Review Article

Molecular Dynamics, Monte Carlo Simulations, and Langevin Dynamics: A Computational Review

Eric Paquet[1,2] and Herna L. Viktor[2]

[1]*Vaccine Program, National Research Council, 1200 Montreal Road, Ottawa, ON, Canada K1A 0R6*
[2]*School of Electrical Engineering and Computer Science, University of Ottawa, 800 King Edward Road,*
  Ottawa, ON, Canada K1N 6N5

Correspondence should be addressed to Eric Paquet; eric.paquet@nrc-cnrc.gc.ca

Macromolecular structures, such as neuraminidases, hemagglutinins, and monoclonal antibodies, are not rigid entities. Rather, they are characterised by their flexibility, which is the result of the interaction and collective motion of their constituent atoms. This conformational diversity has a significant impact on their physicochemical and biological properties. Among these are their structural stability, the transport of ions through the M2 channel, drug resistance, macromolecular docking, binding energy, and rational epitope design. To assess these properties and to calculate the associated thermodynamical observables, the conformational space must be efficiently sampled and the dynamic of the constituent atoms must be simulated. This paper presents algorithms and techniques that address the abovementioned issues. To this end, a computational review of molecular dynamics, Monte Carlo simulations, Langevin dynamics, and free energy calculation is presented. The exposition is made from first principles to promote a better understanding of the potentialities, limitations, applications, and interrelations of these computational methods.

1. Introduction

The ability to properly sample configurational and conformational properties and to subsequently describe at the atomic level the dynamical evolution of complex macromolecular systems has wide application. This research is of paramount importance in the study of macromolecular stability of mutant proteins [1], molecular recognition, ions, and small molecule transportation of the influenza M2 channel [2, 3], protein association, the role of protein flexibility for influenza A RNA binding [4, 5], folding and hydration, influenza neuraminidase inhibitor [6–9], drug resistance [10], enzymatic reactions, folding transitions [11, 12], screening [13], accessibility assessment (see Figure 1), and hemagglutinin fusion peptide [14]. One should also mention multivalent binding mode [15], docking [16], drug (e.g., Oseltamivir and Zanamivir) efficiency against mutants [17, 18], structural biochemistry [19], biophysics, molecular biology, influenza multiple dynamics interactions [20], enzymology, pharmaceutical chemistry [21], biotechnology, rational epitope design [22], computation vaccinology [23],

binding [24], and free energy [25, 26]. For instance, one may wish to calculate the free energy to assess the strength and the stability of the bond in between a monoclonal antibody (mAb) and an antigen, such as the viral hemagglutinin, to quantify the efficiency of the neutralisation process.

This paper presents an algorithmic review from the first principles of Monte Carlo simulation, molecular dynamics, and Langevin dynamics (i.e., techniques that have been shown to address the abovementioned scenario). We focus our attention on the algorithmic aspect, which, within the context of a review, has not received sufficient attention. Our objective is not only to explain the algorithms but also to highlight their potential, limitations, applicability, interrelations, and generalisation in the context of molecular dynamics. To this end, a number of algorithmic approaches are presented in detail, and the pros and cons of each are highlighted. The algorithms are illustrated with examples related to the influenza virus.

This paper is organised as follows. Monte Carlo simulations are reviewed in Section 2. Section 3 is concerned with molecular dynamics in the microcanonical ensemble, that is,
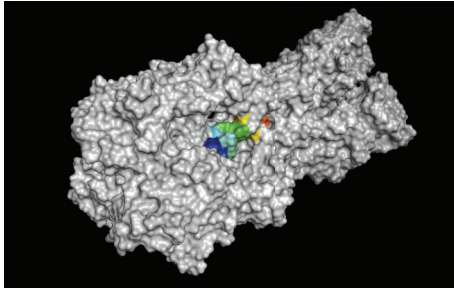
FIGURE 1: Accessibility assessment of a region of the influenza A virus (A/swine/Iowa/15/1930 (H1N1)).

at constant energy. Section 4 extends molecular dynamics to the canonical and the isobaric-isothermal ensemble. Constrained molecular dynamics, hybrid molecular dynamics, and steered molecular dynamics are also presented. Section 5 introduces Langevin and self-guided Langevin dynamics, and Section 6 is concerned with the calculation of the free energy. The application of molecular dynamics to macromolecular docking is addressed in Section 7. Finally, the connection in between molecular dynamics and quantum mechanics (ab initio simulations) is outlined in Section 8. This is followed by a short conclusion.

2. Monte Carlo Simulations

The objective of a Monte Carlo (MC) simulation is to generate an ensemble of representative configurations under specific thermodynamics conditions for a complex macromolecular system [27]. Applying random perturbations to the system generates these configurations. To properly sample the representative space, the perturbations must be sufficiently large, energetically feasible and highly probable. Monte Carlo simulations do not provide information about time evolution. Rather, they provide an ensemble of representative configurations, and, consequently, conformations from which probabilities and relevant thermodynamic observables, such as the free energy, may be calculated. Monte Carlo simulations are not only important on their own right, but they also play a fundamental role when designing complex and hybrid molecular dynamic (MD) algorithms [28].

This section is dedicated to Monte Carlo simulations. In Section 2.1 we review some important notions about Lagrangian and Hamiltonian dynamics, which are pervasive for both Monte Carlo simulations and molecular dynamics. In Section 2.2 we introduce the partition function and the probability density function, as well as the calculation of thermodynamics observable associated with a macromolecule such as the hemagglutinin or the neuraminidase. The partition function is instrumental in computing such observables. In Section 2.3 we explain how to efficiently sample the representative space. For that, we introduce the notions of emission probability, transition probability, acceptance probability, and detailed balance.

Sampling is useful only when performed in realistic experimental conditions. For this reason we explain how to

sample in the canonical ensemble (with a constant number of particles, volume, and temperature) and also in the isothermal-isobaric ensemble (with a constant number of particles, pressure, and volume) in Sections 2.4 and 2.5, respectively. Finally, in Section 2.6 we address the problem of sampling in the presence of numerous minima. This is a problem particularly acute when studying influenza macromolecular structures such as the hemagglutinin and the neuraminidase.

2.1. Lagrangian and Hamiltonian Dynamics or How to Formulate Our Problem. This section presents some important notions about Lagrangian and Hamiltonian dynamics, which are pervasive and recurrent for both MC and MD. Lagrangian and Hamiltonian dynamics provide an ideal framework for the description of complex macromolecular systems, both in Cartesian and generalised coordinates [29]. The Lagrangian is defined as the difference in between the kinetic and the potential energy:

$$\mathscr{L}\left(q^{3N}, \dot{q}^{3N}\right) = \mathscr{K}\left(\dot{q}^{3N}\right) - \mathscr{U}\left(q^{3N}\right), \tag{1}$$

where the kinetic energy is given by

$$\mathscr{K} = \sum_{\alpha=1}^{3N} \frac{1}{2} m_\alpha \dot{q}_\alpha^2, \tag{2}$$

and potential energy $\mathscr{U}(q^{3N})$ is a function of the positions of the constituent atoms. The q^{3N} are the generalised coordinates, and the $\dot{q}^{3N}$ are the generalised velocities. For instance, a generalised coordinate may be a bound length, a bound angle, or a dihedral angle. The space of all generalised coordinates and velocities is called the configuration space. If one introduces the generalised momentum:

$$p^\alpha \equiv \frac{\partial \mathscr{L}}{\partial \dot{q}_\alpha}, \tag{3}$$

one may define the Hamiltonian, which is the Legendre transformation of the Lagrangian:

$$\mathscr{H}\left(q^{3N}, p^{3N}\right) = \sum_{\alpha=1}^{3N} p^\alpha q_\alpha - \mathscr{L}\left(q^{3N}, \dot{q}^{3N}\left(q^{3N}, p^{3N}\right)\right). \tag{4}$$

The Hamiltonian obeys the so-called Hamilton's equations, which is just another formulation of Newton's equations:

$$\frac{d\mathscr{H}}{dt} = 0; \qquad \dot{q}_\alpha = \frac{\partial \mathscr{H}}{\partial p^\alpha}; \qquad \dot{p}^\alpha = -\frac{\partial \mathscr{H}}{\partial q_\alpha}. \tag{5}$$

Throughout the text, we will use both generalised and Cartesian coordinates: the Cartesian coordinates of the i constituent atom are noted $\mathbf{r}_i \equiv [x_i, y_i, z_i]^T$. The set of all Cartesian coordinates is noted as

$$\mathbf{r}^N \equiv \mathbf{r}_1, \dots, \mathbf{r}_N, \tag{6}$$

and the associated differential volume element is expressed as

$$d\mathbf{r}^N \equiv d\mathbf{r}_1 \dots d\mathbf{r}_N. \tag{7}$$

In the next section, we introduce the notions of partition function and probability density function.

2.2. Partition Functions, Probability Density Functions, and Expectation or How to Compute Observables. Partition functions are pervasive to all Monte Carlo simulations. They are required to determine the number of microstates associated with a macromolecule, the probability of occurrence of a specific conformation, and the ensemble averages of observables, such as the enthalpy or thermodynamical quantities like the free energy from which the strength of a bound between a drug, such as Oseltamivir [30], and a viral neuraminidase may be asserted.

The number of microstates may be obtained by

$$\Omega\left(N, V, E\right) \equiv E_0 C_{\{N\}} \int d\mathbf{p}^N d\mathbf{r}^N \delta\left(\mathcal{H}\left(\mathbf{r}^N, \mathbf{p}^N\right) - E\right), \quad (8)$$

where

$$C_{\{N\}} = \frac{1}{h^{3N}\left(N_A! N_B! \cdots\right)} \quad (9)$$

is a quantum factor, which accounts for the indiscernibility of the various atomic species A, B, and so forth, h is the Planck constant, and $\delta(x)$ is Dirac delta function. The function $\Omega(N, V, E)$ counts the number of states of constant energy E in a system. It is directly related to the entropy

$$S\left(N, V, E\right) = k_B \ln \Omega\left(N, V, E\right), \quad (10)$$

where k_B is the Boltzmann constant. The canonical partition function is defined as

$$Z\left(N, V, T\right) \equiv \int d\mathbf{p}^N d\mathbf{r}^N \exp\left[-\beta\mathcal{H}\left(\mathbf{r}^N, \mathbf{p}^N\right)\right], \quad (11)$$

where

$$\beta \equiv \frac{1}{k_B T} \quad (12)$$

and T is the temperature. The canonical partition function is a functional, which is uniquely determined by the Hamiltonian of the corresponding macromolecular system. If the indiscernibility factor is included in the definition, the microcanonical partition function is noted as

$$Q\left(N, V, T\right) \equiv C_{\{N\}} Z\left(N, V, T\right). \quad (13)$$

The probability that the macromolecule is in a state characterised by atomic positions $\mathbf{r}^N$, and atomic momenta $\mathbf{p}^N$ is given by

$$\Pr\left(\mathbf{r}^N, \mathbf{p}^N\right) d\mathbf{p}^N d\mathbf{r}^N = \frac{\exp\left[-\beta\mathcal{H}\left(\mathbf{r}^N, \mathbf{p}^N\right)\right] d\mathbf{p}^N d\mathbf{r}^N}{Z\left(N, V, E\right)}. \quad (14)$$

Consequently, the ensemble average of an observable O (such as the enthalpy) is obtained by weighting the various realisations of the observable by their corresponding probability:

$$\langle O \rangle = \int d\mathbf{p}^N d\mathbf{r}^N O\left(\mathbf{r}^N, \mathbf{p}^N\right) \Pr\left(\mathbf{r}^N, \mathbf{p}^N\right). \quad (15)$$

The uncertainty (standard deviation) associated with the observable is given by

$$\sigma\left(O\right) = \sqrt{\langle O^2 \rangle - \langle O \rangle^2}. \quad (16)$$

Unfortunately, it is not possible to integrate the partition function or to compute the probability directly. This is due to the large number of degrees of freedom. For instance, the *Homo sapiens* influenza hemagglutinin is formed of approximately 23.000 atoms, which means that the partition function must be integrated in a 138.000-dimensional space.

In the next section, we explain how to perform such integration efficiently.

2.3. Stochastic Sampling or How to Sample Efficiently Thermodynamical Quantities. The multidimensional integrals associated with the probability and the partition function may be efficiently calculated with a procedure called Monte Carlo integration. In this approach the integration space is sampled according to a Markovian process and the integral is approximated by the average of the corresponding sampled states. Such an approach is efficient if the sampled states have a high probability of occurrence.

A sufficient, but not necessary, condition for such an efficient sampling to hold is called detailed balance:

$$\Pr\left(\mathcal{I}\right) T\left(\mathcal{I} \longrightarrow \mathcal{J}\right) A\left(\mathcal{I} \longrightarrow \mathcal{J}\right)$$
$$= \Pr\left(\mathcal{J}\right) T\left(\mathcal{J} \longrightarrow \mathcal{I}\right) A\left(\mathcal{J} \longrightarrow \mathcal{I}\right), \quad (17)$$

where $\Pr(\mathcal{I})$ is the probability (emission probability) that the system is in the state $\mathcal{I} \equiv \{\mathbf{r}^N, \mathbf{p}^N\}$, $T(\mathcal{I} \longrightarrow \mathcal{J})$ is the transition probability from state $\mathcal{I}$ to state $\mathcal{J}$, and $A(\mathcal{I} \longrightarrow \mathcal{J})$ is the acceptance probability of such a transition. If we assume that the transition probability is symmetrical

$$T\left(\mathcal{I} \longrightarrow \mathcal{J}\right) = T\left(\mathcal{J} \longrightarrow \mathcal{I}\right), \quad (18)$$

then the detailed balance equation reduces to

$$\frac{A\left(\mathcal{I} \longrightarrow \mathcal{J}\right)}{A\left(\mathcal{J} \longrightarrow \mathcal{I}\right)} = \frac{\Pr\left(\mathcal{J}\right)}{\Pr\left(\mathcal{I}\right)} = \exp\left[-\beta\left(\mathcal{U}\left(\mathcal{J}\right) - \mathcal{U}\left(\mathcal{I}\right)\right)\right]. \quad (19)$$

A possible solution for this equation is

$$A\left(\mathcal{I} \longrightarrow \mathcal{J}\right) = \min\left\{1, \exp\left[-\beta\left(\mathcal{U}\left(\mathcal{J}\right) - \mathcal{U}\left(\mathcal{I}\right)\right)\right]\right\}. \quad (20)$$

This equation is the celebrated Metropolis algorithm [31]. Consequently, each state is defined from the previous one (Markovian process). A transition to a lower energy is always accepted, while a transition to a higher energy is accepted with probability $\exp[-\beta(\mathcal{U}(\mathcal{J}) - \mathcal{U}(\mathcal{I}))]$.

Numerous variations have been designed based on this algorithm [32, 33]. For instance, the local elevation method enhances sampling by adding penalty potential to any state previously sampled (also known as a taboo search algorithm). Although useful, the microcanonical partition function is not realistic from an experimental point of view. Indeed, most observations are performed either at constant volume and temperature (canonical ensemble) or at constant pressure and temperature (isobaric-isothermal ensemble). These distributions are introduced in the next two sections.

2.4. Canonical Ensemble (NVT) Sampling or How to Sample in Realistic Experimental Conditions.

The canonical ensemble is the ensemble associated with the observations made at constant volume and constant temperature. The configurational canonical partition function associated with such an ensemble is obtained by marginalising the momenta in (11):

$$Z(N, V, T) \equiv \int d\mathbf{r}^N \exp\left[-\beta \mathcal{U}\left(\mathbf{r}^N\right)\right]. \quad (21)$$

It may also be defined as

$$Q(N, V, T) \equiv M_{\{N\}} Z(N, V, T), \quad (22)$$

where the constant

$$M_{\{N\}} = \frac{1}{\left(\sqrt{h^2 \beta / 2\pi m_A}\right)^{3N_A} N_A! \left(\sqrt{h^2 \beta / 2\pi m_B}\right)^{3N_B} N_B! \cdots} \quad (23)$$

takes into account the indiscernibility of the constituent atoms. Like in the microcanonical ensemble, the probability of occurrence of a given state $\{\mathbf{r}^N\}$ is equal to

$$\Pr_{NVT}\left(\mathbf{r}^N\right) d\mathbf{r}^N = \frac{\exp\left[-\beta \mathcal{U}\left(\mathbf{r}^N\right)\right] d\mathbf{r}^N}{Z(N, V, T)}. \quad (24)$$

Consequently, the average value of an observable is given by

$$\langle O \rangle = \frac{1}{Z(N, V, T)} \int d\mathbf{r}^N \exp\left[-\beta \mathcal{U}\left(\mathbf{r}^N\right)\right] O\left(\mathbf{r}^N\right). \quad (25)$$

In the case of the canonical partition function, the acceptance probability associated with Monte Carlo method reduces to

$$\begin{aligned} A_{NVT}\left(\mathbf{r}^N \longrightarrow \mathbf{r}'^N\right) \\ = \min\left\{1, \exp\left[-\beta\left(\mathcal{U}\left(\mathbf{r}'^N\right) - \mathcal{U}\left(\mathbf{r}^N\right)\right)\right]\right\}. \end{aligned} \quad (26)$$

From the canonical partition function, it is possible to obtain various thermodynamical quantities such as the Helmholtz free energy:

$$F(N, V, T) = -k_B T \ln Q(N, V, T). \quad (27)$$

Still, most observations are performed at constant pressure and temperature. To address this limitation, the isobaric-isothermal ensemble, as presented in the next section, is introduced.

2.5. Isobaric-Isothermal Ensemble (NPT) Sampling or How to Sample in Even More Realistic Experimental Conditions.

The isobaric-isothermal ensemble is representative of many experimental conditions. From the microcanonical formalism, it is possible to demonstrate that the isobaric-isothermal configurational partition function is equal to

$$Z(N, P, T) \equiv \int dV \exp\left[-\beta PV\right] \int d\mathbf{r}^N \exp\left[-\beta \mathcal{U}\left(\mathbf{r}^N\right)\right]. \quad (28)$$

As usual, if the indiscernibility of the constituent atoms is taken into account, the partition function becomes

$$Q(N, P, T) = \frac{M_{\{N\}}}{V_0} Z(N, P, T). \quad (29)$$

Such a partition function assumes that the deformations of the macromolecular structure are isotropic (the same in all directions). These deformations occur to maintain the pressure constant. If the deformations are anisotropic, the partition function must be modified as follows:

$$Z(N, P, T) = \int d\mathbf{H} Z'(N, P, T, \mathbf{H}) \delta(\det \mathbf{H} - V), \quad (30)$$

where $\mathbf{H}$ is the tensor associated with an elementary parallelepiped volume, which must be integrated (marginalised) over all possible variations of the elementary shape. The probability that a macromolecular system is in a state $\mathbf{r}^N$ is given by

$$\Pr_{NPT}\left(\mathbf{r}^N\right) d\mathbf{r}^N = \frac{\exp\left[-\beta PV\right] \exp\left[-\beta \mathcal{U}\left(\mathbf{r}^N\right)\right] d\mathbf{r}^N}{Z(N, P, T)}. \quad (31)$$

Again, various thermodynamical quantities may be defined from the partition function such as the Gibbs free energy:

$$G(N, P, T) = -k_B T \ln Q(N, P, T). \quad (32)$$

The isobaric-isothermal acceptance probability associated with the Monte Carlo method is

$$\begin{aligned} A_{NPT}&\left(\mathbf{r}^N, V \longrightarrow \mathbf{r}'^N, V'\right) \\ &= \min\left\{1, \exp\left[-\beta\left(\mathcal{U}\left(\mathbf{r}'^N, V'\right) - \mathcal{U}\left(\mathbf{r}^N, V\right)\right)\right]\right. \\ &\qquad \left. \times \exp\left[-\beta P\left(V' - V\right) + N \ln \frac{V'}{V}\right]\right\}. \end{aligned} \quad (33)$$

Irrelevant of the ensemble in which the calculations are performed, the Metropolis algorithm may be impaired by local minima [27]. Indeed, the acceptance probability may become trapped in a local minimum of the potential energy, which may result in an inadequate sampling of the macromolecular states, as seen in the conformational states [32, 33]. This issue is addressed in the following section.

2.6. Sampling and Local Minima or When Temperature May Help to Escape Local Minima.

Many biomolecular processes associated with influenza involve activated processes in which a high-energy barrier exists in between the initial and the final state [34]. In order to efficiently sample the macromolecular states, this type of barrier must be overcome.

An efficient, although computationally expensive, approach to overcome such a barrier is called replica exchange (refer to [34] and, in the same spirit, [35]). These methods involve a certain number of noninteracting simulations, called replicas, which are performed in parallel. Each simulation is characterised by its own temperature:

low temperature simulations tend to explore local minima, while high temperature simulations may overcome energy barriers and consequently move in between local minima. To favour a better exploration of the macromolecular states, the replicas are periodically exchanged (swapped) according to the following acceptance probability:

$$
\begin{aligned}
& A_R\left(\mathcal{I} \longrightarrow \mathcal{J}\right) \\
& = \min \left\{\mathcal{I}, \exp \left[-\left(\beta_{\mathcal{J}} - \beta_{\mathcal{I}}\right)\right.\right. \\
& \left.\left.\times \left(\left.\mathcal{U}\left(\mathcal{J}\right)\right|_{T_{\mathcal{J}}} - \left.\mathcal{U}\left(\mathcal{I}\right)\right|_{T_{\mathcal{I}}}\right)\right]\right\},
\end{aligned}
\tag{34}
$$

where

$$
\beta_{\mathcal{I}} \equiv \frac{1}{k_B T_{\mathcal{I}}}.
\tag{35}
$$

This acceptance probability is similar to the ones introduced before, except for the fact that each state is characterised by its own temperature. Once the exchange is completed, the simulations resume normally until another exchange is performed. The whole procedure allows for a better sampling of the macromolecular states. For instance, this approach has been utilised recently, in conjunction with simulated annealing, for creating the infectious disease model of the H1N1 influenza pandemic [36].

Until now, we have restricted ourselves to symmetrical transition functions. Often, a better sampling may be obtained if a nonsymmetrical sampling function is employed. Let us consider the particular case in which the nonsymmetrical function depends uniquely on the final conformation:

$$
T\left(\mathcal{I} \longrightarrow \mathcal{J}\right) = \pi\left(\mathcal{U}\left(\mathcal{J}\right)\right).
\tag{36}
$$

Then, the acceptance probability becomes

$$
\begin{aligned}
& A\left(\mathcal{I} \longrightarrow \mathcal{J}\right) \\
& = \min \left\{1, \frac{\pi\left(\mathcal{U}\left(\mathcal{I}\right)\right)}{\pi\left(\mathcal{U}\left(\mathcal{J}\right)\right)} \exp \left[-\beta\left(\mathcal{U}\left(\mathcal{J}\right) - \mathcal{U}\left(\mathcal{I}\right)\right)\right]\right\}.
\end{aligned}
\tag{37}
$$

This is the so-called bias sampling algorithm [37], which considerably increases the conformational sampling efficiency of large macromolecular chains [38].

Although MC simulations allow us to sample the most probable macromolecular states, they do not provide us with their temporal evolution. The study of the temporal evolution of a macromolecular state is called molecular dynamics and is the subject of the next section.

3. Molecular Dynamics or When Time Matters

Molecular dynamics studies the temporal evolution of the coordinates and the momenta (the state) of a given macromolecular structure. Such an evolution is called a trajectory. A typical trajectory is obtained by solving Newton's equations. The trajectory is important in assessing numerous time-dependent observables [39] such the accessibility of a given molecular surface [40], the interaction in between a small molecule (e.g., a drug) and the hemagglutinin or the neuraminidase of a given influenza strain, the interaction epitope-paratope in between an antigen (e.g., hemagglutinin) and an antibody (e.g., CR8020), the appearance and disappearance of a particular channel or cavity, and the fusion of the hemagglutinin with a cell membrane (fusion peptide), amongst others.

From an MD trajectory, it is possible to compute a temporal average of an observable by averaging this observable over time along the trajectory:

$$
\overline{O} = \lim_{t \to \infty} \frac{1}{t} \int_0^t d\tau O\left(r\left(\tau\right), \dot{r}\left(\tau\right)\right).
\tag{38}
$$

Although it has never been formally proven (and that it is not always applicable: for instance, when the trajectory is periodic or when the phase space is constituted of disconnected regions), the ergodicity principle is often invoked [41]. The ergodicity principle states that the average over periods of time along a given trajectory of an observable is, at the limit, identical to the ensemble average of this observable as obtained, for instance, from Monte Carlo simulations:

$$
\overline{O} \approx \langle O \rangle.
\tag{39}
$$

As we will see later, ergodicity is instrumental in performing MD simulations in the canonical and isobaric-isothermal ensemble. The next section is devoted to the potential or force field.

3.1. Potential or How to Approximate the Force Field. The choice of a proper potential is of the utmost importance in obtaining accurate molecular dynamics simulations [42]. The potential must be physically sound as well as computationally tractable. An approximate potential may be calculated from quantum mechanics and from the Born-Oppenheimer approximation in which only the positions of the atomic nucleus bonding are considered [42]. The potentials may be divided into bonding potentials and long-range potentials. The bonding potentials involve interaction with two atoms (bound lengths), three atoms (bound angles), and four atoms (dihedral angles). Long-range interactions are associated with the Lennard-Jones potential (van der Waal) and the Columbic potential. The harmonic approximation is utilised for the bonding potentials, which means that solely small displacements are accurately represented. The general form of the potential is

$$
\begin{aligned}
\mathcal{U}\left(\mathbf{r}^N\right) = & \sum_d k_d\left(d - d_0\right)^2 + \sum_S k_S\left(S - S_0\right)^2 \\
& + \sum_\theta k_\theta\left(\theta - \theta_0\right)^2 \\
& + \sum_\chi k_\chi\left(1 + \cos\left(n\chi - \delta\right)\right)
\end{aligned}
$$

$$+ \sum_{\varphi} k_{\varphi} \left(\varphi - \varphi_0 \right)^2$$

$$+ \sum_{i,j} \varepsilon_{ij} \left(\left(\frac{r_{ij}^0}{r_{ij}} \right)^{12} - \left(\frac{r_{ij}^0}{r_{ij}} \right)^6 + \frac{q_i q_j}{\varepsilon_l r_{ij}} \right), \tag{40}$$

where d is the bound length, S is the Urey-Bradley bound length, θ is the bound angle, χ is the dihedral angle, φ is the improper dihedral angle, r_{ij} is the distance in between atom i and j, $k_d, k_S, k_{\theta}, k_{\chi}$, and k_{φ} are constants, $d_0, S_0, \theta_0, \varphi_0$, and r_{ij}^0 are equilibrium positions, ε_{ij} is related to the Lennard-Jones well depth, and ε_l is the effective dielectric constant. Finally, q_i is the partial atomic charge associated with atom i: the partial charge comes from the asymmetrical distribution of the electrons in the chemical bounds. The first term on the last line is the van der Waal interaction (or Lennard-Jones potential), and the last term on the last line is the Columbic interaction. The parameters of the model are determined experimentally and from quantum mechanics. Among the most popular potentials are CHARMM and AMBER [42, 43]. The two differ mostly in the manner in which the parameters are estimated. These potentials may model proteins, lipids, ethers, and carbohydrates, as well as small molecules (e.g., drugs).

The number of interactions involved in long-range interactions rapidly becomes prohibitive. For instance, for the Columbic potential, there are potentially $N!/2!(N-2)!$ interactions, which correspond to approximately a quarter of billon interactions for an influenza hemagglutinin. To reduce the computational burden, their action range is truncated. The truncation should be performed in such a way as not to introduce artificial discontinuities, which may result in computational artefacts.

The next section is concerned with the refinement of experimentally determined macrostructures.

3.2. How to Minimize the Energy of the Conformation or How to Refine Experimentally Determined Structures.

The position of the constituent atoms of a macromolecular structure is usually determined either through X-ray crystallography for the larger structure or through nuclear magnetic resonance (NMR) for the smaller molecules. If only the amino acid sequence is available, the three-dimensional structure may be inferred either from methods based on homology, such as threading, or from ab initio methods, which predict the structure from the sequence alone [44]. Among the larger structures associated with influenza are the hemagglutinin and the neuraminidase. Because a protein has to be crystallised to apply X-ray crystallography, the position of its constituent atoms may be distorted from their natural positions by the crystallisation process. Consequently, bond lengths and bond angles may be distorted and steric clashes in between atoms may occur. Therefore, it is recommended to minimise the potential energy of the macromolecular structure to remediate this deficiency and to create a more realistic structure [45].

The global optimisation of nonlinear functions, such as the potential, is a notoriously difficult problem because of the complexity of the energy landscape and the profusion of local minima [46]. Usually, only local optimisation is performed. Such a minimisation may be achieved through various algorithms [46] such as the steepest descent algorithm, the conjugate gradient algorithm, and the Newton-Raphson method. The first two are based on the gradient, while the latter is based on the Hessian. In most cases a local optimisation is sufficient to refine the structure.

If a global optimisation is suited or required, an approach such as simulated annealing must be utilised [47]. Simulated annealing is an MC method. The position of the atoms is subjected to small random displacements. The acceptance probability of such a displacement is given by

$$A \left(\mathcal{I} \longrightarrow \mathcal{J} \right)$$
$$= \min \left\{ 1, \exp \left[-\beta_k \left(\left. \mathcal{U} \left(\mathcal{J} \right) \right|_{T_k} - \left. \mathcal{U} \left(\mathcal{I} \right) \right|_{T_k} \right) \right] \right\}, \tag{41}$$

where

$$T_k \in \{ T_1, \ldots, T_K \} \mid T_{k+1} < T_k. \tag{42}$$

This means that the temperature acts as a control parameter. Initially, the temperature is high, which implies that transitions from lower to higher energy are allowed with non-negligible probability in being able to escape local minima. Subsequently, the temperature is gradually reduced (cooling) to decrease the occurrence of such a transition. Transitions to lower energy are always accepted. With a proper choice of temperatures, a global optimisation may be achieved. The position of the global minimum associated with the energy landscape may be further refined with local optimisation.

The next section is devoted to the solvation of macromolecules.

3.3. Implicit and Explicit Solvation, Ions, and Poisson-Boltzmann Equation or How to Obtain Realistic Experimental Conditions.

Macromolecules do not exist in isolation. Water molecules and ions surround them. Often, to obtain a realistic simulation, the structure of interest must be solvated. Solvation is a vast and complex subject and we refer the reader to the literature for technical details [4, 44, 45, 48]. This section is devoted to some aspects of solvation, which are particularly relevant to MD.

The solvation may be either implicit or explicit. In the case of an implicit solvation [11], the water molecules are replaced by a potential, which describe their average action while, in the case of an explicit solvation, the macromolecule is surrounded by a solvation box constituted of water molecules. It follows that computers have a limited amount of memory, and thus, the size of this box cannot be infinite. To reduce the number of water molecules, various shapes may be utilised such as cubic, rhombic, dodecahedron truncated octahedron, and sphere. The shape of the solvation box is chosen to minimise the number of molecules required for solvation while maintaining at all times a minimum buffer of solvent. Because of its finite dimensions, the solvation box presents

unnatural boundary effects, which should be minimised. This may be partially achieved with a larger solvation box or with periodic boundary conditions (PBC). For a rectangular solvation box, periodic boundary conditions are defined as

$$O(x_i) = O(x_i + L_x)\big|_{i=1,\ldots,N},$$

$$O(y_i) = O(y_i + L_y)\big|_{i=1,\ldots,N}, \tag{43}$$

$$O(z_i) = O(z_i + L_z)\big|_{i=1,\ldots,N}.$$

Much care must be taken when using periodic boundary conditions to avoid unphysical artefacts. For instance, if the boundary box is too small, the head of a macromolecule, such as the hemagglutinin, may interact with its own tail, which is extremely unrealistic. Also, if Columbic interactions are involved, the system must be electrostatically neutral; otherwise, the total charge becomes infinite due to the endless replication of the system associated with the PBC. Ions could be added to the solvation box to neutralise the system. Even if the latter is neutral, ions, such as sodium and chloride, may be added to reproduce the ionic strength of the solvent in which the macromolecule evolves. Due to the periodic nature of the boundary conditions, duplicate interactions may appear. It is customary to apply the minimum image convention in which such duplicate interactions are not allowed. If the structure is too large, the solvation may be limited to a specific region of interest: for example, a binding site or a channel while implicit solvation may be utilised for the remaining part of the macromolecule. There are various solvation models [49–52]. Generally, their parameters are adjusted to reproduce the enthalpy of vaporisation and density of water.

The solvation box increases the complexity of the simulation. Indeed, most of the computational effort is directed toward simulating the solvent. Nevertheless, dielectric screening, electrostatic effects, and free energy, among others, may only be simulated through explicit solvation, which, consequently, is amply justified [53] though one should notice that explicit solvation does not allow for the simulation of the solvent viscosity.

If the macromolecule is solvated in an ionic solution, the electrostatic potential ϕ may be obtained with a greater accuracy by solving the Poisson-Boltzmann equation:

$$\frac{\partial}{\partial \mathbf{r}} \cdot \left(\varepsilon(\mathbf{r}) \frac{\partial \phi(\mathbf{r})}{\partial \mathbf{r}} \right)$$

$$= -\rho(\mathbf{r}) - \sum_{k=1}^{K} q_i c_i^{\infty} \lambda(\mathbf{r}) \exp(-\beta q_i \phi(\mathbf{r})), \tag{44}$$

where $\rho(\mathbf{r})$ represents the charge density of the solute (the macromolecule), q_i is the ionic charge, c_i^{∞} is the ionic concentration far from the solute, $\varepsilon(\mathbf{r})$ is the dielectric permittivity, and $\lambda(\mathbf{r})$ is an accessibility factor. The Poisson-Boltzmann equation is often used in modelling implicit solvation. It may be solved efficiently with finite element methods (FEM) [54].

In the next section, we show how to solve Newton's equation to obtain the trajectory of a macromolecule.

3.4. Integration of Newton's Equations: When Technical Details Matter. To obtain the trajectory associated with a given macromolecule, one should solve the corresponding Newton's equations. This is appropriate, since the trajectories are assumed to follow the laws of classical mechanics [55]. In this section we present a completely general approach from which most finite difference algorithms may be derived.

Newton's equation and their generalisation, Hamilton's equations, present two important characteristic: they are time-reversible and any infinitesimal volume in phase space (space of all coordinates and momenta) is conserved with time. The latter property is known as the Liouville theorem [56]:

$$d^{3N} p(0) \, d^{3N} q(0) = d^{3N} p(t) \, d^{3N} q(t). \tag{45}$$

This is another formulation of the conservation of the total energy of the system. Any numerical algorithm must enforce these two properties at all times to be physically realistic and consequently relevant. When aiming to derive a finite different algorithm that complies with these requirements, we start with the Liouville operator defined as

$$iL \equiv \sum_{\alpha=1}^{3N} \left[\frac{\partial \mathcal{H}}{\partial p^{\alpha}} \frac{\partial}{\partial q_{\alpha}} - \frac{\partial \mathcal{H}}{\partial q_{\alpha}} \frac{\partial}{\partial p^{\alpha}} \right], \tag{46}$$

where $i = \sqrt{-1}$. This operator allows recasting Hamilton's equations in the form:

$$q^{3N}(t) = \exp[iLt] \, q^{3N}(0), \tag{47}$$

which is known as the Liouville equation. In Cartesian coordinates the Liouville operator becomes

$$iL = iL_1 + iL_2,$$

$$iL_1 \equiv \sum_{i=1}^{N} \frac{\mathbf{p}_i}{m_i} \cdot \frac{\partial}{\partial \mathbf{r}_i}; \qquad iL_2 \equiv \sum_{i=1}^{N} \left(-\frac{\partial \mathcal{U}(\mathbf{r}^N)}{\partial \mathbf{r}_i} \right) \cdot \frac{\partial}{\partial \mathbf{p}_i}. \tag{48}$$

Because the two parts of the Liouville operator do not commute $L_1 L_2 \neq L_2 L_1$, one must rely on the Trotter theorem [5] to develop the argument of the exponential:

$$\exp[iL\Delta t]$$

$$= \exp[i(L_1 + L_2)\Delta t]$$

$$\approx \exp\left[iL_2 \frac{\Delta t}{2}\right] \exp[iL_1 \Delta t] \exp\left[iL_2 \frac{\Delta t}{2}\right] \tag{49}$$

$$+ \mathcal{O}(\Delta t^3).$$

Then, if each exponential is approximated in terms of a truncated Taylor expansion, the Liouville equation becomes

$$\mathbf{r}_i(t + \Delta t)$$

$$= 2\mathbf{r}_i(t) - \mathbf{r}_i(t - \Delta t)$$

$$- \frac{1}{m_i} \frac{\partial \mathcal{U}(\mathbf{r}^N(t))}{\partial \mathbf{r}_i} \Delta t^2 + \mathcal{O}(\Delta t^4)\bigg|_{i=1,\ldots,N}, \tag{50}$$

which is the well-known Verlet algorithm [1]. Note that other common algorithms, such as the Leap Frog algorithm and the reference system propagator algorithm (RESPA), may be derived following a similar approach [6, 7]. The Liouville operator may be partitioned in more than one way. For instance, the potential may be divided into short- and long-range potentials (e.g., Coulombs and van der Waal). Since long-range potentials tend to change more slowly than short-range potentials, a larger time increment may be used for the former. If this procedure is repeated, a multiple time-step algorithm may be designed. Once the trajectory has been calculated, the functional important motions may be separated from the random thermal motion by performing a principal component analysis (PCA), which expresses the dominant modes as linear combinations of the underlying motion [57].

The algorithms described in this section are only valid when the total energy is conserved, that is, in the microcanonical ensemble. In next section we show how to extend this approach to the experimentally more realistic canonical and isobaric-isothermal ensembles.

4. Non-Hamiltonian Molecular Dynamics or How to Reproduce Realistic Experimental Conditions during Molecular Dynamics Simulations

To simulate realistic experimental conditions, the MD simulations must be performed either at constant volume and temperature or at constant pressure and temperature. Unfortunately, the canonical and the isobaric-isothermal ensembles do not conserve the total energy. The conservation of a quantity, such as the volume or the pressure, requires a constant exchange of energy in between the macromolecular system and the surrounding heat bath. Such a process, in which the Liouville theorem is not valid, may be described in terms of non-Hamiltonian molecular dynamics [58], as will be detailed here. Firstly, we describe the general approach, which is subsequently applied to the canonical ensemble.

4.1. General Approach. To apply the abovementioned general method, we must complement the N Cartesian positions and momenta associated with the constituent atoms with M additional generalised coordinates and momenta. The set of all coordinates is collectively denoted by

$$\boldsymbol{\eta} = \left[\mathbf{r}^N \mid \mathbf{p}^N \mid q^M \mid p_q^M \right]^T. \tag{51}$$

It follows that the original Hamiltonian must be modified to include the generalised coordinates as well as an adjustable parameter ℓ:

$$\mathscr{H} \left(\mathbf{r}^N, \mathbf{p}^N \right) \longrightarrow \mathscr{H}' \left(\mathbf{r}^N, \mathbf{p}^N, q^M, p_q^M; \ell \right). \tag{52}$$

Unfortunately, the theory does not specify a particular form for the modified Hamiltonian. Because of the non-Hamiltonian nature of the dynamics, the Liouville theorem does not apply, which means that elementary volume elements are not conserved in phase space. The reason for this

is that the modified dynamics have altered the geometry of the phase space from an Euclidean (flat) geometry to a Riemannian (curved) geometry. In Riemannian geometry, the Liouville theorem becomes [59]

$$\sqrt{\left| \det \mathbf{G} \left(\boldsymbol{\eta} (t) \right) \right|} d\boldsymbol{\eta} (t) = \sqrt{\left| \det \mathbf{G} \left(\boldsymbol{\eta} (0) \right) \right|} d\boldsymbol{\eta} (0), \tag{53}$$

where $\mathbf{G}(\boldsymbol{\eta})$ is the metric associated with the Riemannian space, and det is the determinant of the matrix. We are only interested in the determinant of this matrix since only the latter is involved in the microcanonical partition function associated with the non-Hamiltonian system. The latter may be obtained directly from

$$\sqrt{\left| \det \mathbf{G} \left(\boldsymbol{\eta} (t) \right) \right|} = \exp \left[- \int_0^t \frac{\partial}{\partial \boldsymbol{\eta} (\tau)} \cdot \dot{\boldsymbol{\eta}} (\tau) \, d\tau \right]. \tag{54}$$

The number of microstates associated with the non-Hamiltonian system may be written as

$$\begin{aligned}
&\Omega \left(\mathbf{r}^N, \mathbf{p}^N, q^M, p_q^M; \ell \right) \\
&= \Omega_0 \int d\boldsymbol{\eta} \sqrt{\left| \det \mathbf{G} (\boldsymbol{\eta}) \right|} \prod_{k=1}^M \delta \left(\Lambda_k (\boldsymbol{\eta}) - C_k \right),
\end{aligned} \tag{55}$$

where $d\boldsymbol{\eta} \sqrt{\left| \det \mathbf{G}(\boldsymbol{\eta}) \right|}$ is the invariant volume element in extended phase space and $\Lambda_k(\boldsymbol{\eta}) - C_k$ is a function associated with a conservation law (for instance, the total energy, or the momentum associated with the barycentre of the system). If we integrate or marginalise the additional coordinates and momenta, we obtain a function that depends solely on the coordinates and momenta of the constituent atoms

$$\Omega \left(\mathbf{r}^N, \mathbf{p}^N, q^M, p_q^M; \ell \right) \xrightarrow[\int dq^M dp_q^M]{} \Omega' \left(\mathbf{r}^N, \mathbf{p}^N; \ell \right). \tag{56}$$

The adjustable parameter is chosen in such a way that the marginalised microcanonical partition function is equal to the partition function of interest (for instance, the canonical (constant volume) partition function)

$$\Omega \left(\mathbf{r}^N, \mathbf{p}^N, q^M, p_q^M; \ell \right) \xrightarrow[\int dq^M dp_q^M]{} \Omega' \left(\mathbf{r}^N, \mathbf{p}^N; \ell \right). \tag{57}$$

In the next section, we further clarify these notions by applying the general approach to the canonical ensemble.

4.2. Molecular Dynamics at NVT. In this section we outline the method to obtain the Hamiltonian that describes the canonical ensemble. This Hamiltonian may be substituted in the Liouville operator to obtain a finite difference equation. The modified Hamiltonian, called the Nosé-Hoover chain Hamiltonian [29, 60–62], is defined as

$$\mathscr{H}_{\text{NHC}} = \mathscr{H} \left(\mathbf{r}^N, \mathbf{p}^N \right) + \sum_{i=1}^M \frac{p_{q_i}^2}{2\mu_{q_i}} + \ell \frac{q_1}{\beta} + \sum_{i=2}^M \frac{q_i}{\beta}, \tag{58}$$

where

$$\mathscr{H} \left(\mathbf{r}^N, \mathbf{p}^N \right) = \sum_{i=1}^{N-1} \frac{p_i'^2}{2m_i'} + \frac{p_\odot^2}{2m_\odot} + \mathscr{U} \left(\mathbf{r}'^N \right) \tag{59}$$

is the Hamiltonian of the macromolecule in barycentric coordinates, the primed variables are the barycentric coordinates (relative to the centre of mass), and $\odot$ refers to the barycentre. If we assume that the energy of the modified Hamiltonian and the barycentre momentum are conserved (isolated system), we have, for the microcanonical partition function,

$$
\begin{aligned}
\Omega_{\text{NHC}} \\
= \Omega_0 \int dp_q^M dq^M d\mathbf{p}'^{N-1} dp_\odot d\mathbf{r}'^{N-1} \\
\times \exp\left[(3N-2)\,q_1 + \sum_{i=2}^{M} q_i\right] \delta\left(\mathcal{H}_{\text{NHC}} - C_1\right) \\
\times \delta\left(e^{q_1} p_\Omega - C_2\right).
\end{aligned}
\tag{60}
$$

If we marginalise the additional coordinate and momenta and choose $\ell = 3N$, the marginalised microcanonical partition function becomes equal to the canonical partition function, which means that the modified Hamiltonian describes the dynamics of the canonical ensemble

$$
\Omega'\left(\mathbf{r}^N, \mathbf{p}^N\right)_{\text{NHC}}\Big|_{\ell=3N} \equiv Q(N, V, T).
\tag{61}
$$

In the next section we focus on the constrained molecular dynamics.

4.3. Constrained Molecular Dynamics or How to Reduce the Computational Complexity. Macromolecules, such as the influenza hemagglutinin, possess many degrees of freedom. A typical influenza hemagglutinin is constituted of approximately 23.000 atoms, which means the macromolecules have typically 138.000 degrees of freedom in phase space (atomic positions and momenta). Unsurprisingly, the number of degrees of freedom increases dramatically if the structure is solvated. To reduce the computational complexity, it may be advantageous to impose constraints on certain degrees of freedom. Since the hydrogen atoms are light, they tend to follow the motion of heavier atoms quasi-instantaneously. Consequently, it is customary to fix their bonding length to reduce the computational burden. Let us assume that we have K holonomic (that depends only on time and coordinates) constraints:

$$
\sigma_k(t) = \left\|\mathbf{r}_i(t) - \mathbf{r}_j(t)\right\|^2 - d_k^2 \equiv 0.
\tag{62}
$$

The constraints are enforced in the equations of motion with the method of Lagrange multipliers. Each constraint becomes a potential to which a variable is attached called a Lagrange multiplier λ_k:

$$
m_i \ddot{\mathbf{r}}_i(t) = -\frac{\partial \mathcal{U}\left(\mathbf{r}^N(t)\right)}{\partial \mathbf{r}_i} + \sum_{k=1}^{K} \lambda_k \frac{\partial \sigma_k}{\partial \mathbf{r}_i}\Bigg|_{i=1,\ldots,N}.
\tag{63}
$$

Since the constraints are holonomic, they must be enforced at all times:

$$
\sigma_k(t + \Delta t) \equiv 0.
\tag{64}
$$

If we perform a truncated Taylor development of the previous equation, we may demonstrate that such a condition remains valid if a correction is applied to the Lagrange multiplier at each time step:

$$
\widetilde{\lambda}_k(t + \Delta t) = \widetilde{\lambda}_k(t) + \delta\widetilde{\lambda}_k,
\tag{65}
$$

where $\widetilde{\lambda}_k(t) = (\Delta t^2/2)\lambda_k$ is the value of the Lagrange multiplier at time t, $\delta\widetilde{\lambda}_k$ is the correction, $\widetilde{\lambda}_k(t + \Delta t)$ is the value of the Lagrange multiplier at time $t + \Delta t$, and

$$
\mathbf{A}\delta\widetilde{\boldsymbol{\lambda}} \approx -\boldsymbol{\sigma}\left(\mathbf{r}^N(t + \Delta t)\right),
\tag{66}
$$

which are defined as

$$
\begin{aligned}
\mathbf{A} = \left[A_{lk}\right] = \left[\sum_{i=1}^{N} \frac{1}{m_i} \frac{\partial \sigma_l\left(\mathbf{r}^N(t + \Delta t)\right)}{\partial \mathbf{r}_i} \cdot \frac{\partial \sigma_k\left(\mathbf{r}^N(t)\right)}{\partial \mathbf{r}_i}\right], \\
\delta\widetilde{\boldsymbol{\lambda}} = \left[\delta\widetilde{\lambda}_k\right]; \qquad \boldsymbol{\sigma} = \left[\sigma_k\right].
\end{aligned}
\tag{67}
$$

In this the linear equation governing the corrections to the Lagrange multipliers in which $\mathbf{r}^N(t)$ are the *constrained* atomic coordinates at time t. Consequently, to obtain the correction, one must solve the linear equation associated with the correction.

Algorithms distinguish themselves by the approach they use to solve this equation. For the RATTLE algorithm, this equation is solved analytically [63]. For the SHAKE algorithm [64], only the diagonal elements of the $\mathbf{A}$ matrix are considered to reduce the complexity of the calculation, while for the MSHAKE algorithm [65] the whole matrix is solved with the LU decomposition. If the constraints are periodical, they must be expressed in terms of quaternions in which case the algorithm becomes QSHAKE [66].

In the next section, we consider a hybrid approach based on both molecular dynamics and Monte Carlo simulations.

4.4. Hybrid Monte Carlo Dynamics or How to Wisely Accelerate the Dynamics. In order to obtain realistic molecular dynamics simulations, the time increment must be kept small. Usually, the value of the time increment is chosen around 1 femtosecond (10^{-15} seconds), which is typically one to two orders of magnitude smaller than the time scale associated with the fastest molecular event, the bound-length vibration [48]. If one desires to explore slow molecular events, the time required to run the simulation may become rapidly prohibitive. For instance, the rotation of buried side chains takes about 10^{-4} to 1 second to complete, while an helix-coil transition may take as long as 10^4 seconds. The situation might be even worse with macromolecular docking. Naturally, the time increment could be increased but the corresponding trajectory becomes rapidly unphysical.

This problem may be partially solved with an approach called hybrid Monte Carlo (HMC) simulation [28, 67]. As indicated by its name, HMC is a combination of MD and MC. The time evolution of the macromolecule is calculated with standard MD but with larger time increments. The outcome

of a time iteration is accepted only if it is physically feasible. The feasibility of the outcome is assessed with a standard MC acceptance probability rule:

$$
\begin{aligned}
A \left(\mathbf{r}^N (t), \mathbf{p}^N (t) \longrightarrow \mathbf{r}^N (t + \Delta t), \mathbf{p}^N (t + \Delta t) \right) \\
= \min \left\{ 1, \exp \left[-\beta \left(\mathscr{H} \left(\mathbf{r}^N (t + \Delta t), \mathbf{p}^N (t + \Delta t) \right) \right. \right. \right. \\
\left. \left. \left. - \mathscr{H} \left(\mathbf{r}^N (t), \mathbf{p}^N (t) \right) \right) \right] \right\}.
\end{aligned}
\tag{68}
$$

Note that there is an importance subtlety associated with HMC. Because Newton's equations are time reversible (any trajectory may be reversed by inverting the direction of the motion), the transition probability must also be so [68]. That implies that the transition probability must be chosen so that

$$
\begin{aligned}
T \left(\mathbf{r}^N (t), \mathbf{p}^N (t) \longrightarrow \mathbf{r}^N (t + \Delta t), \mathbf{p}^N (t + \Delta t) \right) \\
= T \left(\mathbf{r}^N (t + \Delta t), -\mathbf{p}^N (t + \Delta t) \longrightarrow \mathbf{r}^N (t), -\mathbf{p}^N (t) \right).
\end{aligned}
\tag{69}
$$

Otherwise, the HMC may generate unphysical results.

In the next section, we explore other ways that may be used to assess slow processes.

4.5. Computational Alchemy or How to Virtually Explore Experimentally Hidden Processes. Molecular dynamics is not limited to real physical potentials and equilibrium processes. Introducing an external, nonphysical, time-dependent, and possibly position-dependent potential may accelerate the dynamics of a slow process. This is called steered molecular dynamics or computational alchemy [69]. For instance, a force may be applied to a hemagglutinin to study potential conformational changes or to a drug (ligand) to analyse the docking process in between a drug and a neuraminidase. An example of how this could be achieved is by applying a random force (both in terms of amplitude and direction) to the barycentre of the ligand [70]. If the displacement of the ligand is above a certain threshold, the force is reapplied without modification. Otherwise, a new random force is generated. The whole process is repeated until the desired outcome is obtained. Such an approach is often combined with an immersive virtual environment to better visualise the outcome of the simulation [69]. In the next section, we explore an alternative to molecular dynamics and Monte Carlo simulations, namely, stochastic dynamics.

4.6. Molecular Dynamics and Influenza. There have been numerous applications of molecular dynamics in studies related to the influenza virus. For instance, molecular dynamics has been utilised for rational drug design for the influenza neuraminidase [7]. It has been instrumental to illustrate that the electrostatic funnel directs binding of Tamiflu to the influenza N1 neuraminidase [8]. Using molecular dynamics, it has been revealed that HR1039 is a potent inhibitor of the 2009 A (H1N1) influenza neuraminidase [9]. In addition, molecular dynamics has been valuable to show the role played by the BM2 channel in proton conductance and drug resistance for the influenza virus B [10]. Using molecular

dynamics has made possible the design of inhibitor targeting drug-resistant mutants of the influenza A virus M2. Finally, conformational analysis of peptides and lipids of the influenza hemagglutinin fusion peptide in micelles and bilayer would not have been possible without the aid of molecular dynamics [14].

5. Langevin Dynamics, Self-Guide Langevin Dynamics, and Self-Guided Molecular Dynamics: Toward a Better Sampling of the Conformational Space

The dynamics of a macromolecular system is entirely determined by the potential $\mathscr{U}(\mathbf{r}^N)$ associated with the process. For computational and practical reasons, this potential is virtually always an approximation of the real physical potential. Stochastic (random) dynamics attempts to bridge the gap in between the real and the approximate potential. Stochastic dynamics does not attempt to define the real potential but a general correction, which is independent on the particular details of the real potential. In other words, stochastic dynamics attempt to take into account the neglected degrees of freedom to obtain more realistic simulations, especially in the case of conformational sampling [71].

In this framework, it is assumed that each particle is under the influence of the potential $\mathscr{U}(\mathbf{r}^N)$ and of a heat bath formed by the remaining $N - 1$ particles. Under general conditions using the Mori-Zwanzig theory [72] (one may obtain the same result by assuming that each particle is minimally coupled to a harmonic heat bath formed from the other particles), it is possible to demonstrate that the particles obey the generalised Langevin equation (GLE) [73]:

$$
\begin{aligned}
m_i \ddot{\mathbf{r}}_i (t) = -\frac{\partial \mathscr{U} \left(\mathbf{r}^N (t) \right)}{\partial \mathbf{r}_i} \\
- \int_0^t d\tau \dot{\mathbf{r}}_i (\tau) \gamma_i (t - \tau) + \xi_i (t) \bigg|_{i=1,\dots,N},
\end{aligned}
\tag{70}
$$

where $\gamma_i(t - \tau)$ is the memory kernel or dynamic friction and $\xi_i(t)$ is a stochastic term with mean:

$$
\langle \xi_i (t) \rangle = 0
\tag{71}
$$

and covariance:

$$
\langle \xi_i (0) \xi_j (t) \rangle = 2 m_i k_B \gamma_i T \delta (t) \delta_{ij} \mathbf{I}.
\tag{72}
$$

Consequently, the correction is formed of two terms: a friction term, which introduces an artificial viscosity, and a stochastic term, which takes into account the unknown nature of the correction. One should notice that the frictional term depends on the previous history of the trajectory (Markovian process). The friction term is important in obtaining realistic simulations, as it takes into account the viscosity of the solvent (a feature, which is absent from both MD and MC). If one assumes that the frictional term is

constant (no history), one obtains the celebrated Langevin equation (LE):

$$\dot{\mathbf{p}}_i(t) = -\frac{\partial \mathcal{U}\left(\mathbf{r}^N(t)\right)}{\partial \mathbf{r}_i} - \gamma_i \dot{\mathbf{p}}_i(t) + \boldsymbol{\xi}_i(t)\Bigg|_{i=1,\dots,N}. \qquad (73)$$

In that particular case, γ_i is simply called the frictional or damping term. The Langevin equation improves conformational sampling over standard molecular dynamics.

Conformational sampling may still be further improved if the trajectory history is reintroduced into the model [71]. To take history into account, a history-dependent guiding term $\boldsymbol{\Gamma}_i(t)$ is added to the Langevin equation:

$$\dot{\mathbf{p}}_i(t) = -\frac{\partial \mathcal{U}\left(\mathbf{r}^N(t)\right)}{\partial \mathbf{r}_i} - \gamma_i \dot{\mathbf{p}}_i(t) + \boldsymbol{\xi}_i(t) + \lambda_i \boldsymbol{\Gamma}_i(t)\Bigg|_{i=1,\dots,N}. \qquad (74)$$

This term could be defined in various manners. In self-guided Langevin dynamics (SGLD), the history-dependent guiding term is defined as the time average of the momentum over the last ℓ iterations:

$$\begin{aligned}
&{}^{\text{SGLD}}\boldsymbol{\Gamma}_i(t) \\
&= \gamma_i \overline{\mathbf{p}_i(t)}\Big|_{t \in [t-\ell\Delta t, t]} \\
&= \left(1 - \frac{\Delta t}{\ell \Delta t}\right)\boldsymbol{\Gamma}_i(t - \Delta t) + \frac{\Delta t}{\ell \Delta t}\gamma_i m_i \dot{\mathbf{r}}_i\left(t - \frac{\Delta t}{2}\right).
\end{aligned} \qquad (75)$$

When, for self-guided molecular dynamics (SGMD), the history-dependent guiding term is defined as the time average of the potential plus its self-time average,

$$\begin{aligned}
&{}^{\text{SGMD}}\boldsymbol{\Gamma}_i(t) \\
&= \overline{-\frac{\partial \mathcal{U}\left(\mathbf{r}^N(t)\right)}{\partial \mathbf{r}_i} + \lambda_i \boldsymbol{\Gamma}_i(t)}\Bigg|_{t \in [t-\ell\Delta t, t]} \\
&= \left(1 - \frac{\Delta t}{\ell \Delta t}\right)\boldsymbol{\Gamma}_i(t - \Delta t) \\
&\quad + \frac{\Delta t}{\ell \Delta t}\left(-\frac{\partial \mathcal{U}\left(\mathbf{r}^N(t)\right)}{\partial \mathbf{r}_i} + \lambda_i \boldsymbol{\Gamma}_i(t)\right).
\end{aligned} \qquad (76)$$

The guiding term is unphysical (as opposed to the memory kernel) and does not conserve energy. To reestablish energy conservation, an additional term $\varsigma(t)\mathbf{p}_i(t)$, proportional to the momentum, must be added to the SGLD equation:

$$\begin{aligned}
\dot{\mathbf{p}}_i(t) = {}&-\frac{\partial \mathcal{U}\left(\mathbf{r}^N(t)\right)}{\partial \mathbf{r}_i} - \gamma_i \dot{\mathbf{p}}_i(t) \\
&+ \boldsymbol{\xi}_i(t) + \lambda_i \boldsymbol{\Gamma}_i(t) - \varsigma(t)\mathbf{p}_i(t)\Big|_{i=1,\dots,N}.
\end{aligned} \qquad (77)$$

This term ensures the conservation of energy, providing that the running constant $\varsigma(t)$ obeys

$$\varsigma(t) = \frac{\sum_{i=1}^{N} \lambda_i \boldsymbol{\Gamma}_i(t) \cdot \dot{\mathbf{r}}_i(t)}{\sum_{i=1}^{N} \mathbf{p}_i(t) \cdot \dot{\mathbf{r}}_i(t)}. \qquad (78)$$

These equations may be easily integrated if they are recast under the form of Wiener processes [74]. The integration requires two independent Gaussian random variables.

Langevin dynamics has been used in many applications. For instance it has been instrumental to unravel the bilayer conformation of the fusion peptide of the influenza virus hemagglutinin [75]. Further, the use of Langevin dynamics aided researchers to characterise the loop dynamics and ligand recognition in human- and avian-type influenza neuraminidase [76].

In the next section, we address complexes stability such as an interaction between a drug and neuraminidase or that between monoclonal antibodies and hemagglutinin.

6. Free Energy, Binding, and Complexes or a Measure of Stability

The interaction in between a small molecule (drug) or an antibody with an antigen (hemagglutinin or neuraminidase) may result in a complex. It follows that, often, the stability of such a complex must be assessed. Indeed, if the complex is stable, the small molecule or the antibody may efficiently neutralise the corresponding antigen. Free energy [53] is a measure of such stability. Consequently, the calculation of free energy is addressed in the next section.

6.1. Thermodynamic Integration. The thermodynamics integration method [77] requires two states to determine the free energy: an unbound state $\mathcal{I}$ and a bound state $\mathcal{J}$. From them, a parametric potential is defined as a linear superposition of the two potentials:

$$\mathcal{U}\left(\mathbf{r}^N; \lambda\right) = (1 - \lambda)\mathcal{U}_{\mathcal{I}}\left(\mathbf{r}^N\right) + \lambda\mathcal{U}_{\mathcal{J}}\left(\mathbf{r}^N\right). \qquad (79)$$

A corresponding parametric partition function is also defined as follows:

$$Z(N, V, T; \lambda) = \int d\mathbf{r}^N \exp\left[-\beta\mathcal{U}\left(\mathbf{r}^N; \lambda\right)\right]. \qquad (80)$$

Such a procedure is called adiabatic switching. The parametric derivative of the free energy may be obtained from the definition of the free energy:

$$\frac{\partial F(N, V, T; \lambda)}{\partial \lambda} = -\frac{1}{\beta}\frac{\partial}{\partial \lambda}\ln Z(N, V, T; \lambda). \qquad (81)$$

It follows that the difference of free energy in between the bound and the unbound state may be evaluated from

$$F_{\mathcal{J}} - F_{\mathcal{I}} = \int_0^1 d\lambda \left\langle \frac{\partial \mathcal{U}\left(\mathbf{r}^N; \lambda\right)}{\partial \lambda} \right\rangle_{\mathcal{U}(\mathbf{r}^N; \lambda)}. \qquad (82)$$

In the next section, we introduce a method based on irreversible work.

6.2. Nonequilibrium Method. In this section we describe an approach inspired from steered molecular dynamics (Section 4.5). As opposed to thermodynamic integration,

the nonequilibrium method does not require the bound state. This state is reached as a result of the irreversible work effectuated by the synthetic time-dependent Hamiltonian. The nonequilibrium method is based on Jarzinski's equality and irreversible work [78]. Jarzinski's equality may be derived from a time-dependent Hamiltonian ensemble average and the Liouville theorem. From this equality, the difference in free energy in between a bound and an unbound state may be calculated from the ensemble average of the irreversible work applied to the system to bring the unbound state into a bound state:

$$F_{\mathcal{J}} - F_{\mathcal{I}} = -kT \ln \left\langle \exp\left[-\beta \mathcal{W}_{\mathcal{I}\mathcal{J}}\left(\mathbf{r}^N, \mathbf{p}^N\right)\right]\right\rangle_{\mathcal{I}}, \quad (83)$$

where $\mathcal{W}_{\mathcal{I}\mathcal{J}}(\mathbf{r}^N, \mathbf{p}^N)$, the irreversible work, is given by

$$\mathcal{W}_{\mathcal{I}\mathcal{J}}\left(\mathbf{r}^N, \mathbf{p}^N\right) = \int_{t_{\mathcal{I}}}^{t_{\mathcal{J}}} d\tau \frac{\partial \mathcal{H}_{\mathcal{I}\mathcal{J}}\left(\mathbf{r}^N, \mathbf{p}^N, t\right)}{\partial \tau}, \quad (84)$$

and where the time-dependent Hamiltonian $\mathcal{H}_{\mathcal{I}\mathcal{J}}(\mathbf{r}^N, \mathbf{p}^N, t)$ is equal to the sum of the unbound macromolecular Hamiltonian $\mathcal{H}_{\mathcal{I}}(\mathbf{r}^N, \mathbf{p}^N)$ and of an external arbitrary time-dependent potential $\mathcal{U}_{\mathcal{I}\mathcal{J}}(\mathbf{r}^N; t)$, which brings the macromolecular system from an unbound state $\mathcal{I}$ to a bound state $\mathcal{J}$:

$$\mathcal{H}_{\mathcal{I}\mathcal{J}}\left(\mathbf{r}^N, \mathbf{p}^N, t\right) = \mathcal{H}_{\mathcal{I}}\left(\mathbf{r}^N, \mathbf{p}^N\right) + \mathcal{U}_{\mathcal{I}\mathcal{J}}\left(\mathbf{r}^N; t\right). \quad (85)$$

The ensemble average of the irreversible work is performed with the unbound Hamiltonian:

$$\begin{aligned}
&\left\langle \exp\left[-\beta \mathcal{W}_{\mathcal{I}\mathcal{J}}\left(\mathbf{r}^N, \mathbf{p}^N\right)\right]\right\rangle_{\mathcal{I}} \\
&= \left(\int d\mathbf{p}^N d\mathbf{r}^N \exp\left[-\beta \mathcal{W}_{\mathcal{I}\mathcal{J}}\left(\mathbf{r}^N, \mathbf{p}^N\right)\right]\right. \\
&\qquad \left.\times \exp\left[-\beta \mathcal{H}_{\mathcal{I}}\left(\mathbf{r}^N, \mathbf{p}^N\right)\right]\right) \\
&\quad \times \left(Z_{\mathcal{I}}\left(N, V, T\right)\right)^{-1}.
\end{aligned} \quad (86)$$

The next two sections present methods, which evaluate the free energy based on a subset of their generalised coordinates.

6.3. Blue Moon Ensemble Approach.

Thermodynamic integration and the nonequilibrium method both exploit all the coordinates and momenta associated with the macromolecular system to evaluate the free energy. Correspondingly, one may evaluate the free energy from a subset called the reaction coordinates (e.g., bound length and bound angle) of the generalised coordinates such as the generalised coordinates directly associated with the binding process. This considerably reduces the complexity of the calculation but introduces a certain level of approximation, as most coordinates are neglected. In order not to clutter the notation, we will use only one reaction coordinate while keeping in mind that the generalisation of more than one coordinates is immediate.

The probability that a reaction coordinate q_α has a value s is

$$\Pr(s) = \frac{\int d\mathbf{p}^N d\mathbf{r}^N \exp\left[-\beta \mathcal{H}\right] \delta\left(q_\alpha\left(\mathbf{r}^N\right) - s\right)}{Z(N, V, T)}, \quad (87)$$

where

$$q_\alpha = q_\alpha\left(\mathbf{r}^N\right) \quad (88)$$

is the representation of the reaction coordinates in terms of the Cartesian coordinates. The free energy associated with this coordinate is by definition

$$F(s) = -k_B T \ln \Pr(s). \quad (89)$$

If we marginalise (integrate) the momenta and if we express the potential in terms of the generalised coordinates, we obtain for the free energy [79]

$$\begin{aligned}
F(q) = F(s_i) \\
+ \int_{s_i}^{q} \left\langle \frac{\partial \mathcal{U}\left(\mathbf{r}^N\left(q^{3N}\right)\right)}{\partial q_\alpha} - k_B T \frac{\partial \ln \det \mathbf{J}}{\partial q_\alpha}\right\rangle_q ds,
\end{aligned} \quad (90)$$

where

$$\begin{aligned}
\langle O \rangle_q &= \frac{1}{Z(N, V, T)\left\langle \delta\left(q_\alpha - s\right)\right\rangle} \\
&\quad \times \int dq^{3N} O \\
&\quad \times \exp\left[-\beta\left(\mathcal{U}\left(\mathbf{r}^N\left(q^{3N}\right)\right)\right.\right. \\
&\qquad\qquad \left.\left. - k_B T \ln \det \mathbf{J}\right)\right] \delta\left(q_\alpha - s\right),
\end{aligned} \quad (91)$$

and where the Jacobian matrix of the transformation is

$$\mathbf{J} = \frac{\partial \mathbf{r}^N}{\partial q^{3N}}. \quad (92)$$

The main disadvantage of this approach is that all Cartesian coordinates must be converted into generalised coordinates, although only a subset of them (the reactions coordinates) are effectively exploited in the calculation of the free energy. This inefficiency is addressed in the next section.

6.4. Umbrella Sampling and Weighted Histogram Methods.

When using umbrella sampling [80, 81] and the weighted histogram method [82] for the blue moon ensemble approach, the free energy is estimated from the reaction coordinates. The range (codomain and image) associated with the reaction coordinates is divided into a set of windows or intervals. A reaction coordinate distribution is associated with each window. The unbiased estimator for the probability of occurrence of a reaction coordinate *within a given window* is assumed to be the product of a biased Gaussian probability distribution and a biased correction based on an umbrella potential:

$$\begin{aligned}
&\Pr(q) \\
&\approx \frac{1}{\sqrt{2\pi\sigma_k^2}} \exp\left[-\frac{(q - \bar{q}_k)^2}{2\sigma_k^2}\right] \exp\left[-\beta W\left(q, s^{(k)}\right)\right],
\end{aligned} \quad (93)$$

where

$$W\left(q, s^{(k)}\right) = \frac{1}{2}\kappa\left(q - s^{(k)}\right)^2 \qquad (94)$$

is the harmonic umbrella potential and $s^{(k)}$ is the centre of a given window, while $\bar{q}_k$ and σ_k^2 are, respectively, the average and the variance of all the realisations of the reaction coordinates within window k. By definition, the derivative of the free energy associated with a particular window is equal to

$$\frac{dF_k}{dq} = -\frac{k_B T}{\Pr(q)}\frac{d\Pr(q)}{dq} = \frac{k_B T}{\sigma_k^2}\left(q - \bar{q}_k\right) - \kappa\left(q - s^{(k)}\right). \qquad (95)$$

It is assumed that the derivative of the total free energy is a combination of the derivatives of the windows' free energy:

$$\frac{dF}{dq} = \sum_{k=1}^{K} C_k\left(q\right)\frac{dF_k}{dq}, \qquad (96)$$

and that the weighting coefficients are normalised

$$\sum_{k=1}^{K} C_k\left(q\right) = 1. \qquad (97)$$

The full free energy profile may be extracted with numerical integration. In the next section, we describe another approach for the calculation of the free energy, which introduces a bias potential to better sample the reaction coordinates.

6.5. Metadynamics. As opposed to the previous methods, which were based on a single molecular dynamics simulation, the metadynamics approach involved two molecular dynamics simulations: the usual one in Cartesian coordinates under the influence of the real physical potential and a second one, the meta-simulation, which is performed under the influence of a biased potential. The bias potential is defined as

$$\mathcal{U}_G\left(\mathbf{r}^N\right)$$
$$= W\sum_{t=\tau_G, 2\tau_G, \dots} \exp\left[-\sum_{\alpha=1}^{M}\frac{\left(q_\alpha\left(\mathbf{r}^N(t)\right) - q_\alpha\left(\mathbf{r}_G^N(t)\right)\right)^2}{2\Delta q^2}\right], \qquad (98)$$

where W is a constant, τ_G is a time interval; $\mathbf{r}^N(t)$ is the time evolution (trajectory) of the complete set of Cartesian coordinates up to time t under the action of the real physical potential $\mathcal{U}(\mathbf{r}^N)$, and $\mathbf{r}_G^N(t)$ is the trajectory of the system under the influence of both the physical potential and the biased potential $\mathcal{U}(\mathbf{r}^N) + \mathcal{U}_G(\mathbf{r}^N)$. From an analysis based on the Langevin equation, it is possible to demonstrate that the free energy may be obtained from

$$F\left(q^M\right) \approx W\sum_{t=\tau_G, 2\tau_G, \dots} \exp\left[-\sum_{\alpha=1}^{M}\frac{\left(q_\alpha - q_\alpha\left(\mathbf{r}_G^N(t)\right)\right)^2}{2\Delta q^2}\right], \qquad (99)$$

in which the reaction coordinates are expressed in a function of the metacoordinates. The proof of this equation is beyond the scope of this review and may be found in [83]. As for umbrella sampling, only the reaction coordinates need to be expressed in terms of the Cartesian coordinates.

The use of free energy calculations has provided insights into the susceptibility of antiviral drugs against the E119G mutant of the 2009 influenza A (H1N1) neuraminidase [26]. It has also been utilized to show the impact of calcium on the N1 influenza neuraminidase [25].

In the next section, we explain how molecular dynamics may be combined with macromolecular docking to assert the dynamics of a macromolecular complex, such as a viral hemagglutinin, when interacting with a broadly neutralizing monoclonal antibody.

7. Binding and Docking or Computational Pharmacology and Vaccinology

From the outset, one may assume that the relative pose of the constituent of a complex may be obtained through molecular dynamics without resorting to macromolecular docking. Although this is, in principle, feasible, it is usually not realistic in practice. Indeed, the time scale associated with macromolecular docking is usually out of reach for most molecular dynamics simulations [48]. Larger time steps may be taken if a hybrid Monte Carlo approach is followed (Section 4.4), but most of the time, the increase is insufficient. For this reason macromolecular docking remains the favourite method to determine the relative pose in between a receptor and a ligand. (Note that macromolecular docking is beyond the scope of this paper, but a comprehensive algorithmic review may be found in [84]).

The abovementioned observation does not imply that MD is irrelevant for macromolecular docking. On the contrary, once the pose in between the receptor and the ligand has been determined through macromolecular docking, the dynamics of the resulting complex may be explored with MD simulations. Various dynamical aspects of the complex may be studied. These include the study of the conformational space associated with the receptor [85], binding path analysis [56], examining the structural changes associated with induced fit, studying the flexibility of the docking region, analysing transient binding, and aiding in structural refinement of the complex [86] as well as the design of drugs [87] and antibodies (see Figure 2) with optimal kinetic properties. MD may also be employed prior to macromolecular docking to generate conformations, which become the initial point for static docking. This is particularly beneficial if the ligand binds to conformations that rarely occur (relax complex method). Nevertheless, standard MD should not be utilised to explore low frequency motions associated with large conformational transformations. This is due to its limited ability to explore conformational change. Langevin dynamics may be more suited in that particular case. Also, from a computational point of view, multiple copies of the ligand may simultaneously interact with the receptor to search, in parallel, for the best docking region. There should be no
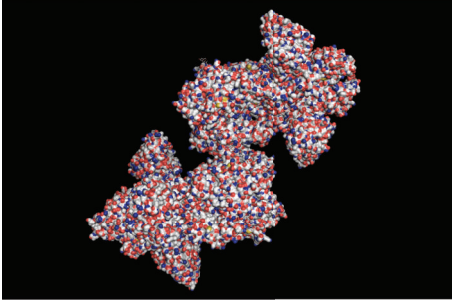
FIGURE 2: Molecular surface of influenza hemagglutinin (A/Viet Nam/1203/2004 (H5N1)) in complex with a broadly neutralizing antibody F10.

interaction in between the copies. Molecular dynamics may also be utilised to evaluate the structural accessibility of a highly conserved epitope [4].

In the next section, we briefly explore the relationship in between quantum mechanics and molecular dynamics.

8. Path Integral or When Quantum Mechanics Meets Molecular Dynamics

The calculations that we have performed so far are based on classical and semiclassical approximations of quantum mechanics. Though being valid, these approximations are not as accurate as their quantum-mechanical counterpart [88]. However, quantum-mechanical (QM) methods are often computationally prohibitive. For that reason, quantum-mechanical methods are restricted to a small number of coordinates of interest (reaction coordinates) and the residual coordinates are analysed with molecular dynamics (the so-called hybrid quantum-mechanical molecular dynamics). Quantum mechanical methods are outside the scope of this review. More details may be found in [88]. Here, we limit ourselves to the connection in between quantum-mechanical methods and molecular dynamics. Quantum-mechanical methods are also based on the partition function formalism, which in the quantum case takes the form:

$$Q(N, V, T) = \text{tr} \exp\left[-\beta \widehat{\mathcal{H}}\right], \tag{100}$$

where tr is the trace operator and $\widehat{\mathcal{H}}$ is the quantum Hamiltonian operator, which may be obtained from the classical Hamiltonian as follows:

$$\mathcal{H}\left(\mathbf{r}_i, \frac{\hbar}{i} \frac{\partial}{\partial \mathbf{r}_i}\right) \longrightarrow \widehat{\mathcal{H}}. \tag{101}$$

By extensively using the following identities:

$$\int dx \, |x\rangle \langle x| = \widehat{I}; \qquad \int dp \, |p\rangle \langle p| = \widehat{I};$$

$$\langle x \mid p \rangle = \exp\left[\frac{\hbar}{i} px\right], \tag{102}$$

and by making a certain number of approximations (Born-Oppenheimer approximation, Boltzmann statistics, and adiabatic approximation [83], etc.), it may be demonstrated that the partition function

$$Q(N, V, T) = \oint \mathscr{D}\mathbf{r}_1(\tau) \cdots \mathscr{D}\mathbf{r}_N(\tau)$$

$$\times \exp\left[-\frac{1}{\hbar} \int_0^{\beta\hbar} d\tau \frac{1}{2} \sum_{i=1}^{N} m_i \dot{\mathbf{r}}_i^2(\tau) + \mathscr{U}\left(\mathbf{r}^N(\tau)\right)\right] \tag{103}$$

is equal to the path integral over all possible closed paths of the exponential of the classical Hamiltonian. The notation $\mathscr{D}\mathbf{r}_i(\tau)$ means that we must integrate $\mathbf{r}_i(\tau)$ over all possible paths or trajectories. The resulting approach is known as path integral molecular dynamics [89]. Equation (103) involves only the atomics coordinates, but it is possible to introduce the conjugate momenta without altering the partition function by adding a harmonic potential to the Hamiltonian:

$$Q(N, V, T)$$

$$= \oint \mathscr{D}\mathbf{r}_1(\tau) \cdots \mathscr{D}\mathbf{r}_N(\tau) \mathscr{D}\mathbf{p}_1(\tau) \mathscr{D}\mathbf{p}_N(\tau)$$

$$\times \exp\left[-\frac{1}{\hbar} \int_0^{\beta\hbar} d\tau \sum_{i=1}^{N} \frac{\mathbf{p}_i^2}{2m_i'}(\tau)\right] \tag{104}$$

$$\times \exp\left[-\frac{1}{\hbar} \int_0^{\beta\hbar} \sum_{i=1}^{N} \frac{1}{2} m_i \omega_K^2 \mathbf{r}_i^2(\tau) + \mathscr{U}\left(\mathbf{r}^N(\tau)\right)\right],$$

where

$$m_i' = \frac{m_i K}{(2\pi\hbar)^2}; \quad \omega_K = \frac{\sqrt{K}}{\beta\hbar}; \quad K \longrightarrow \infty. \tag{105}$$

The two equations are entirely equivalent. If we develop the notation, we obtain

$$Q(N, V, T)$$

$$\approx \int \prod_{i=1}^{N} d\mathbf{r}_i^{(1)} \cdots d\mathbf{r}_i^{(K)} d\mathbf{p}_i^{(1)} \cdots d\mathbf{p}_i^{(K)}$$

$$\times \exp\left[-\beta \sum_{k=1}^{K} \sum_{i=1}^{N} \frac{\left(\mathbf{p}_i^{(k)}\right)^2}{2m_i'} + \frac{1}{2} m_i \omega_K^2 \left(\mathbf{r}_i^{(k+1)} - \mathbf{r}_i^{(k)}\right)^2 \right.$$

$$\left. + \frac{1}{K} \mathscr{U}\left(\mathbf{r}_1^{(k)} \cdots \mathbf{r}_N^{(k)}\right)\right]\Bigg|_{\mathbf{r}_i^{(K+1)} = \mathbf{r}_i^{(1)} \forall i}. \tag{106}$$

This path integral may be solved with the following infinite set of molecular dynamics equations:

$$\dot{\mathbf{r}}_i^{(k)} = \frac{\mathbf{p}_i^{(k)}}{m_i'}\bigg|_{k=1,\ldots,\infty},$$

$$\dot{\mathbf{p}}_i^{(k)} = -m_i\omega_K^2\left(2\mathbf{r}_i^{(k)} - \mathbf{r}_i^{(k+1)} - \mathbf{r}_i^{(k-1)}\right) \tag{107}$$

$$-\frac{1}{K}\frac{\partial \mathscr{U}\left(\mathbf{r}_1^{(k)}\cdots\mathbf{r}_N^{(k)}\right)}{\partial \mathbf{r}_i^{(k)}}\bigg|_{k=1,\ldots,\infty}.$$

Consequently, QM (at the approximation level indicated earlier) may be formulated in terms of an infinite (in practice, large) number of MD equations.

These equations may also be solved with an MC approach in which the position is sampled from a Gaussian distribution (the harmonic distribution in the path integral), and the new position is accepted or rejected according to a Metropolis rule based on the potential $\mathscr{U}$. The sampling is made difficult by the fact that the positions are not statistically independent. To make them independent, the positions are expended in terms of their Fourier transform or their staging variables. When the staging variables are used, the partition function becomes

$$Q\left(N, V, T\right)$$

$$\approx \int \prod_{i=1}^{N} d\mathbf{u}_i^{(1)}\cdots d\mathbf{u}_i^{(K)} d\mathbf{p}_i^{(1)}\cdots d\mathbf{p}_i^{(K)}$$

$$\times \exp\left[-\beta\sum_{k=1}^{K}\sum_{i=1}^{N}\frac{\left(\mathbf{p}_i^{(k)}\right)^2}{2m_i^{\prime(k)}} + \frac{1}{2}m_i^{(k)}\omega_K^2\left(\mathbf{u}_i^{(k)}\right)^2\right.$$

$$\left.+\frac{1}{K}\mathscr{U}\left(\mathbf{r}_1^{(k)}\left(\mathbf{u}_1\right),\ldots,\mathbf{r}_N^{(k)}\left(\mathbf{u}_N\right)\right)\right]\bigg|_{\mathbf{u}_i^{(K+1)}=\mathbf{u}_i^{(1)}\forall i},$$

$$\tag{108}$$

where

$$\mathbf{r}_i^{(1)} = \mathbf{u}_i^{(1)},$$

$$\mathbf{r}_i^{(k)} = \mathbf{u}_i^{(1)} + \sum_{l=k}^{K}\frac{k-1}{l-1}\mathbf{u}_i^{(k)}, \quad k = 2,\ldots,K,$$

$$\tag{109}$$

$$m_i^{(1)} = 0; \quad m_i^{(k)} = \frac{k}{k-1}m_i, \quad k = 2,\ldots,K,$$

$$m_i^{\prime(1)} = m_i; \qquad m_i^{\prime(k)} = m_i^{(k)}.$$

As a result, each normal mode $\mathbf{u}_i^{(k)}$ may be sampled from an independent Gaussian distribution and the positions may be reconstructed from their corresponding staging transformations. For instance, path integrals have been instrumental to underline the importance of quantum effects in the influenza neuraminidase [90].

9. Conclusions

The conformational space and the dynamics associated with macromolecules and macromolecular complexes, as well as their structural stability, may be explored and asserted through molecular dynamics and Monte Carlo simulations. Although full knowledge of the algorithms involved is not absolutely required, to obtain meaningful simulations, it is impressible to understand the conditions, approximations, and hypotheses on which they are based, as well as their limitations and potentialities.

Although most simulations are currently limited to individual structures or small organisms and complexes (a few millions atoms), it is conceivable that in the near future large viruses, such as the influenza, may be entirely simulated in silico over long periods of time. That would open the door to an essentially in silico design and lead to a better assessment of toxicity, specificity, and efficiency. Such a phenomenon has occurred before in aerodynamics in which computational fluid dynamics has essentially replaced wind tunnel experiments in the design of large aeroplane fuselages and wings.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

[1] G. Morra and G. Colombo, "Relationship between energy distribution and fold stability: insights from molecular dynamics simulations of native and mutant proteins," *Proteins: Structure, Function and Genetics*, vol. 72, no. 2, pp. 660–672, 2008.

[2] R. Alhadeff, D. Assa, P. Astrahan, M. Krugliak, and I. T. Arkin, "Computational and experimental analysis of drug binding to the Influenza M2 channel," *Biochimica et Biophysica Acta—Biomembranes*, vol. 1838, no. 4, pp. 1068–1073, 2014.

[3] H. Dong, G. Fiorin, W. F. Degrado, and M. L. Klein, "Exploring histidine conformations in the M2 channel lumen of the influenza a virus at neutral pH via molecular simulations," *Journal of Physical Chemistry Letters*, vol. 4, no. 18, pp. 3067–3071, 2013.

[4] C. F. Reboul, G. R. Meyer, B. T. Porebski, N. A. Borg, and A. M. Buckle, "Epitope flexibility and dynamic footprint revealed by molecular dynamics of a pMHC-TCR complex," *PLoS Computational Biology*, vol. 8, no. 3, Article ID e1002404, 2012.

[5] B. Tarus, C. Chevalier, C. A. Richard, B. Delmas, C. Primo, and A. Slama-Schwok, "Molecular dynamics studies of the nucleoprotein of influenza a virus: role of the protein flexibility in RNA binding," *PLoS ONE*, vol. 7, no. 1, Article ID e30038, 2012.

[6] R. E. Amaro, R. V. Swift, L. Votapka, W. W. Li, R. C. Walker, and R. M. Bush, "Mechanism of 150-cavity formation in influenza neuraminidase," *Nature Communications*, vol. 2, no. 1, article 388, 2011.

[7] L. Le, "Incorporating molecular dynamics simulations into rational drug design: a case study on influenza a neuraminidases," in *Bioinformatics*, H. Pérez-Sánchez, Ed., chapter 7, pp. 159–184, InTech, 2012.

[8] L. Le, E. H. Lee, D. J. Hardy, T. N. Truong, and K. Schulten, "Molecular dynamics simulations suggest that electrostatic funnel directs binding of Tamiflu to influenza N1 neuraminidases," *PLoS Computational Biology*, vol. 6, no. 9, Article ID e1000939, 2010.

[9] Y.-T. Wang, C.-H. Chan, Z.-Y. Su, and C.-L. Chen, "Homology modeling, docking, and molecular dynamics reveal HR1039 as a potent inhibitor of 2009 A(H1N1) influenza neuraminidase," *Biophysical Chemistry*, vol. 147, no. 1-2, pp. 74–80, 2010.

[10] Y. Zhang, H. Shen, M. Zhang, and G. Li, "Exploring the proton conductance and drug resistance of BM2 channel through molecular dynamics simulations and free energy calculations at different pH conditions," *Journal of Physical Chemistry B*, vol. 117, no. 4, pp. 982–988, 2013.

[11] N.-J. Deng, W. Zheng, E. Gallicchio, and R. M. Levy, "Insights into the dynamics of HIV-1 protease: a kinetic network model constructed from atomistic simulations," *Journal of the American Chemical Society*, vol. 133, no. 24, pp. 9387–9394, 2011.

[12] J. Wang, C. Ma, G. Fiorin et al., "Molecular dynamics simulation directed rational design of inhibitors targeting drug-resistant mutants of influenza A virus M2," *Journal of the American Chemical Society*, vol. 133, no. 32, pp. 12834–12841, 2011.

[13] S. S. Chang, H. J. Huang, and C. Y. C. Chen, "High performance screening, structural and molecular dynamics analysis to identify H1 inhibitors from TCM Database@Taiwan," *Molecular BioSystems*, vol. 7, no. 12, pp. 3366–3374, 2011.

[14] P. Lagüe, B. Roux, and R. W. Pastor, "Molecular dynamics simulations of the influenza hemagglutinin fusion peptide in micelles and bilayers: conformational analysis of peptide and lipids," *Journal of Molecular Biology*, vol. 354, no. 5, pp. 1129–1141, 2005.

[15] C. Siebena, R. Kappelb Zhuc, A. Wozniakd et al., "Influenza virus binds its host cell using multiple dynamic interactions," *Proceedings of the National Academy of Sciences*, vol. 109, no. 34, pp. 13626–13631, 2012.

[16] Z. Yang, Y. Nie, G. Yang, Y. Zu, Y. Fu, and L. Zhou, "Synergistic effects in the designs of neuraminidase ligands: analysis from docking and molecular dynamics studies," *Journal of Theoretical Biology*, vol. 267, no. 3, pp. 363–374, 2010.

[17] T. Rungrotmongkol, M. Malaisree, N. Nunthaboot, P. Sompornpisut, and S. Hannongbua, "Molecular prediction of oseltamivir efficiency against probable influenza a (H1N1-2009) mutants: molecular modeling approach," *Amino Acids*, vol. 39, no. 2, pp. 393–398, 2010.

[18] D. Pan, H. Sun, C. Bai et al., "Prediction of zanamivir efficiency over the possible 2009 Influenza A (H1N1) mutants by multiple molecular dynamics simulations and free energy calculations," *Journal of Molecular Modeling*, vol. 17, no. 10, pp. 2465–2473, 2011.

[19] G. G. Dodson, D. P. Lane, and C. S. Verma, "Molecular simulations of protein dynamics: new windows on mechanisms in biology," *EMBO Reports*, vol. 9, no. 2, pp. 144–150, 2008.

[20] C. Sieben, C. Kappel, R. Zhu et al., "Influenza virus binds its host cell using multiple dynamic interactions," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 109, no. 34, pp. 13626–13631, 2012.

[21] D. Vlachakis, A. Karozou, and S. Kossida, "3D molecular modelling study of the H7N9 RNA-dependent RNA polymerase as an emerging pharmacological target," *Influenza Research and Treatment*, vol. 2013, Article ID 645348, 9 pages, 2013.

[22] C. Peri, P. Gagni, F. Combi et al., "Rational epitope design for protein targeting," *ACS Chemical Biology*, vol. 8, no. 2, pp. 397–404, 2013.

[23] D. R. Flower, K. Phadwal, I. K. MacDonald, P. V. Coveney, M. N. Davies, and S. Wan, "T-cell epitope prediction and immune complex simulation using molecular dynamics: state of the art and persisting challenges," *Immunome Research*, vol. 6, no. 2, article S4, 2010.

[24] D. Xu, E. I. Newhouse, R. E. Amaro et al., "Distinct glycan topology for avian and human sialopentasaccharide receptor analogues upon binding different hemagglutinins: a molecular dynamics perspective," *Journal of Molecular Biology*, vol. 387, no. 2, pp. 465–491, 2009.

[25] M. Lawrenz, J. Wereszczynski, R. Amaro, R. Walker, A. Roitberg, and J. A. McCammon, "Impact of calcium on N1 influenza neuraminidase dynamics and binding free energy," *Proteins*, vol. 78, no. 11, pp. 2523–2532, 2010.

[26] P. Pan, L. Li, Y. Li, D. Li, and T. Hou, "Insights into susceptibility of antiviral drugs against the E119G mutant of 2009 influenza A (H1N1) neuraminidase by molecular dynamics simulations and free energy calculations," *Antiviral Research*, vol. 100, no. 2, pp. 356–364, 2013.

[27] K. A. Fichthorn and W. H. Weinberg, "Theoretical foundations of dynamical Monte Carlo simulations," *The Journal of Chemical Physics*, vol. 95, no. 2, pp. 1090–1096, 1991.

[28] S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth, "Hybrid Monte Carlo," *Physics Letters B*, vol. 195, no. 2, pp. 216–222, 1987.

[29] G. J. Martyna, D. J. Tobias, and M. L. Klein, "Constant pressure molecular dynamics algorithms," *The Journal of Chemical Physics*, vol. 101, no. 5, pp. 4177–4189, 1994.

[30] M. Shu, Z. Lin, Y. Zhang, Y. Wu, H. Mei, and Y. Jiang, "Molecular dynamics simulation of oseltamivir resistance in neuraminidase of avian influenza H5N1 virus," *Journal of Molecular Modeling*, vol. 17, no. 3, pp. 587–592, 2011.

[31] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equation of state calculations by fast computing machines," *The Journal of Chemical Physics*, vol. 21, no. 6, pp. 1087–1092, 1953.

[32] D. Frenkel, G. C. A. M. Mooij, and B. Smit, "Novel scheme to study structural and thermal properties of continuously deformable molecules," *Journal of Physics: Condensed Matter*, vol. 4, no. 12, pp. 3053–3076, 1992.

[33] J. I. Siepmann and D. Frenkel, "Configurational-bias Monte Carlo: a new sampling scheme for flexible chains," *Molecular Physics*, vol. 75, pp. 59–70, 1992.

[34] P. Minary, M. E. Tuckerman, and G. J. Martyna, "Dynamical spatial warping: a novel method for the conformational sampling of biophysical structure," *SIAM Journal on Scientific Computing*, vol. 30, no. 4, pp. 2055–2083, 2007.

[35] T. Huber and W. F. van Gunsteren, "SWARM-MD: searching conformational space by cooperative molecular dynamics," *Journal of Physical Chemistry A*, vol. 102, no. 29, pp. 5937–5943, 1998.

[36] J. Zhang, "An effective simulated annealing refined replica exchange Markov chain Monte Carlo algorithm for the infectious disease model of H1N1 influenza pandemic," *World Journal of Modelling and Simulation*, vol. 7, no. 1, pp. 29–39, 2011.

[37] J. I. Siepmann and D. Frenkel, "Configurational-bias Monte Carlo: a new sampling scheme for flexible chains," *Molecular Physics*, vol. 75, pp. 59–70, 1992.

[38] J. P. Ulmschneider and W. L. Jorgensen, "Monte Carlo backbone sampling for polypeptides with variable bond angles and dihedral angles using concerted rotations and a Gaussian bias," *The Journal of Chemical Physics*, vol. 118, no. 9, pp. 4261–4271, 2003.

[39] K. Lindorff-Larsen, P. Maragakis, S. Piana, M. P. Eastwood, R. O. Dror, and D. E. Shaw, "Systematic validation of protein force fields against experimental data," *PLoS ONE*, vol. 7, no. 2, Article ID e32131, 2012.

[40] A. K. Harrisa, J. R. Meyersona, Y. Matsuokac et al., "Structure and accessibility of HA trimers on intact 2009 H1N1 pandemic influenza virus to stemregion-specific neutralizing antibodies," *Proceedings of the National Academy of Sciences*, vol. 110, pp. 4592–4597, 2009.

[41] S. Toxvaerd, "Hamiltonians for discrete dynamics," *Physical Review E*, vol. 50, no. 3, pp. 2271–2274, 1994.

[42] K. Vanommeslaeghe, E. Hatcher, C. Acharya et al., "CHARMM general force field: a force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields," *Journal of Computational Chemistry*, vol. 31, no. 4, pp. 671–690, 2010.

[43] B. R. Brooks, C. L. Brooks III, A. D. Mackerell Jr. et al., "CHARMM: the biomolecular simulation program," *Journal of Computational Chemistry*, vol. 30, no. 10, pp. 1545–1614, 2009.

[44] D. Baker and A. Sali, "Protein structure prediction and structural genomics," *Science*, vol. 294, no. 5540, pp. 93–96, 2001.

[45] A. Özen, T. Haliloğlu, and C. A. Schiffer, "Dynamics of preferential substrate recognition in HIV-1 protease: redefining the substrate envelope," *Journal of Molecular Biology*, vol. 410, no. 4, pp. 726–744, 2011.

[46] J. Nocedal and S. J. Wright, *Numerical Optimisation*, Springer, New York, NY, USA, 2006.

[47] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, pp. 671–680, 1983.

[48] R. O. Dror, M. O. Jensen, D. W. Borhani, and D. E. Shaw, "Exploring atomic resolution physiology on a femtosecond to millisecond timescale using molecular dynamics simulations," *Journal of General Physiology*, vol. 135, no. 6, pp. 555–562, 2010.

[49] M. Feig and C. L. Brooks III, "Recent advances in the development and application of implicit solvent models in biomolecule simulations," *Current Opinion in Structural Biology*, vol. 14, no. 2, pp. 217–224, 2004.

[50] W. Im, S. Bernèche, and B. Roux, "Generalized solvent boundary potential for computer simulations," *The Journal of Chemical Physics*, vol. 114, no. 7, pp. 2924–2937, 2001.

[51] E. Gallicchio and R. M. Levy, "GBNP, an analytic implicit solvent model suitable for molecular dynamics simulations and high-resolution modeling," *Journal of Computational Chemistry*, vol. 25, no. 4, pp. 479–499, 2004.

[52] M. W. Mahoney and W. L. Jorgensen, "A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions," *The Journal of Chemical Physics*, vol. 112, no. 20, pp. 8910–8922, 2000.

[53] C. Chipot and A. Pohorille, Eds., *Free Energy Calculations*, Springer, Berlin, Germany, 2007.

[54] T. J. Chung, *Computational Fluid Dynamics*, Cambridge University Press, New York, NY, USA, 2010.

[55] M. Tuckerman, B. J. Berne, and G. J. Martyna, "Reversible multiple time scale molecular dynamics," *The Journal of Chemical Physics B*, vol. 380, pp. 665–677, 1992.

[56] C. Dellago, P. G. Bolhuis, and P. L. Geissler, "Transition path sampling," in *Advances in Chemical Physics*, vol. 132, chapter 1, 2002.

[57] A. E. Garcia, "Large-amplitude non-linear motions in proteins," *Physical Review Letters*, vol. 68, Article ID 2696, 1992.

[58] M. E. Tuckerman, Y. Liu, G. Ciccotti, and G. J. Martyna, "Non-Hamiltonian molecular dynamics: Generalizing Hamiltonian phase space principles to non-Hamiltonian systems," *The Journal of Chemical Physics*, vol. 115, no. 4, pp. 1678–1702, 2001.

[59] G. S. Ezra, "On the statistical mechanics of non-Hamiltonian systems: the generalized Liouville equation, entropy, and time-dependent metrics," *Journal of Mathematical Chemistry*, vol. 35, no. 1, pp. 29–53, 2004.

[60] S. Nosé, "A unified formulation of the constant temperature molecular dynamics methods," *The Journal of Chemical Physics*, vol. 81, no. 1, pp. 511–519, 1984.

[61] Y. Liu and M. E. Tuckerman, "Generalized Gaussian moment thermostatting: a new continuous dynamical approach to the canonical ensemble," *The Journal of Chemical Physics*, vol. 112, no. 4, pp. 1685–1700, 2000.

[62] G. J. Martyna, M. L. Klein, and M. Tuckerman, "Nosé-Hoover chains: the canonical ensemble via continuous dynamics," *The Journal of Chemical Physics*, vol. 97, no. 4, pp. 2635–2643, 1992.

[63] H. C. Andersen, "Rattle: a "velocity" version of the shake algorithm for molecular dynamics calculations," *Journal of Computational Physics*, vol. 52, no. 1, pp. 24–34, 1983.

[64] J.-P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen, "Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes," *Journal of Computational Physics*, vol. 23, no. 3, pp. 327–341, 1977.

[65] V. Kraeutler, W. F. van Gunsteren, and P. H. Huenenberger, "A fast SHAKE algorithm to solve distance constraint equtions for small molecules in molecular dynamics simulations," *Journal of Computational Chemistry*, vol. 22, pp. 501–508, 2001.

[66] T. R. Forester and W. Smith, "SHAKE, rattle, and roll: efficient constraint algorithms for linked rigid bodies," *Journal of Computational Chemistry*, vol. 19, no. 1, pp. 102–111, 1998.

[67] J. A. Izaguirre and S. S. Hampton, "Shadow hybrid Monte Carlo: an efficient propagator in phase space of macromolecules," *Journal of Computational Physics*, vol. 200, no. 2, pp. 581–604, 2004.

[68] G. J. Martyna, M. E. Tuckerman, D. J. Tobias, and M. L. Klein, "Explicit reversible integrators for extended systems dynamics," *Molecular Physics*, vol. 87, no. 5, pp. 1117–1157, 1996.

[69] B. K. Mai, M. H. Viet, and M. S. Li, "Top leads for swine influenza A/H1N1 virus revealed by steered molecular dynamics approach," *Journal of Chemical Information and Modeling*, vol. 50, no. 12, pp. 2236–2247, 2010.

[70] M. Klvana, M. Pavlova, T. Koudelakova et al., "Pathways and mechanisms for product release in the engineered haloalkane dehalogenases explored using classical and random acceleration molecular dynamics simulations," *Journal of Molecular Biology*, vol. 392, no. 5, pp. 1339–1356, 2009.

[71] X. Wu and B. R. Brooks, "Self-guided Langevin dynamics simulation method," *Chemical Physics Letters*, vol. 381, no. 3-4, pp. 512–518, 2003.

[72] R. W. Zwanzig, "Nonlinear generalized Langevin equations," *Journal of Statistical Physics*, vol. 9, no. 3, pp. 215–220, 1973.

[73] S. Izvekov and G. A. Voth, "Modeling real dynamics in the coarse-grained representation of condensed phase systems," *Journal of Chemical Physics*, vol. 125, no. 15, Article ID 151101, 2006.

[74] E. Vanden-Eijnden and G. Ciccotti, "Second-order integrators for Langevin equations with holonomic constraints," *Chemical Physics Letters*, vol. 429, no. 1–3, pp. 310–316, 2006.

[75] Q. Huang, C.-L. Chen, and A. Herrmann, "Bilayer conformation of fusion peptide of influenza virus hemagglutinin: a molecular dynamics simulation study," *Biophysical Journal*, vol. 87, no. 1, pp. 14–22, 2004.

[76] R. E. Amaro, X. Cheng, I. Ivanov, D. Xu, and J. A. McCammon, "Characterizing loop dynamics and ligand recognition in human- and avian-type influenza neuraminidases via generalized born molecular dynamics and end-point free energy calculations," *Journal of the American Chemical Society*, vol. 131, no. 13, pp. 4702–4709, 2009.

[77] J. G. Kirkwood, "Statistical mechanics of fluid mixtures," *The Journal of Chemical Physics*, vol. 3, no. 5, pp. 300–313, 1935.

[78] S. Park, F. Khalili-Araghi, E. Tajkhorshid, and K. Schulten, "Free energy calculation from steered molecular dynamics simulations using Jarzynski's equality," *The Journal of Chemical Physics*, vol. 119, no. 6, pp. 3559–3566, 2003.

[79] M. Sprik and G. Ciccotti, "Free energy from constrained molecular dynamics," *The Journal of Chemical Physics*, vol. 109, no. 18, pp. 7737–7744, 1998.

[80] J. Kästner and W. Thiel, "Bridging the gap between thermodynamic integration and umbrella sampling provides a novel analysis method: umbrella integration," *Journal of Chemical Physics*, vol. 123, no. 14, Article ID 144104, 2005.

[81] G. M. Torrie and J. P. Valleau, "Nonphysical sampling distributions in Monte Carlo free-energy estimation: umbrella sampling," *Journal of Computational Physics*, vol. 23, no. 2, pp. 187–199, 1977.

[82] S. Kumar, D. Bouzida, R. H. Svendsen, P. A. Kollman, and J. M. Rosenber, "The weighted histogram analysis method for free-energy calculations on biomolecules. 1. The method," *Journal of Computational Chemistry*, vol. 13, p. 1011, 1992.

[83] A. Laio, A. Rodriguez-Fortea, F. L. Gervasio, M. Ceccarelli, and M. Parrinello, "Assessing the accuracy of metadynamics," *Journal of Physical Chemistry B*, vol. 109, no. 14, pp. 6714–6721, 2005.

[84] E. Paquet and H. L. Viktor, "Macromolecular structure comparison and docking: an algorithmic Review," *Current Pharmaceutical Design*, vol. 19, no. 12, pp. 2183–2193, 2013.

[85] M. Raab and I. Tvaroška, "The binding properties of the H5N1 influenza virus neuraminidase as inferred from molecular modeling," *Journal of Molecular Modeling*, vol. 17, no. 6, pp. 1445–1456, 2011.

[86] C. T. T. Su, C. Schönbach, and C.-K. Kwoh, "Molecular docking analysis of 2009-H1N1 and 2004-H5N1 influenza virus HLA-B*4405-restricted HA epitope candidates: implications for TCR cross-recognition and vaccine development," *BMC Bioinformatics*, vol. 14, supplement 2, article S21, 2013.

[87] H. Liu and X. Yao, "Molecular basis of the interaction for an essential subunit PA-PB1 in influenza virus RNA polymerase: Insights from molecular dynamics simulation and free energy calculation," *Molecular Pharmaceutics*, vol. 7, no. 1, pp. 75–85, 2010.

[88] M. Dominik and J. Hutter, *Ab Initio Molecular Dynamics: Basic Theory and Advanced Methods*, Cambridge University Press, Cambridge, UK, 2009.

[89] M. Ceriotti, M. Parrinello, T. E. Markland, and D. E. Manolopoulos, "Efficient stochastic thermostatting of path integral molecular dynamics," *Journal of Chemical Physics*, vol. 133, no. 12, Article ID 124104, 2010.

[90] F. Perrucio, L. Ridder, and A. J. Mulholland, "Quantum-mechanicl/molecular-mechanical methods in medicinal chemistry," in *Quamtum Medicinal Chemistry*, P. Carloni and F. Alber, Eds., Wilely, Weinheim, Germany, 2003.