouya, I., Lundborg, M., Rotskoff, G., Wesén, B., Kasson, P.M. et al. (2015) Molecular simulation workflows as parallel algorithms: the execution engine of copernicus, a distributed high-performance computing platform. *J. Chem. Theory Comput.* **11**, 2600–2608 <https://doi.org/10.1021/acs.jctc.5b00234>
- 44 Doerr, S., Harvey, M.J., Noé, F. and De Fabritiis, G. (2016) HTMD: high-Throughput molecular dynamics for molecular discovery. *J. Chem. Theory Comput.* **12**, 1845–1852 <https://doi.org/10.1021/acs.jctc.6b00049>
- 45 Husic, B.E. and Pande, V.S. (2018) Markov state models: from an Art to a science. *J. Am. Chem. Soc.* **140**, 2386–2396 <https://doi.org/10.1021/jacs.7b12191>
- 46 Chodera, J.D. and Noé, F. (2014) Markov state models of biomolecular conformational dynamics. *Curr. Opin. Struct. Biol.* **25**, 135–144 <https://doi.org/10.1016/j.sbi.2014.04.002>
- 47 Juraszek, J., Vreede, J. and Bolhuis, P.G. (2012) Transition path sampling of protein conformational changes. *Chem. Phys.* **396**, 30–44 <https://doi.org/10.1016/j.chemphys.2011.04.032>
- 48 Bello-Rivas, J.M. and Elber, R. (2015) Exact milestoneing. *J. Chem. Phys.* **142**, 094102 <https://doi.org/10.1063/1.4913399>
- 49 Tribello, G.A. and Gasparotto, P. (2019) Using dimensionality reduction to analyze protein trajectories. *Front. Mol. Biosci.* **6**, 46 <https://doi.org/10.3389/fmolb.2019.00046>
- 50 Doerr, S., Ariz-Extreme, I., Harvey, M.J. and De Fabritiis, G. (2017) Dimensionality reduction methods for molecular simulations. *arXiv* <https://arxiv.org/abs/1710.10629v2>
- 51 Barducci, A., Bonomi, M. and Parrinello, M. (2011) Metadynamics. *WIREs Comput. Mol. Sci.* **1**, 826–843 <https://doi.org/10.1002/wcms.31>
- 52 Bussi, G. and Laio, A. (2020) Using metadynamics to explore complex free-energy landscapes. *Nat. Rev. Phys.* **2**, 200–212 <https://doi.org/10.1038/s42254-020-0153-0>
- 53 Bussi, G., Laio, A. and Tiwary, P. (2020) Metadynamics: A Unified Framework for Accelerating Rare Events and Sampling Thermodynamics and Kinetics. In *Handbook of Materials Modeling: Methods: Theory and Modeling* (Andreoni, W. and Yip, S., eds), pp. 565–595, Springer International Publishing, Cham
- 54 Rohrdanz, M.A., Zheng, W. and Clementi, C. (2013) Discovering mountain passes via torchlight: methods for the definition of reaction coordinates and pathways in complex macromolecular reactions. *Annu. Rev. Phys. Chem.* **64**, 295–316 <https://doi.org/10.1146/annurev-physchem-040412-110006>
- 55 Sidky, H., Chen, W. and Ferguson, A.L. (2020) Machine learning for collective variable discovery and enhanced sampling in biomolecular simulation. *Mol. Phys.* **118**, e1737742 <https://doi.org/10.1080/00268976.2020.1737742>
- 56 Noé, F. and Clementi, C. (2017) Collective variables for the study of long-time kinetics from molecular trajectories: theory and methods. *Curr. Opin. Struct. Biol.* **43**, 141–147 <https://doi.org/10.1016/j.sbi.2017.02.006>
- 57 Palazzesi, F., Valsson, O. and Parrinello, M. (2017) Conformational entropy as collective variable for proteins. *J. Phys. Chem. Lett.* **8**, 4752–4756 <https://doi.org/10.1021/acs.jpcllett.7b01770>
- 58 Granata, D., Camilloni, C., Vendruscolo, M. and Laio, A. (2013) Characterization of the free-energy landscapes of proteins by NMR-guided metadynamics. *Proc. Natl Acad. Sci. U.S.A.* **110**, 6817 <https://doi.org/10.1073/pnas.1218350110>
- 59 Spiwok, V., Lipovová, P. and Králová, B. (2007) Metadynamics in essential coordinates: free energy simulation of conformational changes. *J. Phys. Chem. B* **111**, 3073–3076 <https://doi.org/10.1021/jp068587c>
- 60 Mendels, D., Piccini, G., Brotzakis, Z.F., Yang, Y.I. and Parrinello, M. (2018) Folding a small protein using harmonic linear discriminant analysis. *J. Chem. Phys.* **149**, 194113 <https://doi.org/10.1063/1.5053566>
- 61 Bonati, L., Rizzi, V. and Parrinello, M. (2020) Data-driven collective variables for enhanced sampling. *J. Phys. Chem. Lett.* **11**, 2998–3004 <https://doi.org/10.1021/acs.jpcllett.0c00535>
- 62 Noé, F. and Nüske, F. (2013) A variational approach to modeling slow processes in stochastic dynamical systems. *Multiscale Model. Simul.* **11**, 635–655 <https://doi.org/10.1137/110858616>
- 63 Nüske, F., Keller, B.G., Pérez-Hernández, G., Mey, A.S.J.S. and Noé, F. (2014) Variational approach to molecular kinetics. *J. Chem. Theory Comput.* **10**, 1739–1752 <https://doi.org/10.1021/ct4009156>
- 64 Prinz, J.-H., Wu, H., Sarich, M., Keller, B., Senne, M., Held, M. et al. (2011) Markov models of molecular kinetics: generation and validation. *J. Chem. Phys.* **134**, 174105 <https://doi.org/10.1063/1.3565032>
- 65 Pérez-Hernández, G., Paul, F., Giorgino, T., De Fabritiis, G. and Noé, F. (2013) Identification of slow molecular order parameters for Markov model construction. *J. Chem. Phys.* **139**, 015102 <https://doi.org/10.1063/1.4811489>
- 66 Molgedey, L. and Schuster, H.G. (1994) Separation of a mixture of independent signals using time delayed correlations. *Phys Rev Lett.* **72**, 3634–3637 <https://doi.org/10.1103/PhysRevLett.72.3634>
- 67 Naritomi, Y. and Fuchigami, S. (2011) Slow dynamics in protein fluctuations revealed by time-structure based independent component analysis: the case of domain motions. *J. Chem. Phys.* **134**, 065101 <https://doi.org/10.1063/1.3554380>

- 68 Schwantes, C.R. and Pande, V.S. (2013) Improvements in Markov state model construction reveal many non-native interactions in the folding of NTL9. *J. Chem. Theory Comput.* **9**, 2000–2009 <https://doi.org/10.1021/ct300878a>
- 69 Sultan M, M. and Pande, V.S. (2017) tICA-Metadynamics: accelerating metadynamics by using kinetically selected collective variables. *J. Chem. Theory Comput.* **13**, 2440–2447 <https://doi.org/10.1021/acs.jctc.7b00182>
- 70 McCarty, J. and Parrinello, M. (2017) A variational conformational dynamics approach to the selection of collective variables in metadynamics. *J. Chem. Phys.* **147**, 204109 <https://doi.org/10.1063/1.4998598>
- 71 Schwantes, C.R. and Pande, V.S. (2015) Modeling molecular kinetics with tICA and the Kernel Trick. *J. Chem. Theory Comput.* **11**, 600–608 <https://doi.org/10.1021/ct5007357>
- 72 Harrigan, M.P. and Pande, V.S. (2017) Landmark Kernel tICA for conformational dynamics. *bioRxiv* <https://doi.org/10.1101/123752>
- 73 Mardt, A., Pasquali, L., Wu, H. and Noé, F. (2018) VAMPnets for deep learning of molecular kinetics. *Nat. Commun.* **9**, 5 <https://doi.org/10.1038/s41467-017-02388-1>
- 74 Wehmeyer, C. and Noé, F. (2018) Time-lagged autoencoders: deep learning of slow collective variables for molecular kinetics. *J. Chem. Phys.* **148**, 241703 <https://doi.org/10.1063/1.5011399>
- 75 Sultan, M.M., Wayment-Steele, H.K. and Pande, V.S. (2018) Transferable neural networks for enhanced sampling of protein dynamics. *J. Chem. Theory Comput.* **14**, 1887–1894 <https://doi.org/10.1021/acs.jctc.8b00025>
- 76 Hernández, C.X., Wayment-Steele, H.K., Sultan, M.M., Husic, B.E. and Pande, V.S. (2018) Variational encoding of complex dynamics. *Phys. Rev. E* **97**, 062412 <https://doi.org/10.1103/PhysRevE.97.062412>
- 77 Schöberl, M., Zabarar, N. and Koutsourelakis, P.-S. (2019) Predictive collective variable discovery with deep Bayesian models. *J. Chem. Phys.* **150**, 024109 <https://doi.org/10.1063/1.5058063>
- 78 Scherer, M.K., Husic, B.E., Hoffmann, M., Paul, F., Wu, H. and Noé, F. (2019) Variational selection of features for molecular kinetics. *J. Chem. Phys.* **150**, 194108 <https://doi.org/10.1063/1.5083040>
- 79 Husic, B.E. and Noé, F. (2019) Deflation reveals dynamical structure in nondominant reaction coordinates. *J. Chem. Phys.* **151**, 054103 <https://doi.org/10.1063/1.5099194>
- 80 Tiwary, P. and Berne, B.J. (2016) Spectral gap optimization of order parameters for sampling complex molecular systems. *Proc. Natl Acad. Sci. U.S.A.* **113**, 2839 <https://doi.org/10.1073/pnas.1600917113>
- 81 Lemke, T. and Peter, C. (2019) Encodermap: dimensionality reduction and generation of molecule conformations. *J. Chem. Theory Comput.* **15**, 1209–1215 <https://doi.org/10.1021/acs.jctc.8b00975>
- 82 Brandt, S., Sittel, F., Ernst, M. and Stock, G. (2018) Machine learning of biomolecular reaction coordinates. *J. Phys. Chem. Lett.* **9**, 2144–2150 <https://doi.org/10.1021/acs.jpcllett.8b00759>
- 83 Hansen, H.S. and Hünenberger, P.H. (2010) Using the local elevation method to construct optimized umbrella sampling potentials: calculation of the relative free energies and interconversion barriers of glucopyranose ring conformers in water. *J. Comput. Chem.* **31**, 1–23 <https://doi.org/10.1002/jcc.21253>
- 84 Huggins, D.J., Biggin, P.C., Dämgen, M.A., Essex, J.W., Harris, S.A., Henchman, R.H. et al. (2019) Biomolecular simulations: from dynamics and mechanisms to computational assays of biological activity. *WIREs Comput. Mol. Sci.* **9**, e1393 <https://doi.org/10.1002/wcms.1393>
- 85 Roe, D.R., Bergonzo, C. and Cheatham, T.E. (2014) Evaluation of enhanced sampling provided by accelerated molecular dynamics with Hamiltonian replica exchange methods. *J. Phys. Chem. B* **118**, 3543–3552 <https://doi.org/10.1021/jp4125099>
- 86 Qi, R., Wei, G., Ma, B. and Nussinov, R. (2018) Replica exchange molecular dynamics: a practical application protocol with solutions to common problems and a peptide aggregation and self-Assembly example. *Methods Mol. Biol.* **1777**, 101–119 [https://doi.org/10.1007/978-1-4939-7811-3\\_5](https://doi.org/10.1007/978-1-4939-7811-3_5)
- 87 Buchete, N.-V. and Hummer, G. (2008) Peptide folding kinetics from replica exchange molecular dynamics. *Phys. Rev. E* **77**, 030902 <https://doi.org/10.1103/PhysRevE.77.030902>
- 88 Stelzl, L.S., Kells, A., Rosta, E. and Hummer, G. (2017) Dynamic histogram analysis To determine free energies and rates from biased simulations. *J. Chem. Theory Comput.* **13**, 6328–6342 <https://doi.org/10.1021/acs.jctc.7b00373>
- 89 Stelzl, L.S. and Hummer, G. (2017) Kinetics from replica exchange molecular dynamics simulations. *J. Chem. Theory Comput.* **13**, 3927–3935 <https://doi.org/10.1021/acs.jctc.7b00372>
- 90 Fukunishi, H., Watanabe, O. and Takada, S. (2002) On the Hamiltonian replica exchange method for efficient sampling of biomolecular systems: application to protein structure prediction. *J. Chem. Phys.* **116**, 9058–9067 <https://doi.org/10.1063/1.1472510>
- 91 Liu, P., Kim, B., Friesner, R.A. and Berne, B.J. (2005) Replica exchange with solute tempering: a method for sampling biological systems in explicit water. *Proc. Natl Acad. Sci. U.S.A.* **102**, 13749 <https://doi.org/10.1073/pnas.0506346102>
- 92 Hamelberg, D., Mongan, J. and McCammon, J.A. (2004) Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. *J. Chem. Phys.* **120**, 11919–11929 <https://doi.org/10.1063/1.1755656>
- 93 Huber, T., Torda, A.E. and van Gunsteren, W.F. (1994) Local elevation: a method for improving the searching properties of molecular dynamics simulation. *J. Comput. Aid Mol. Des.* **8**, 695–708 <https://doi.org/10.1007/BF00124016>
- 94 Grubmüller, H. (1995) Predicting slow structural transitions in macromolecular systems: conformational flooding. *Phys. Rev. E* **52**, 2893–2906 <https://doi.org/10.1103/PhysRevE.52.2893>
- 95 Laio, A. and Parrinello, M. (2002) Escaping free-energy minima. *Proc. Natl Acad. Sci. U.S.A.* **99**, 12562 <https://doi.org/10.1073/pnas.202427399>
- 96 Hruska, E., Abella, J.R., Nüske, F., Kavrakli, L.E. and Clementi, C. (2018) Quantitative comparison of adaptive sampling methods for protein dynamics. *J. Chem. Phys.* **149**, 244119 <https://doi.org/10.1063/1.5053582>
- 97 Wu, H., Mardt, A., Pasquali, L. and Noe, F. (2018) *Deep generative Markov state models. Proceedings of the 32nd International Conference on Neural Information Processing Systems*, Curran Associates Inc., Montréal, Canada, pp. 3979–3988
- 98 Noé, F., Olsson, S., Köhler, J. and Wu, H. (2019) Boltzmann generators: sampling equilibrium states of many-body systems with deep learning. *Science* **365**, eaaw1147 <https://doi.org/10.1126/science.aaw1147>
- 99 Tuckerman, M.E. (2019) Machine learning transforms how microstates are sampled. *Science* **365**, 982 <https://doi.org/10.1126/science.aay2568>
- 100 Olsson, S. and Noé, F. (2019) Dynamic graphical models of molecular kinetics. *Proc. Natl Acad. Sci. U.S.A.* **116**, 15001 <https://doi.org/10.1073/pnas.1901692116>

- 101 Stone, J.E., Gullingsrud, J. and Schulten, K. (2001) *A system for interactive molecular dynamics simulation. Proceedings of the 2001 symposium on Interactive 3D graphics: Association for Computing Machinery*, pp. 191–194
- 102 Dreher, M., Piuze, M., Turki, A., Chavent, M., Baaden, M., Férey, N. et al. (2013) Interactive molecular dynamics: scaling up to large systems. *Procedia Comput. Sci.* **18**, 20–29 <https://doi.org/10.1016/j.procs.2013.05.165>
- 103 O'Connor, M., Deeks, H.M., Dawn, E., Metatla, O., Roudaut, A., Sutton, M. et al. (2018) Sampling molecular conformations and dynamics in a multiuser virtual reality framework. *Sci. Adv.* **4**, eaat2731 <https://doi.org/10.1126/sciadv.aat2731>
- 104 Jordi, J.-J., Philip, T., Michael, O.C., Salome, L., Rebecca, S., David, G. et al. (2020) A Virtual Reality Ensemble Molecular Dynamics Workflow to Study Complex Conformational Changes in Proteins
- 105 Bonomi, M., Bussi, G., Camilloni, C., Tribello, G.A., Banáš, P., Barducci, A. et al. (2019) Promoting transparency and reproducibility in enhanced molecular simulations. *Nat. Methods* **16**, 670–673 <https://doi.org/10.1038/s41592-019-0506-8>
- 106 Hedges, L., Mey, A.S.J.S., Laughton, C.A., Gervasio, F.L., Mulholland, A.J., Woods, C.J. et al. (2019) BiosimSpace: an interoperable Python framework for biomolecular simulation. *J. Open Source Softw.* **4**, 1831 <https://doi.org/10.21105/joss.01831>
- 107 Kirkpatrick, S., Gelatt, C.D. and Vecchi, M.P. (1983) Optimization by simulated annealing. *Science* **220**, 671 <https://doi.org/10.1126/science.220.4598.671>
- 108 Pan, A.C., Weinreich, T.M., Piana, S. and Shaw, D.E. (2016) Demonstrating an order-of-magnitude sampling enhancement in molecular dynamics simulations of complex protein systems. *J. Chem. Theory Comput.* **12**, 1360–1367 <https://doi.org/10.1021/acs.jctc.5b00913>
- 109 Marinari, E. and Parisi, G. (1992) Simulated tempering: a new monte carlo scheme. *Europhys. Lett.* **19**, 451–458 <https://doi.org/10.1209/0295-5075/19/6/002>
- 110 Zhang, C. and Ma, J. (2010) Enhanced sampling and applications in protein folding in explicit solvent. *J. Chem. Phys.* **132**, 244101 <https://doi.org/10.1063/1.3435332>
- 111 Nymeyer, H. (2008) How efficient is replica exchange molecular dynamics? An analytic approach. *J. Chem. Theory Comput.* **4**, 626–636 <https://doi.org/10.1021/ct7003337>
- 112 Bernardi, R.C., Melo, M.C.R. and Schulten, K. (2015) Enhanced sampling techniques in molecular dynamics simulations of biological systems. *Biochim. Biophys. Acta* **1850**, 872–877 <https://doi.org/10.1016/j.bbagen.2014.10.019>
- 113 MacCallum, J.L., Muniyat, M.I. and Gaalswyk, K. (2018) Online optimization of total acceptance in Hamiltonian replica exchange simulations. *J. Phys. Chem. B* **122**, 5448–5457 <https://doi.org/10.1021/acs.jpcc.7b11778>
- 114 Nadler, W. and Hansmann, U.H.E. (2008) Optimized explicit-solvent replica exchange molecular dynamics from scratch. *J. Phys. Chem. B* **112**, 10386–7 <https://doi.org/10.1021/jp805085y>
- 115 Predescu, C., Predescu, M. and Ciobanu, C.V. (2005) On the efficiency of exchange in parallel tempering Monte Carlo simulations. *J. Phys. Chem. B* **109**, 4189–4196 <https://doi.org/10.1021/jp045073+>
- 116 Nadler, W., Meinke, J.H. and Hansmann, U.H.E. (2008) Folding proteins by first-passage-times-optimized replica exchange. *Phys. Rev. E* **78**, 061905 <https://doi.org/10.1103/PhysRevE.78.061905>
- 117 Prakash, M.K., Barducci, A. and Parrinello, M. (2011) Replica temperatures for uniform exchange and efficient roundtrip times in explicit solvent parallel tempering simulations. *J. Chem. Theory Comput.* **7**, 2025–2027 <https://doi.org/10.1021/ct200208h>
- 118 Li, H., Li, G., Berg, B.A. and Yang, W. (2006) Finite reservoir replica exchange to enhance canonical sampling in rugged energy surfaces. *J. Chem. Phys.* **125**, 144902 <https://doi.org/10.1063/1.2354157>
- 119 Lyman, E., Ytreberg, F.M. and Zuckerman, D.M. (2006) Resolution exchange simulation. *Phys Rev Lett.* **96**, 028105 <https://doi.org/10.1103/PhysRevLett.96.028105>
- 120 Okur, A., Roe, D.R., Cui, G., Hornak, V. and Simmerling, C. (2007) Improving convergence of replica-exchange simulations through coupling to a high-temperature structure reservoir. *J. Chem. Theory Comput.* **3**, 557–568 <https://doi.org/10.1021/ct600263e>
- 121 Henriksen, N.M., Roe, D.R. and Cheatham, T.E. (2013) Reliable oligonucleotide conformational ensemble generation in explicit solvent for force field assessment using reservoir replica exchange molecular dynamics simulations. *J. Phys. Chem. B* **117**, 4014–4027 <https://doi.org/10.1021/jp400530e>
- 122 Rhee, Y.M. and Pande, V.S. (2003) Multiplexed-Replica exchange molecular dynamics method for protein folding simulation. *Biophys. J.* **84**, 775–786 [https://doi.org/10.1016/S0006-3495\(03\)74897-8](https://doi.org/10.1016/S0006-3495(03)74897-8)
- 123 Maragliano, L. and Vanden-Eijnden, E. (2006) A temperature accelerated method for sampling free energy and determining reaction pathways in rare events simulations. *Chem. Phys. Lett.* **426**, 168–175 <https://doi.org/10.1016/j.cplett.2006.05.062>
- 124 Wang, L., Friesner, R.A. and Berne, B.J. (2011) Replica exchange with solute scaling: a more efficient version of replica exchange with solute tempering (REST2). *J. Phys. Chem. B* **115**, 9431–9438 <https://doi.org/10.1021/jp204407d>
- 125 Lee, M.S. and Olson, M.A. (2010) Protein folding simulations combining self-guided langevin dynamics and temperature-based replica exchange. *J. Chem. Theory Comput.* **6**, 2477–2487 <https://doi.org/10.1021/ct100062b>
- 126 Wu, X., Hodocsek, M. and Brooks, B.R. (2012) Replica exchanging self-guided langevin dynamics for efficient and accurate conformational sampling. *J. Chem. Phys.* **137**, 044106 <https://doi.org/10.1063/1.4737094>
- 127 Grubmüller, H., Heymann, B. and Tavan, P. (1996) Ligand binding: molecular mechanics calculation of the streptavidin-biotin rupture force. *Science* **271**, 997 <https://doi.org/10.1126/science.271.5251.997>
- 128 Izrailev, S., Stepaniants, S., Balseira, M., Oono, Y. and Schulten, K. (1997) Molecular dynamics study of unbinding of the avidin-biotin complex. *Biophys. J.* **72**, 1568–1581 [https://doi.org/10.1016/S0006-3495\(97\)78804-0](https://doi.org/10.1016/S0006-3495(97)78804-0)
- 129 Park, S. and Schulten, K. (2004) Calculating potentials of mean force from steered molecular dynamics simulations. *J. Chem. Phys.* **120**, 5946–5961 <https://doi.org/10.1063/1.1651473>
- 130 Torrie, G.M. and Valleau, J.P. (1977) Nonphysical sampling distributions in monte carlo free-energy estimation: umbrella sampling. *J. Comput. Phys.* **23**, 187–199 [https://doi.org/10.1016/0021-9991\(77\)90121-8](https://doi.org/10.1016/0021-9991(77)90121-8)
- 131 Ostermeir, K. and Zacharias, M. (2014) Hamiltonian replica-exchange simulations with adaptive biasing of peptide backbone and side chain dihedral angles. *J. Comput. Chem.* **35**, 150–158 <https://doi.org/10.1002/jcc.23476>
- 132 Xu, C., Wang, J. and Liu, H. (2008) A Hamiltonian replica exchange approach and its application to the study of side-chain type and neighbor effects on peptide backbone conformations. *J. Chem. Theory Comput.* **4**, 1348–1359 <https://doi.org/10.1021/ct7003534>



- 133 Zhang, C. and Ma, J. (2012) Folding helical proteins in explicit solvent using dihedral-biased tempering. *Proc. Natl Acad. Sci. U.S.A.* **109**, 8139 <https://doi.org/10.1073/pnas.1112143109>
- 134 Affentranger, R., Tavernelli, I. and Di Iorio, E.E. (2006) A novel Hamiltonian replica exchange MD protocol to enhance protein conformational space sampling. *J. Chem. Theory Comput.* **2**, 217–228 <https://doi.org/10.1021/ct050250b>
- 135 Hritz, J. and Oostenbrink, C. (2008) Hamiltonian replica exchange molecular dynamics using soft-core interactions. *J. Chem. Phys.* **128**, 144121 <https://doi.org/10.1063/1.2888998>
- 136 Itoh, S.G., Okumura, H. and Okamoto, Y. (2010) Replica-exchange method in van der Waals radius space: overcoming steric restrictions for biomolecules. *J. Chem. Phys.* **132**, 134105 <https://doi.org/10.1063/1.3372767>
- 137 Hritz, J. and Oostenbrink, C. (2007) Optimization of replica exchange molecular dynamics by fast mimicking. *J. Chem. Phys.* **127**, 204104 <https://doi.org/10.1063/1.2790427>
- 138 Christen, M. and van Gunsteren, W.F. (2006) Multigraining: an algorithm for simultaneous fine-grained and coarse-grained simulation of molecular systems. *J. Chem. Phys.* **124**, 154106 <https://doi.org/10.1063/1.2187488>
- 139 Liu, P. and Voth, G.A. (2007) Smart resolution replica exchange: an efficient algorithm for exploring complex energy landscapes. *J. Chem. Phys.* **126**, 045106 <https://doi.org/10.1063/1.2408415>
- 140 Cheng, X., Cui, G., Hornak, V. and Simmerling, C. (2005) Modified replica exchange simulation methods for local structure refinement. *J. Phys. Chem. B* **109**, 8220–8230 <https://doi.org/10.1021/jp045437y>
- 141 Wang, Y. and Brooks, C.L. (2019) Enhanced sampling applied to modeling allosteric regulation in transcription. *J. Phys. Chem. Lett.* **10**, 5963–5968 <https://doi.org/10.1021/acs.jpcclett.9b02226>
- 142 Fajer, M., Hamelberg, D. and McCammon, J.A. (2008) Replica-Exchange accelerated molecular dynamics (REXAMD) applied to thermodynamic integration. *J. Chem. Theory Comput.* **4**, 1565–1569 <https://doi.org/10.1021/ct800250m>
- 143 Zacharias, M. (2008) Combining elastic network analysis and molecular dynamics simulations by Hamiltonian replica exchange. *J. Chem. Theory Comput.* **4**, 477–487 <https://doi.org/10.1021/ct7002258>
- 144 Vreede, J., Wolf, M.G., de Leeuw, S.W. and Bolhuis, P.G. (2009) Reordering hydrogen bonds using Hamiltonian replica exchange enhances sampling of conformational changes in biomolecular systems. *J. Phys. Chem. B* **113**, 6484–6494 <https://doi.org/10.1021/jp809641j>
- 145 Ceriotti, M., Brain, G.A.R., Riordan, O. and Manolopoulos, D.E. (2012) The inefficiency of re-weighted sampling and the curse of system size in high-order path integration. *Proc. R. Soc. A* **468**, 2–17 <https://doi.org/10.1098/rspa.2011.0413>
- 146 Shen, T. and Hamelberg, D. (2008) A statistical analysis of the precision of reweighting-based simulations. *J. Chem. Phys.* **129**, 034103 <https://doi.org/10.1063/1.2944250>
- 147 Hoof, R.W.W., van Eijck, B.P. and Kroon, J. (1992) An adaptive umbrella sampling procedure in conformational analysis using molecular dynamics and its application to glycol. *J. Chem. Phys.* **97**, 6690–6694 <https://doi.org/10.1063/1.463947>
- 148 Mezei, M. (1987) Adaptive umbrella sampling: self-consistent determination of the non-Boltzmann bias. *J. Comput Phys.* **68**, 237–248 [https://doi.org/10.1016/0021-9991\(87\)90054-4](https://doi.org/10.1016/0021-9991(87)90054-4)
- 149 Marsili, S., Barducci, A., Chelli, R., Procacci, P. and Schettino, V. (2006) Self-healing umbrella sampling: a non-equilibrium approach for quantitative free energy calculations. *J. Phys. Chem. B* **110**, 14011–3 <https://doi.org/10.1021/jp062755j>
- 150 Bartels, C. and Karplus, M. (1997) Multidimensional adaptive umbrella sampling: applications to main chain and side chain peptide conformations. *J. Comput. Chem.* **18**, 1450–1462 [https://doi.org/10.1002/\(SICI\)1096-987X\(199709\)18:12<1450::AID-JCC3>3.0.CO;2-I](https://doi.org/10.1002/(SICI)1096-987X(199709)18:12<1450::AID-JCC3>3.0.CO;2-I)
- 151 Raiteri, P., Laio, A., Gervasio, F.L., Micheletti, C. and Parrinello, M. (2006) Efficient reconstruction of complex free energy landscapes by multiple walkers metadynamics. *J. Phys. Chem. B* **110**, 3533–3539 <https://doi.org/10.1021/jp054359r>
- 152 Bonomi, M. and Parrinello, M. (2010) Enhanced sampling in the well-tempered ensemble. *Phys. Rev. Lett.* **104**, 190601 <https://doi.org/10.1103/PhysRevLett.104.190601>
- 153 Piana, S. and Laio, A. (2007) A bias-exchange approach to protein folding. *J. Phys. Chem. B* **111**, 4553–4559 <https://doi.org/10.1021/jp067873i>
- 154 Pfaendtner, J. and Bonomi, M. (2015) Efficient sampling of high-dimensional free-energy landscapes with parallel bias metadynamics. *J. Chem. Theory Comput.* **11**, 5062–5067 <https://doi.org/10.1021/acs.jctc.5b00846>
- 155 Bussi, G., Gervasio, F.L., Laio, A. and Parrinello, M. (2006) Free-Energy landscape for  $\beta$  hairpin folding from combined parallel tempering and metadynamics. *J. Am. Chem. Soc.* **128**, 13435–13441 <https://doi.org/10.1021/ja062463w>
- 156 Camilloni, C., Provasi, D., Tiana, G. and Broglio, R.A. (2008) Exploring the protein G helix free-energy surface by solute tempering metadynamics. *Proteins* **71**, 1647–1654 <https://doi.org/10.1002/prot.21852>
- 157 Deighan, M., Bonomi, M. and Pfaendtner, J. (2012) Efficient simulation of explicitly solvated proteins in the well-tempered ensemble. *J. Chem. Theory Comput.* **8**, 2189–2192 <https://doi.org/10.1021/ct300297t>
- 158 Barducci, A., Bussi, G. and Parrinello, M. (2008) Well-Tempered metadynamics: a smoothly converging and tunable free-energy method. *Phys. Rev. Lett.* **100**, 020603 <https://doi.org/10.1103/PhysRevLett.100.020603>
- 159 Dama, J.F., Rotskoff, G., Parrinello, M. and Voth, G.A. (2014) Transition-tempered metadynamics: robust, convergent metadynamics via on-the-fly transition barrier estimation. *J. Chem. Theory Comput.* **10**, 3626–3633 <https://doi.org/10.1021/ct500441q>
- 160 Dickson, B.M. (2015)  $\mu$ -tempered metadynamics: artifact independent convergence times for wide hills. *J. Chem. Phys.* **143**, 234109 <https://doi.org/10.1063/1.4937939>
- 161 Gil-Ley, A. and Bussi, G. (2015) Enhanced conformational sampling using replica exchange with collective-variable tempering. *J. Chem. Theory Comput.* **11**, 1077–1085 <https://doi.org/10.1021/ct5009087>
- 162 Dama, J.F., Hocky, G.M., Sun, R. and Voth, G.A. (2015) Exploring valleys without climbing every peak: more efficient and forgiving metadynamics via robust on-the-fly bias domain restriction. *J. Chem. Theory Comput.* **11**, 5638–5650 <https://doi.org/10.1021/acs.jctc.5b00907>
- 163 Invernizzi, M. and Parrinello, M. (2020) Rethinking metadynamics: from bias potentials to probability distributions. *J. Phys. Chem. Lett.* **11**, 2731–2736 <https://doi.org/10.1021/acs.jpcclett.0c00497>
- 164 Valsion, O. and Parrinello, M. (2014) Variational approach to enhanced sampling and free energy calculations. *Phys. Rev. Lett.* **113**, 090601 <https://doi.org/10.1103/PhysRevLett.113.090601>

- 165 Bonati, L., Zhang, Y.-Y. and Parrinello, M. (2019) Neural networks-based variationally enhanced sampling. *Proc. Natl Acad. Sci. U.S.A.* **116**, 17641 <https://doi.org/10.1073/pnas.1907975116>
- 166 Zhang, J., Yang, Y.I. and Noé, F. (2019) Targeted adversarial learning optimized sampling. *J. Phys. Chem. Lett.* **10**, 5791–5797 <https://doi.org/10.1021/acs.jpcllett.9b02173>
- 167 Doerr, S. and De Fabritiis, G. (2014) On-the-Fly learning and sampling of ligand binding by high-throughput molecular simulations. *J. Chem. Theory Comput.* **10**, 2064–2069 <https://doi.org/10.1021/ct400919u>
- 168 Preto, J. and Clementi, C. (2014) Fast recovery of free energy landscapes via diffusion-map-directed molecular dynamics. *Phys. Chem. Chem. Phys.* **16**, 19181–19191 <https://doi.org/10.1039/C3CP54520B>
- 169 Chiavazzo, E., Covino, R., Coifman, R.R., Gear, C.W., Georgiou, A.S., Hummer, G. et al. (2017) Intrinsic map dynamics exploration for uncharted effective free-energy landscapes. *Proc. Natl Acad. Sci. U.S.A.* **114**, E5494 <https://doi.org/10.1073/pnas.1621481114>
- 170 Ribeiro, J.M.L., Bravo, P., Wang, Y. and Tiwary, P. (2018) Reweighted autoencoded variational Bayes for enhanced sampling (RAVE). *J. Chem. Phys.* **149**, 072301 <https://doi.org/10.1063/1.5025487>
- 171 Shamsi, Z., Cheng, K.J. and Shukla, D. (2018) Reinforcement learning based adaptive sampling: REAPing rewards by exploring protein conformational landscapes. *J. Phys. Chem. B* **122**, 8386–8395 <https://doi.org/10.1021/acs.jpcc.8b06521>
- 172 Chen, W. and Ferguson, A.L. (2018) Molecular enhanced sampling with autoencoders: on-the-fly collective variable discovery and accelerated free energy landscape exploration. *J. Comput. Chem.* **39**, 2079–2102 <https://doi.org/10.1002/jcc.25520>