



HAL
open science

Advanced simulation techniques for the thermodynamic and kinetic characterization of biological systems

Carlo Camilloni, Fabio Pietrucci

► **To cite this version:**

Carlo Camilloni, Fabio Pietrucci. Advanced simulation techniques for the thermodynamic and kinetic characterization of biological systems. *Advances in Physics: X*, 2018, 3 (1), pp.UNSP 1477531. 10.1080/23746149.2018.1477531 . hal-01884567

HAL Id: hal-01884567

<https://hal.sorbonne-universite.fr/hal-01884567>

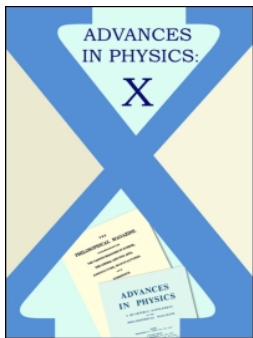
Submitted on 1 Oct 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



Advanced simulation techniques for the thermodynamic and kinetic characterization of biological systems

Carlo Camilloni & Fabio Pietrucci

To cite this article: Carlo Camilloni & Fabio Pietrucci (2018) Advanced simulation techniques for the thermodynamic and kinetic characterization of biological systems, *Advances in Physics: X*, 3:1, 1477531, DOI: [10.1080/23746149.2018.1477531](https://doi.org/10.1080/23746149.2018.1477531)

To link to this article: <https://doi.org/10.1080/23746149.2018.1477531>



© 2018 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 15 Sep 2018.



Submit your article to this journal [↗](#)





Article views: 528



View Crossmark data [↗](#)

Advanced simulation techniques for the thermodynamic and kinetic characterization of biological systems

Carlo Camilloni ^a and Fabio Pietrucci ^b

^aDipartimento di Bioscienze, Università degli Studi di Milano, Milan, Italy; ^bInstitut de Minéralogie, de Physique des Matériaux et de Cosmochimie, IMPMC, Muséum National d'Histoire Naturelle, UMR CNRS 7590, IRD, Sorbonne Université, Paris, France

ABSTRACT

This review discusses successful strategies and key open problems in the kinetic and thermodynamic characterization of complex biomolecular systems by computer simulations. The main focus is on established techniques and emerging trends in the fields of enhanced sampling and of kinetic models, as applied to biological problems ranging from protein folding and conformational dynamics to protein–protein and protein–ligand interaction. We address especially the following questions: How to choose a computational approach suited to a particular problem? What are the strengths and limitations of alternative approaches? What is the current accuracy of thermodynamic and kinetic predictions? What are today's open challenges and promising development directions? Towards the aim of accurately reproducing and interpreting experimental results, we briefly discuss hybrid approaches that combine together theoretical and experimental information.

ARTICLE HISTORY

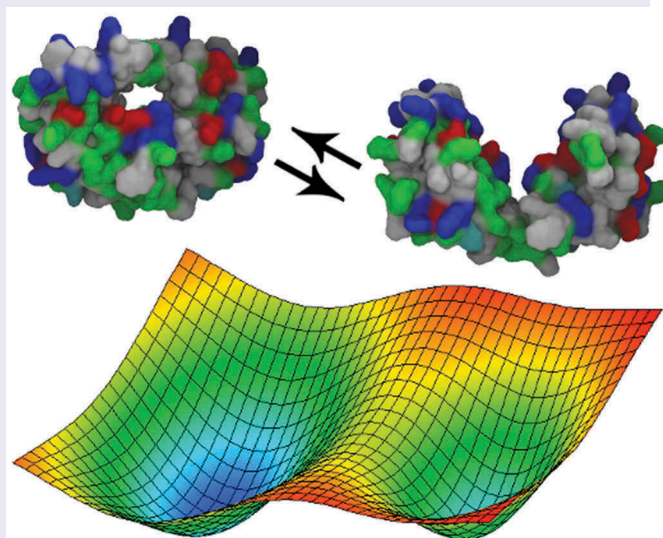
Received 15 March 2018
Accepted 7 May 2018

KEYWORDS

Molecular simulation;
molecular dynamics;
enhanced sampling;
free-energy landscapes;
kinetic models; biophysics

PACS

87.10.Tf 87.15.ap 07.05.Tp
87.15.A-



CONTACT Fabio Pietrucci  fabio.pietrucci@sorbonne-universite.fr; Carlo Camilloni  carlo.camilloni@uni-mi.it

© 2018 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abbreviations: CV: Collective variable. DHAM: Dynamic histogram analysis method. DoF: Degree of freedom. H-REX: Hamiltonian replica exchange. M&M: Metadynamic metainference. MetaD: Metadynamics. MD: Molecular dynamics. MSM: Markov state model. NMR: Nuclear magnetic resonance. PBMetaD: Parallel-bias metadynamics. pMaxEnt: principle of maximum entropy. PT: Parallel tempering. RECT: Replica exchange with collective-variable tempering. RMSD: Root mean square deviation. SGOOP: Spectral gap optimization of order parameters. TICA: Time-lagged independent component analysis. TRAM: Transition-based reweighing analysis method. US: Umbrella sampling. WHAM: Weighted histogram analysis method. WTMetaD: Well-tempered metadynamics.

Introduction: key problems

In recent years, molecular dynamics (MD) simulation techniques have been applied to a large and growing number of problems in biophysics and biochemistry, providing crucial qualitative and quantitative insight complementary to experiments. The three main reasons for this success are the increasing accuracy of physics-based force fields to model interactions between atoms, the growing system size and timescale that are accessible to simulation on modern computer facilities, and the progress in theoretical and numerical approaches to the *in silico* study of thermodynamics and kinetics.

The present review focuses on these latter methods that allow to explore the relevant configurations of a system and to estimate the corresponding equilibrium probabilities and transition times. In this context, the term ‘enhanced sampling’ emphasizes the need to accelerate the collection of geometric and energetic information across the configurational space of biomolecules. Indeed, the direct application of MD or Monte Carlo techniques is inadequate to most transformation processes, such as those involving conformational changes and binding/unbinding of partner molecules [1]; these so-called rare events display a slow kinetics, ranging from microseconds to minutes and beyond, due to the presence of bottlenecks separating the different metastable states of the system.

Such bottlenecks, from the viewpoint of free-energy landscapes traced as a function of some collective variable (CV) such as gyration radius, number of contacts, dihedral angles and so on, typically correspond to free-energy barriers (see e.g. Figure 1). The free-energy landscape point of view is indeed very intuitive from a molecular perspective, because it allows to reduce the dimensionality of a problem sometimes leading to identify its key ingredients. This has led to the development of a number of techniques focused on enhancing the sampling of equilibrium properties,

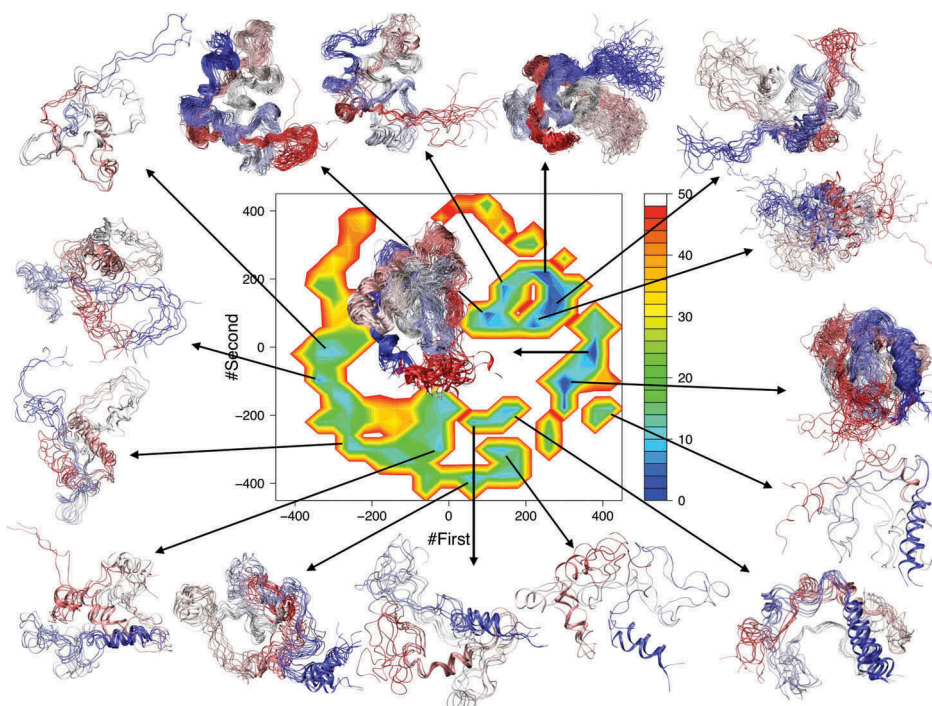


Figure 1. A free-energy landscape (in kJ/mol) for the denatured state of Frataxin, adapted from ref. [253]. The two collective variables for the projection were determined making use of Sketchmaps [104]. Multiple conformational substrates, with their relative free energies and displaying different extent of secondary structure are clearly visible.

often disregarding kinetic information. In parallel, other methods have tried to focus on enhancing the sampling from a more kinetic perspective, focusing on reactive pathways.

The focus of this review, rather than going into the details of many methods, for which we refer the reader to a number of other reviews [2–9], is more that of highlighting a common thread: contamination between methods is fundamental to understand what are the key ingredients to perform statistically sound simulations. Essentially, the methods presented here are seldom used in their original form: in most of the cases they are used in combination with other methods. In particular, developments in enhanced sampling of equilibrium properties are learning from developments in enhanced sampling of kinetic properties and vice-versa, thus making less and less important such distinctions.

One should also not forget that sampling is only the preliminary step. It is the instrument we need to perform accurate numerical experiments (i.e. with error bars and reproducible). Once one can perform such experiments, it becomes possible to pinpoint the limitations of the overall simulation machinery and to improve it. This is what is going on with force-fields development

in the last few years, leading to stronger predictive power. Given intrinsic limitations of force-fields, additional accuracy can be gained making use of statistical inference tools. The last section of the review is focused on such developments. MD-based methods grow quickly in number and fashion more and more specific names: still, the guiding principle and final aim remains that of performing accurate *in silico* experiments.

The problem of thermodynamics: learning from kinetics

Given the Boltzmann configuration distribution $P(\mathbf{R}) \propto \exp[-U(\mathbf{R})/k_B T]$, where k_B is the Boltzmann constant, one can identify three ways to modify it and increase the probability of observing high-energy conformations and cross barriers: (1) By increasing the temperature T ; (2) By modifying the potential $U(\mathbf{R})$, that is the force field; and (3) By adding an external potential $V(\mathbf{R})$. Generally speaking, these strategies correspond, among other techniques, to (1) Simulated and Parallel Tempering [10–12] (PT); (2) Hamiltonian Replica-Exchange [13,14] (H-REX) and Thermodynamic Integration 15 and (3) Umbrella Sampling [16] (US) (and also H-REX for terms beyond the force field). Importantly, $V(\mathbf{R})$ is generally a function of a CV and can be in principle a time-dependent function $V(CV(\mathbf{R}), t)$ (cf. Metadynamics [17] (MetaD) and Steering-MD [18,19], among the others).

PT is probably the simplest enhanced sampling technique. By using only one parameter, the temperature, it is possible to increase the diffusion as well as the probability of exploring high-energy conformations. Multiple replicas of a system are run in parallel at different temperatures and exchanges among replicas are tried and accepted or rejected using Metropolis Monte Carlo. H-REX variants allow modifying parameters different from the temperature but works following the same scheme (cf. Fig. 2). In general, for replica-exchange methods, the probability of exchange between two replicas i and j is given by: $\min\left(1, \frac{P_i(\mathbf{R}_j)P_j(\mathbf{R}_i)}{P_i(\mathbf{R}_i)P_j(\mathbf{R}_j)}\right)$. The factors affecting the efficiency of PT and H-REX methods are the frequency of exchange and the probability of exchange, together determining the round-trip time (i.e. the time needed for one replica to move across the parameter space). Most common setups are based on high-exchange frequency and a uniform average probability of exchange between 10 and 40% [20–25]. In particular, it was shown that the higher the frequency of exchange the better [25] this resulted in additional variants of PT pushing the boundaries towards the infinite swapping limit [26].

In PT, the probability of exchange is determined by the overlap in the distribution of the potential energy. As a consequence, the number of replicas needed to cover a given range of temperatures with a given average probability of exchange scales with $\sqrt{N_{atoms}}$, making it suboptimal for typical biological

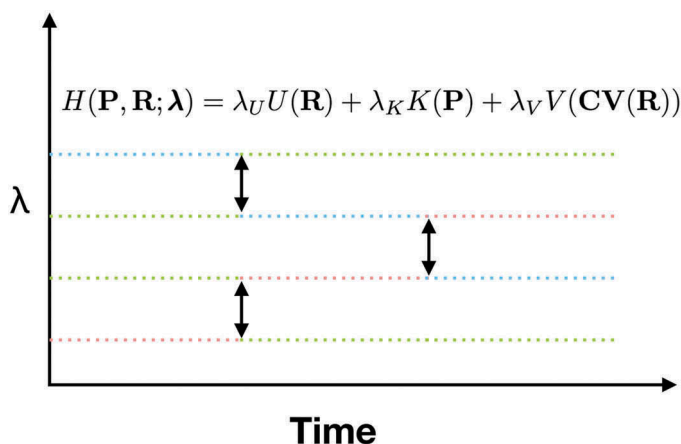


Figure 2. Replica-exchange simulations allow running multiple replicas of a system by varying the force-field, the temperature and/or adding an external potential. Exchanges between replicas are attempted periodically and accepted or rejected following Metropolis' rule. Algorithmically, one can exchange either configurations or Hamiltonians. In the first case, one has discontinuous trajectories of a thermodynamics ensemble, in the latter continuous trajectories not belonging to any thermodynamics ensemble. While the first are the actual result of the simulation, the latter are useful to assess convergence, i.e. proper sampling of the conformational space by each walker.

systems solvated in large boxes of water. Nonetheless, given its simplicity, PT is a widely used method to study protein folding and for the exhaustive search of the conformational space of short peptides, as well as for sampling other large conformational changes [12,27–30]. In the case of H-REX, there is not a general rule for the scaling of the number of replicas with system size. For example, in Solute Tempering [31], the energy of each replica m is scaled with respect to a reference temperature as: $E_m(\mathbf{R}) = E_p(\mathbf{R}) + \frac{\beta_0}{\beta_m} E_{ww}(\mathbf{R}) + \frac{\beta_0 + \beta_m}{2\beta_m} E_{pw}(\mathbf{R})$ (with $\beta_i = 1/k_B T_i$), where the energy is given by the sum of the energy of the solute E_p , the energy of the solvent E_{ww} and that of the solute-solvent interaction E_{pw} . As a consequence, the number of replicas needed to cover the same range of rescaling goes with $\sqrt{N_{solute}}$, the number of atoms in the solute. We remark that, using parameters that are more complex than the temperature makes less intuitive the assessment of whether a method is efficient in speeding up a process of interest; see for example Ref. [32] where the limitations of Solute Tempering are discussed. Following the spirit of Solute Tempering and trying to overcome its limitations, other H-REX schemes have been introduced [14,33–36]. A very elegant approach to optimize the number of replicas needed in PT and H-REX (as well as for simulated tempering) comes from MetaD (see later). By using the potential energy as CV (or in general the potential energy term relevant for H-REX), in the so-called Well-Tempered Ensemble [37], it is possible to increase

the fluctuations of the potential energy without changing its average, thus increasing the acceptance rate.

H-REX is not limited to scaling force-field terms but can also be used to add an external potential of increasing strength (also defined along a $CV(\mathbf{R})$) (cf. Fig. 2). Examples goes from increasing secondary structure propensities in sampling the conformational space of peptides and proteins [38], to increasing the repulsion between ligand and receptor to facilitate the unbinding and the search of more binding sites, in ligand binding problems [36,39]. Essentially, while PT can be considered an 'unbiased' method of enhanced sampling, the choice of the best variant of H-REX is determined by the specific process of interest and by our own understanding of the latter. The sampling of a reference (unbiased) replica can be readily used to obtain probability distributions of conformational parameters, while the sampling of the other replicas can be reweighed making use of techniques like the Weighted Histogram Analysis Method (WHAM) [40–42]. Importantly, in PT and H-REX simulations, convergence should be evaluated both for each single walker in the parameter space (i.e. continuous trajectories), as well as for the discontinuous trajectories sampled for a given parameter. Convergence of the former is usually harder to achieve, and is consequently more relevant, than the latter.

From PT to H-REX, one introduces his/her own understanding of the process to boost the sampling where it is needed. CV-based methods try to exploit even more some preliminary knowledge about a system. In the following, we will focus on the most recent advances in these latter methods, and in particular on MetaD. Indeed, in our opinion, MetaD has been responsible for spurring many developments of which other CV-based methods have benefited, in part because of its sensitivity to the choice of the CV. Indeed, many of the trends and problems highlighted in the following for MetaD are also valid for other CV-based methods like US [16], steering MD [18,19], the adaptive biasing force method [43,44], adiabatic free-energy dynamics [45,46], and temperature-accelerated MD [47].

CV-based methods can be seen as a tentative answer to the following problem: If we could know the free energy as a function of one or more CVs useful to describe a process, then we could sample along them optimally. The key point is that given a constant bias $V(CV(\mathbf{R}))$ it is possible to sample a system with a modified probability $P'(\mathbf{R}) \propto \exp[-(U(\mathbf{R}) + V(CV(\mathbf{R}))) / k_B T]$ and eventually recover the original probability distribution. US was and still is the method of choice to find such optimum biasing potential [48] and is usually applied in problems like conformational changes and ligand binding [49,50]. In its original implementation, the biasing potential was guessed and refined iteratively in order to increase the probability of sampling unlikely conformations [16]. This approach is not particularly easy without any prior knowledge

on the shape of the free energy, the height of the barriers and the relative weights of different states. In order to make the approach usable, the reconstruction was split by sampling the conformational space around a fixed number of points (windows) along the chosen CVs. The sampling of the windows is eventually merged by WHAM [51]. While this approach scales optimally, the number of windows and their separation cannot be determined a priori and should be such that there is always significant overlap between the sampling of neighbour windows, a property that is determined by the local shape of the underlying free energy. Nonetheless, additional sampling can always be added a posteriori. Attractive for its simplicity, US, as all CV-based methods, assumes that while directly sampling N degrees of freedom (DoF) is too time consuming, sampling $N - 1$ (or $N - n$ with $n \ll N$) is feasible. The validity of this assumption in real simulations can be misjudged by the fact that, given some overlap between the CV distribution in neighbour windows, WHAM will always converge to a solution, irrespectively of the global convergence of the sampling [42,43]. Indeed, most of the times US is used in combination with very simple CVs like distances. One should always verify by block analysis, autocorrelation analysis, and, most importantly, comparisons among multiple windows that the sampling in other unbiased directions is actually converged before applying WHAM [52]. In order to improve the convergence of the sampling, US is often combined with replica-exchange [53–55] (for example running PT for each window or exchanging between neighbour windows) and MetaD [56], among others.

With the introduction of MetaD [17], the iterative approach of building a bias $V(CV(\mathbf{R}))$ in the spirit of the original US is back on the spotlight. In MetaD, a biasing potential is built over time as the accumulation of kernel functions (typically Gaussians), colloquially named hills, that are deposited in such a way to keep track and discourage the visiting of already-visited conformations defined as a function of few CVs. As for US, the assumption is that enhancing the sampling along few DoF should be enough to converge all the others in the time of the simulation. In the case of MetaD, if this assumption does not hold, hysteresis effects will make it impossible for the history-dependent bias to converge [57]. The possibility to visualize such effects provides a very useful diagnostic tool. MetaD thus shifted the problem from determining $V(CV(\mathbf{R}))$ towards the determination of relevant CVs able not only to discriminate between different states, but also to approximate the transition mechanisms between them. This has driven a dramatic increase in the variety and complexity of CVs employed, well beyond traditional distances, angles and root mean square deviations (RMSD) with respect to known configurations, and thus making a service for all other methods based on CVs. For example, for biological systems, the effort resulted in the development of CVs to follow paths in conformational space [58], CVs specific for proteins [59,60] or nucleic acids [61],

CVs for interfacial water molecules [62], CVs for experimental observables [63–65], for shapes [66], and more (cf. the manual of the PLUMED code for more examples [67,68]).

In the last 15 years, MetaD has seen a considerable progress and many biological applications [57,69,70]: From a method based on heuristic assumptions [17,71], it is now well established in its formalism [72–74] (cf. Ref. [75] for a review). Running a MetaD simulation requires the choice of few parameters in addition to the choice of the CVs:

- (1) A target effective free-energy (that is the sum of the system free-energy along the CVs plus the MetaD bias) [76–80]. In the case of well-tempered MetaD (WTMetaD) [76], this is the free energy scaled by a constant factor. Therefore, the only parameter to choose is the scaling factor (or bias factor, e.g. a bias factor of 10 means that the effective free energy will be that of the system scaled down by ten times). In other cases, the target distribution can be different from a scaled free energy and could for example match a probability distribution obtained from an experiment [79,80];
- (2) The initial deposition rate, i.e. the ratio between the height of the Gaussian and the frequency of addition (again in the case of WTMetaD, since the height is quickly scaled down it is often reasonable to have a high initial rate, for example, of the order of a fraction of $k_B T$ per picosecond);
- (3) The widths of the Gaussian for the different CVs, or better, by using multivariate Gaussian [81], either the timescale for the estimate of the fluctuations or the typical atomic displacement (i.e. one single number irrespectively of the number of CVs).

Overall, Metadynamics is robust with respect to the choice of the above parameters, while one open issue is the treatment of the boundaries [82–84] for multidimensional MetaD, even if in WTMetaD the problem might be mitigated with respect with MetaD with fixed Gaussian height. Most importantly, it is generally possible to reweigh a MetaD sampling and obtain back the unbiased statistics for biased and unbiased DoF [74,81,85–87]. In particular, in the case of WTMetaD, one can reweigh the sampling making use of the US weight, i.e. the weight of each configuration as given by the bias at the end of the simulation [81].

We note that the concept of a target equilibrium state has recently brought to the development of a new method, called variationally enhanced sampling, where the bias is based on a functional form meant to variationally reach a desired target [88], possibly allowing a more efficient boost of the sampling in many dimensions.

The performances of all CV-based methods, as well as their chances to achieve a satisfactory convergence within a given amount of computing resources, are significantly dependent from the choice of the CVs [89]. Suboptimal choices can be alleviated by combining MetaD with PT or H-REX [90–93]. Particularly interesting is the case of Bias-Exchange MetaD [83,91]: N replicas of a system (typically up to 10) are simulated in parallel, each replica being biased along a different CV, with exchanges of bias attempted at regular time intervals (according to Metropolis' rule). In this way, each replica is able to explore in an efficient way the N -dimensional space spanned by the ensemble of all CVs, by sequentially changing the bias direction in this space. After having verified the convergence of the bias profile corresponding to each CV, data from all replicas can be combined together by means of WHAM to reconstruct the N dimensional free-energy landscape (originally, a neutral, unbiased, replica was included to simplify the analysis) [87], employing, e.g. the METAGUI graphical interface [94,95]. As a consequence, the choice of the CVs is less critical, since it becomes possible to use as many CVs as the number of replicas one can afford to simulate, a key advantage in studying complex biomolecular processes involving different types of conformational changes, molecular recognition, water degrees of freedom, etc. [62,63,96–99].

An alternative approach to employ many CVs in an efficient way is provided by replica exchange with collective-variable tempering (RECT) [93] in this case, several replicas are evolved in parallel, each replica being biased by the sum of N independent WMetaD bias potentials acting along different CVs, with different bias factors in the different replicas. Since the biased CVs are not necessarily orthogonal, each bias potential can have a complicated effect on probability distribution along the other CVs: as a consequence, an unbiased replica must be used to accumulate the unbiased sampling. In Parallel-Bias MetaD (PBMetaD) [100], multiple unidimensional MetaD are performed at the same time on a single simulation but the Gaussians are scaled to account for the hidden correlations (cf. Fig. 3). Consequently, each single MetaD converges to the correct WMetaD free energy. Heuristically, efficiency can be kept at an optimum level by increasing the bias factor by $\sqrt{N_{CV}}$ with respect to that of a standard WMetaD. Additional parallelization can be trivially obtained through multiple walkers MetaD [101], and/or using the whole arsenal of possible hybrid methods. For example, one could easily devise a PT-Well-Tempered-Ensemble-PBMetaD, where many CVs are biased by PBMetaD including the potential energy (Well-Tempered Ensemble [37]) and an optimal number of replicas are run in parallel at different temperatures as in parallel-tempering MetaD [90].

Even by using a large number of CVs, relevant slow DoF can still escape from the list and the intuition. Furthermore, using too many CVs can become

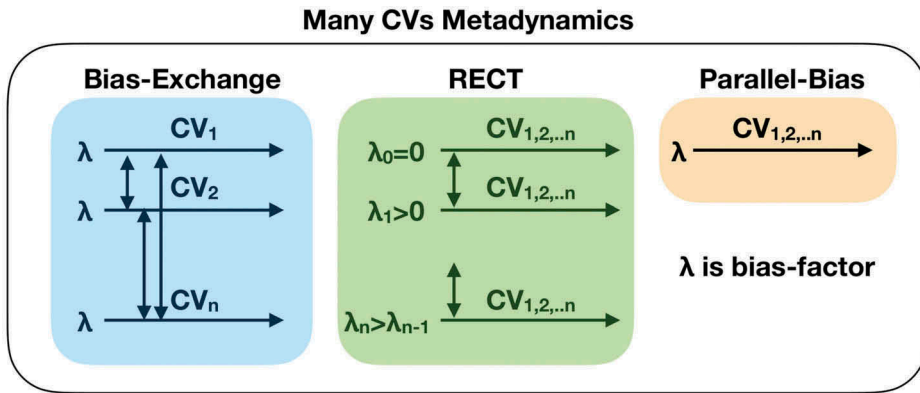


Figure 3. Metadynamics variants that allow increasing the number of CVs employed in a simulations. In Bias-Exchange MetaD, an H-REX scheme is employed to couple the sampling obtained by many independent unidimensional MetaD. The final sampling is obtained by merging that of all replicas making use of WHAM or other reweighing schemes. Replica exchange with collective-variable tempering is again an H-REX scheme where each replica is biased by many unidimensional bias potentials, with different replicas featuring different bias factors, including a neutral unbiased replica to extract the final sampling. In Parallel Bias MetaD, a single replica is biased by many unidimensional MetaD. The rate of deposition is tuned to keep into account for cross-biasing of non-orthogonal CVs. The final sampling is obtained reweighing for the accumulated bias.

computationally demanding. How can one alleviate this issue and simplify or standardize the choice of the CVs? Ideally, it should be possible, either iteratively or on-the-fly, to identify those DoF that are slow with respect to the timescale accessible by standard MD. One possibility is that of extracting information from preliminary simulations, for example making use of dimensionality reduction techniques such as principal component analysis [102], Isomap [103] or Sketchmaps [104] (e.g. Figure 1). Alternatively, if the initial and final structures of a transformation process are known and well defined, a path variable could be optimized on-the-fly [105]. These methods try to indirectly solve an additional problem: If we could identify the slowest CVs of a process we could use this information to optimally enhance the sampling. Recent methods try to directly face this problem: in the case of SGOOP, spectral gap optimization of order parameters, [106] starting from a CV defined as the linear combination of simpler CVs it is possible to optimize the coefficients in order to maximize the spectral gap. Interestingly, the method can be used to iteratively move towards a better and better CV. What we think is very relevant in this direction are two complementary strategies [107,108] that have been recently proposed. Both are based on a recent advancement in the field of conformational dynamics, namely time-lagged-independent component analysis (TICA) [109,110] (of notice that TICA has also been used in combination with US [111]). TICA is a signal analysis procedure, related to principal component analysis, where one

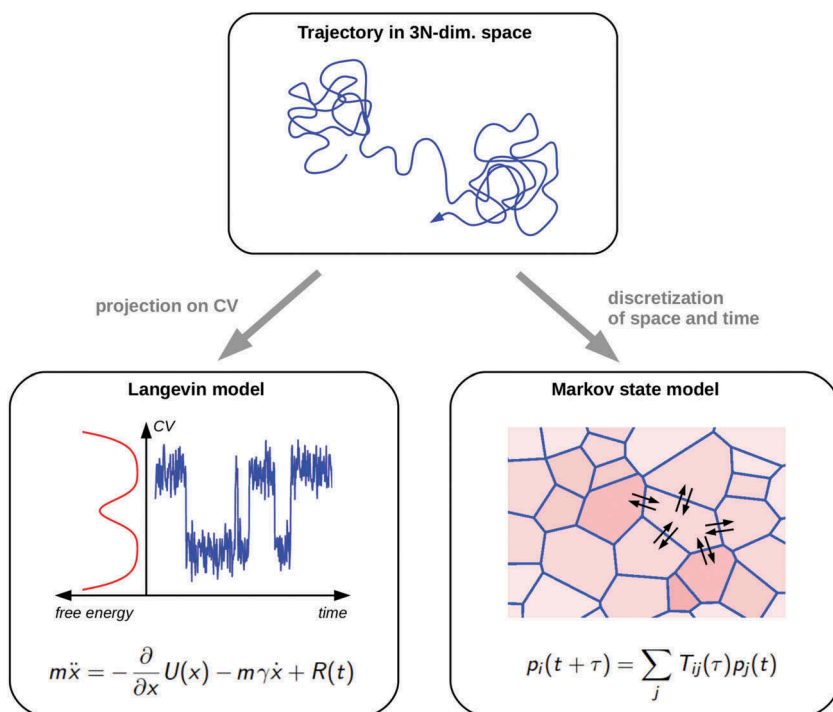


Figure 4. Illustration of the two most common approaches to the construction of kinetic models starting from dynamical trajectories in the full $3N$ -dimensional space. Projection on one (or a few) CV leads to stochastic differential equations (Langevin equations) containing the gradient of the potential of mean force (here indicated as $U(x)$) as well as friction and noise terms modelling the effect of the ‘bath’ coordinates. Discretization of configuration space and introduction of a lag time τ lead to Markov state models where the population of different configurational clusters (here depicted as cells) evolves in time based on a matrix of transition probabilities.

identifies directions with the slowest autocorrelation times. In the first approach [107], closer in spirit to the SGOOP method, a first CV is defined as a linear combination of a basis set of simpler CVs (e.g. the dihedral angles of a protein). Then a WTMetaD simulation is performed to get a preliminary estimate of the free energy and of the direction to overcome a free-energy barrier; after reweighing, the TICA eigenvalues and eigenvectors are obtained. From the characteristic time of the eigenvectors, one has the indication of how many CVs are needed to sample the system given the available computing time. In the second approach [108], a preliminary sampling by multiple MD simulations in the spirit of Markov State Model analysis is performed and projected again on a basis set of simple CVs, while TICA is once again used to determine optimal CVs to sample the system accurately. A relevant observation is that one should be able to bias all the eigenvectors associated with a timescale longer or comparable to the one affordable by standard MD; this goes towards a better formalization of how many CVs should be biased so that

all other DoF can be sampled at convergence in the available simulations time. From this point of view we think it is appealing the combination of these approaches with multiple-CVs methods like Bias Exchange MetaD, RECT or PBMetaD. Note that a set of CVs that is optimal according to the preceding criteria, is expected to be optimal in particular for studying the kinetic properties. The last years saw also the beginning of the use of machine learning techniques aimed at finding both the optimal CVs and the optimal bias at once [112–115], so that new interesting developments and applications are expected in the near future (see also ref. [116].).

Given the level of efficiency and sophistication reached today by enhanced sampling techniques, we believe that the time is ripe to address some open challenges. First, the reproducibility of simulations should be improved: in this respect, an important role can be played by libraries such as COLVAR [117], PLUMED [67,68] and SSAGES [118], making enhanced sampling protocols available on multiple MD engines (a good practice consists in sharing the input in the supporting material of papers or on platforms like GitHub). A related problem is reaching agreement in the community around robust methods to estimate errors: unfortunately many computational works lack error estimates, and in several cases there is no consensus yet, even in principle, on the best statistical tools to employ. Finally, our understanding and quantification of the kinetic properties of biomolecular systems – the subject of next section – is often insufficient: as discussed above, advances in this direction can be highly beneficial to improve sampling and free-energy calculation methods.

Theoretical and computational approaches to kinetics

As explained in the previous section, the reconstruction of free-energy landscapes is nowadays a mature field, where a number of successful strategies have been highly optimized and, in most cases, theoretically formalized. An important current challenge consists in devising systematic approaches to predict the kinetics of complex biological systems in an affordable and accurate way. Clearly, thermodynamics can be obtained from kinetics, but not vice versa, and a complete *in silico* picture of a biophysical process includes a detailed account of kinetics. The latter is particularly important in order to compare with experiments, examples being temperature-jump and fluorescence techniques for protein folding [119] or surface plasmon resonance for protein-ligand interaction [120,121].

A deep connection between free-energy landscapes discussed in the previous section and kinetic concepts is represented by the committor (or commitment) probability function, i.e. the probability $p_B(\mathbf{R})$, in a two-state system, that a given atomic configuration \mathbf{R} will evolve to reach state B before reaching state A

[2,122,123]. An example popular in biophysics is $p_{\text{fold}}(\mathbf{R})$, indicating the probability to reach the folded state of a globular protein [122]. Even if the committor function cannot be estimated explicitly in the whole configuration space due to computational cost (multiple trajectories must be evolved from each \mathbf{R} until ending in a metastable state), it nevertheless provides a definition of ideal reaction coordinate for drawing free-energy landscapes, and it is well suited as a benchmark tool: e.g. a structure belongs to the transition state ensemble if it is committed to A and B with equal probabilities [2,89,123,124].

In a few, ideal situations, mean first passage times (i.e. inverse rates) connecting metastable states can be directly measured from long MD simulations displaying an ergodic behaviour. An example is the conformational dynamics of few-amino acid peptides, as reconstructed from trajectories on the microsecond timescale, nowadays easily feasible. Remarkably, for small proteins close to the melting temperature (where $\Delta G = 0$ between folded and unfolded states), reversible folding/unfolding millisecond trajectories in explicit solvent could be generated using a specifically designed supercomputer, Anton [125,126]. Unfortunately, in more general cases and for a vast range of biophysical processes, kinetics remains a difficult and prohibitively expensive simulation target. Other fields, such as materials science and solution chemistry, face a similar unsatisfactory condition.

Two main theoretical frameworks serve as conceptual basis for existing numerical approaches to kinetics [127–129]. In the first framework, projection of the high-dimensional dynamics of a complex system on a low-dimensional manifold of CVs (often only one) leads to dynamical equations of the Langevin kind, i.e. stochastic differential equations including a deterministic force (the gradient of the free energy), a friction and a random force (both conveying the influence of the neglected degrees of freedom) [130–132]. In the second framework, the configuration space is partitioned into discrete sets of structures, called clusters or microstates, connected by transitions that typically are assumed memory-less. This approach results into Markov state models, complemented with mathematical tools to analyse them [9,133]. Even if the two theoretical frameworks are interrelated, in practice they lead to different numerical techniques.

Markov state models represent a widespread approach to kinetics [133–138]. Typically, the approach is applied to protein conformational dynamics, and it follows four steps:

- (1) The configuration space of the system is explored, to some extent, using an atomistic simulation technique (e.g. MD);
- (2) The explored configuration space is partitioned into hundreds of clusters (sets of similar structures);

- (3) Local transition rates between pairs of connected clusters are estimated, obtaining a kinetic matrix;
- (4) Metastable states are identified as kinetically connected basins (sets of clusters linked by fast internal transitions), and transition rates between them can be compared with experiments.

The first step is where the computational cost, often very large, is concentrated. In most cases, a sufficiently long MD trajectory repeatedly crossing among all the relevant metastable states is unfeasible, and, besides, it might not be an efficient way of exploiting computer resources. The main idea is that global equilibration, i.e. observation of steady probabilities and currents over the whole configuration space, can be substituted by local equilibration, region by region, which is easier to achieve. MSM analysis is very well suited to the kind of parallel independent simulated trajectories which can be generated ('high throughput') by distributed computing and GPUs [133,139]. Enhanced sampling techniques can be invoked to choose in an effective way the starting points of such trajectories [9]. For instance, one can pick them from a preliminary inexpensive MetaD simulation that explores a wide portion of the configuration space [140].

In the second step, the configuration space is subdivided into small parts, clustering together sets of structures on the basis, usually, of a distance metric for structural similarity. Common choices are the RMSD of Cartesian coordinates, distance in dihedral angles space or contact-map space and principal components of some of the previous coordinates. A wealth of effective clustering algorithms is available to this purpose [141–143]. This is a critical step, since the partitioning, ideally, should group together only structures connected by fast transitions, and not structures whose interconversion implies long time scales. Unfortunately, the relationship between distance metrics in configuration space and time-domain distances is unclear [144–148]. Both the spatial discretization and the choice of an observational lag time are the source of systematic errors: in this context, it has been proposed that instead of partitioning the whole configuration space, Markov state models can be constructed including only a limited number of core sets, formed by small regions surrounding free-energy minima, allowing to better distinguish real transitions between metastable states from fast recrossing [134,149]. Furthermore, an interesting development consists in the use of TICA [109,110], explicitly seeking an optimal space of slow coordinates for spatial discretization (see also previous section). In this direction, machine learning approaches are now under investigation to learn optimal coordinates [115].

In the third step, transition rates between nearby clusters are estimated, obtaining a kinetic matrix and, hence, an explicit master equation [134]. In the simplest approach, rates are estimated directly as the inverse of the

mean first passage time, i.e. by dividing the number of observed $i \rightarrow j$ transitions by the total time spent in the starting cluster i .

In the fourth step, a few kinetic basins – each one formed by quickly interconverting microstates – are identified out of the set of clusters. Basins are usually assigned well-defined ‘labels’ (e.g. the native state, the unfolded state, a misfolded state) and, in favourable cases, can be experimentally observed. Available techniques for basin identification include the analysis of the sign structure of the eigenvectors of the kinetic matrix, clustering algorithms, and hidden Markov state models [150]. A simple approach, once the attractors of the kinetic basins are identified, consists in explicitly performing a committor analysis: a large number of random jump sequences (obtained with kinetic Monte Carlo or stochastic simulation methods [151,152]) are started from each cluster, and the latter is assigned to the attractor that is reached first in the majority of cases [94,98]. Finally, transition rates between kinetic basins – typically much slower than rates between adjacent microstates – can be estimated either by analysing the kinetic matrix or, again, using kinetic Monte Carlo, and associated to experimental observables like folding times or protein-ligand association/dissociation rates. Available software packages for the estimation, validation and analysis of Markov state models include PyEMMA [153], HTMD [154], MSMBuilder [155] and METAGUI [94,95].

To date, most Markov state models were targeted to the folding of small proteins [156,157] (and, recently, also conformational changes in disordered proteins [158] and in RNA oligonucleotides [159]). However, a growing field of application is the binding of small ligands to proteins, a problem of great interest for drug discovery [120,160–164]. Still, many important biomolecular processes have time scales beyond the second and remain hard to tackle with Markov state models based on standard MD. Examples are protein folding far from the melting temperature, as well as many protein–drug/protein–protein interaction or protein aggregation processes. In this respect, an alternative source of data for Markov state models construction is represented by enhanced sampling simulations. As shown in Ref. [83,87], relatively cheap bias-exchange simulations (cf. Fig. 3) can be employed for this purpose. In this approach, multiple replicas of the system are biased along different CVs, capturing the relevant degrees of freedom, and the results are combined to reconstruct a high-dimensional (discretized) free-energy landscape by WHAM [87] (see previous section). Transition rates between adjacent bins in the landscape are estimated employing an explicit formula containing free energy differences and the diffusion tensor, obtained by discretizing the Smoluchowsky equation [87,165,166]. At this point, the Markov state model can be analysed as explained above. The approach has been applied to challenging problems including the folding of globular and intrinsically

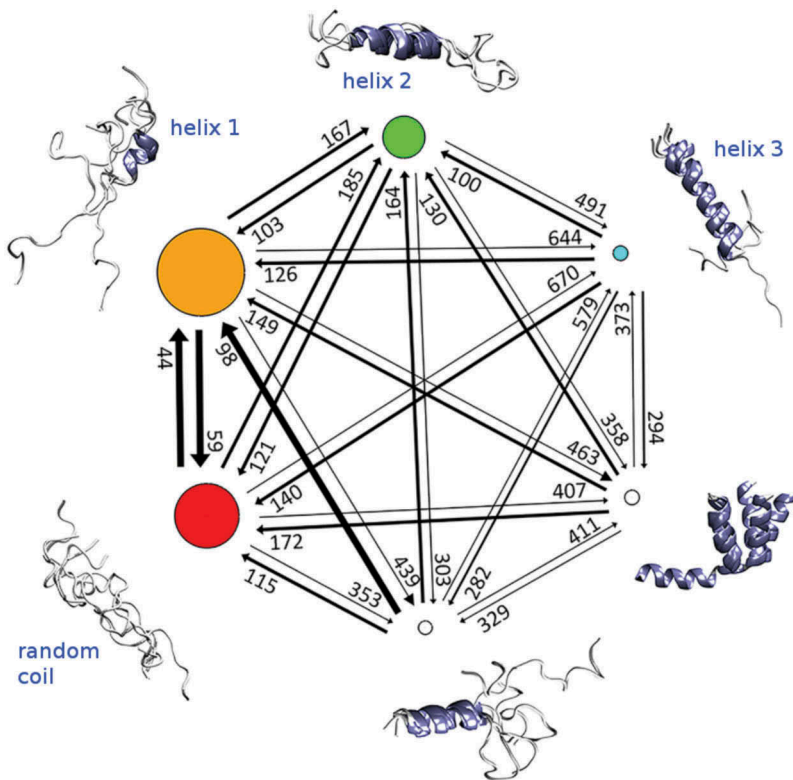


Figure 5. Interconversion kinetics between the conformational states of the C-terminal domain of Sendai virus nucleoprotein, an intrinsically disordered protein. Disc areas are proportional to equilibrium populations, while mean first passage times (in ns) are indicated on the arrows. The kinetic model has been constructed starting from all-atom explicit-solvent simulations, exploiting parallel-tempering metadynamics [90] in the well-tempered ensemble [37], combined with a bin-based Markov state model based on the Smoluchowski equation [87]. The figure is adapted from Ref. [169].

disordered proteins [87,121,167] (see Fig. 5), ligand-protein (un)binding kinetics [62], and drug permeation across membranes [168].

It is interesting to note that, recently, the world of free-energy calculations and the world of Markov state models have been further connected by the introduction of trajectory reweighing schemes like the dynamic histogram analysis method (DHAM) [170,171] and the transition-based reweighing analysis method (TRAM) [172], exploiting also replica-exchange MD simulations [173,174]. Moreover, within the approximations of transition state theory, techniques like conformational flooding [175], hyperdynamics [176] or MetaD [177,178] can be exploited to estimate kinetic rates – provided the transition state region remains bias free – properly accounting for the boosting effect of the bias. This approach led to interesting results on ligand binding kinetics and thermodynamics [179,180]. In an alternative approach, the ranking of unbinding rates for a series of congeneric ligands is inferred

based on simulations subject to a scaled (smoothed) potential energy surface, to which restraints are added to enforce protein stability [181].

In a class of approaches sharing some analogies with Markov state models, kinetic rates are reconstructed by means of a large number of relatively short MD trajectories that, taken collectively, reconstruct the progression of the transition pathways going from an initial to a final state. The region of configuration space connecting the two metastable states is subdivided into a number of 'slices' by suitable interfaces, and trajectories are repeatedly initiated at one interface and ended once they reach the next interface. The collected statistics is analysed, usually in terms of a Markov model, to access long-time kinetic information. Such techniques include forward flux sampling [182], milestoning [183], transition interface sampling [184] and non-equilibrium umbrella sampling [185]. Typically, a CV is used to define interfaces, and the computational cost of the protocol can often be large. In principle, even if individual trajectories evolve in an unbiased manner, the specific algorithm used to select their starting configurations might introduce a form of bias in the simulations. Note that in its original formulation, transition path sampling aimed at the direct generation of continuous pathways joining the end states, starting from an initial guess, exploiting a Markov chain Monte Carlo procedure called shooting [2,186]. The software packages OpenPathSampling [187] and PyRETIS [188] have been specifically developed to perform and analyse path sampling simulations.

A different approach to kinetics is based on a projection of the high-dimensional atomic dynamics on a low-dimensional space of CVs, resulting in Langevin stochastic differential equations [132,189–191]. The latter are widely acknowledged as a key conceptual tool to discuss activated processes, even if their application to make quantitative predictions on specific systems is less widespread than Markov state models (probably, also due to the lack of widespread and flexible software tools). With respect to the latter, the Langevin approach has a more intuitive physics-based form and avoids the approximations represented by the discretization of configuration space and by the choice of a lag time. The price to pay is the effort, often non-trivial, to identify one (or a few) CV, meant to approximate the ideal reaction coordinate. An important advantage is the possibility to adopt different types of stochastic equations: for instance, even if the overdamped Langevin equation (corresponding to the Smoluchowski Fokker–Planck equation [132,189]) is the most employed form in the biophysical community, inertial and non-Markovian effects can also be included in Langevin equations when necessary [131,192–194].

Langevin equations have been shown to represent a flexible and powerful tool for the study of a range of complex systems, from proteins to chemical reactions in solution [195–197]. Given its limited computational

cost, numerical integration of Langevin equations simplifies the extraction of mean first passage times, committor functions, etc., according to well-established mathematical procedures [2,118,123,124,198]. In numerical applications, the stochastic equation, e.g. for the conformational dynamics of small peptides, can be obtained from long equilibrium simulations or from enhanced sampling approaches. The drift, friction and noise fields can be directly obtained as trajectory averages, or employing self-consistent maximum-likelihood approaches [166,199–203]. In Ref. [194], an efficient data-driven Langevin model is constructed on-the-fly starting from local averages over the original MD trajectory, demonstrating that, already for a simple 9-residue peptide in solution, five CVs and the inclusion of small-damping effects can be necessary to accurately reproduce the observed hierarchy of time scales.

How accurate are *in silico* predictions and how to improve them?

In the last years, large efforts have been devoted to assess the accuracy of the most widespread force fields for the simulation of biomolecules. Several studies evaluated explicit-solvent atomistic models by generating an ensemble of equilibrium configurations of short peptides and globular proteins in the native state, comparing the results with NMR observables including chemical shifts, scalar couplings, residual dipolar couplings, nuclear Overhauser enhancements and relaxation order parameters [204–210]. In these works, employing MD trajectories ranging from the sub-microsecond to the millisecond timescale (sometimes in combination with enhanced sampling methods), the most recent force fields were found to accurately reproduce experimental structures, while still missing a quantitative agreement with NMR observables. Folding times and free energies, as well as melting temperatures, are, in general, accurately reproduced, whereas both folding enthalpies and heat capacities tend to be underestimated [210].

Furthermore, unfolded states of proteins as observed in simulations often display an incorrect balance of secondary structure types, and typically they are excessively compact compared to, e.g. small angle X ray scattering experiments: these problems are particularly relevant to intrinsically disordered proteins, and were the focus of several recent force-field assessment and improvement studies [210–213]. Among possible reasons, it has been pointed out the lack of an accurate description of protein–water interactions and an underestimation of water dispersion forces: both problems have been targeted with specific corrections that significantly improve the agreement with experiments [214–216], even if results are still force-field dependent.

Importantly, in the case of the simulation of RNA structures, even if the corresponding atomistic force fields traditionally lack the accuracy of their protein counterparts, significant improvements have been recently reported [217,218]. Finally, we only mention that coarse grained models

can be an effective alternative to fully atomistic ones when it is necessary to simulate significantly larger system sizes and time scales: this is a broad and very active field of research that goes beyond the scope of the present review.

The fact that, clearly, discrepancies between simulation and experiments can arise both from force-field limitations and from sampling limitations underlines, once again, the important role of enhanced sampling approaches in molecular simulations [219]. The interplay between the two types of limitations is crucial also in the difficult field of protein-ligand interactions [6,220,221]: one of the major open challenges of biomolecular simulations, both from the theoretical point of view and for its practical implications, remains the accurate and affordable prediction of the binding free energy and kinetic rates of protein-protein and protein-drug complexes.

Several attempts were done to critically compare the performance of different free-energy calculation methods [139,222–228]. This is a non-trivial task, which becomes even more difficult if kinetic properties are the target. In this context, software tools that implement a range of different enhanced sampling approaches can facilitate the comparisons [67,68,117,229]. It is desirable to enlarge the set of agreed benchmark systems of higher complexity than customary 2–10 residue peptides, including also protein-protein [230] and protein-ligand association cases (see e.g. the SAMPL challenge [231]), better approximating the rich behaviour observed in realistic systems of interest. In the case of folding, the set of extensive trajectories generated with the Anton supercomputer [125] recently played a constructive role in the development and assessment of sampling methods.

Generally speaking, one can anyway wonder whether transferable force-fields can be quantitatively and equally accurate for a large class of different systems. What usually happens is that a simulation is performed, agreement with some data is tested, and then a choice is done between using the results (for prediction or interpretation of experiments) or, alternatively, repeating the simulation with a different force field hoping for the best. From a statistical perspective, one can try to tackle the problem of the accuracy of a simulation before running it, or try to reweigh a posteriori the simulation to improve the agreement with the available experimental data (cf. Figure 6). The question to ask is, following Bayes, given our prior knowledge and some available data, what is the best model that describes the data? Our prior knowledge is a state-of-the-art force-field, the data are usually equilibrium experimental observables like those available from nuclear magnetic resonance, fluorescence, small angles X-ray scattering, etc. [232–235].

Reweighting approaches can always be used a posteriori without changing the original results and are usually computationally inexpensive. On the other hand, they are strongly affected by the original sampling. If relevant states have not been sampled, reweighting will not make them

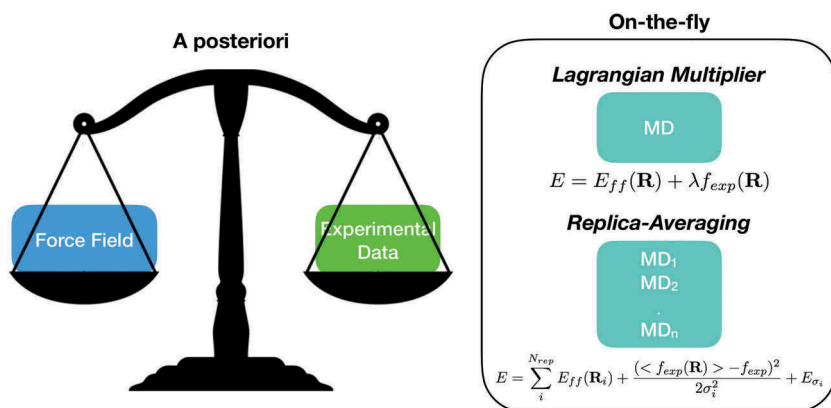


Figure 6. Hybrid schemes for the integration of experimental and theoretical information. A posteriori reweighing schemes allow to modify the weights of the sampling resulting from a MD simulation, while on-the-fly schemes can integrate directly experimental data in the energy function employed in the simulation. Among on-the-fly schemes, in the case of the MaxEnt method based on the solution of the Lagrangian multiplier[235], the energy results from the energy of the force-field E_{ff} and a linear contribution from experimental observables. In the case of replica-averaging methods like MetaInference [244], multiple replicas of the systems are simulated in parallel and the energy is given by E_{ff} plus an harmonic coupling with experimental data whose strength is related to an estimated error σ_i .

appear. Alternatively, there are hybrid methods based on the direct inclusion of additional energy terms related to the experimental information, than can in principle overcome intrinsic limitation of force-fields functional forms [234].

The problem one wants to solve is what is the minimum perturbation of the force-field that will result in an ensemble of conformations in agreement with an equilibrium experimental data. The minimum perturbation is an important ingredient because it allows to include only the experimental information and nothing more. This can be theoretically formulated using the principle of maximum entropy (pMaxEnt) [236]. Pitera and Chodera [237] showed how to implement pMaxEnt by a linear bias applied on a single simulation with an unknown force constant (the Lagrangian multiplier) that can be obtained iteratively or on-the-fly during the simulation [228,237–239]. Alternatively, instead of searching for the Lagrangian multiplier, one can perform multiple simulations in parallel and couple them with an harmonic potential centred on the value of the equilibrium observable of interest, the so called replica-averaging approach [240–242]. Particularly important is the ability of such methods to take into account errors, affecting the experimental data as well as the predictors used for their back-calculation. Again, both linear, single replica-approaches, as well as replica-averaging approaches can be extended to consider errors [239,243,244]. Note that, by using these

methods, it is possible to obtain results that depend less and less from the initial choice of the force field [245–247].

Single-replica as well as replica-averaging approaches can also be coupled with enhanced sampling techniques [235,248,249], the former usually making use of H-REX with an unbiased neutral replica, the latter by directly biasing all the replicas. For example, in the case of Metadynamic Metainference (M&M) [249], the sampling is typically enhanced by PBMetaD, accumulating the biases from all the replicas as in multiple walkers MetaD. This allows calculating weighted averages for the used observables, whose coupling with the experimental values (the errors) is sampled by Monte Carlo. These methods allow obtaining an equilibrium sampling of processes not accessible by standard MD, in agreement with the available experimental knowledge within some reasonable errors. Furthermore, they also provide an indication of the contribution of different sources of experimental data to the quality of the simulation, and can provide feedbacks about the quality of the experimental data themselves. Last but not least, these techniques make the dependence of the results on the initial choices less important, rendering overall more robust all the simulation process. MaxEnt approaches (with and without replicas) as well as Metainference are available and well documented in PLUMED (cf. EDS, ENSEMBLE, MAXENT and METAINFERENCE in the PLUMED manual), including a large number of methods to calculate experimental observables provided in the PLUMED-ISDB module [250].

Hybrid methods have been successfully employed in a number of problems to obtain results in quantitative agreement with experimental data, mostly NMR, including the characterization of the dynamics of folded [240,251,252] and unfolded/disordered proteins [253–257], in molecular recognition [258,259], membrane proteins [260,261] and nucleic acid dynamics [239,262].

An approach based on MaxEnt has also been proposed to reweigh and optimize a Markov State Model [263] to match experimental equilibrium observables. In addition to equilibrium observables, the Maximum Caliber approach [264] (cf. Ref. [265,267] for a review) can be used to infer distributions of trajectories in agreement with kinetic quantities [266,268]. A recent suggestion on how to couple Maximum Caliber with Markov state models can be found in Ref. [269], while a replica-averaging implementation has been proposed in Ref. [270]. This particular direction can be very promising in pushing the boundaries of out-of-equilibrium MD simulations.

Perspectives

Molecular modellers are living in an exciting time. Computing power is increasing and MD engines are getting better and better at exploiting it [271–276]. These improvements are effectively multiplied by using the computational approaches to free-energy landscapes and kinetic networks discussed in the previous sections, often reducing by several orders of magnitude the timescale needed to simulate rare events. A trend in method development, that hopefully will continue in the future, consists in combining thermodynamic and kinetic concepts more and more tightly together.

The first result is an increasing ability to obtain insight into biological problems, with an ever growing complementarity between simulations and experiments. A second, significant result, consists in the critical assessment of force-fields and their overall improvement to the benefit of the whole community. Whenever force-fields are not accurate enough, statistical inference, through hybrid methods for the inclusion of experimental data in MD simulations or reweighing, allows to increase the accuracy and the predictive power of simulations. Altogether, this is allowing simulations to move from a descriptive technique, ‘useful at least for theoretical chemists’ (quoting the joke in the introduction to the chemistry 2013 Nobel lecture) to a predictive and quantitative *in silico* platform for biomolecular experiments.

Acknowledgements

We thank Massimiliano Bonomi, Giovanni Bussi, and Toni Giorgino for useful comments and discussions.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Carlo Camilloni  <http://orcid.org/0000-0002-9923-8590>

Fabio Pietrucci  <http://orcid.org/0000-0002-4892-2667>

References

- [1] K. Teilum, J.G. Olsen and B.B. Kragelund, *Cell. Mol. Life Sci.* 66 (2009) 2231–2247. doi:10.1007/s00018-009-0014-6.
- [2] P.G. Bolhuis, D. Chandler, C. Dellago and P.L. Geissler, *Annu. Rev. Phys. Chem.* 53 (2002) 291–318.
- [3] D.M. Zuckerman, *Ann. Rev. Biophys.* 40 (2011) 41–62.

- [4] C. Abrams and G. Bussi, *Entropy* 16 (2013) 163–199.
- [5] M. Luitz, R. Bomblies, K. Ostermeir and M. Zacharias, *J. Phys.: Condens. Matter* 27 (2015) 323101.
- [6] A. Perez, J.A. Morrone, C. Simmerling and K.A. Dill, *Curr. Opin. Struct. Biol.* 36 (2016) 25–31.
- [7] D.L. Mobley and M.K. Gilson, *Ann. Rev. Biophys.* 46 (2017) 531–558.
- [8] T.J. Harpole and L. Delemotte, *Biochim. Biophys. Acta* 1860 (2018) 909–926. doi:10.1016/j.bbame.2017.10.033.
- [9] B.E. Husic and V.S. Pande, *J. Am. Chem. Soc.* 140 (2018) 2386–2396. doi:10.1021/jacs.7b12191.
- [10] E. Marinari and G. Parisi, *Europhys. Lett.* 19 (1992) 451. doi:10.1209/0295-5075/19/6/002.
- [11] U.H. Hansmann, *Chem. Phys. Lett.* 281 (1997) 140–150. doi:10.1016/S0009-2614(97)01198-6.
- [12] Y. Sugita and Y. Okamoto, *Chem. Phys. Lett.* 314 (1999) 141–151. doi:10.1016/S0009-2614(99)01123-9.
- [13] A. Mitsutake, Y. Sugita and Y. Okamoto, *Peptide Sci* 60 (2001) 96–123. doi:10.1002/1097-0282(2001)60:2<96::AID-BIP1007>3.0.CO;2-F.
- [14] R. Affentranger, I. Tavernelli and E.E. Di Iorio, *J. Chem. Theory Comput.* 2 (2006) 217–228. doi:10.1021/ct600235y.
- [15] M.J. Mitchell and J.A. McCammon, *J. Comput. Chem.* 12 (1991) 271–275. doi:10.1002/jcc.540120218.
- [16] G.M. Torrie and J.P. Valleau, *J. Comput. Phys.* 23 (1977) 187–199. doi:10.1016/0021-9991(77)90121-8.
- [17] A. Laio and M. Parrinello, *Proc. Natl. Acad. Sci. U.S.A.* 99 (2002) 12562–12566. doi:10.1073/pnas.202427399.
- [18] H. Grubmüller, B. Heymann and P. Tavan, *Science* 271 (1996) 997–999. doi:10.1126/science.271.5251.997.
- [19] C. Jarzynski, *Phys. Rev. Lett.* 78 (1997). 2690. doi:10.1103/PhysRevLett.78.2690.
- [20] C. Predescu, M. Predescu and C.V. Ciobanu, *J. Phys. Chem. B* 109 (2005) 4189–4196. doi:10.1021/jp0519870.
- [21] D. Sindhikara, Y. Meng and A.E. Roitberg, *J. Chem. Phys.* 128 (2008) 024103. doi:10.1063/1.2816560.
- [22] E. Rosta and G. Hummer, *J. Chem. Phys.* 131 (2009) 165102.
- [23] D.J. Sindhikara, D.J. Emerson and A.E. Roitberg, *J. Chem. Theory Comput.* 6 (2010) 2804–2808. PMID: 26616081.
- [24] M.K. Prakash, A. Barducci and M. Parrinello, *J. Chem. Theory Comput.* 7 (2011) 2025–2027.
- [25] P. Dupuis, Y. Liu, N. Plattner and J.D. Doll, *Multiscale Model. Simul.* 10 (2012) 986–1022.
- [26] T.-Q. Yu, J. Lu, C.F. Abrams and E. Vanden-Eijnden, *Proc. Natl. Acad. Sci. U.S.A.* 113 (2016) 11744–11749.
- [27] D.A.C. Beck, G.W.N. White and V. Daggett, *J. Struct. Biol.* 157 (2007) 514–523.
- [28] A. De Simone, A. Zagari and P. Derreumaux, *Biophys. J.* 93 (2007) 1284–1292.
- [29] A. De Simone, L. Esposito, C. Pedone and L. Vitagliano, *Biophys. J.* 95 (2008) 1965–1973.
- [30] P. Kührová, A. De Simone, M. Otyepka, R. Best and R.B. Best, *Biophys. J.* 102 (2012) 1897–1906.
- [31] P. Liu, B. Kim, R.A. Friesner and B.J. Berne, *Proc. Natl. Acad. Sci. U.S.A.* 102 (2005) 13749–13754. doi:10.1073/pnas.0506346102.

- [32] X. Huang, M. Hagen, B. Kim, R.A. Friesner, R. Zhou and B.J. Berne, *J. Phys. Chem. B* 111 (2007) 5405–5410.
- [33] S. Kannan and M. Zacharias, *Proteins* 66 (2007) 697–706.
- [34] S.L.C. Moors, S. Michielssens and A. Ceulemans, *J. Chem. Theory Comput.* 7 (2011) 231–237.
- [35] L. Wang, R.A. Friesner and B.J. Berne, *J. Phys. Chem. B* 115 (2011) 9431–9438.
- [36] M.P. Luitz and M. Zacharias, *J. Chem. Inf. Model.* 54 (2014) 1669–1675.
- [37] M. Bonomi and M. Parrinello, *Phys. Rev. Lett.* 104 (2010) 190601.
- [38] R. Meloni, C. Camilloni and G. Tiana, *J. Chem. Theory Comput.* 10 (2014) 846–854.
- [39] K. Ostermeir and M. Zacharias, *PLoS ONE* 12 (2017) e0172072–17.
- [40] A. Ferrenberg and R. Swendsen, *Phys. Rev. Lett.* 63 (1989) 1195–1198.
- [41] S. Kumar, J.M. Rosenberg, D. Bouzida, R.H. Swendsen and P.A. Kollman, *J. Comput. Chem.* 13 (1992) 1011–1021.
- [42] M.R. Shirts and J.D. Chodera, *J. Chem. Phys.* 129 (2008) 124105–124111.
- [43] T. Lelièvre, M. Rousset and G. Stoltz, *J. Chem. Phys.* 126 (2007) 134111–134119.
- [44] L. Zheng and W. Yang, *J. Chem. Theory Comput.* 8 (2012) 810–823.
- [45] L. Rosso, P. Minář, Z. Zhu and M.E. Tuckerman, *J. Chem. Phys.* 116 (2002) 4389–4402.
- [46] J.B. Abrams and M.E. Tuckerman, *J. Phys. Chem. B* 112 (2008) 15742–15757.
- [47] L. Maragliano and E. Vanden-Eijnden, *Chem. Phys. Lett.* 426 (2006) 168–175.
- [48] J. Kästner, *WIREs Comput Mol Sci* 1 (2011) 932–942.
- [49] S. Bernèche and B. Roux, *Nature* 414 (2001) 73–77.
- [50] Y. Deng and B. Roux, *J. Phys. Chem. B* 113 (2009) 2234–2246.
- [51] B. Roux and M. Souaille, *Comput. Phys. Commun.* 135 (2001) 40–57.
- [52] F. Zhu and G. Hummer, *J. Comput. Chem.* 33 (2012) 453–465.
- [53] S. Auer and D. Frenkel, *J. Chem. Phys.* 120 (2004) 3015–3029.
- [54] B. Roux, W. Jiang, Y. Luo and L. Maragliano, *J. Chem. Theory Comput.* 8 (2012) 4672–4680.
- [55] H. Kokubo, T. Tanaka and Y. Okamoto, *J. Chem. Theory Comput.* 9 (2013) 4660–4671.
- [56] V. Babin, C. Roland, T. Darden and C. Sagui, *J. Chem. Phys.* 125 (2006) 204909. doi:10.1063/1.2393236.
- [57] A. Laio and F.L. Gervasio, *Rep. Prog. Phys.* 71 (2008) 126601.
- [58] D. Branduardi, F.L. Gervasio and M. Parrinello, *J. Chem. Phys.* 126 (2007) 054103.
- [59] F. Pietrucci and A. Laio, *J. Chem. Theory Comput.* 5 (2009) 2197–2201.
- [60] M. Bonomi, F.L. Gervasio, G. Tiana, D. Provasi, R.A. Broglia and M. Parrinello, *Biophys. J.* 93 (2007) 2813–2821.
- [61] S. Bottaro, F. Di Palma and G. Bussi, *Nucleic Acids Res* 42 (2014) 13306–13314.
- [62] F. Pietrucci, F. Marinelli, P. Carloni and A. Laio, *J. Am. Chem. Soc.* 131 (2009) 11811–11818.
- [63] D. Granata, C. Camilloni, M. Vendruscolo and A. Laio, *Proc. Natl. Acad. Sci. U.S.A.* 110 (2013) 6817–6822.
- [64] D. Kimanius, I. Pettersson, G. Schluckebier, E. Lindahl and M. Andersson, *J. Chem. Theory Comput.* 11 (2015) 3491–3498.
- [65] F. Palazzesi, O. Valsson and M. Parrinello, *J. Phys. Chem. Lett* 8 (2017) 4752–4756.
- [66] J. Vymtal and J. Vondrášek, *J. Phys. Chem. A* 115 (2011) 11455–11465.
- [67] M. Bonomi, D. Branduardi, G. Bussi, C. Camilloni, D. Provasi, P. Raiteri, D. Donadio, F. Marinelli, F. Pietrucci, R.A. Broglia and M. Parrinello, *Comput. Phys. Commun.* 180 (2009) 1961–1972.

- [68] G.A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni and G. Bussi, *Comput. Phys. Commun.* 185 (2014) 604–613.
- [69] A. Barducci, M. Bonomi and M. Parrinello, *WIREs Comput. Mol. Sci.* 1 (2011) 826–843.
- [70] L. Sutto, S. Marsili and F.L. Gervasio, *WIREs Comput. Mol. Sci.* 2 (2012) 771–779.
- [71] A. Laio, A. Rodriguez-Fortea, F.L. Gervasio, M. Ceccarelli and M. Parrinello, *J. Phys. Chem. B* 109 (2005) 6714–6721.
- [72] G. Bussi, A. Laio and M. Parrinello, *Phys. Rev. Lett.* 96 (2006) 090601.
- [73] J.F. Dama, M. Parrinello and G.A. Voth, *Phys. Rev. Lett.* 112 (2014) 240602.
- [74] P. Tiwary and M. Parrinello, *J. Phys. Chem. B* 119 (2015) 736–742.
- [75] O. Valssson, P. Tiwary and M. Parrinello, *Annu. Rev. Phys. Chem.* 67 (2016) 159–184.
- [76] A. Barducci, G. Bussi and M. Parrinello, *Phys. Rev. Lett.* 100 (2008) 020603.
- [77] S. Singh, -C.-C. Chiu and J.J. De Pablo, *J. Stat. Phys.* 145 (2011) 932–945.
- [78] J.F. Dama, G. Rotskoff, M. Parrinello and G.A. Voth, *J. Chem. Theory Comput.* 10 (2014) 3626–3633.
- [79] A.D. White, J.F. Dama and G.A. Voth, *J. Chem. Theory Comput.* 11 (2015) 2451–2460. PMID: 26575545
- [80] F. Marinelli and J.D. Faraldo-Gómez, *Biophys. J.* 108 (2015) 2779–2782.
- [81] D. Branduardi, G. Bussi and M. Parrinello, *J. Chem. Theory Comput.* 8 (2012) 2247–2254.
- [82] Y. Crespo, F. Marinelli, F. Pietrucci and A. Laio, *Phys. Rev. E* 81 (2010) 055701.
- [83] F. Baftizadeh, P. Cossio, F. Pietrucci and A. Laio, *Curr. Phys. Chem.* 2 (2012) 79–91.
- [84] M. McGovern and J. De Pablo, *J. Chem. Phys.* 139 (2013) 084102.
- [85] G. Tiana, *Eur. Phys. J. B* 63 (2008) 235–238.
- [86] M. Bonomi, A. Barducci and M. Parrinello, *J. Comput. Chem.* 30 (2009) 1615–1621.
- [87] F. Marinelli, F. Pietrucci, A. Laio and S. Piana, *PLoS Comput. Biol.* 5 (2009) e1000452.
- [88] O. Valssson and M. Parrinello, *Phys. Rev. Lett.* 113 (2014) 090601.
- [89] F. Pietrucci, *Rev. Phys.* 2 (2017) 32–45.
- [90] G. Bussi, F.L. Gervasio, A. Laio and M. Parrinello, *J. Am. Chem. Soc.* 128 (2006) 13435–13441.
- [91] S. Piana and A. Laio, *J. Phys. Chem. B* 111 (2007) 4553–4559.
- [92] C. Camilloni, D. Provasi, G. Tiana and R.A. Broglia, *Proteins* 71 (2008) 1647–1654.
- [93] A. Gil-Ley and G. Bussi, *J. Chem. Theory Comput.* 11 (2015) 1077–1085.
- [94] X. Biarnés, F. Pietrucci, F. Marinelli and A. Laio, *Comput. Phys. Commun.* 183 (2012) 203–211.
- [95] T. Giorgino, A. Laio and A. Rodriguez, *Comput. Phys. Commun.* 217 (2017) 204–209.
- [96] A. Kranjc, S. Bongarzone, G. Rossetti, X. Biarnés, A. Cavalli, M.L. Bolognesi, M. Roberti, G. Legname and P. Carloni, *J. Chem. Theory Comput.* 5 (2009) 2565–2573.
- [97] P. Cossio, A. Trovato, F. Pietrucci, F. Seno, A. Maritan and A. Laio, *PLoS Comput. Biol.* 6 (2010) e1000957.
- [98] F. Baftizadeh, X. Biarnés, F. Pietrucci, F. Affinito and A. Laio, *J. Am. Chem. Soc.* 134 (2012) 3886–3894. doi:10.1021/ja210826a.
- [99] F. Pietrucci, A.V. Vargiu and A. Kranjc, *Sci. Rep.* 5 (2015) 18555. doi:10.1038/srep18555.
- [100] J. Pfandtner and M. Bonomi, *J. Chem. Theory Comput.* 11 (2015) 5062–5067. doi:10.1021/acs.jctc.5b00846.
- [101] P. Raiteri, A. Laio, F.L. Gervasio, C. Micheletti and M. Parrinello, *J. Phys. Chem. B* 110 (2006) 3533–3539. doi:10.1021/jp054359r.

- [102] L. Sutto, M. D'Abramo and F.L. Gervasio, *J. Chem. Theory Comput.* 6 (2010) 3640–3646. doi:10.1021/ct100413b.
- [103] V. Spiwok and B. Králová, *J. Chem. Phys.* 135 (2011) 224504. doi:10.1063/1.3660208.
- [104] M. Ceriotti, G.A. Tribello and M. Parrinello, *Proc. Natl. Acad. Sci. U.S.A.* 108 (2011) 13023–13028.
- [105] G. Daz Leines and B. Ensing, *Phys. Rev. Lett.* 109 (2012) 020601. doi:10.1103/PhysRevLett.109.020601.
- [106] P. Tiwary and B.J. Berne, *Proc. Natl. Acad. Sci. U.S.A.* 113 (2016) 2839–2844. doi:10.1073/pnas.1600917113.
- [107] J. McCarty and M. Parrinello, *J. Chem. Phys.* 147 (2017) 20410–20419. doi:10.1063/1.4998598.
- [108] M. Sultan and V.S. Pande, *J. Chem. Theory Comput.* 13 (2017) 2440–2447. doi:10.1021/acs.jctc.7b00182.
- [109] C.R. Schwantes and V.S. Pande, *J. Chem. Theory Comput.* 9 (2013) 2000–2009. doi:10.1021/ct4005992.
- [110] G. Pérez-Hernández, F. Paul, T. Giorgino, G. De Fabritiis and F. Noé, *J. Chem. Phys.* 139 (2013) 015102. doi:10.1063/1.4811489.
- [111] S. Jo, D. Suh, Z. He, C. Chipot and B. Roux, *J. Phys. Chem. B* 120 (2016) 8733–8742. doi:10.1021/acs.jpcc.6b04525.
- [112] R. Galvelis and Y. Sugita, *J. Chem. Theory Comput.* 13 (2017) 2489–2500. doi:10.1021/acs.jctc.7b00188.
- [113] M.M. Sultan, H.K. Wayment-Steele and V.S. Pande, *J. Chem. Theory Comput.* 14 (2018) 1887–1894. doi:10.1021/acs.jctc.8b00025.
- [114] J.M. Lamim Ribeiro, P. Bravo Collado, Y. Wang and P. Tiwary, *J. Chem. Phys.* 149 (2018) 072301. doi:10.1063/1.5025487.
- [115] C. Wehmeyer and F. Noé, *J. Chem. Phys.* 148 (2018) 241703. doi:10.1063/1.5011399.
- [116] A. Pérez, G. Martínez-Rosell and G. De Fabritiis, *Curr. Opin. Struct. Biol.* 49 (2018) 139–144. doi:10.1016/j.sbi.2018.02.004.
- [117] G. Fiorin, M.L. Klein and J. Henin, *Mol. Phys.* 111 (2013) 3345–3362. doi:10.1080/00268976.2013.813594.
- [118] P. Metzner, C. Schütte and E. Vanden-Eijnden, *J. Chem. Phys.* 125 (2006) 084110. doi:10.1063/1.2335447.
- [119] B. Nölting, *Protein Folding Kinetics: Biophysical Methods*, Springer Science & Business Media, Berlin, 2005.
- [120] N. Ferruz and G. De Fabritiis, *Mol. Inform.* 35 (2016) 216–226. doi:10.1002/minf.201501018.
- [121] M. Bernetti, A. Cavalli and L. Mollica, *Med Chem Comm* 8 (2017) 534–550. doi:10.1039/C6MD00581K.
- [122] R. Du, V.S. Pande, A. Grosberg, T. Tanaka and E.I. Shakhnovich, *J. Chem. Phys.* 108 (1998) 334. doi:10.1063/1.475393.
- [123] E. Weinan and E. Vanden-Eijnden, *Annu. Rev. Phys. Chem.* 61 (2010) 391–420. doi:10.1146/annurev.physchem.040808.090412.
- [124] R.B. Best and G. Hummer, *Proc. Natl. Acad. Sci. U.S.A.* 102 (2005) 6732–6737. doi:10.1073/pnas.0408098102.
- [125] K. Lindorff-Larsen, S. Piana, R.O. Dror and D.E. Shaw, *Science* 334 (2011) 517–520. doi:10.1126/science.1208351.
- [126] S. Piana, K. Lindorff-Larsen and D.E. Shaw, *Proc. Natl. Acad. Sci. U.S.A.* 109 (2012) 17845–17850. doi:10.1073/pnas.1201811109.
- [127] P. Hänggi, P. Talkner and M. Borkovec, *Rev. Mod. Phys.* 62 (1990) 251–341.

- [128] N.G. Van Kampen, *Stochastic Processes in Physics and Chemistry*, Elsevier, Amsterdam, Vol. 1 1992.
- [129] D. Givon, R. Kupferman and A. Stuart, *Nonlinearity* 17 (2004) R55. doi:10.1088/0951-7715/17/6/R01.
- [130] D.T. Gillespie, *Am. J. Phys.* 64 (1996) 1246–1257. doi:10.1119/1.18387.
- [131] R. Zwanzig, *Nonequilibrium Statistical Mechanics*, Oxford University Press, New York, 2001.
- [132] W.T. Coffey, Y.P. Kalmykov and J. Waldron, *The Langevin Equation: With Applications to Stochastic Problems in Physics, Chemistry and Electrical Engineering*, World Scientific, Singapore, 2004.
- [133] J.-H. Prinz, H. Wu, M. Sarich, B. Keller, M. Senne, M. Held, J.D. Chodera, C. Schütte and F. Noé, *J. Chem. Phys.* 134 (2011) 174105. doi:10.1063/1.3565032.
- [134] N.-V. Buchete and G. Hummer, *J. Phys. Chem. B* 112 (2008) 6057–6069.
- [135] V.S. Pande, K. Beauchamp and G.R. Bowman, *Methods* 52 (2010) 99–105. doi:10.1016/j.jymeth.2010.06.002.
- [136] G.R. Bowman, V.S. Pande and F. Noé, *An Introduction to Markov State Models and Their Application to Long Timescale Molecular Simulation*, Springer Science & Business Media, Berlin, Vol. 797 2013.
- [137] J.D. Chodera and F. Noé, *Curr. Opin. Struct. Biol.* 25 (2014) 135–144. doi:10.1016/j.sbi.2014.04.002.
- [138] C.R. Schwantes, R.T. McGibbon and V.S. Pande, *J. Chem. Phys.* 141 (2014) 09B201_1. doi:10.1063/1.4895044.
- [139] R.D. Malmstrom, C.T. Lee, A.T. Van Wart and R.E. Amaro, *J. Chem. Theory Comput.* 10 (2014) 2648–2657. doi:10.1021/ct500394t.
- [140] M. Biswas, B. Lickert and G. Stock, *J. Phys. Chem. B* 122 (2018) 5508–5514.
- [141] J. Shao, S.W. Tanner, N. Thompson and T.E. Cheatham, *J. Chem. Theory Comput.* 3 (2007) 2312–2334. doi:10.1021/ct700119m.
- [142] A. Rodriguez and A. Laio, *Science* 344 (2014) 1492–1496. doi:10.1126/science.1242072.
- [143] F. Sittel and G. Stock, *J. Chem. Theory Comput.* 12 (2016) 2426–2435. doi:10.1021/acs.jctc.5b01233.
- [144] O.M. Becker and M. Karplus, *J. Chem. Phys.* 106 (1997) 1495–1517. doi:10.1063/1.473299.
- [145] P. Cossio, A. Laio and F. Pietrucci, *Phys. Chem. Chem. Phys.* 13 (2011) 10421–10425. doi:10.1039/c1cp21236b.
- [146] R.T. McGibbon and V.S. Pande, *J. Chem. Theory Comput.* 9 (2013) 2900–2906. doi:10.1021/ct4005992.
- [147] G. Hummer and A. Szabo, *J. Phys. Chem. B* 119 (2014) 9029–9037. doi:10.1021/jp508375q.
- [148] F. Noé and C. Clementi, *J. Chem. Theory Comput.* 11 (2015) 5002–5011. doi:10.1021/acs.jctc.5b00553.
- [149] C. Schütte, F. Noé, J. Lu, M. Sarich and E. Vanden-Eijnden, *J. Chem. Phys.* 134 (2011) 05B609. doi:10.1063/1.3590108.
- [150] F. Noé, H. Wu, J.-H. Prinz and N. Plattner, *J. Chem. Phys.* 139 (2013) 184114. doi:10.1063/1.4828816.
- [151] D.T. Gillespie, *Annu. Rev. Phys. Chem.* 58 (2007) 35–55. doi:10.1146/annurev.physchem.58.032806.104637.
- [152] A.F. Voter, *Introduction to the kinetic Monte Carlo method*, in *Radiation Effects in Solids. NATO Science Series*, K.E. Sickafus, E.A. Kotomin and B.P. Uberuaga, eds, Vol. 235 Springer, Berlin, 2007, 1–23.

- [153] M.K. Scherer, B. Trendelkamp-Schroer, F. Paul, G. Pérez-Hernández, M. Hoffmann, N. Plattner, C. Wehmeyer, J.-H. Prinz and F. Noé, *J. Chem. Theory Comput.* 11 (2015) 5525–5542. doi:10.1021/acs.jctc.5b00743.
- [154] S. Doerr, M. Harvey, F. Noé and G. De Fabritiis, *J. Chem. Theory Comput.* 12 (2016) 1845–1852. doi:10.1021/acs.jctc.6b00049.
- [155] M.P. Harrigan, M.M. Sultan, C.X. Hernández, B.E. Husic, P. Eastman, C.R. Schwantes, K.A. Beauchamp, R.T. McGibbon and V.S. Pande, *Biophys. J.* 112 (2017) 10–15. doi:10.1016/j.bpj.2016.10.042.
- [156] V.A. Voelz, G.R. Bowman, K. Beauchamp and V.S. Pande, *J. Am. Chem. Soc.* 132 (2010) 1526–1528. doi:10.1021/ja9090353.
- [157] V.A. Voelz, M. Jäger, S. Yao, Y. Chen, L. Zhu, S.A. Waldauer, G.R. Bowman, M. Friedrichs, O. Bakajin, L.J. Lapidus, S. Weiss and V.S. Pande, *J. Am. Chem. Soc.* 134 (2012) 12565–12577. doi:10.1021/ja302528z.
- [158] N. Stanley, S. Esteban-Martn and G. De Fabritiis, *Nat. Commun.* 5 (2014) 5272. doi:10.1038/ncomms5972.
- [159] G. Pinamonti, J. Zhao, D.E. Condon, F. Paul, F. Noe, D.H. Turner and G. Bussi, *J. Chem. Theory Comput.* 13 (2017) 926–934. doi:10.1021/acs.jctc.6b00982.
- [160] I. Buch, T. Giorgino and G. De Fabritiis, *Proc. Natl. Acad. Sci. U.S.A.* 108 (2011) 10184–10189. doi:10.1073/pnas.1103547108.
- [161] M.J. Harvey and G. De Fabritiis, *Drug Discov. Today* 17 (2012) 1059–1062. doi:10.1016/j.drudis.2012.03.017.
- [162] N. Ferruz, M.J. Harvey, J. Mestres and G. De Fabritiis, *J. Chem. Inf. Model.* 55 (2015) 2200–2205. doi:10.1021/acs.jcim.5b00453.
- [163] N. Plattner and F. Noé, *Nat. Commun* 6 (2015) 7653.
- [164] M. De Vivo and A. Cavalli, *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 7 (2017) e1320. doi:10.1002/wcms.1320.
- [165] D. Bicout and A. Szabo, *J. Chem. Phys.* 109 (1998) 2325–2338. doi:10.1063/1.476800.
- [166] G. Hummer, *New J. Phys.* 7 (2005) 34. doi:10.1088/1367-2630/7/1/034.
- [167] N. Todorova, F. Marinelli, S. Piana and I. Yarovsky, *J. Phys. Chem. B* 113 (2009) 3556–3564. doi:10.1021/jp809776v.
- [168] Z. Ghaemi, D. Alberga, P. Carloni, A. Laio and G. Lattanzi, *J. Chem. Theory Comput* 12 (2016) 4093–4099. doi:10.1021/acs.jctc.5b01126.
- [169] M. Bernetti, M. Masetti, F. Pietrucci, M. Blackledge, M.R. Jensen, M. Recanatini, L. Mollica and A. Cavalli, *J. Phys. Chem. B* 121 (2017) 9572–9582. doi:10.1021/acs.jpcc.7b05568.
- [170] E. Rosta and G. Hummer, *J. Chem. Theory Comput.* 11 (2014) 276–285. doi:10.1021/ct500719p.
- [171] L.S. Stelzl, A. Kells, E. Rosta and G. Hummer, *J. Chem. Theory Comput.* 13 (2017) 6328–6342. doi:10.1021/acs.jctc.7b00373.
- [172] H. Wu, F. Paul, C. Wehmeyer and F. Noé, *Proc. Natl. Acad. Sci. U.S.A.* (2016) E3221–E3230. doi:10.1073/pnas.1525092113.
- [173] L.S. Stelzl and G. Hummer, *J. Chem. Theory Comput.* 13 (2017) 3927–3935. doi:10.1021/acs.jctc.7b00372.
- [174] N.-V. Buchete and G. Hummer, *Phys. Rev. E* 77 (2008) 030902. doi:10.1103/PhysRevE.77.030902.
- [175] H. Grubmüller, *Phys. Rev. E* 52 (1995) 2893. doi:10.1103/PhysRevE.52.2893.
- [176] A.F. Voter, *Phys. Rev. Lett.* 78 (1997) 3908–3911. doi:10.1103/PhysRevLett.78.3908.
- [177] P. Tiwary and M. Parrinello, *Phys. Rev. Lett.* 111 (2013) 230602. doi:10.1103/PhysRevLett.111.230602.

- [178] M. Salvalaglio, P. Tiwary and M. Parrinello, *J. Chem. Theory Comput.* 10 (2014) 1420–1425. doi:10.1021/ct500394t.
- [179] P. Tiwary, V. Limongelli, M. Salvalaglio and M. Parrinello, *Proc. Natl. Acad. Sci. U.S.A.* 112 (2015) E386–E391. doi:10.1073/pnas.1424461112.
- [180] Y. Wang, J.M. Martins and K. Lindorff-Larsen, *Chem. Sci.* 8 (2017) 6466–6473. doi:10.1039/c7sc01787a.
- [181] L. Mollica, S. Decherchi, S.R. Zia, R. Gaspari, A. Cavalli and W. Rocchia, *Sci. Rep.* 5 (2015) 11539. doi:10.1038/srep11539.
- [182] R.J. Allen, P.B. Warren and P.R. Ten Wolde, *Phys. Rev. Lett.* 94 (2005) 018104. doi:10.1103/PhysRevLett.94.107601.
- [183] A.K. Faradjian and R. Elber, *J. Chem. Phys.* 120 (2004) 10880–10889. doi:10.1063/1.1738640.
- [184] T.S. Van Erp and P.G. Bolhuis, *J. Comput. Phys.* 205 (2005) 157–181. doi:10.1016/j.jcp.2004.11.003.
- [185] A. Dickson and A.R. Dinner, *Annu. Rev. Phys. Chem.* 61 (2010) 441–459. doi:10.1146/annurev.physchem.012809.103433.
- [186] P. Bolhuis and C. Dellago, *Eur. Phys. J. Special Topics* 224 (2015) 2409–2427. doi:10.1140/epjst/e2015-02419-6.
- [187] J.-H. Prinz, D.W. Swenson, J. Chodera and P. Bolhuis OpenPathSampling. <http://openpathsampling.org>, 2017
- [188] A. Lervik, E. Riccardi and T.S. Van Erp, *J. Comput. Chem.* 38 (2017) 2439–2451. doi:10.1002/jcc.24900.
- [189] H. Risken, *The Fokker-Planck Equation*, Springer, Berlin, 1996, 63–95.
- [190] K. Schulten and I. Kosztin, *Lectures in Theoretical Biophysics*, University of Illinois, Urbana, 2000.
- [191] M. Tuckerman, *Statistical Mechanics: Theory and Molecular Simulation*, Oxford University Press, New York, 2010.
- [192] E. Darve, J. Solomon and A. Kia, *Proc. Natl. Acad. Sci. U.S.A.* 106 (2009) 10884–10889. doi:10.1073/pnas.0902633106.
- [193] H.S. Lee, S.-H. Ahn and E.F. Darve, MRS Online Proc. Library Archive. 1753 (2015) mrsf14-1753-nn09-03 doi:10.1557/opl.2015.185.
- [194] N. Schaudinnus, B. Bastian, R. Hegger and G. Stock, *Phys. Rev. Lett.* 115 (2015) 050602. doi:10.1103/PhysRevLett.115.050602.
- [195] R.F. Grote and J.T. Hynes, *J. Chem. Phys.* 73 (1980) 2715–2732. doi:10.1063/1.440485.
- [196] A. Nitzan, *Chemical Dynamics in Condensed Phases: Relaxation, Transfer and Reactions in Condensed Molecular Systems*, Oxford university press, New York, 2006.
- [197] N. Schaudinnus, B. Lickert, M. Biswas and G. Stock, *J. Chem. Phys.* 145 (2016) 184114. doi:10.1063/1.4954660.
- [198] B. Peters, *Annu. Rev. Phys. Chem.* 67 (2016) 669–690. doi:10.1146/annurev-physchem-040215-112215.
- [199] G. Hummer, A.E. Garca and S. Garde, *Phys. Rev. Lett.* 85 (2000) 2637. doi:10.1103/PhysRevLett.85.2637.
- [200] C. Micheletti, G. Bussi and A. Laio, *J. Chem. Phys.* 129 (2008) 074105. doi:10.1063/1.3026364.
- [201] R.B. Best and G. Hummer, *Phys. Chem. Chem. Phys.* 13 (2011) 16902–16911. doi:10.1039/c1cp21236b.
- [202] R. Meloni, C. Camilloni and G. Tiana, *Phys. Rev. E* 94 (2016) 052406. doi:10.1103/PhysRevE.94.052406.
- [203] A. Ghysels, R.M. Venable, R.W. Pastor and G. Hummer, *J. Chem. Theory Comput.* 13 (2017) 2962–2976. doi:10.1021/acs.jctc.7b00039.

- [204] R.B. Best, N.-V. Buchete and G. Hummer, *Biophys. J.* 95 (2008) L07–L09. doi:10.1529/biophysj.108.132696.
- [205] O.F. Lange, D. Van Der Spoel and B.L. De Groot, *Biophys. J.* 99 (2010) 647–655. doi:10.1016/j.bpj.2010.04.062.
- [206] K.A. Beauchamp, Y.-S. Lin, R. Das and V.S. Pande, *J. Chem. Theory Comput.* 8 (2012) 1409–1414. doi:10.1021/ct2007814.
- [207] K. Lindorff-Larsen, P. Maragakis, S. Piana, M.P. Eastwood, R.O. Dror and D.E. Shaw, *PLoS One* 7 (2012) e32131. doi:10.1371/journal.pone.0032131.
- [208] J. Huang and A.D. MacKerell, *J. Comput. Chem.* 34 (2013) 2135–2145. doi:10.1002/jcc.23420.
- [209] S. Piana, K. Lindorff-Larsen and D.E. Shaw, *Proc. Natl. Acad. Sci. U.S.A.* 110 (2013) 5915–5920. doi:10.1073/pnas.1218321110.
- [210] S. Piana, J.L. Klepeis and D.E. Shaw, *Curr. Opin. Struct. Biol.* 24 (2014) 98–105. doi:10.1016/j.sbi.2013.12.006.
- [211] J. Henriques, C. Cragnell and M. Skepo, *J. Chem. Theory Comput.* 11 (2015) 3420–3431. doi:10.1021/ct501178z.
- [212] S. Rauscher, V. Gapsys, M.J. Gajda, M. Zweckstetter, B.L. De Groot and H. Grubmüller, *J. Chem. Theory Comput.* 11 (2015) 5513–5524. doi:10.1021/acs.jctc.5b00736.
- [213] F. Palazzesi, M.K. Prakash, M. Bonomi and A. Barducci, *J. Chem. Theory Comput.* 11 (2015) 2–7. doi:10.1021/ct500718s.
- [214] R.B. Best, W. Zheng and J. Mittal, *J. Chem. Theory Comput.* 10 (2014) 5113–5124. doi:10.1021/ct500394t.
- [215] S. Piana, A.G. Donchev, P. Robustelli and D.E. Shaw, *J. Phys. Chem. B* 119 (2015) 5113–5123. doi:10.1021/acs.jpcc.5b00144.
- [216] J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B.L. De Groot, H. Grubmüller and A. MacKerell, *Nat. Methods* 14 (2017) 71–73. doi:10.1038/nmeth.4067.
- [217] J. Sponer, G. Bussi, M. Krepl, P. Banas, S. Bottaro, R.A. Cunha, A. Gil-Ley, G. Pinamonti, S. Poblete, P. Jurecka, N.G. Walter and M. Otyepka, *Chem. Rev.* (2018) doi:10.1021/acs.chemrev.7b00427. in press.
- [218] D. Tan, S. Piana, R.M. Dirks and D.E. Shaw, *Proc. Natl. Acad. Sci. U.S.A.* 115 (2018) E1346–E1355. doi:10.1073/pnas.1713027115.
- [219] Z.A. Levine and J.-E. Shea, *Curr. Opin. Struct. Biol.* 43 (2017) 95–103. doi:10.1016/j.sbi.2016.11.006.
- [220] L. Wang et al., *J. Am. Chem. Soc.* 137 (2015) 2695–2703. doi:10.1021/ja511978y.
- [221] Z. Gaieb, S. Liu, S. Gathiaka, M. Chiu, H. Yang, C. Shao, V.A. Feher, W.P. Walters, B. Kuhn, M.G. Rudolph, S.K. Burley, M.K. Gilson and R.E. Amaro, *J. Comput. Aid. Mol. Des.* 32 (2018) 1–20. doi:10.1007/s10822-017-0088-4.
- [222] X. Huang, G.R. Bowman and V.S. Pande, *J. Chem. Phys.* 128 (2008) 205106. doi:10.1063/1.2908251.
- [223] H. Huang, E. Ozkirimli and C.B. Post, *J. Chem. Theory Comput.* 5 (2009) 1304–1314. doi:10.1021/ct900333c.
- [224] F. Pietrucci, L. Mollica and M. Blackledge, *J. Phys. Chem. Letters* 4 (2013) 1943–1948. doi:10.1021/jz4007806.
- [225] M.A. Cuendet and M.E. Tuckerman, *J. Chem. Theory Comput.* 10 (2014) 2975–2986. doi:10.1021/ct500394t.
- [226] D. Bochicchio, E. Panizon, R. Ferrando, L. Monticelli and G. Rossi, *J. Chem. Phys.* 143 (2015) 144108. doi:10.1063/1.4932159.
- [227] A.C. Pan, T.M. Weinreich, S. Piana and D.E. Shaw, *J. Chem. Theory Comput.* 12 (2016) 1360–1367. doi:10.1021/acs.jctc.5b00913.

- [228] Q. Wei, W. Zhao, Y. Yang, B. Cui, Z. Xu and X. Yang, *Chem. Phys. Chem* 19 (2018) 690–702. doi:10.1002/cphc.201701241.
- [229] H. Sidky et al., *J. Chem. Phys.* 148 (2018) 044104. doi:10.1063/1.5008853.
- [230] T. Vreven, I.H. Moal, A. Vangone, B.G. Pierce, P.L. Kastiritis, M. Torchala, R. Chaleil, B. Jiménez-Garca, P.A. Bates, J. Fernandez-Recio, A.J.J. Bonvin and Z. Wang, *J. Mol. Biol.* 427 (2015) 3031–3041. doi:10.1016/j.jmb.2015.07.016.
- [231] J. Yin, N.M. Henriksen, D.R. Slochower, M.R. Shirts, M.W. Chiu, D.L. Mobley and M.K. Gilson, *J. Comput. Aid. Mol. Des.* 31 (2017) 1–19.
- [232] W. Boomsma, K. Lindorff-Larsen and J. Ferkinghoff-Borg, *PLoS Comput. Biol.* 10 (2014) e1003406.
- [233] E. Ravera, L. Sgheri, G. Parigi and C. Luchinat, *Phys. Chem. Chem. Phys.* 18 (2016) 5686–5701.
- [234] M. Bonomi, G.T. Heller, C. Camilloni and M. Vendruscolo, *Curr. Opin. Struct. Biol.* 42 (2017) 106–116.
- [235] A. Cesari, S. Reißer and G. Bussi, *Comput.* 6 (2018) 15.
- [236] E.T. Jaynes, *Phys. Rev.* 106 (1957) 620.
- [237] J.W. Pitera and J.D. Chodera, *J. Chem. Theory Comput.* 8 (2012) 3445–3451.
- [238] A.D. White and G.A. Voth, *J. Chem. Theory Comput.* 10 (2014) 3023–3030.
- [239] A. Cesari, A. Gil-Ley and G. Bussi, *J. Chem. Theory Comput.* 12 (2016) 6192–6200.
- [240] K. Lindorff-Larsen, R.B. Best, M.A. Depristo, C.M. Dobson and M. Vendruscolo, *Nature* 433 (2005) 128–132.
- [241] B. Roux and J. Weare, *J. Chem. Phys.* 138 (2013) 084107.
- [242] A. Cavalli, C. Camilloni and M. Vendruscolo, *J. Chem. Phys.* 138 (2013) 094112.
- [243] G. Hummer and J. Köfinger, *J. Chem. Phys.* 143 (2015) 243150.
- [244] M. Bonomi, C. Camilloni, A. Cavalli and M. Vendruscolo, *Sci. Adv.* 2 (2016) e1501177.
- [245] K.A. Beauchamp, V.S. Pande and R. Das, *Biophys. J.* 106 (2014) 1381–1390.
- [246] M. Tiberti, E. Papaleo, T. Bengtsen, W. Boomsma and K. Lindorff-Larsen, *PLoS Comput. Biol.* 11 (2015) e1004415–16.
- [247] T. Löhner, A. Jussupow and C. Camilloni, *J. Chem. Phys.* 146 (2017) 165102–165111.
- [248] C. Camilloni, A. Cavalli and M. Vendruscolo, *J. Chem. Theory Comput.* 9 (2013) 5610–5617.
- [249] M. Bonomi, C. Camilloni and M. Vendruscolo, *Sci. Rep.* 6 (2016) 31232.
- [250] M. Bonomi and C. Camilloni, *Bioinformatics* 33 (2017) 3999–4000.
- [251] A. De Simone, M. Vendruscolo, B. Richter and X. Salvatella, *J. Am. Chem. Soc.* 131 (2009) 3810–3811.
- [252] C. Camilloni, P. Robustelli, A. De Simone, A. Cavalli and M. Vendruscolo, *J. Am. Chem. Soc.* 134 (2012) 3968–3971.
- [253] C. Camilloni, D. Bonetti, A. Morrone, R. Giri, C.M. Dobson, M. Brunori, S. Gianni and M. Vendruscolo, *Sci. Rep.* 6 (2016) 28285. doi:10.1038/srep28285.
- [254] C. Camilloni and M. Vendruscolo, *J. Am. Chem. Soc.* 136 (2014) 8982–8991.
- [255] L.D. Cabrita, A.M.E. Cassaignau, H.M.M. Launay, C.A. Waudby, T. Wlodarski, C. Camilloni, M.-E. Karyadi, A.L. Robertson, X. Wang, A.S. Wentink, L.S. Goodsell, C. A. Woolhead, M. Vendruscolo, C.M. Dobson and J. Christodoulou, *Nat. Struct. Mol. Biol.* 23 (2016) 278–285.
- [256] P. Kukic, Y. Pustovalova, C. Camilloni, S. Gianni, D.M. Korzhnev and M. Vendruscolo, *J. Am. Chem. Soc.* 139 (2017) 6899–6910.
- [257] G. Hultqvist, E. Åberg, C. Camilloni, G.N. Sundell, E. Andersson, J. Dogan, C.N. Chi, M. Vendruscolo and P. Jemth, *eLife* 6 (2017) 79.

- [258] C. Camilloni, A.B. Sahakyan, M.J. Holliday, N.G. Isern, F. Zhang, E.Z. Eisenmesser and M. Vendruscolo, *Proc. Natl. Acad. Sci. U.S.A.* 111 (2014) 10203–10208.
- [259] K.D. Brewer et al., *Nat. Struct. Mol. Biol.* 22 (2015) 555–564.
- [260] A. De Simone, M. Gustavsson, R. Montalvao, L. Shi, G. Veglia and M. Vendruscolo, *Biochemistry* 52 (2013) 6684–6694.
- [261] M. Sanz-Hernández, V.V. Vostrikov, G. Veglia and A. De Simone, *Sci. Rep.* 6 (2016) 1–9.
- [262] A.N. Borkar, M.F. Bardaro, C. Camilloni, F.A. Aprile, G. Varani and M. Vendruscolo, *Proc. Natl. Acad. Sci. U.S.A.* 113 (2016) 7171–7176.
- [263] S. Olsson, H. Wu, F. Paul, C. Clementi and F. Noé, *Proc. Natl. Acad. Sci. U.S.A.* 114 (2017) 8265–8270.
- [264] E.T. Jaynes, *Ann. Rev. Phys. Chem.* 31 (1980) 579–601.
- [265] S. Pressé, K. Ghosh, J. Lee and K.A. Dill, *Rev. Mod. Phys.* 85 (2013) 1115.
- [266] P.D. Dixit, J. Wagoner, C. Weistuch, S. Pressé, K. Ghosh and K.A. Dill, *J. Chem. Phys.* 148 (2018) 010901.
- [267] H. Wan, G. Zhou and V.A. Voelz, *J. Chem. Theory Comput.* 12 (2016) 5768–5776.
- [268] G. Zhou, G.A. Pantelopulos, S. Mukherjee and V.A. Voelz, *Biophys. J.* 113 (2017) 785–793.
- [269] P.D. Dixit and K.A. Dill, *J. Chem. Theory Comput.* 14 (2018) 1111–1119.
- [270] R. Capelli, G. Tiana and C. Camilloni, *J. Chem. Phys.* 148 (2018) 184114. doi:10.1063/1.5030339.
- [271] J.C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé and K. Schulten, *J. Comput. Chem.* 26 (2005) 1781–1802.
- [272] K.J. Bowers, E. Chow, H. Xu, R.O. Dror, M.P. Eastwood, B.A. Gregersen, J.L. Klepeis, I. Kolossvary, M.A. Moraes, F.D. Sacerdoti, J.K. Salmon, Y. Shan and D.E. Shaw Scalable algorithms for molecular dynamics simulations on commodity clusters. Proceedings of the 2006 ACM/IEEE Conference on Supercomputing, SC'06. New York, USA, 2006; p 84
- [273] M.J. Harvey, G. Giupponi and G. De Fabritiis, *J. Chem. Theory Comput.* 5 (2009) 1632–1639.
- [274] R. Salomon-Ferrer, D.A. Case and R.C. Walker, *WIREs Comput Mol Sci* 3 (2012) 198–210.
- [275] M.J. Abraham, T. Murtola, R. Schulz, S. Páll, J.C. Smith, B. Hess and E. Lindahl, *SoftwareX* 1-2 (2015) 19–25.
- [276] P. Eastman, J. Swails, J.D. Chodera, R.T. McGibbon, Y. Zhao, K.A. Beauchamp, L.-P. Wang, A.C. Simmonett, M.P. Harrigan, C.D. Stern, R.P. Wiewiora, B.R. Brooks and V.S. Pande, *PLoS Comput. Biol.* 13 (2017) e1005659–17.