

Molecular simulations of biomolecules: some useful guidelines

Irene Marzuoli and Franca Fraternali

Abstract In studying biological processes and focusing on the molecular mechanisms at the basis of these, Molecular Dynamics simulations (MD) have demonstrated to be a very useful tool for the past 50 years. This suite of computational methods calculates the time dependent evolution of a molecular system using physics-based first principles. In this chapter, we give a brief introduction to the theory and practical use of Molecular Dynamics simulations, highlighting the different models and algorithms that have been developed to tackle specific problems, with a special focus on classical force fields. Some examples of how simulations have been used in the past will help the reader in discerning their power, limitations and significance.

1 Introduction

Molecular Dynamics simulations have been rightly defined as the ‘Computational Microscope’ [1, 2] as they offer otherwise inaccessible insights into the molecular

Irene Marzuoli

Randall Center for Cell and Molecular Biophysics, King’s College London, UK e-mail:
irene.marzuoli@kcl.ac.uk

Franca Fraternali

Randall Center for Cell and Molecular Biophysics, King’s College London, UK e-mail:
franca.fraternali@kcl.ac.uk

details underlying conformational changes of proteins and nucleic acids. Computational methods and tools based on MD are routinely applied in structural biology to quantitatively characterise the dynamics and thermodynamics of proteins and their complexes. The increasing computational power available, and the flexibility of the algorithms that can be implemented on different platforms, have made possible to access molecular dynamical properties inaccessible to experiments. These techniques and the associated force fields are also commonly used in the process of structure determination from NMR data, and theoretical structure prediction [3, 4]. Modelling and simulating a biological system consists in describing its components and their mutual interactions, and implementing the laws of physics to reproduce its natural evolution. A quantum mechanic description would be more accurate but expensive to achieve for large systems. To facilitate the task, several simplified models have been devised, each most suitable to investigate particular cases. In particular, a large class of models focuses on a classical mechanics description of the dynamics: for increasing sizes of the systems and longer time spans described, the classical approximations will be more than accurate, and also the only possible representation which is computationally affordable.

Understanding the methodology of classical Molecular Dynamics (MD) provides the interpretative key with which simulations must be designed, run and interpreted in each specific case. We offer here some general background and directions on which are the used algorithms and their applicability in some selected scenarios that we consider representative and challenging.

2 Materials

2.1 Extracting the Dynamics of a Molecular System: the Time Evolution Algorithm

In a classical MD framework, Newton's second law of motion rules the dynamics, stating that the acceleration \mathbf{a} that a particle is subject to at time t , depends on the total force \mathbf{F} acting on the particle itself and on its mass m (bold denotes vectorial quantities):

$$\mathbf{F}(t) = m \cdot \mathbf{a}(t) . \quad (1)$$

As the acceleration $\mathbf{a}(t)$ is the second derivative of the position $\mathbf{r}(t)$ with respect to time, given the initial position and velocity of the particle ($\mathbf{r}(t_0)$, $\mathbf{v}(t_0)$), their temporal evolution can be computed integrating $\mathbf{a}(t) = \mathbf{F}(t)/m$ as follow:

$$\mathbf{v}(t) = \mathbf{v}(t_0) + \int_{t_0}^t \frac{\mathbf{F}(t')}{m} dt' ; \quad (2)$$

$$\mathbf{r}(t) = \mathbf{r}(t_0) + \int_{t_0}^t \mathbf{v}(t') dt' + \int_{t_0}^t \int_{t_0}^{t'} \frac{\mathbf{F}(t')}{m} dt'' dt' . \quad (3)$$

In the case of complex biomolecular systems with many particles and multiple interactions acting between them, it is impossible to integrate analytically Equation 2, while a different and feasible approach consists in discretising Equation 1. The idea is to consider very short time steps of length Δt , so that in such intervals the forces are (almost) constant, and the integration of Equation 2 becomes trivial. A careful choice of the values to integrate allows to reduce the approximations derived from such approach. For example, choosing the velocity value at time $t_0 + \Delta t/2$ (and not at t_0) decreases the error down to orders of $(\Delta t)^4$ (rather than $(\Delta t)^2$). This framework is at the basis of the so-called leap-frog algorithm, which is used in the vast majority of MD engines:

$$\mathbf{v}\left(t_0 + \frac{\Delta t}{2}\right) = \mathbf{v}\left(t - \frac{\Delta t}{2}\right) + \frac{\mathbf{F}(t)}{m} \Delta t ; \quad (4)$$

$$\mathbf{r}(t_0 + \Delta t) = \mathbf{r}(t_0) + \mathbf{v}\left(t_0 + \frac{\Delta t}{2}\right) \Delta t . \quad (5)$$

This algorithm can thus “solve” every possible Newton equation, at the expenses of some precision.

2.2 Thermostats and barostats: rescuing the approximation limit

As the integration procedure is not exact, specific algorithms have been developed to compensate for the necessary approximations and to more realistically reproduce the simulation’s conditions of choice, for example Temperature and Pressure.

Most experiments are conducted under constant temperature, so that it is desirable to reproduce this condition in simulations. To set up a temperature, at the beginning of the simulations particles can be given random initial velocities distributed according to the Maxwell-Boltzmann curve, which describe velocities of atoms of a noble gas at temperature T . The velocity of each of them will be influenced by the specific interactions occurring in the system but, in a constant temperature environment, the total average kinetic energy (proportional to T) must remain constant. Even in absence of any dissipative term in the dynamics, the approximations performed in the MD algorithm make this quantity to drift away from its initial value, therefore to ensure the temperature is maintained throughout the simulation, thermostat algorithms have been devised.

The principle behind a thermostat consists in rescaling the velocity of one or few selected particles at fixed interval of times, to restore the correct average kinetic energy. This fixed interval of time must not correspond to the timestep itself, as the goal of a thermostat is to maintain the average temperature, and not its value at all times, as fluctuations are allowed in natural systems. Moreover, it is strongly recommended to couple solute and solvent to separate baths, to ensure that both maintain the correct temperature. Indeed, it is possible that the energy exchange between solute and water (or other components) is not perfect, due to different conditions adopted for their simulations, like cut-offs or restraints. The most used

thermostat algorithms are the Berendsen [5], Nosé-Hoover [6, 7], Andersen [8] and velocity rescale [9], which differ in the quantities selected for such rescaling.

Another macroscopic condition one wish to maintain is either the volume or the pressure of the system. While maintaining the volume constant is easy (and, combined with constant temperature, gives the NVT ensemble), pressure regulation (i.e. maintaining a NPT ensemble) requires a barostat. Pressure is directly proportional to the average quantity of motion exchanged between the particles and the walls of the box they are confined to, which depends on the frequency of collision and thus on the extent of the box. Barostats change the box size to regulate the pressure. It has to be noticed that most MD simulations are run under periodic boundary conditions, i.e. a particle which exits from the simulation box during a move is brought back on the opposite side of the box, leaving the box density constant. This mimics the presence of an infinite number of equivalent boxes one next to the other, and alleviates the finite size effects that arise when simulating small systems. In this scenario, particles are not bouncing on the box walls, rather a virtual pressure is computed from the velocities of the ones trespassing the box boundaries during a move. Similarly as for thermostats, the coupling frequency of barostats must be larger than the time step. Usually all the box dimensions are rescaled by the same amount. However, in the case of anisotropic systems like lipids, to maintain the correct pressure, the directions parallel to the membrane plane can be rescaled separately with respect to the one perpendicular to it. Also for pressure coupling several algorithms can be used: the Berendsen [5], Parrinello-Rahman [10], or Martyna-Tuckerman-Tobias-Klein (MTTK) [11] (to be used in conjunction with the Nosé-Hoover thermostat), which differ in the way they approach the desired pressure (e.g. exponentially, in an oscillatory way, etc.).

2.3 Force fields

2.3.1 Force field definition

Once the equations of motion and the control algorithms are set up, the next challenge is represented by modelling the forces and thus the potential energy function. Force fields for classical MD simulations usually rely on the breakdown of interactions into several, independent terms, identified on an empirical physical basis. We report here the functional form adopted for the GROMOS force field [12, 13] as implemented in the GROMACS MD engine [14–16], as an example of the structure of a classical force field.

Covalent (bonded) interactions Covalent interactions are modelled with potential energy terms representing bond stretching, angle bending, improper and proper dihedral angles torsion. The functional form of the potential energy function for each of them aims at a simplified, semi-classical description of the sub atomic motion of molecules. Often, it is modelled as a harmonic-like vibration around the equilibrium position, regulated by a constant with suitable energy units.

Type	Eq. pos.	Const.	[Const.]	Functional form
Bond	b_{ij}	k_{ij}^b	$\frac{\text{kJ}}{\text{mol m}^2}$	$V_b(\mathbf{r}_{ij}) = \frac{1}{4} k_{ij}^b (\mathbf{r}_{ij} ^2 - b_{ij}^2)^2$
Angle	θ_{ijk}^0	k_{ijk}^θ	$\frac{\text{kJ}}{\text{mol}}$	$V_a(\theta_{ijk}) = \frac{1}{2} k_{ijk}^\theta (\cos(\theta_{ijk}) - \cos(\theta_{ijk}^0))^2$
Dihedral	ϕ_{ijkl}^0	k_{ijkl}^ϕ	$\frac{\text{kJ}}{\text{mol rad}^2}$	$V_d(\phi_{ijkl}) = k_{ijkl}^\phi (1 + \cos(n \phi_{ijkl} - \phi_{ijkl}^0))$
Improper	ξ_{ijkl}^0	k_{ijkl}^ξ	$\frac{\text{kJ}}{\text{mol}}$	$V_{id}(\xi_{ijkl}) = \frac{1}{2} k_{ijkl}^\xi (\xi_{ijkl} - \xi_{ijkl}^0)^2$

(6)

In the GROMOS force field, this translates in the equations displayed in Table 6, where for proper dihedrals, the convention states that ϕ_{ijkl} is the angle between the (i, j, k) and (j, k, l) planes; with i, j, k , and l four subsequent atoms, for example along a protein backbone. A value of zero for ϕ_{ijkl} corresponds to a *cis* configuration and π to a *trans*. The integer n denotes the number of equally spaced energy minima available in a 360° turn. The same conventions hold for improper dihedrals ξ_{ijkl} , which are used to ensure ring planarity and control the chirality of tetrahedric centres.

It must be noticed that these types of potentials can not model the rupture of a bond: for this, more sophisticated descriptions are needed.

Non bonded interactions Non bonded interactions include the short range Pauli repulsion, the “mid”-range van der Waals attraction, and the long range electrostatic term.

The first two can be modelled together by a Lennard-Jones potential. Its functional form, describing the interaction between two neutral atoms at distance r , models the long range dispersion with a r^6 behaviour typical of the dipole-dipole interactions found in noble gases (London dispersion forces), while the Pauli term is represented by a r^{12} behaviour to ease the computation in relation with the previous one:

$$V_{LJ}(r) = 4\epsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right]. \quad (7)$$

Two parameters, ϵ and σ , tune the interaction strength and the equilibrium distance. They are fitted against experimental data and are specific of each pair of atoms species.

The Coulomb energy between two charges q_1 and q_2 at distance r is represented by the Coulomb law:

$$V_C(r) = \frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{\epsilon_r r_{ij}} \quad (8)$$

with ϵ_0 the dielectric constant of vacuum and ϵ_r the relative dielectric constant, introduced to properly take into account the screening provided by the material surrounding the object, as polarisability is not included in this description.

The treatment of non-bonded interactions requires particular care because of their long range nature: in every point of the simulation box many forces from distant atoms are acting at the same time, making the prediction of the outcome difficult. The van der Waals forces decay fast, therefore the tail of their functional can be cut after a threshold distance with little impact on the outcome; while Coulomb interactions, with their slower decay, must be taken into account throughout the whole simulation box. Many algorithms have been devised to efficiently compute them, like the Particle Mesh Ewald [17] or the Reaction Field [18] approaches.

Finally, all biomolecular force fields, and in particular their van der Waals interactions, are parametrised to describe systems at room temperature, therefore simulations performed at substantially different temperatures must be interpreted carefully.

2.3.2 Force fields: classifications

Many force fields for classical MD simulations adopt a functional form equal or similar to the one described above. Their difference lies in the number of degrees of freedom modelled, in a hierarchy of descriptions proceeding from detailed to coarse (Figure 1). Three possible classes of descriptions are:

- all-atoms force fields, where all the atoms are presented in the description, and represented as spheres of variable size according to their van der Waals radius (e.g. proportional to σ in a Lennard-Jones model). Examples of all-atoms force fields are AMBER [23–25], CHARMM [26–28] and OPLS all-atom [29].
- united atoms force fields, similar to the previous ones but where non-polar hydrogens are incorporated in the heavy atom they are bonded to. The “united atom” is given a new σ parameter and increased mass according to how many hydrogen it includes. The GROMOS force field [12, 13] follows this philosophy, and OPLS has also a united atom version [30].
- coarse grain force fields, which group together in one unique bead few atoms, to reduce the number of variables to compute. The clustered atoms are such that their mutual distances are expected to vary little with respect to the ample movements of components of the system far away from each other (which will be grouped in different beads). The MARTINI [31–33] and SIRAH [34, 35] force fields belong to this category.

We now give a more detailed insight in the characteristics and parametrisation strategies of an atomistic and a coarse grain force field among the ones mentioned.

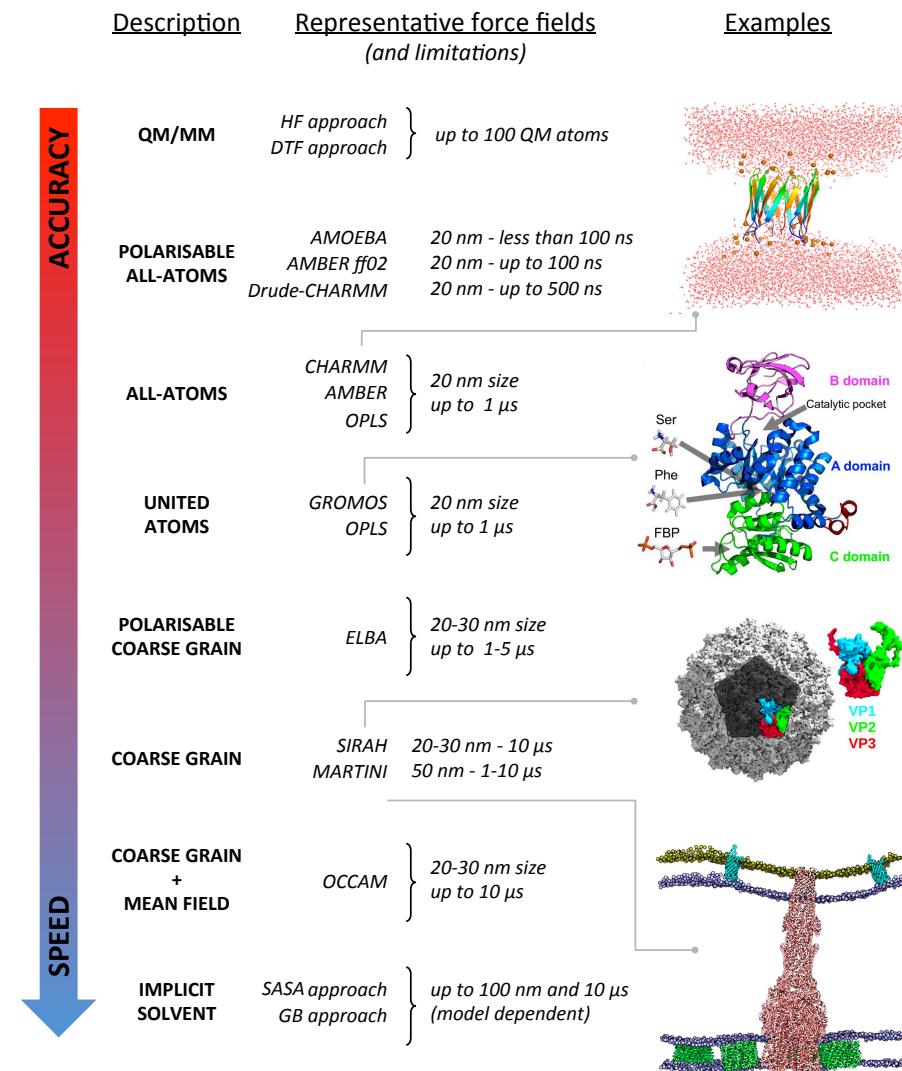


Fig. 1 List of most popular simulation force field for biomolecules, ordered from detailed to coarse (reference to the relative papers in Section 2.3.2). On the left, snapshot of notable systems simulated with the force fields CHARMM (adapted with permission from [19]). Copyright (2017) American Chemical Society; GROMOS (adapted with permission from [20]); SIRAH (adapted with permission from [21]). Copyright (2017) American Chemical Society and MARTINI (adapted with permission from [22]). Copyright (2017) Elsevier).

2.3.3 The GROMOS force field

All-atoms and united atoms force field are parametrised against experimental values. While for the all-atom force fields AMBER and CHARMM the parametrisation is based on ab initio quantum mechanics calculations refined against experimental data [23–27], the united atom GROMOS force field relies on the reproduction of free enthalpies of solvation and heat of vaporization of small molecules at physiological temperatures and pressures [13, 36, 37]. This procedure sets not only the constant of the bonded interactions, but also the partial charges of the atoms inside a molecule: as no electrons are included for the sake of efficiency, their redistribution across atoms which are bonded is modelled through fractional charges assigned to each atom (while the total charge of a molecule must sum to an integer). Moreover, it is assumed that the parametrisation performed for small moieties can be transferred to a larger compound including these moieties. This limits the number of chemical groups to be described in order to simulate biomolecules.

In every MD simulation, the description of water is crucial. Out of the many water model proposed, the GROMOS parametrisation has been performed with a flexible simple point charge (SPC [38]) model. This description represents water as a three atoms molecule, with a negative charge on the oxygen and a positive complementary one on the two hydrogen atoms, and allows flexible hydrogen-oxygen bonds. This model reproduces correctly the density and dielectric permittivity of water.

The improvement of computational techniques and reparametrisation strategies prompts the periodical release of newer versions of force fields. Accordingly, the latest version of the GROMOS force field, version 54a8 [13], has been released in 2011.

2.3.4 The MARTINI force field

The MARTINI force field is a popular coarse grain description of biological molecules [31–33]: developed originally with a focus on lipids, it has been then extended to include proteins, small ligands and DNA/RNA molecules.

MARTINI opts for a four-to-one approach, i.e. four heavy atoms are grouped in one bead. The number of bead types has been kept to the minimum necessary to represent biological molecules. They are organised systematically in polar, non-polar, apolar, or charged, and each type has a number of subtypes with increasing polarity to differentiate the chemical nature of the underlying atomistic structures. This systematic approach can be easily transferred to new compounds, without the need of introducing new bead types. The only exception is represented by rings molecules, where a two-to-one approach is needed to maintain the circular topology.

Two approaches are possible to develop a coarse grain description: parameters can be fit directly to global quantities derived from experiments, in a top-down approach as the one followed in GROMOS; or fitting the coarse-grain simulations results to the ones from atomistic simulations, in a bottom-up approach. The MARTINI force field chooses a top-down approach to parametrise non-bonded interactions, tuning them against experimental partitioning free energies between polar and apolar phases, while bonded interactions are derived from reference all-atom information, in a bottom-up approach.

The four-to-one mapping implies that the amino acid backbone is represented by one bead only, preventing the description of directional bonds which are key to reproduce the secondary structure. The bonded parameters partially account for this, favouring for each residue type the backbone conformation in which it is most likely found (as computed from the Protein Data Bank - PDB [39]). When this is not sufficient, the protein can be constrained around a given structure through an elastic network model approach (ElNeDyn [40]). However, both the backbone parametrisation and the use of ElNeDyn imply that conformational changes in the structure are penalised and therefore not well sampled in MARTINI simulations.

The MARTINI force field provides two water models. The standard one groups four water molecules in one bead only, loosing the polarisability typical of water, the effect of which is partially restored with the use of a high dielectric constant. The polarisable water model [41] maps instead four water molecule to a single “inflated” water, i.e. a three-beads molecule with the same shape of a single molecule, but expanded, and a charge splitting which can account for the water dipole.

Backmapping techniques Coarse-grain descriptions are very effective in reproducing long time scales; however, to retrieve finer details after such extensive exploration, backmapping techniques have been designed to obtain atomistic configurations from the coarse-grain ones [42]. The easy transfer between the two resolution, gave rise to many multiscale studies applied to biomolecular systems [43].

2.4 Beyond a classical atomistic framework

Without entering into the details, we want to bring to the reader attention two possible refinements of the aforementioned models, and two computational strategies which on the contrary speed up the calculations at the expenses of the loss of some details.

Regarding the accuracy of simulations, it must be noticed that none of the force field mentioned above takes into account polarisability, i.e. the displacement of electrons with respect to the nucleus, as a consequence of the surrounding electrostatic environment, because electrons and nucleus of an atom are modelled as a unique object. Specific force fields have been modelled to include this effect, on top of atomistic descriptions, as in the AMOEBA [44, 45], Drude polarisable CHARMM [46] or AMBERff02 [47], or on coarse grain descriptions, as in the ELBA force field [48]. Polarisability does improve the accuracy of simulations, but it can significantly slow down simulations. Going beyond the classical approximation, for biological processes governed by quantum mechanics - such as photosynthesis, DNA mutation processes or some enzymatic activities - many semi classical hybrid techniques have been developed [49]. They combine computational quantum mechanical modelling

methods, such as Density Functional Theory (DFT) or Hartree-Fock computations (HF) [50], with classical Molecular Dynamics to gain the accuracy of a quantum description in the region of interest and the speed up of a classical one in the surrounding areas.

Tackling instead efficiency issues, an implicit solvent model can be used to speed up simulations. The solvent is represented as a continuous medium, as opposed to explicit models which include all its particles [51]. Models of implicit solvent can be based on different assumptions: for example the solute-solvent interactions can be taken as proportional to the solvent-accessible surface area (SASA) of every particle of solute [52–55], or instead can be derived from a solution of the Poisson-Boltzmann equation governing the charge density in a material, for example in the form of the Generalised Born equation [56] which is valid under particularly simple conditions. Another speed up technique is constituted by hybrid particle-field algorithms. The idea is to treat non-bonded interactions through a mean field approach, where atoms/beads move in the field generated by the others. The field does not need to be updated at every time step, as it is a collective and thus slowly evolving variable; moreover, for each particle only the interaction with the field, and not with all the neighbour particles needs to be computed, reducing the computations effort further. This approach has been employed with a coarse grain description of polymers and biological molecules in the OCCAM software [57].

Finally, further strategies are possible to enhance the sampling performed by a simulation in the case that the one obtained by the natural evolution of the system would exceed the computational time available. As a non comprehensive list of these techniques, we mention replica-exchange algorithms [58] which combine together multiple simulations held at different conditions, local potential-energy elevation (or metadynamics) [59, 60] which avoids the re-sampling of already visited conformations adding an energy penalty to them, umbrella sampling [61] which reconstructs free-energy barriers from simulations held at specific values of the coordinate along which the barrier exists, or finally the simple use of higher temperature to overcome energy barriers [62].

2.5 System setup

To set up a simulation, the desired resolution, i.e. force field, must be chosen according to the system to simulate. A few examples on this will be given in Section 3; however, for any choice of parameters, some common “ingredients” and steps are necessary to prepare the system:

- **structural file:** a pdb file (or any structural file format suitable with the MD engine of choice) which contains only the elements to be simulated, with a correct format of names and positions. The choice of the initial conformation is particularly important, especially for atomistic resolution which can access limited time scales and thus is likely to sample conformations in the vicinity of the initial position.
- **topology file:** given a structural file, every MD engine has dedicated commands to retrieve the list of parameters for bonded atoms. The presence of exotic residues or a non standard network of bonds might need case-to-case manual parametrisation.
- **simulation box:** for simulations of a protein, it is good practice to place it within a box large enough to avoid interference of the protein with its periodic images, for example allowing a minimum distance between protein and box of at least the cut-off chosen for non-bonded interactions. However, it must be considered that the protein might adopt extended conformations during the simulation, therefore one should find a compromise between a sufficient box size and the cost of increasing the effective number of particles to calculate. In the case of lipids instead, to reproduce an infinite membrane and account for lateral tension, the box can be chosen exactly as big as the membrane patch (in the directions parallel to its plane), so that periodic images merge together.
- **solvation:** the prepared box is filled with molecules of the water model of choice, except in the spaces already occupied by the solute.
- **ions:** ions are added in replacement of a suitable number of water molecules to neutralise the charge of the system. Additional ones can be added to reach the

experimental ion concentration (though there will be a slight imbalance between the positive and negative species due to the counter ions already added).

- **energy minimisation:** the system prepared according to the steps above might have several atom clashes. This can derive from the initial protein/membrane structure, or from imperfect packing of the solvent around the solute. To alleviate this, an energy minimisation is performed prior to the dynamical run. Usually, a protocol restraining the the solute and letting relax solvent positions in the first instance, followed by a second run that releases the solute is used in these cases.
- **equilibration:** even after energy minimisation, the system prepared is far away from the ideal equilibrium configuration. For this reason, several rounds of simulations with specific conditions are run before the final production. The specific protocol depends on the system and the force field, however for proteins in solution it is quite common to start with NVT ensemble runs with position restraints at increasing temperatures, followed by NPT runs at the same temperatures but without restraints. For lipid system, the simulated ensemble is preferably NPT, as a NVT ensemble with fixed box size would allow unphysical penetration of water in the bilayer if it shrinks. The equilibration then consists in a temperature increase protocol only. In some cases, for example for the MARTINI force field with standard water, it is not possible to simulate low temperatures, thus the equilibration consists in a short production employing different thermostat and barostat with respect to the production. Some of these protocols are more suitable for approaching efficiently the correct temperature or pressure value from a configuration far from equilibrium, while others for maintaining the correct ensemble properties. Generally, every force field is provided with validated equilibration procedures.

3 Methods: relevant examples

One of the challenges of MD simulations consists in choosing the most suitable granularity of the description, together with the choice of the system to simulate. We list here a few examples which might convey better the idea of suitable setup for systems of current interest.

Simulations of protein and their interactions Simulations of protein have been crucial to understand small molecular details of their structure and functioning, which influence their macroscopic behaviour.

A challenging subject in the field of enzymatic regulation is constituted by allosteric regulation, as the pathways involved are of difficult exploration. Molecular dynamics simulations can shed light on the residues responsible for transferring the information [63]. For example, atomistic simulations of pre-assembled pyruvate kinase M2 (PKM2) tetramers elucidated the mechanism of allosteric activation by fructose 1,6-bisphosphate (FBP) [20]. In this work, the correlated motions extracted from simulations of the apo and FBP bound tetramer were processed within an information-theoretical framework, allowing for the extraction of specific residue hubs. These were experimentally demonstrated to be responsible of the communication between orthosteric and allosteric pockets.

Another particularly interesting concept explored by MD is the one of *ensemble*: proteins adopt a variety of shapes in solution, which can differ significantly from the crystal-structures available. This flexibility is proven by techniques as small angle X-ray scattering (SAXS) or nuclear magnetic resonance (NMR), which results cannot be explained by a single-conformation scenario [64–66]. As such, simulations can provide “snapshots” of all these conformations and their mutual interchange. Some of these changes are crucial for the protein function, as in the case of prions, proteins which become pathogenic under conversion from an α -helix rich form to a β -sheet isoform which is prone to aggregation. Atomistic MD simulations of the prion protein showed the (reversible) unfolding of the protein between the two structural states [67], starting from the helical crystal structure available. Moreover, they identified

in the C-terminus region the trigger of the conformational change, corresponding to a region which is enriched in pathogenic mutations. The results were then used to investigate the aggregation process after the conversion, again identifying critical residues promoting it [68]. Interestingly, the water distribution density around the prion molecule was demonstrated to have a role in identifying possible spots prone to aggregation [69, 70]. Such examples show that one should consider the solvent, or the forces induced by the protein surrounding media, as an important determinant in the conformational equilibria and recognition processes. Assembly simulations applied to amyloid forming peptides has provided recently a map of all the relevant states of the assembly process [71], suggesting that the constant improvement of computational power will make this possible also for more complex systems like the prion in a near future.

Simulations of model membranes Membrane simulations had a key role in elucidating the mechanical and dynamical characteristics of these important organelles clarifying for example how the lipids composing the membrane influence its fluidity [72, 73], or elucidating the interactions between lipids and surface proteins, transmembrane ones or membrane active peptides, such as antimicrobial ones [74–76].

The last point in particular raised the interest in simulating bacterial membranes. As these objects are very complex, simplified models can be used to approach the problem. Very often, these models use only one or two lipid species to represent membranes, e.g. bacterial ones contains neutral phospholipids with a percentage of negatively charged ones [19, 77–79] as key characteristic distinguishing them from mammal membranes, which possess only zwitterionic phospholipids, with the occasional inclusion of cholesterol, deemed important in achieving the flexibility typical of mammal membranes [19, 72, 77–79]. Because of their simplicity, these systems are extensively used also in experiments [80–82], making a direct comparison with simulations possible. Nevertheless, attempts to model more accurately cell membranes have been pursued [83].

For membrane systems, the atomistic level is suitable to reproduce single components of the membrane, e.g. the isolated inner membrane of Gram-negative bacteria

[84], even if atomistic computations started in recent years to be affordable also for larger systems. For example, atomistic simulation of the outer membrane of Gram negative bacteria combined with the peptoglycan layer (which is positioned between the two membranes) elucidated how their distance is variable, thanks to the presence of Braun's lipoproteins which act as a bridge between the two, and can bring them closer by bending and tilting [22]. Moreover, the permeability of membranes to ions and small compounds needs to be assessed at the atomistic level to get sufficient accuracy, and because of the intensive task, often enhanced MD techniques such as umbrella sampling are employed [85].

Coarse grain descriptions are instead the most suitable to represent the full bacterial envelope, especially in the case of Gram-negative bacteria, as the inclusion of all its elements results in large systems.

Accordingly, MARTINI simulations have been able to reproduce the behaviour of several transmembrane proteins which spanned both the inner and outer membrane [86] and to model all the different components of the Gram-negative cell envelope [83]. Similarly, the ability of coarse grain simulations to investigate very large systems makes them suitable to assess elastic properties of membranes, as they can access low frequency undulations with little influence from finite size effects [87].

4 Notes: validation and challenges of MD simulations

Validation of MD simulations is performed by comparison with experiments: the same properties obtained experimentally are computed from the MD trajectory as well, and the two compared. If these are correctly reproduced, it is usually assumed that the simulation is sampling the correct ensemble of states and then one can identify in the simulation the determinants responsible for the experimental outcome of interest, which are not accessible by the experiment itself.

The comparison however is not always easy: often the experimentally measured quantity is an average in time and/or space (for example Circular Dichroism spectra

or SAXS profiles of a peptide in solution) and many different combinations of computationally derived structure ensembles can produce compatible results. It is still challenging to compare experimentally derived ensembles and the ones derived by molecular simulations. Indeed the extracted average properties may be different, so it is important to understand which are the relevant ones playing a role in the measured ensemble before attempting comparisons.

Thus, in the validation of MD outcomes, it is important to have a critical attitude and to interpret the result within the validity of the approximation [88]. Indeed, agreement may arise from either a simulation that reflects correctly the experimental system; but also when the property examined is insensitive to the details of the simulated trajectory, or it can be a result of compensation of errors, which is more likely to occur for systems with a high number of degrees of freedom. Similarly, disagreement may hint at an error in the simulation (either in the model, the implementation, or simply the estimated simulation's convergence) or an error in the interpretation and/or conditions of experimental set-up (either in the result itself or its interpretation), so that both must be carefully checked to improve a convergence in the agreement. Some apparently negative results may suggest or stimulate new experimental settings to validate the hypotheses one was set to test [89, 90].

Moreover, simulations still suffer from the limited computational time available: most of the times, the real experimental system is simply too large to be reproduced and the time scale of the process too long. Simulations are thus confined to explore a restricted space, implying that the initial conditions must be chosen carefully to optimise the search and avoid any bias which might persist for the whole length of the simulation. The use of enhanced MD techniques can increase the chances of sampling relevant states, however it introduces a bias which must be removed or properly accounted for in the interpretation of the results [91–95]. Finally, one should keep in mind that the force fields used are far from optimal, partly because they rely on approximate functional forms, and partly because it is difficult to find experimental observables measured with the desired resolution able to discriminate between sets of parameters.

Nevertheless, it remains important to note the contribution that Molecular Dynamics simulations have played in disclosing important details behind biophysical processes and in unravelling molecular details not accessible to experiments.

References

- [1] Lee EH, et al (2009) Discovery Through the Computational Microscope. *Structure* 17(10):1295–1306
- [2] Dror RO, et al (2012) Biomolecular Simulation: A Computational Microscope for Molecular Biology. *Annual Review of Biophysics* 41(1):429–452
- [3] Vogel A, and Huster D (2017) Combining NMR Spectroscopy and Molecular Dynamics Simulation to Investigate the Structure and Dynamics of Membrane-Associated Proteins. Springer, Cham, pp 311–350
- [4] Heo L, and Feig M (2018) Experimental accuracy in protein structure refinement via molecular dynamics simulations. *Proceedings of the National Academy of Sciences of the United States of America* 115(52):13276–13281
- [5] Berendsen HJC, et al (1984) Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics* 81(8):3684–3690
- [6] Nosé S, and Klein M (1983) Constant pressure molecular dynamics for molecular systems. *Molecular Physics* 50(5):1055–1076
- [7] Hoover WG (1985) Canonical dynamics: Equilibrium phase-space distributions. *Physical Review A* 31(3):1695–1697
- [8] Andersen HC (1980) Molecular dynamics simulations at constant pressure and/or temperature. *The Journal of Chemical Physics* 72(4):2384–2393
- [9] Bussi G, Donadio D, and Parrinello M (2007) Canonical sampling through velocity rescaling. *The Journal of Chemical Physics* 126(1):014101
- [10] Parrinello M, and Rahman A (1981) Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics* 52(12):7182–7190

- [11] Martyna GJ, et al (1996) Explicit reversible integrators for extended systems dynamics. *Molecular Physics* 87(5):1117–1157
- [12] Oostenbrink C, et al (2004) A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6. *Journal of Computational Chemistry* 25(13):1656–1676
- [13] Schmid N, et al (2011) Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *European Biophysics Journal* 40(7):843–856
- [14] Berendsen H, van der Spoel D, and van Drunen R (1995) GROMACS: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications* 91(1-3):43–56
- [15] Abraham MJ, et al (2015) GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 1-2:19–25
- [16] Abraham MJ, et al (2018) GROMACS User Manual version 2016. www.gromacs.org
- [17] Essmann U, et al (1995) A smooth particle mesh Ewald method. *The Journal of Chemical Physics* 103(19):8577–8593
- [18] Tironi IG, et al (1995) A generalized reaction field method for molecular dynamics simulations. *The Journal of Chemical Physics* 102(13):5451–5459
- [19] Lipkin R, Pino-Angeles A, and Lazaridis T (2017) Transmembrane Pore Structures of β -Hairpin Antimicrobial Peptides by All-Atom Simulations. *The Journal of Physical Chemistry B* 121(39):9126–9140
- [20] Macpherson JA, et al (2019) Functional cross-talk between allosteric effects of activating and inhibiting ligands underlies PKM2 regulation. *eLife* 8
- [21] Machado MR, González HC, and Pantano S (2017) MD Simulations of Virus-like Particles with Supra CG Solvation Affordable to Desktop Computers. *Journal of Chemical Theory and Computation* 13(10):5106–5116
- [22] Samsudin F, et al (2017) Braun's Lipoprotein Facilitates OmpA Interaction with the Escherichia coli Cell Wall. *Biophysical Journal* 113(7):1496–1504

- [23] Maier JA, et al (2015) ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of Chemical Theory and Computation* 11(8):3696–3713
- [24] Dickson CJ, et al (2014) Lipid14: The Amber Lipid Force Field. *Journal of Chemical Theory and Computation* 10(2):865–879
- [25] Wang J, et al (2004) Development and testing of a general amber force field. *Journal of Computational Chemistry* 25(9):1157–1174
- [26] MacKerell AD, et al (1998) All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins . *The Journal of Physical Chemistry B* 102(18):3586–3616
- [27] Klauda JB, et al (2010) Update of the CHARMM All-Atom Additive Force Field for Lipids: Validation on Six Lipid Types. *The Journal of Physical Chemistry B* 114(23):7830–7843
- [28] Huang J, and MacKerell AD (2013) CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *Journal of Computational Chemistry* 34(25):2135–2145
- [29] Jorgensen WL, and Tirado-Rives J (1988) The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *Journal of the American Chemical Society* 110(6):1657–1666
- [30] William L Jorgensen, et al (1996) Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids
- [31] Marrink SJ, et al (2007) The MARTINI force field: Coarse grained model for biomolecular simulations. *Journal of Physical Chemistry B* 111(27):7812–7824
- [32] Monticelli L, et al (2008) The MARTINI Coarse-Grained Force Field: Extension to Proteins. *Journal of chemical theory and computation* 4(5):819–34
- [33] de Jong DH, et al (2013) Improved Parameters for the Martini Coarse-Grained Protein Force Field. *Journal of Chemical Theory and Computation* 9(1):687–

697

- [34] Machado MR, et al (2018) The SIRAH force field 2.0: Altius, Fortius, Citius. bioRxiv p 436774
- [35] Barrera EE, Machado MR, and Pantano S (2019) Fat SIRAH: Coarse-grained phospholipids to explore membrane-protein dynamics. bioRxiv p 627570
- [36] Oostenbrink C, et al (2005) Validation of the 53A6 GROMOS force field. European Biophysics Journal 34(4):273–284
- [37] Reif MM, Winger M, and Oostenbrink C (2013) Testing of the GROMOS Force-Field Parameter Set 54A8: Structural Properties of Electrolyte Solutions, Lipid Bilayers, and Proteins. Journal of Chemical Theory and Computation 9(2):1247–1264
- [38] Berendsen H, et al (1981) Interaction models for water in relation to protein hydration. In: Pullman B (ed) Intermolecular Forces, D. Reidel Publishing Company, Dordrecht, pp 331–342
- [39] Berman HM, et al (2000) The protein data bank. Nucleic Acids Research, 28: 235-242
- [40] Periole X, et al (2009) Combining an Elastic Network With a Coarse-Grained Molecular Force Field: Structure, Dynamics, and Intermolecular Recognition. Journal of Chemical Theory and Computation 5(9):2531–2543
- [41] Yesylevskyy SO, et al (2010) Polarizable Water Model for the Coarse-Grained MARTINI Force Field. PLoS Computational Biology 6(6):e1000810
- [42] Wassenaar TA, et al (2015) Computational Lipidomics with *insane* : A Versatile Tool for Generating Custom Membranes for Molecular Simulations. Journal of Chemical Theory and Computation 11(5):2144–2155
- [43] Lee OS, Cho V, and Schatz GC (2012) Modeling the Self-Assembly of Peptide Amphiphiles into Fibers Using Coarse-Grained Molecular Dynamics. Nano Letters 12(9):4907–4913
- [44] Ren P, and Ponder JW (2003) Polarizable Atomic Multipole Water Model for Molecular Mechanics Simulation. Jurnal of Physical Chemistry B 24(107):5933–5947

- [45] Ponder JW, et al (2010) Current Status of the AMOEBA Polarizable Force Field. *The Journal of Physical Chemistry B* 114(8):2549–2564
- [46] Anisimov VM, et al (2005) Determination of Electrostatic Parameters for a Polarizable Force Field Based on the Classical Drude Oscillator. *Journal of Chemical Theory and Computation* 1(1):153–168
- [47] Cieplak P, Caldwell J, and Kollman P (2001) Molecular mechanical models for organic and biological systems going beyond the atom centered two body additive approximation: aqueous solution free energies of methanol and N-methyl acetamide, nucleic acid base, and amide hydrogen bonding and chloroform/water partition coefficients of the nucleic acid bases. *Journal of Computational Chemistry* 22(10):1048–1057
- [48] Orsi M, and Essex JW (2011) The ELBA Force Field for Coarse-Grain Modeling of Lipid Membranes. *PLoS ONE* 6(12):e28637
- [49] Ahmadi S, et al (2018) Multiscale modeling of enzymes: QM-cluster, QM/MM, and QM/MM/MD: A tutorial review. *International Journal of Quantum Chemistry* 118(9):e25558
- [50] Shao Y, et al (2015) Advances in molecular quantum chemistry contained in the Q-Chem 4 program package. *Molecular Physics* 113(2):184–215
- [51] Kleinjung J, and Fraternali F (2014) Design and application of implicit solvent models in biomolecular simulations. *Current Opinion in Structural Biology* 25:126–134
- [52] Fraternali F, and van Gunsteren W (1996) An Efficient Mean Solvation Force Model for Use in Molecular Dynamics Simulations of Proteins in Aqueous Solution. *Journal of Molecular Biology* 256(5):939–948
- [53] Kleinjung J, et al (2003) Thermal unfolding simulations of apo-calmodulin using leap-dynamics. *Proteins: Structure, Function, and Bioinformatics* 50(4):648–656
- [54] Kleinjung J, et al (2012) Implicit Solvation Parameters Derived from Explicit Water Forces in Large-Scale Molecular Dynamics Simulations. *Journal of Chemical Theory and Computation* 8(7):2391–2403

- [55] Fornili A, et al (2012) Protein–Water Interactions in MD Simulations: POPC/POPCSCOMP Solvent Accessibility Analysis, Solvation Forces and Hydration Sites. Springer, New York, NY, pp 375–392
- [56] Zhu J, Alexov E, and Honig B (2005) Comparative Study of Generalized Born Models: Born Radii and Peptide Folding. *The Journal of Physical Chemistry B* 109(7):3008–3022
- [57] Milano G, and Kawakatsu T (2009) Hybrid particle-field molecular dynamics simulations for dense polymer systems. *The Journal of Chemical Physics* 130(21):214106
- [58] Okamoto Y (2004) Generalized-ensemble algorithms: enhanced sampling techniques for Monte Carlo and molecular dynamics simulations. *Journal of Molecular Graphics and Modelling* 22(5):425–439
- [59] Huber T, Torda AE, and van Gunsteren WF (1994) Local elevation: A method for improving the searching properties of molecular dynamics simulation. *The Journal of Physical Chemistry A* 101(33):5926–5930
- [60] Laio A, and Parrinello M (2002) Escaping free-energy minima. *Proceedings of the National Academy of Sciences of the United States of America* 99(20):12562–6
- [61] Torrie G, and Valleau J (1977) Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics* 23(2):187–199
- [62] Kirkpatrick S, Gelatt CD, and Vecchi MP (1983) Optimization by Simulated Annealing. *Science* 220(4598):671–680
- [63] Pandini A, et al (2012) Detection of allosteric signal transmission by information-theoretic analysis of protein dynamics. *The FASEB Journal* 26(2):868–881
- [64] Bonomi M, et al (2017) Principles of protein structural ensemble determination. *Current Opinion in Structural Biology* 42:106–116
- [65] Kikhney AG, and Svergun DI (2015) A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. *FEBS Letters*

- Letters 589(19PartA):2570–2577
- [66] Kleckner IR, and Foster MP (2011) An introduction to NMR-based approaches for measuring protein dynamics. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* 1814(8):942–968
 - [67] Chakroun N, et al (2013) Decrypting Prion Protein Conversion into a β -Rich Conformer by Molecular Dynamics. *Journal of Chemical Theory and Computation* 9(5):2455–2465
 - [68] Collu F, et al (2018) Probing the early stages of prion protein (PrP) aggregation with atomistic molecular dynamics simulations. *Chemical Communications* 54(57):8007–8010
 - [69] De Simone A, et al (2005) Prion and water: tight and dynamical hydration sites have a key role in structural stability. *Proceedings of the National Academy of Sciences of the United States of America* 102(21):7535–40
 - [70] De Simone A, et al (2006) Water molecules as structural determinants among prions of low sequence identity. *FEBS Letters* 580(10):2488–2494
 - [71] Sengupta U, Carballo-Pacheco M, and Strodel B (2019) Automated Markov state models for molecular dynamics simulations of aggregation and self-assembly. *The Journal of Chemical Physics* 150(11):115101
 - [72] Risselada HJ, and Marrink SJ (2008) The molecular face of lipid rafts in model membranes. *Proceedings of the National Academy of Sciences of the United States of America* 105(45):17367–72
 - [73] Song C, de Groot BL, and Sansom MS (2019) Lipid Bilayer Composition Influences the Activity of the Antimicrobial Peptide Dermcidin Channel. *Biophysical Journal* 116(9):1658–1666
 - [74] Leontiadou H, Mark AE, and Marrink SJ (2006) Antimicrobial Peptides in Action. *Journal of the American Chemical Society* 128(37):12156–12161
 - [75] Ulmschneider JP (2017) Charged Antimicrobial Peptides Can Translocate across Membranes without Forming Channel-like Pores. *Biophysical Journal* 113(1):73–81

- [76] Sun D, Forsman J, and Woodward CE (2015) Atomistic Molecular Simulations Suggest a Kinetic Model for Membrane Translocation by Arginine-Rich Peptides. *The Journal of Physical Chemistry B* 119(45):14413–14420
- [77] Wang Y, et al (2012) Comparative molecular dynamics simulations of the antimicrobial peptide CM15 in model lipid bilayers. *Biochimica et Biophysica Acta (BBA) - Biomembranes* 1818(5):1402–1409, NIHMS150003
- [78] Zhao L, et al (2018) Molecular Dynamics Simulations of Human Antimicrobial Peptide LL-37 in Model POPC and POPG Lipid Bilayers. *International Journal of Molecular Sciences* 19(4):1186
- [79] Chen C, et al (2019) Simulation-guided rational *de novo* design of a small pore-forming antimicrobial peptide. *Journal of the American Chemical Society* p jacs.8b11939
- [80] Castelletto V, et al (2016) Structurally plastic peptide capsules for synthetic antimicrobial viruses. *Chem Sci* 7(3):1707–1711
- [81] Tang M, and Hong M (2009) Structure and mechanism of β -hairpin antimicrobial peptides in lipid bilayers from solid-state NMR spectroscopy. *Molecular BioSystems* 5(4):317
- [82] Glukhov E, et al (2005) Basis for selectivity of cationic antimicrobial peptides for bacterial versus mammalian membranes. *The Journal of biological chemistry* 280(40):33960–7
- [83] Khalid S, Piggot TJ, and Samsudin F (2019) Atomistic and Coarse Grain Simulations of the Cell Envelope of Gram-Negative Bacteria: What Have We Learned? *Accounts of Chemical Research* 52(1):180–188
- [84] Piggot TJ, Holdbrook DA, and Khalid S (2011) Electroporation of the *E. coli* and *S. Aureus* Membranes: Molecular Dynamics Simulations of Complex Bacterial Membranes. *The Journal of Physical Chemistry B* 115(45):13381–13388
- [85] Carpenter TS, Parkin J, and Khalid S (2016) The Free Energy of Small Solute Permeation through the *Escherichia coli* Outer Membrane Has a Distinctly

- Asymmetric Profile. *The Journal of Physical Chemistry Letters* 7(17):3446–3451
- [86] Hsu PC, et al (2017) It Is Complicated: Curvature, Diffusion, and Lipid Sorting within the Two Membranes of *Escherichia coli*. *The Journal of Physical Chemistry Letters* 8(22):5513–5518
- [87] Fowler PW, et al (2016) Membrane stiffness is modified by integral membrane proteins. *Soft Matter* 12:7792
- [88] van Gunsteren WF, Dolenc J, and Mark AE (2008) Molecular simulation as an aid to experimentalists. *Current Opinion in Structural Biology* 18(2):149–153
- [89] Gonçalves MB, et al (2013) Structural prediction of a rhodamine-based biosensor and comparison with biophysical data. *Phys Chem Chem Phys* 15(6):2177–2183
- [90] Meißner RH, et al (2014) Computational Prediction of Circular Dichroism Spectra and Quantification of Helicity Loss upon Peptide Adsorption on Silica. *Langmuir* 30(12):3487–3494
- [91] Bernardi RC, Melo MCR, and Schulten K (2015) Enhanced sampling techniques in molecular dynamics simulations of biological systems. *Biochimica et biophysica acta* 1850(5):872–7
- [92] Best RB, Clarke J, and Karplus M (2005) What Contributions to Protein Side-chain Dynamics are Probed by NMR Experiments? A Molecular Dynamics Simulation Analysis. *Journal of Molecular Biology* 349(1):185–203
- [93] Barducci A, Bonomi M, and Parrinello M (2010) Linking Well-Tempered Metadynamics Simulations with Experiments. *Biophysical Journal* 98(9):L44–L46
- [94] Barducci A, Bonomi M, and Parrinello M (2011) Metadynamics. *Wiley Interdisciplinary Reviews: Computational Molecular Science* 1(5):826–843
- [95] Mills M, and Andricioaei I (2008) An experimentally guided umbrella sampling protocol for biomolecules. *The Journal of chemical physics* 129(11):114101