[Frontiers in Bioscience 19, 578-604, January 1, 2014]

Molecular mechanics and dynamics: numerical tools to sample the configuration space

Matteo Masetti[1], Walter Rocchia[2]

[1]Department of Pharmacy and Biotechnology, Alma Mater Studiorum – Università di Bologna, via Belmeloro 6, 40126 Bologna, Italy, [2]Drug Discovery and Development, Istituto Italiano di Tecnologia, via Morego 30, 16163 Genova, Italy

TABLE OF CONTENTS

1. ABSTRACT

Molecular simulation is increasingly used in many theoretical as well applicative fields in both Life and Material sciences. It relies on the capability of estimating the forces acting both inter- and intra-molecularly, providing configurations of minimum energy and sampling the configurational space consistently with the main statistical ensembles. In this context, the engine that approximates the temporal evolution and the force field that expresses the instantaneously occurring forces, *i.e.* Molecular Dynamics and Molecular Mechanics, respectively, are the crucial bases that need to be known and handled. Here, the fundamentals of these tools are provided, with particular attention to numerical and simulative aspects.

2. INTRODUCTION: SAMPLING THE CONFIGURATIONAL SPACE

A correct sampling of the phase space, consistent with the current thermodynamic ensemble, is instrumental to the estimation of free energy differences and profiles that, in turn, are key quantities for the description of the physical processes of practical interest, such as protein-ligand binding (1). Different ensembles correspond to different experimental conditions and the most common are *microcanonical*, *canonical* and *isothermal-isobaric*. In the microcanonical ensemble (NVE), volume, number of particles and overall energy are conserved, in the canonical one (NVT) thermodynamic temperature is controlled in place of energy. Finally, in the isothermal-isobaric ensemble (NPT) temperature, number of particles and

pressure are controlled. For most applications, what is most important is to have an estimate of the probabilistic weight of the configurations of a system. For sake of clarity, we underline that we attribute to the term "configuration" its usual meaning in physics, that is an arrangement of N atoms in the Cartesian space described by the vector $\{x_1, y_1, z_1, x_2, y_2, z_2, \ldots, x_N, y_N, z_N\}$. This should not be confused with the meaning of "configuration" in chemistry, which refers to the absolute arrangements of atoms in a molecule provided by their chemical bonds.

Regardless of the specific ensemble, configurational sampling can be performed via Monte Carlo (MC) methods, by suitable integration of the phase space sampling done through molecular dynamics (MD), and by means of hybrid variants of the two.

The basic idea behind MC is to randomly generate configurations of the system compliant with the probability density function corresponding to the considered ensemble. MC moves can be performed in coordinate space only due to the possibility to separate potential from kinetic energy contributions to the partition function. Within the Metropolis formulation, the MC algorithm creates a Markov chain of states by making at each step a *trial move* (i.e. randomly drawing a new configuration state) and accepting or rejecting the move so as that, in the end, the final relative occupation of the states corresponds to the ratio of their statistical weights in the considered ensemble (2). Differently, MD methods attempt to solve the equations of motion supposedly mimicking the actual dynamics of a real system (3-4). The simplest set of equations are the Hamilton equations, where energy is conserved and therefore they are suited to sample the microcanonical ensemble. In order to recover the canonical distribution of microstates, or possibly to sample other ensembles, the equations of motion must be properly adapted. This may involve deterministic or stochastic corrections that account for the interaction with the environment. The algorithms that "correct" the Hamilton equations and that are used to sample ensembles other than the microcanonical one, are usually referred to as *thermostats* or *barostats* according to whether they aim at controlling the thermodynamic temperature or pressure, respectively. It should be noted that, under the ergodic hypothesis, when using MD the ensemble averages are actually calculated in the form of time averages (5). Similarly to what has been said about the MC method, and for the same reasons, also MD is most often used to sample points in the configurational space, rather than in the whole phase space. With some caveats, MD is also used to estimate correlation functions and corresponding observables such as the diffusion coefficient (6).

MD methods became much more popular than MC in the community dedicated to the simulation of biological systems. This is mainly due to the fact that, when dealing with a system of several tens of thousands atoms, the probability of randomly performing a trial move compatible with the correct ensemble is so low that MC becomes computationally unaffordable. In spite of this, the importance of the role of MC in computer assisted drug design should not be underestimated: virtually all the docking applications are founded on MC background, and relative binding free energies are routinely estimated by means of MC frameworks, also because they adopt a simplified representation, e.g. implicit solvent, with much less degrees of freedom. However, the present work will primarily deal with MD based methods.

It is worth pointing out that while, according to the principle of *ensemble equivalence*, in the thermodynamic limit – that is in the limit of an infinitely large system – the average values calculated for the basic thermodynamic properties are consistent among different ensembles, fluctuations (defined as the absolute deviation from the average value) are not. Thus, care must be taken in the choice of the proper statistical ensemble whenever properties depending upon *fluctuation averages* instead of the averages themselves are derived, and also when the simulated system is so small that the thermodynamic limit is far from being reached. A conceptually similar problem concerns the possibility to calculate entropy related properties, such as the free energy or the chemical potential, by making use of standard sampling techniques.

In this work, we will first provide the basic numeric algorithms underlying MD, then we will focus on *molecular mechanics*, that is on the models (*force fields*) that provide the interatomic forces used throughout the dynamics.

3. MD: NUMERICAL INTEGRATION OF THE EQUATIONS OF MOTIONS

The basic idea behind MD is to study the time evolution of a microscopic system by solving the Newton's equations of motion:

$$\mathbf{f}_i = m_i \ddot{\mathbf{r}}_i = -\frac{\partial V_{\mathrm{FF}}(\mathbf{r})}{\partial \mathbf{r}_i} \qquad 1)$$

where $\mathbf{f}_i$ is the net force acting on the i^{th} ($i = 1, \ldots, N$) atom having mass m_i, the bold face identifies vectors and tensors, and the n^{th} over-dot notation stands for n differentiations of positions over time. For convenience, while the formal description of the phase space envisions positions and momenta as independent variables, we will often consider the description in terms of positions and velocities. These are two equivalent approaches that lead to the Hamilton and the Eulero-Lagrange sets of equations, respectively. Eq. (1) represents a set of $3N$ coupled second order differential equations: once initial conditions are specified, positions and velocities for all the particles may be calculated by integration over time. Initial conditions for positions are often taken from experimental data, when available, such as crystallographic structures. The corresponding conditions for velocities are suitably drawn from their probability distribution, which is known for the most common statistical ensembles.

For isolated systems, the forces acting on the atoms are only due to their interactions with the other particles constituting the system itself. As we will detail

later on, in the classical MD simulation, the physics underlying the atomic interactions is described by means of a properly parameterized scalar potential energy function ($V_{FF}(\mathbf{r})$ in Eq. (1)), and the model arising by such a representation is referred to as the *force field*. Whenever the forces can be calculated as the gradient of such a potential energy function, the mechanical system is said to be *conservative* and the total energy (the sum of the potential and kinetic energy) is a constant of the motion. Since forces depend upon the particle's positions, Eq. (1) is an example of a many-body problem which can be analytically solved only for extremely simple systems. In any other case, numerical methods must be employed, unavoidably leading to approximate solutions with respect to the analytical trajectory. By using finite differences methods, the integration of the equations of motion is performed at discrete time intervals δt, called *time steps*, by means of algorithms referred to as *integrators*. The foundation for any integrator scheme is the Taylor expansion of the coordinates around the time:

$$\mathbf{r}_i(t \pm \delta t) = \mathbf{r}_i(t) \pm \dot{\mathbf{r}}_i(t)\delta t + \frac{1}{2}\ddot{\mathbf{r}}_i(t)\delta t^2 + \pm \frac{1}{6}\dddot{\mathbf{r}}_i(t)\delta t^3 + \mathcal{O}(\delta t^4) \quad (2)$$

For simplicity, in Eq. (2) the expansion is truncated at the fourth order. Similarly, the *order* of an integrator indicates the smallest term in the time step expansion not included explicitly. It represents a measure of the precision of the algorithm (higher order schemes introduce a lower amount of numerical error). The order is an important feature for an integrator, however when choosing an algorithm, many other intertwined aspects must be considered. From a purely mechanical perspective, the *time reversal symmetry* is the minimal requirement that should be satisfied. This means that at any point in time, if the momenta of all the particles change sign, the same trajectory must be covered back leading to the exact initial condition of positions and velocities. It is important to stress that this is not concerned with the finite precision which intrinsically affects numerical schemes: some algorithms satisfy the time reversibility requirement only in the limit of an infinitely small time step.

An ideal integrator should preserve conservation laws as well. If the Hamiltonian is a constant of motion, the ability of an integrator to preserve it is referred to as *stability*. Clearly, numerical precision and, most importantly, the magnitude of the time step play a critical role in determining the stability of an algorithm. At equal time steps, high order schemes naturally provide better short term energy conservation with respect to low order schemes, but this trend is not necessarily preserved at long time scales, since other aspects come into play (see below). Conversely, a stable algorithm is able to tolerate a more aggressive time stepping without returning a significant energy drift. The latter feature is often referred to as the *robustness* of an algorithm. Finally, the *accuracy* of an integrator is defined as the divergence of the numerical trajectory from the analytical one, and it is determined by most of the factors considered so far. Due to the unavoidable introduction of errors during the numerical integration of the equations of motion, most of the algorithms used in MD lead to an exponential divergence of the calculated trajectory with respect to the analytical solution. This is a case of *Lyapunov instability*. Despite the fact that it is unfeasible to follow the "real" trajectory for a many-body system, such instability is not as problematic as it might appear since the ultimate aim is to generate sample configurations consistent with the equilibrium probability density so to be able to calculate average properties rather than depicting the detailed time evolution of a system.

Concerning the statistical mechanical aspects of the integration of the equations of motion, it must be stressed that, according to the Liouville theorem, in a real Hamiltonian dynamics the volume of phase space occupied by a "cloud" of replicas of the system is conserved. This means that the points in phase space corresponding to a given value of constant energy encompass a certain volume that is kept constant during the time evolution of the system. As a consequence, any integrator that does not preserve the phase space volume implicitly violates the energy conservation laws. Usually, non-reversible algorithms are also non-volume preserving, leading to long term energy drift issues. Integrators that are both time reversible and volume preserving are said to be *symplectic*, and symplecticity is actually the minimal requirement needed for an integrator to conserve energy.

In terms of *efficiency*, the calculation of the forces represents the real bottleneck for any MD code. Consequently, a well performing integrator is one involving the smallest number of force evaluations per iteration. Lastly, the choice of the time step is always dictated by a tradeoff between the accuracy of the numerical trajectory and computational resources. Clearly, to some extent, robust algorithms allow the use of larger time steps. However, as a general rule, to obtain acceptable energy conservation, the time step should be chosen to be at least one order of magnitude smaller than the fastest period of motion contemplated by the physics of the problem. When simulating biological systems, the highest frequency oscillations are represented by bonds involving hydrogen atoms, which have a period of approximately 10 fs. As a consequence, the time step used to perform biological simulations should be smaller than 1 fs, inevitably leading to a very hard sampling of the phase space. This is why some MD engines apply suitable constraints on hydrogen trajectories so as to avoid the calculation of the highest frequency forces, leading to the possibility of using a time step of 2 fs. In the following sections, the basic ingredients for some simple integrator schemes will be provided along with some notes covering the time step issue.

3.1. Discretization strategies

3.1.1. Verlet algorithm

The Verlet algorithm (or position-Verlet) represents the prototype for all the integrators based on the direct solution of the equations of motion. By summing the expressions ($t + \delta t$) and ($t - \delta t$) of the Taylor expansion

reported in Eq. (2) one obtains the Verlet equation for advancing positions (7):

$$\mathbf{r}_i(t+\delta t) = 2\mathbf{r}_i(t) - \mathbf{r}_i(t-\delta t) + \frac{\mathbf{f}_i(\mathbf{r}(t))}{m_i}\delta t^2 + \\ + \mathcal{O}(\delta t^4) \quad 3)$$

As it may be noticed from Eq. (3), velocities are not explicitly present in the Verlet algorithm. Still, they are needed to calculate the kinetic energy and other properties depending on them. By subtracting the expressions of the Taylor expansions of Eq. (2), it is possible to derive a useful relation to recover velocities:

$$\dot{\mathbf{r}}_i(t) = \frac{\mathbf{r}_i(t+\delta t) - \mathbf{r}_i(t-\delta t)}{2\delta t} + \mathcal{O}(\delta t^2) \quad 4)$$

However, from Eq. (4) it is clear that velocities at time t are available only once the positions have reached the next time step ($t + \delta t$). Beyond its simplicity, the Verlet algorithm has many appealing features: it is time reversible, stable (then it is also symplectic) and moderately robust. The main drawbacks of the Verlet algorithm are both the relatively poor precision (in the right hand side of Eq. (3) terms of different order in the time-step expansion are mixed together), and the definitely uncomfortable handling of velocities.

3.1.2. Leap-frog

The idea behind the leap-frog algorithm is to explicitly handle velocities, and it is based on the possibility to evaluate them at each half time step. From the derivation of the Verlet algorithm (Eq. (3)), after calculation of the forces at time t, the velocities may be updated according to (8)

$$\dot{\mathbf{r}}_i(t+\delta t/2) = \dot{\mathbf{r}}_i(t-\delta t/2) + \ddot{\mathbf{r}}_i(t)\delta t \quad 5)$$

then, the positions are advanced by using the following relation derived from Eq. (5):

$$\mathbf{r}_i(t+\delta t) = \mathbf{r}_i(t) + \dot{\mathbf{r}}_i(t+\delta t/2)\delta t \quad 6)$$

An advantage of the leap-frog algorithm over the Verlet integrator is a greater precision. However, the calculated positions and velocities are shifted by a factor $\delta t/2$. In order to calculate the total energy, the current velocities at time t must be re-derived:

$$\dot{\mathbf{r}}_i(t) = \frac{\dot{\mathbf{r}}_i(t+\delta t/2) + \dot{\mathbf{r}}_i(t-\delta t/2)}{2} \quad 7)$$

3.1.3. Velocity-Verlet

The velocity-Verlet algorithm is an integrator able to store positions, velocities and accelerations at the same time t (9). Actually, velocities are evaluated at each half time step as in the leap-frog algorithm, but since they are evaluated twice per iteration, the synchronization with positions is ensured. The advancement in position directly follows the Taylor expansion of Eq. (2):

$$\mathbf{r}_i(t+\delta t) = \mathbf{r}_i(t) + \dot{\mathbf{r}}_i(t)\delta t + \frac{1}{2}\ddot{\mathbf{r}}_i(t)\delta t^2 \quad 8)$$

Then, the velocities are propagated according to:

$$\dot{\mathbf{r}}_i(t+\delta t/2) = \dot{\mathbf{r}}_i(t) + \frac{1}{2}\ddot{\mathbf{r}}_i(t)\delta t \quad 9)$$

$$\dot{\mathbf{r}}_i(t+\delta t) = \dot{\mathbf{r}}_i(t+\delta t/2) + \frac{1}{2}\ddot{\mathbf{r}}_i(t+\delta t)\delta t \quad 10)$$

In between the velocities half steps, the evaluation of the forces at time ($t + \delta t$) is performed. Then, the net velocity update is obtained:

$$\dot{\mathbf{r}}_i(t+\delta t) = \dot{\mathbf{r}}_i(t) + [\ddot{\mathbf{r}}_i(t) + \ddot{\mathbf{r}}_i(t+\delta t)]\,\delta t/2 \quad 11)$$

The velocity-Verlet integrator is stable and well behaved, and, in addition, high order variants have been developed in order to increase its precision.

All the Verlet-like integrators considered are efficient algorithms, involving just one force evaluation per iteration. A scheme summarizing their details and differences in propagating positions, velocities and accelerations is depicted in Figure 1.

It is clear that, especially in case of large scale systems, using a time step smaller than 1 fs makes extremely long, if not impossible, an efficient sampling of the phase space, thus severely hampering the obtainment of converged ensemble averages estimates in a reasonable amount of computer time. Without claiming neither any completeness nor a thorough description (that would be out of the purpose of the present work), it is nonetheless worth giving the reader some flavor about some among the most widely used computational solutions aimed at increasing the simulated time at a given computational cost.

The first solution comes from the observation that in conventional MD simulations, most of the time is spent to calculate the forces between every pair of atoms composing the system. The simplest way to deal with this problem in case of short range potentials is to use distance cutoffs in order to limit the force evaluation to the nearest particles. However, to do so the information on which particles are located within the cutoff radius is needed, and the effort to calculate it at each time step is almost comparable to that required to perform the full force evaluation. By using the Verlet neighbor list the cutoff radius r_c is surrounded by an additional layer r_l, and the list of the neighboring atoms is updated only every 10 or 20 time steps (7). Thus, between updates, only the atoms included within the distance corresponding to the external layer are checked, hence saving computational time. The method is reliable as long as the external layer is large enough so that between consecutive updates no atom enters nor leaves the cutoff sphere where the interactions are calculated. In other words, during the update period, the particles displacements must be always smaller than $|r_l - r_c|$.

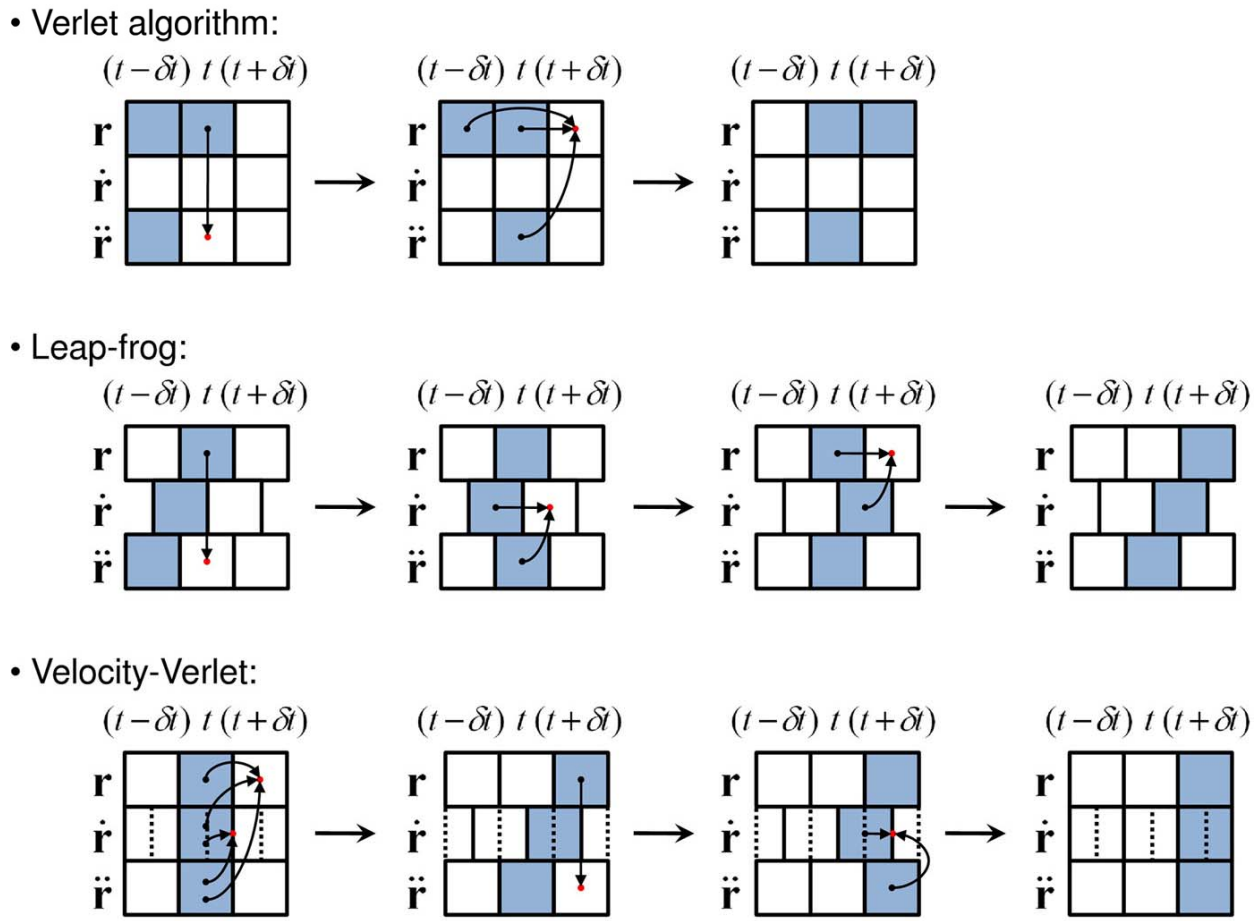

Figure 1. Pictorial scheme of the advancement in position, velocities and accelerations for the Verlet-like integrator algorithms (adapted from (5)). The available, stored, variables at a given MD step are shown as blue boxes, whereas red dots represent the quantities to be calculated. The connections between these latter and those required to perform the calculation are depicted by arrows. In the first step of the *Verlet integrator* cycle, positions at time t and $(t - \delta t)$ are available, as well as accelerations at time $(t - \delta t)$. From positions at time t, forces and, consequently, accelerations are calculated. The cycle is advanced by updating the acceleration array. Then, also positions are progressed according to Eq. (3), leading to the new starting configuration for the next cycle. Velocities are never stored in this algorithm. Within the *Leap-Frog* algorithm, velocities are stored at each half time step, and this is illustrated in figure by a shift of a factor $\delta t/2$ in the velocity boxes with respect to position and acceleration boxes. The cycle starts with stored positions, velocities, and accelerations at the time step t, $(t - \delta t/2)$, and $(t - \delta t)$, respectively. As in the Verlet integrator, the first step of the cycle involves a force evaluation to calculate the accelerations at time t, and the concurrent update of the accelerations of the previous step. Then, in the two following steps, velocities and positions are advanced according to Eq. (5) and (6), respectively. Finally, the position array is updated, recovering the initial configuration of the cycle. Lastly, in the *Velocity-Verlet integrator*, velocities are again shifted by a factor of $\delta t/2$ compared to positions and accelerations, but since they are evaluated twice per cycle, they are alternately synchronized with the advancement of positions and accelerations. This double behavior of velocities is depicted in the figure as solid boxes centered at half time steps and dotted boxes at full steps. Positions, velocities, and accelerations at the time t are available at the beginning of the cycle. First, positions are advanced to $(t + \delta t)$ according to Eq. (8), and at the same time the first evaluation of velocities (to be advanced to $(t + \delta t/2)$) takes place (Eq. (9)). Then, the force evaluation leads to the accelerations at the time $(t + \delta t)$, whereas the second evaluation of velocities (Eq. 10) ensures their temporary alignment with full steps. The original configuration of the cycle is finally achieved by discarding the velocities at half steps. Notably, all the algorithms discussed here involve only one force evaluation per cycle.

A more elegant way to reduce the computational demand is based upon the fact that in general different interactions evolve in time with different rates. Taking advantage from this understanding, Tildesley and coworkers developed a multiple time stepping algorithm where the total force acting on the atoms is split into two components that are evaluated at different time frequencies (10). The primary force component is due to the closest neighbor atoms located on a sphere of radius r_1, whereas the secondary involves atoms included between r_1 and an external shell of radius r_2. In correspondence of the slowest time step, the forces due to the primary and secondary

layers are explicitly evaluated as well as the time derivatives of the secondary forces up to a given order. Then, primary forces are explicitly calculated at the shortest time step, and secondary components are estimated using a Taylor expansion exploiting the previously computed time derivatives.

An alternative formulation based on the classification of forces upon the interaction type was provided by Martyna and colleagues (11). Accordingly, exploiting the Liouville formalism, a reversible and volume preserving multiple time stepping algorithm was derived (reversible reference system propagation algorithm, r-RESPA).

Another class of methods that is commonly used in order to increase the sampling efficiency is based upon the idea of eliminating the fastest oscillators (which usually correspond to covalent bonds involving hydrogen atoms) from the numerical integration of the equations of motion by appropriately constraining some degrees of freedom. By doing so, it is possible to safely scale up the time step without incurring into energy conservation problems. Within the SHAKE algorithm, which has been originally developed for the Verlet scheme, the equations of motion are solved while at the same time the imposed constraints are satisfied (12). A velocity-Verlet variant of the method, usually referred to as the RATTLE algorithm, was also derived (13). By constraining the bond stretching for bonds involving hydrogen atoms, the time step can be increased approximately of an order of magnitude. With these methods, time steps of 1.5 to 2.0 fs can be safely employed. It is important to note that for most of the applications the gain in speed comes with a negligible loss of accuracy in the overall representation.

3.2. Numerical treatment of long-range interactions

As it will be shown in more detail later in this work, there can be terms in the potential energy presented in Eq. (1) that connect remote as well as closer parts of the overall system. These are termed "long-range" interactions and mostly arise from electrostatics, as shown by the behavior of Coulomb law describing forces between charged entities. While in principle this does not involve any problem, the computational repercussions of calculating forces among all the interacting centers of a system can be daunting, as it scales as the square of these centers.

The most naïve solution to this issue is to put a cutoff on these forces, however, it can be shown that this approach is appropriate only when the long distance potential energy decay is faster than r^{-3}, which is not the case for the electrostatic potential generated by charges nor by that of dipoles. The consequences of this crude approximation are described by Steinbach and Brooks (14). There are alternatives that reduce the computational cost of the simulation while preserving the long-range character of these forces. Among them, we would like to mention the Ewald summation, the particle-mesh based methods and the fast multipole methods (15-16). Their detailed description is out of the scope of this work, we will limit ourselves to illustrate the basic underlying ideas of the first two and of their widespread contamination, the *particle mesh Ewald* method, and to discuss their use in the context of the simulation of biological systems.

Under the assumption that system is overall neutral and periodic the *Ewald summation* technique consists in adding, and subtracting, a Gaussianly distributed screening charge at each charge's site, assumed to be point-like. The overall amount of each screening charge is equal and opposite to the co-centered original point charge. In this way, the forces exerted by the screened point charges have now become short-range and a distance cutoff can be applied without serious repercussions. On the other hand, the charges that we have to subtract in order to compensate for the added screening generate a potential that, due to the periodicity of the system, can be conveniently calculated in the Fourier domain. *Ad hoc* procedures have to be taken so as to avoid self-interactions. Full details, and the scaling order of the algorithm, $O(N^{3/2})$, are described in the dedicated literature (17-20).

The *particle-particle/particle-mesh* based approaches try to improve the Ewald summation by modifying the calculation occurring on the Fourier space. They exploit the fact that the electrostatic interactions can be very conveniently calculated in the Fourier reciprocal space if the charges are located on a regular grid. In this case the good scaling, $O(N\log N)$, provided by the Fast Fourier Transform algorithm can be exploited (21). In the first of these schemes, described by Hockney and Eastwood, all the charges in the physical space were mapped onto the grid, leading to good performance but not very high precision (22). Since then, the method has been improved by splitting the calculation in a short-range and a long-range contribution, similarly to what is done in the Ewald summation method and originating the so-called Particle Mesh Ewald technique (23-24). Overall, the latter method is the most frequently used and most convenient for medium sized system, while better scaling alternatives, such as the fast multipoles approach, start showing their merits only for large systems (16). It is worth pointing out that the underlying hypotheses of neutrality and periodicity are not the most suitable for the simulation of biomolecular systems. More recently, *ad hoc* approaches have been devised specifically for the simulation of biomolecules in aqueous solution, that do not need either neutrality or periodicity but they are not so widely used (25-26).

3.3. Non-hamiltonian dynamics: thermostats and barostats

In the canonical ensemble, the thermodynamic temperature is kept fixed at a specified value by the presence of a thermal bath, which however permits fluctuations in the instantaneous kinetic energy. In order to simulate a closed system where energy is allowed to flow in and out, in the heat form, suitable algorithms named *thermostats* are employed. Although the aim of a thermostat is to control the temperature of the system, the analogy with the experimental heat bath is not necessarily strict. Indeed, thermostats are algorithms designed to properly alter the Newton equations of motion in order to

attain the canonical distribution of microstates, even though no heat flow is actually simulated. To reproduce the correct canonical ensemble statistics is clearly a more stringent requirement than simply controlling the temperature, and for the same reason not all the methods afterward reported should be rigorously called thermostats.

By altering the equations of motion, the dynamics of the system is unavoidably affected as well. Considering a macroscopic system in contact with a heat bath at the temperature T_0, the rate of change on the average temperature can be expressed as (27):

$$\frac{\mathrm{d}T(t)}{\mathrm{d}t} = \zeta^{-1}\left(T_0 - T(t)\right) \tag{12}$$

where ζ is the system specific temperature relaxation time, which is a function of the ratio between the isochoric (constant volume) heat capacity and the thermal conductivity of the system. Thus, provided that the trajectory achieved is continuous, the thermostats that are able at best to represent the dynamics of the system are those inducing fluctuations rate of the order of ζ (27).

3.3.1. Velocity rescaling methods

Velocity rescaling is the simplest way to control temperature even though it is not the most convenient approach to be used (28). The idea is to rescale velocities by a factor of λ at every given number of time steps so that the instantaneous temperature $T(t)$ given by the instantaneous kinetic energy provides the desired equilibrium temperature value. From energy equipartition principle, it is possible to express this temperature difference as:

$$\begin{aligned}\Delta T &= T_0 - T(t) = \\ &= \frac{1}{2}\sum_i^N \frac{2}{3}\frac{m_i\left(\lambda\dot{\mathbf{r}}_i\right)^2}{Nk_B} - \frac{1}{2}\sum_i^N \frac{2}{3}\frac{m_i\dot{\mathbf{r}}_i^2}{Nk_B} = \\ &= \left(\lambda^2 - 1\right)T(t)\end{aligned} \tag{13}$$

Thus, the rescaling factor is derived as:

$$\lambda = \sqrt{1 + \left(\frac{T_0 - T(t)}{T(t)}\right)} = \sqrt{\frac{T_0}{T(t)}} \tag{14}$$

It must be kept in mind that instantaneous and thermodynamic temperature are different quantities: the former is a mechanical quantity, whereas the latter instantiates a macroscopic concept. As it has been previously discussed, to obtain the canonical distribution of states, the instantaneous temperature fluctuations are necessary. Unfortunately, in the limit of rescaling velocities at each step according to Eq. (14), the kinetic energy results to be constrained to its average value with a null variance, which contradicts the statistical mechanical behavior of the canonical ensemble.

The weak coupling approach introduced by Berendsen and coworkers is similar in spirit to the velocity rescaling method (29). However, in this case the temperature is allowed to relax to the desired macroscopic value at a rate determined by a coupling parameter τ. From Eq. (12) the rate of the kinetic temperature change is thus expressed as:

$$\frac{\mathrm{d}T(t)}{\mathrm{d}t} = \frac{1}{\tau}\left(T_0 - T(t)\right) \tag{15}$$

Considering a temperature difference evaluated at each time step δt, from Eq. (13) the following relation is achieved:

$$\frac{\mathrm{d}T(t)}{\mathrm{d}t} = \frac{1}{\delta t}\left(\lambda^2 - 1\right)T(t) \tag{16}$$

Thus, equating (15) and (16) the rescaling factor for the Berendsen thermostat is readily obtained:

$$\lambda = \sqrt{1 + \frac{\delta t}{\tau}\left(\frac{T_0 - T(t)}{T_0}\right)} \tag{17}$$

Note that, depending on the specific integrator used, the instantaneous temperature can be calculated in a previous instant than the current one. For the special case of $\tau = \delta t$ (tight coupling) the velocity rescaling method is recovered, whereas for values of $\tau \to \infty$ (infinitely loose coupling) the scaling factor approaches the unity, leading to a microcanonical sampling. Consequently, τ behaves as an empirical parameter whose choice somehow affects kinetic energy fluctuations; values of $\tau \approx 0.1$ ps are typically used in MD simulations of condensed-phase systems. As a matter of fact, even though the trajectory obtained is continuous, the canonical distribution in general cannot be rigorously achieved with this scheme, and for this reason more sophisticated methods are advisable.

3.3.2. Stochastic velocity reassignment

With the Andersen thermostat, the temperature is controlled by periodically reassigning the velocity of randomly chosen particles according to the desired Maxwell-Boltzmann distribution (30). Thus, by applying stochastic collisions to the atoms of the system, a virtual heat bath is actually mimicked. Closely resembling MC moves, these collisions allow the system to jump between different constant energy surfaces. In between successive velocity reassignments, positions and velocities evolve in time according to the Hamiltonian dynamics, and the sampling in a constant energy surface is recovered. If a proper collision frequency is chosen, in the limit of an infinitely long sampling, the Boltzmann distribution is ensured by the stochastic nature of the reassignment.

3.3.3. Extended system methods

The idea of the thermostat introduced by Nosé, and later reformulated by Hoover, is to consider the thermal bath as an integral part of the system (31-33). To do so, the system is extended by means of a pair of accessory "conjugated" variables that should account for the heat flow with the thermal bath. The original detailed formulation is a bit involved but, more recently, a simpler

one has been devised; according to this latter, the new equations of motions read (27):

$$\dot{\mathbf{r}}_i = \frac{\mathbf{p}_i}{m_i}; \quad \mathbf{p}_i = \mathbf{f}_i - \chi \mathbf{p}_i \tag{18}$$

This expression is very similar to a traditional Hamiltonian dynamics with the addition of a viscous frictional force with the important difference that here χ is not a positive constant but a further dynamical variable that can have any value in the real number set. Its role is to either inject to or withdraw energy from the system depending on the instantaneous temperature. Its dynamics is determined by:

$$\dot{\chi} = -\frac{1}{\tau_{\mathrm{NH}}^2}\frac{T_0 - T(t)}{T_0} \tag{19}$$

Consistently with the intuitive behavior of this parameter, it increases if the current instantaneous temperature of the system is higher than the target one, and *vice versa*. τ_{NH} is a coupling parameter that controls how quick the correction of the kinetic energy acts. If it is large the energy flow is slow (loose coupling) leading to small temperature fluctuations. The dynamical system built in this way admits a conserved quantity in the extended space but it is no longer Hamiltonian in the sense that what is conserved is not the total energy. In conclusion, given an ergodic and conservative system, the Nosé-Hoover thermostat provides a deterministic dynamics, with a continuous trajectory that samples a canonical distribution of microstates. Improvements over the original definition, such as the so called Nosé-Hoover chains, were developed to overcome the non-ergodic behavior of the Nosé-Hoover thermostat in the context of small or stiff systems or of systems at low temperatures (34).

3.3.4. Stochastic dynamics

Given the complexity introduced by the Nosé-Hoover chains, it is worth wondering whether it is possible to sample the canonical distribution with simpler schemes. The most natural way to proceed is to use the Langevin equation of motion (35):

$$m_i \ddot{\mathbf{r}}_i = \mathbf{f}_i - \gamma_i m_i \dot{\mathbf{r}}_i + \mathbf{R}_i \tag{20}$$

The right hand side in Eq. (20) includes: the forces on the atoms as derived from the potential energy function, a frictional term and a stochastic force $\mathbf{R}_i$. The frictional coefficient γ (which has the units of time^{-1}) has positive values and controls the damping of the thermostat, whereas the stochastic forces obey to a Gaussian distribution with zero mean and no correlation in time. It can be shown that the Langevin equations of motion sample the canonical distribution and, even though its dynamics is not deterministic, they produce a substantially smooth trajectory, in contrast to the Andersen thermostat. For those reasons, and its intrinsic simplicity, Langevin dynamics has recently become a very popular standard to control temperature in MD simulations.

3.3.5. Potential pitfalls in the application of thermostats

It is worth mentioning two problems that can arise when applying a thermostat to molecular systems involving distinct sets of variables with either (*i*) very different characteristic frequencies or (*ii*) very different heating rates. Then, the coupling of all the degrees of freedom to the same thermostat may lead to different effective temperatures for the mentioned distinct subsets, due to a too slow exchange of kinetic energy between them. A typical example is the so-called "hot solvent – cold solute problem" in simulations of biomolecular macromolecules. Because the solvent is more significantly affected by algorithmic noise (e.g., due to the use of an electrostatic cutoff), the coupling of the whole system to a single thermostat may cause the average solute temperature to be significantly lower than the average solvent temperature. A possible solution to this problem is to couple separately the solute and solvent degrees of freedom to two different thermostats. Another problem arises when the thermostat acts directly to the absolute atomic velocities without removing the velocity of the center of mass. In this case, the linear and angular momenta may be not conserved and the thermostat could lead to violation of energy equipartition (36).

3.3.7. Barostats

Realizing that the pressure is controlled by scaling the volume of the system under investigation, similarly to the temperature baths, analogue algorithms for pressure control may be derived. Along with the thermostat, a Berendsen variant for the pressure coupling was developed whereas Andersen proposed a barostat based on an extended system formulation (29). Depending upon the physics of the system, an isotropic or anisotropic rescaling of the box axis may be chosen. Eventually, with the aim of studying solid state phase transitions, it is worth mentioning that Parrinello and Rahman extended the Andersen formulation to allow the box to change its shape as well as size (37-38).

4. DESCRIBING THE ATOMIC INTERACTIONS: MOLECULAR MECHANICS AND FORCE FIELDS

Modeling molecular interactions is an important and long standing goal of physics which goes back to the late XIX century when the early phenomenological theories of real gases were developed (39). A detailed understanding of the atomic and molecular driving forces became possible after the advent of the theory of quantum mechanics (QM). In particular the Schrödinger equation (1926) provided a theoretically powerful albeit in practice extraordinarily complex route to calculate geometries and other molecular properties. At the atomic scale, if no nuclear decays are considered, the only relevant physical force is the *electrostatic interaction* between nuclei and electrons. The Schrödinger equation explicitly takes into account these interactions but, due to its inherent computational complexity, several approximations must be taken. Within the largely employed Born-Oppenheimer (BO) approximation (1927), the nuclear and electronic problems are separately treated. This is justified by the fact that electrons are particles much lighter (roughly 2000 times) and faster than nuclei, hence it is reasonable to assume that in most cases electrons instantaneously rearrange their configuration over a given nuclear

geometry. In the first step of the BO approximation, the electronic Schrödinger equation is solved, under the assumption that nuclei are fixed in a given conformation, usually the equilibrium one. In the second step, the result of the previous one is used to derive an external field, the *potential energy surface* (PES), for a Schrödinger equation containing only the nuclei. Unfortunately, even with the currently available computational resources, and despite the simplification provided by the BO approximation, this kind of approach remains a cumbersome task and many other approaches to study the quanto-mechanical behavior of molecules have been devised. Another useful tool is provided by the Density Functional Theory (DFT), where the simpler, but still daunting, problem of finding the electronic density of a molecular system, especially in its ground state, is faced (40). From that density, all physical observables at the ground state can be derived. DFT can be used, for instance, in the context of biomolecular simulations, especially to describe enzymatic reactions, that have an inherent quanto-mechanical nature. Also DFT, however, is too computationally demanding do describe the time evolution of many systems of practical interest and is out of the scope of this work.

A much less costly alternative to model molecular geometries is represented by building classical interaction potentials between atoms, known as *force fields* or molecular mechanics. Still relying upon the BO approximation, force fields bypass the issue of solving the Schrödinger equation by expressing the potential energy as a function of the nuclear coordinates only, where the dependence upon electrons is implicitly treated by a suitable parameterization. As a consequence, this kind of approach can hardly describe the formation or breaking of covalent bonds nor any other mechanism where the knowledge of electronic structure or of its rearrangement is explicitly required (but see references for reactive force fields shown below).

A force field is univocally defined only once both the *functional form* of the energy and the *parameters* are chosen. Distinct force fields may share the same functional form but differ in the employed parameters, thus diverging results may be expected. A common founding assumption for force fields is the concept of *transferability*, which is the possibility to treat in the same way (i.e. with the same set of parameters) identical chemical groups located in different molecular environments. Thus, for each structural unit, parameters are generally derived by maximizing the agreement of the force field outcome with either available experimental data or more costly QM-based calculations on small representative molecules. From a practical standpoint, the transferability concept is implemented in force fields in the form of *atom types* by which the information concerning the atomic number, hybridization and local environment are specified and stored. As a general rule, the larger is the number of atom types provided, the greater is the average accuracy achievable against a diverse set of compounds. From this perspective, force fields are usually classified as *general* or *specific*, depending on the design and hence the target of their use. Accordingly, Amber (41), CHARMM (42), OPLS (43), and GROMOS (44), are the most popular specific force fields used to simulate proteins in condensed-phase environment, whereas MM2 (45), MM3 (46-48), MMFF (49-51), and Dreiding (52), are examples of general force fields usually employed to calculate properties of small organic molecules in gas-phase. As already mentioned, since electrons are treated implicitly, no chemical reactions (where redistributions of the electronic density over nuclei occur) can be actually simulated with force fields (with the exception of reactive force fields (53), and transition states force fields (54), which will not be covered here). The information concerning the mutual connectivity among atoms defines the *topology* for the given system, which is (normally) kept constant during molecular mechanics calculations. The level of accuracy used to describe the underlying chemistry of the problem under investigation is usually referred to as the *representation*. For reasons of computational efficiency, earlier force fields used to treat explicitly only the hydrogen atoms bound to hetero atoms (United Atom representation, UA) (55-58). As computer power has grown over the years, All Atoms (AA) extensions have been developed as well (43, 59-60). Recently, the need to speed up the calculations especially in the field of polymers and biological membranes where large systems are involved and a very detailed atomistic description is generally not needed, prompted several groups to adopt a so called Coarse Grain (CG) representation (61-62). Within CG models, groups of atoms (typically 3 or 4) are joint together and modeled as single units called *beads*.

The aim of the present section is to provide a general overview regarding the most commonly employed force fields for the study of protein-ligand interactions. For an exhaustive excursus on the protein force fields evolution, the reader is redirected to the literature (63). Here, some peculiar aspects regarding the functional form and the parameterization procedure will be covered mainly in the context of the most popular force fields designed for protein modeling (especially, Amber, CHARMM, OPLS and GROMOS), but gas-phase force field characteristics will be occasionally discussed too. Then, some intrinsic issues related to the use of simple functional forms and cheap but not optimal representations will be addressed in the light of the ability to reproduce selected interactions frequently encountered in protein-ligand binding. Finally, the topic of polarization will be briefly discussed.

4.1. The functional form in additive force fields

The functional form of the force field specifies the way different contributions to the potential energy of the system are taken into account as a function of the atomic (i.e. nuclear) coordinates. Since several types of force fields have been developed over the years with both different purposes and distinct ranges of applicability, only the fundamental aspects of the functional form will be covered here. Without any claim of completeness, a general functional form for a force field may be expressed as follows:

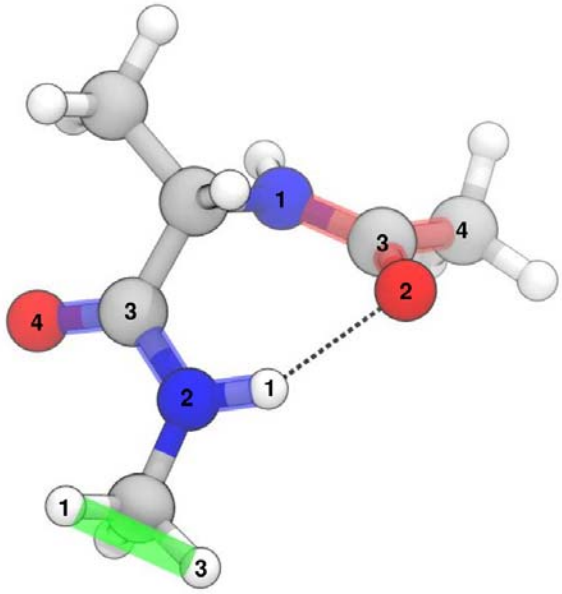

Figure 2. Schematic picture of atomic interactions and their relative topological relationships in a typical force field model. The molecular model of the alanine dipeptide in the C7eq conformation is taken as a reference compound. In the Figure, an example of a proper dihedral, an improper dihedral, and a Urey-Bradley interaction are highlighted in blue, red, and green, respectively. For each of these bonded contributions, the relative topological relationships of the atoms involved are illustrated by a progressive numeral index (the starting index is arbitrary). A non-bonded contribution between atoms separated by more than four bonds, in this case an intermolecular hydrogen bond, is represented by a dotted line.

$$V_{\mathrm{FF}} = \left(\sum_{1-2} V_{\mathrm{stretch}} + \sum_{1-3} V_{\mathrm{bend}} + \sum_{1-4} V_{\mathrm{tors}} \right)_{\mathrm{b}} + \left(\sum_{\mathrm{pairs}} \left(V_{\mathrm{elect}} + V_{\mathrm{vdW}} \right) \right)_{\mathrm{nb}} \quad 21)$$

As reported in Eq. (21), from a topological point of view, it is convenient to group the single contributions of a force field as *bonded* (b) and *non-bonded* (nb) terms. Clearly, this separation strictly applies only as long as a non-reactive force field description is considered. Accordingly, bonded terms are in turn composed by the *stretching* (1-2 interactions, i.e. between an atom and its topological nearest neighbors) and the *bending* (1-3 interactions) contributions which are evaluated over all bonds and angles, respectively. In addition, the *torsional* term describing the 1-4 interactions over dihedrals of bound atoms is also considered as a bonded term. The second group of contributions concerns the non-bonded interactions, which are represented by the electrostatic and van der Waals terms. Differently from most bonded contributions, in the simplest force field models non-bonded terms are treated as pair-wise additive forces acting among the atoms of the system. In this summation, all the atoms involved in a bonded 1-2 or 1-3 relationship are excluded, whereas for the 1-4 interactions a special non-bonded treatment is usually applied (see later). The topological relationships among atoms and the respective force field interactions are summarized in Figure 2.

Since a force field is intended to be used in conjunction with either optimization techniques or sampling schemes such as MD in which forces evaluations are performed, an intrinsic requirement of the functional form is to be at least once (and possibly twice) differentiable with respect to the atomic coordinates.

In the following subsections, the many terms constituting the functional form of the force fields will be discussed and commented.

4.1.1. Bonded terms

The bonded contributions to a force field are constituted by the stretching, the bending and the torsional terms.

4.1.1.1. Stretching term

Since the energy of chemical bonds goes asymptotically to zero upon infinite elongation, in order to attain a realistic stretching behavior with force fields, 1-2 interactions should be in principle described by means of suitable anharmonic function of the atomic coordinates. A valuable common expression for the potential energy is the Morse function (64):

$$V_l = D\left(1 - \mathrm{e}^{-\sqrt{\frac{k}{2D}}(l-l^0)}\right)^2 \quad 22)$$

Where l is the independent variable, that is the actual bond length. In order to describe a bond stretching by means of Eq. (22), three parameters are thus needed: (*i*) the dissociation energy D, which describes the well depth; (*ii*) the bond constant k; and (*iii*) the reference bond length l^0. Taken together, D and k determine the curvature of the energy function around the minimum, which is located in correspondence of l^0 (see Figure 3A).

Although appealing, for several reasons the Morse function is rarely used in force fields. First of all, three parameters need to be derived for each bond type making the parameterization procedure an uneasy task. Moreover, at large distances the Morse function is associated to small restoring forces, potentially leading to pathological behaviors either during the optimization of highly distorted structures or when performing MD simulations (52). Eventually, from a purely computational point of view, it is always advisable to avoid employing functions containing the expensive exponential term (see Eq. (22)). It is important to realize that whenever a force field description is adopted, a correct chemical behavior of bond break and formation is generally not needed. Instead, for a given topology, it is rather more interesting to correctly reproduce the shape of the potential energy function in proximity of the minimum. Thus, the energy function for a bond stretching may be expressed as a Taylor expansion around the reference length l^0:

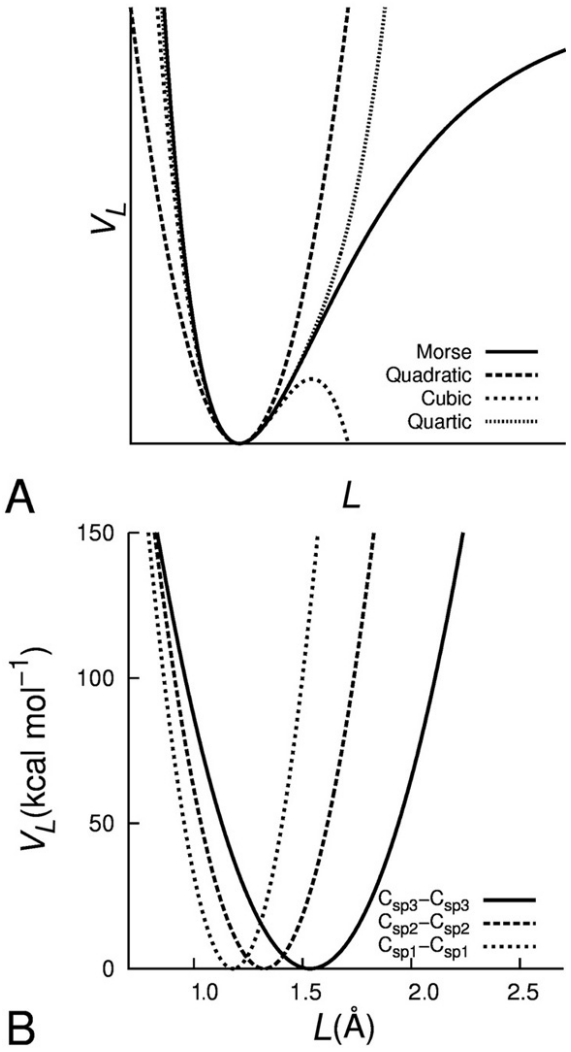

Figure 3. Comparison of different functional forms used to describe the bond stretching (panel A). Panel B: typical behavior of the bond stretching curves described as quadratic polynomial functions for a series of selected atom types. Both relationships are considered with respect to interatomic distance L.

$$V_l = V_l(0) + \frac{\partial V_l}{\partial l}\left(l - l^0\right) + \frac{1}{2}\frac{\partial^2 V_l}{\partial l^2}\left(l - l^0\right)^2 + \\ + \sum_{n=3}^{\infty} \frac{1}{n!}\frac{\partial^n V_l}{\partial l^n}(l - l^0)^n \qquad 23)$$

In Eq. (23), $V_l(0)$ represents the offset of the energy at the minimum that it is usually taken to be zero. Since the expansion is performed around a minimum, the term involving the first derivative is null. Realizing that the second derivative of the energy with respect to the coordinates corresponds to the force constant of the bond, if higher order terms are omitted, Eq. (23) reduces to Hooke's law:

$$V_l = \frac{1}{2}\frac{\partial^2 V_l}{\partial l^2}\left(l - l^0\right)^2 = \frac{k}{2}\left(l - l^0\right)^2 \qquad 24)$$

The obtained expression represents the simplest functional form that can be used to describe bonds in force fields, and a reasonable approximation of the bottom of the bond dissociation curve can be achieved if a proper parameterization is provided. It is important to note that using Eq. (24) only two parameters (the force constant k and the reference bond length l^0) must be derived for each pair of atom types. In order to achieve a better agreement with the Morse function, for moderate deviation around the reference value, higher order terms may be introduced in the expansion. As an example, both the MM3 (46-48) and the MMFF (49-51) force fields employ a polynomial expansion of the fourth order of the form:

$$V_l = \frac{k}{2}\left(l - l^0\right)^2\left[1 + k_c(l - l^0) + \frac{7}{12}k_c^2(l - l^0)^2\right] \qquad 25)$$

Where k_c is the cubic anharmonicity constant which is determined once for the whole force field. Since k_c is usually negative, the expansion truncated at the third term is associated to the undesirable behavior that the energy goes asymptotically to minus infinity with the distance, and for this reason a quartic polynomial expansion is normally used (see Figure 3A).

4.1.1.2. Bending term

Similarly to the stretching contribution, the bending term is usually described in force fields with polynomial expansions of different order, and similar considerations apply. However, the conceptual difference between the two interactions should not be overlooked, as it occasionally emerges for peculiar geometries (this is the case, for instance, of molecules having a divalent atom in the position "2" of the topology such as alcohols and esters, where an energy maximum is expected at bending values of π). In these cases, the correct angular behavior may be accounted for by either using polynomial expansions truncated at the third order enforced by special boundary conditions (MMFF (49)) or alternatively using periodic terms similar to those employed to treat torsional angles (Dreiding (52)) (see below). More frequently, the bending term is simply accounted for by a polynomial expansion of the second order equivalent to Eq. (24).

The bonded interactions such as the stretching and the bending terms are usually parameterized against experimental infrared data or either experimental or computational geometries obtained at a higher level of theory (QM calculations). It is important to point out that the geometrical parameter l^0 has been termed reference value, and it should not be confused with the equilibrium length (the same considerations are valid for the bending analogue). In fact, the reference value is the actual bond length against which the force field is parameterized, whereas the equilibrium length is the bond length corresponding to an energy minimum in a molecular

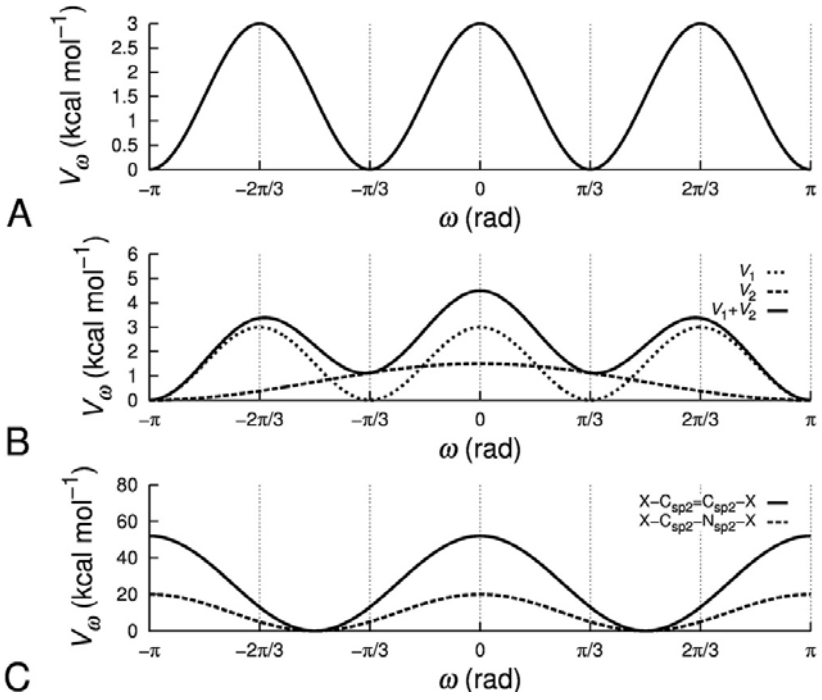

Figure 4. Typical torsional energy for the ethane (A) and butane (B) molecules. For the latter, the contributions of each term in the expansion (V_1: (3/2)·(1+cos(3ω)) and V_2: (1.5/2)·(1+cos(ω))) are also shown. Panel C: comparison of the torsional energy between a double bond such as that of ethylene and an amide partial double bond.

environment when *all* the other force terms are acting. Since parameters belonging to different terms are inter-related, meaning that they are optimized in such a way to reproduce at best experimental or theoretical geometries when considered together, the force field parameterization becomes necessarily an iterative and complex multi-objective procedure. Different force fields adopt different strategies to preserve as much as possible the transferability of their parameters throughout different environments.

Taken as a whole, the stretching and the bending terms can be considered as "stiff" interactions, in the sense that their contribution to the overall energy of the system is large even when small displacements around the reference values are considered (see Figure 3B). This is not surprising, since they represent strong bonding interactions, which allow to consider groups of atoms as distinct molecular entities. In this respect, the "softer" torsional terms (taken together with the non-bonded interactions) permit a much larger number of conformations. Their complex balance is responsible for the relative conformational equilibrium of molecules and therefore their value is more informative than that of the "harder" terms.

4.1.1.3. Torsional terms

Considering a group of atoms in a 1-4 topological relationship, the dihedral angle is the one defined by the planes passing from atoms 1, 2, 3 and 2, 3, 4. However, depending upon the relative connectivity of the four atoms, two distinct types of torsional terms may be envisioned: the *proper* and the *improper* torsions. The former term is used whenever the atoms are linked as in a chain, whereas the latter applies when branched structures (by convention around the third atom of the sequence) are involved, and they are also known as *out-of-plane bending* term (see Figure 2). Because of their intrinsic periodicity, proper torsions are usually expanded in a cosine series:

$$V_\omega = \sum_{i=1}^{N} \frac{V_i}{2} [1 + \cos(n_i \cdot \omega - \alpha_i)] \quad (26)$$

where V is the energetic barrier, n is the periodicity (or multiplicity), α is the phase, and ω is the actual dihedral angle belonging to the i^{th} term of the expansion. The summation runs over the N terms of the series (the occasional need for high order terms will be clarified later). The periodicity defines the number of minima (or maxima) located in the interval [−π; +π], whereas the phase characterizes the offset of the function along the x axis: a minimum at ω = 0 is obtained when α = π, whereas a maximum is achieved by using a phase of 0 radians. It is worth noting that the unity added to the cosine function in Eq. (26) has the effect to shift the function along the energy axis in order that only positive values of energy are considered (similarly to the offset encountered for the stretching and bending terms). Whenever along the central atoms of a proper torsion (atoms 2 and 3) a symmetry axis is found (such as in ethane), the cosine expansion is truncated at the first order. In any other case, higher order terms are used in order to finely tune the energy profile. A comparison of the torsional term for ethane and butane is shown in Figures 4A and 4B, respectively. Each i^{th} order is characterized by a consistent set of V_i, n_i and α_i parameters.

From this perspective, the term "energetic barrier" used for V_i is somewhat misleading since each order contributes to the global energy curve. Having defined a general form for the torsional term, it is important to understand the different contexts where such contributions are employed in a force field. First, proper torsional terms are used whenever there is the need to enforce peculiar symmetries around atoms in position 2 and 3, typically when a non-negligible π bonding contribution is involved in the chemical bond between the central atoms such in the case of alkenes or amides (Figure 4C). Clearly, for such examples the proper torsion accounts for the central bond twisting, but it is not sufficient to describe planar geometries observed at the 2 and 3 centers. For this purpose, improper torsions may be employed, and different functional forms can be envisioned. A periodic function similar to Eq. (26) where for geometrical reasons the periodicity and the phase are respectively always 2 and π, is usually adopted, such as in the Amber and OPLS-AA (which was initially built over the bonded Amber potentials) force fields, whereas CHARMM and GROMOS rather employ a simple harmonic functional form.

It must be considered that proper torsional terms are defined for all the atoms involved in a 1-4 relationship, no matter the nature of the central bond. It is therefore legitimate wondering why they are actually used when a σ bond (i.e. a linear bond among two atoms) is involved. The reason for that has to be found in the influence of the external atoms (at the 1 and 4 positions) onto the dihedral energy profile. In this context, torsional terms may be considered as useful corrections required to reproduce the proper torsional behavior once the non-bonded interactions are taken into account. In this respect, the already considered case of ethane is emblematic. In ethane, the proper torsion around the central carbon atoms is needed to account for hyperconjugation effects which are responsible to stabilize the staggered conformation over the eclipsed one more than it would be if only steric effects were involved (65). It is then not surprising that in the MM2, (45), and MM3, (46-48), force fields different orders of the cosine expansion of Eq. (26) are associated to a distinct physical interpretation, including electronegativity, hyperconjugation, conjugation and steric effects. Because of their corrective meaning, torsional terms are usually fitted after the derivation of all the other bonded and non-bonded contributions. From a parameterization point of view, dihedrals are usually tuned in order to reproduce the torsional energy profile of prototypical molecules obtained via QM calculations.

4.1.1.4. Cross terms

Additionally to the bonded terms considered so far, some force fields (such as MM3, (46-48), and MMFF, (49-51)) include further terms that are suited to describe the coupling behavior between internal degrees of freedom. Cross terms, sometimes also referred to as *non-diagonal terms* to be distinguished by the usual uncoupled (or *diagonal*) contributions to the potential energy, are often used in force fields aimed at reproducing molecular vibrations in addition to energies and geometries. In principle, all the combinations of the internal degrees of freedom contributions should be included, however most often only few of them are necessary to achieve a reasonable accuracy, and typically only pair-wise cross terms are considered. These terms are usually expressed as products of Taylor expansions of the energy as a function of the uncoupled degrees of freedom, truncated at the first order. As an example, the stretch-stretch cross term between the two distances l_1 and l_2 takes the following form:

$$V_{(l_1,l_2)} = \frac{k_{(l_1,l_2)}}{2}\left[(l_1 - l_1^0)(l_2 - l_2^0)\right] \quad 27)$$

where l_1^0 and l_2^0 are the two reference distances and $k(l_1, l_2)$ the force constant.

Similarly, the most important contribution, represented by the stretch-bend term, is expressed as follows:

$$V_{(\theta,l_1,l_2)} = \frac{k_{(\theta,l_1,l_2)}}{2}(\theta - \theta^0)\left[(l_1 - l_1^0) + (l_2 - l_2^0)\right] \quad 28)$$

where θ is the angle, θ^0 its reference value, and the remaining symbols have the usual meaning. Eq. (28) describes the coupled energetic contribution between the stretching of two bonds and the variation of the angle between them. In other words, it accounts for the fact that the angle reduces or increases upon symmetric elongation or shrinkage of the two bonds, respectively.

In general, in condensed-phase force fields, those contributions are not used. An important exception is represented by the CHARMM force field where the stretch-bend term for atoms in a terminal angle (1-3 relationship) is treated with the so called Urey-Bradley potential (42):

$$V_{(l_{1-3})} = \frac{k_{(l_{1-3})}}{2}(l_{1-3} - l_{1-3}^0)^2 \quad 29)$$

The Urey-Bradley potential is equivalent in spirit to that in Eq. (28), where the coupling between the bond stretching and bending is implicitly described by a spring acting on the atoms lying in relative topological positions "1" and "3" (see Figure 2). Moreover, starting from the CHARMM22 release, the same force field also adopts a torsion-torsion cross term to better describe the potential energy as a function of the coupled backbone φ/ψ angles, which is implemented as a corrective tabulated potential (CMAP) (66).

It is useful to conclude this brief overview of bonded terms by highlighting the fact that force fields are usually classified as "class I" if harmonic and diagonal bonded terms are involved, and as "class II" if non-harmonic and off-diagonal terms are rather considered. Condensed-phase force fields for proteins definitely belong to the first group.

4.1.2. Non bonded terms

The non-bonded terms of a force field include the electrostatics and the van der Waals contributions.

Although bonded terms are usually well characterized, meaning that their performance is generally satisfactory under the employed approximations, accounting for the non-bonded interactions is considerably more problematic, and most of the efforts in the active research in force fields are spent to improve the models used for their representation (55). The main source of complexity in describing non-bonded interactions relies on their many-body nature. Adopting the simplest force field model, non-bonded interactions are treated assigning to each particle of the system electrostatic (partial charges) and van der Waals parameters. Accordingly, atoms are considered as *interaction sites* and the non-bonded energy of the system should be expressed as (5, 67):

$$V_{\mathrm{nb}} = \sum_i V_1(\mathbf{r}_i) + \sum_i \sum_{j \neq i} V_2(\mathbf{r}_i, \mathbf{r}_j) + \\ + \sum_i \sum_{j \neq i} \sum_{k \neq j \neq i} V_3(\mathbf{r}_i, \mathbf{r}_j, \mathbf{r}_k) + \ldots \qquad 30)$$

where the first term on the right represents the effect of an external field (which is usually zero), whereas the remaining terms account for the interactions among particles. The expansion shown in Eq. (30) is usually truncated at the second term, meaning that only pair-wise additive potentials are considered. This approximation is largely adopted since the increase in computational cost upon inclusion of higher order terms is generally not justified. However, by considering only pair potentials, all the many-body effects are definitely lost, and for this reason force fields referring to this model are said to be *additive*. It must be emphasized that, even within a pair-wise approximation, from a computational standpoint non-bonded terms represent the most demanding terms of a force field. Indeed, while the computational cost due to the bonded terms scales almost linearly with the system size, the non-bonded part evaluated as pair potential scales, in principle, with the square power of the number of the atoms of the system since all the possible pairs of interacting sites have to be considered. For short range interactions, such as van der Waals forces, the use of a cutoff can largely reduce the computational burden. In contrast, long range interactions need *ad hoc* procedures, such as those described in Sect. 3.2.

The pair-wise additive approximation results in the impossibility to quantitatively account for those interactions where polarization and cooperativity (which are many-body in nature) play a major role. For this reason, additive force fields are also intrinsically *non-polarizable* force fields. However, part of the many-body effects may be incorporated in an average way in additive force fields with a proper parameterization of the non-bonded contributions leading to the so called *effective pair potentials* (67-68).

4.1.2.1. Electrostatic term

Electrostatic interactions between molecules (or between distinct moieties of the same molecule) arise as a consequence of the uneven electronic distribution around nuclei, and, in the vast majority of additive force field models, they are accounted for by assigning point charges (q_i) to each interaction site. Thus, the electrostatic interaction between the i and j pair is evaluated by the Coulomb potential:

$$V_{\mathrm{elect}} = \frac{q_i q_j}{4\pi\epsilon_0\epsilon_r\, r_{ij}} \qquad 31)$$

where r_{ij} is the distance separating the charges, and ε_r is the relative dielectric constant of the medium, which in standard MD simulation is taken to be 1.

Although different classifications may be found in the scientific literature and in textbooks, for sake of simplicity it is possible to distinguish three main classes of partial charges depending on the adopted derivation philosophy: (*i*) topological charges, (*ii*) charges derived from QM calculations, and (*iii*) empirical charges. *Topological charges* are not dependent on the molecular geometry, but rather on atomic connectivity. They are usually adopted by molecular modeling software employed in drug design, where a large chemical variability must be routinely accounted for in a fast and automatic way. As an example, in the popular Gasteiger-Marsili scheme, charges are calculated based upon an inter-molecular charge balance that is a function of the electronegativities (experimentally measurable quantities) of the directly linked atoms (more details concerning this charge derivation scheme will be provided in the context of the Fluctuating Charge polarizable model, see below) (69). A similar approach is also adopted by the MMFF force field (49, 70).

A different charge derivation philosophy relies upon properties calculated by *QM methods*. The definition of partial charge is not univocal, and there is no well defined observable that can be associated to it. Therefore, its derivation is somewhat an arbitrary procedure that largely depends upon the specific properties that one wishes to reproduce. From this perspective, charge derivation procedures based upon electronic population analysis, where the molecular electronic density is partitioned and assigned to the atomic nuclei so that a fractional number of electrons are associated to each interaction site, should be mentioned. Many methods have been devised with this purpose, such as the Mulliken analysis, or the partition schemes based upon the Bader's theory of atoms in molecules (71-72). However, a common perception in the scientific community is that charges derived by population analysis are in general inadequate to be used in force fields. Indeed, because of their derivation philosophy, partial charges of this kind mostly carry *intra*-molecular information regarding the electronic distribution, but they are by no means suited to properly describe the *inter*-molecular properties of molecules, which is exactly what needs to be accounted for by the non bonded terms of the force fields. Conversely, the latter feature can be reproduced by deriving charges from the molecular ElectroStatic Potential (ESP), that is the potential generated outside the molecule by the QM charge distribution. Many methods have been developed with this purpose. Generally, a fitting procedure is employed in order to obtain the set of

partial charges that best reproduce the QM-ESP sampled at specific points in space located around the considered molecule. This is in general done by minimizing a cost function (χ^2_{ESP}) defined as the sum of squares of the differences between the QM-ESP (V_i^{QM}) and the ESP generated by using the partial point charges via Coulomb law (V_i^{FF}):

$$\chi^2_{\mathrm{ESP}} = \sum_i^N \left(V_i^{\mathrm{QM}} - V_i^{\mathrm{FF}} \right)^2 = \\ = \sum_i^N \left(V_i^{\mathrm{QM}} - \sum_A^M \frac{q_A}{r_{iA}} \right)^2 \qquad 32)$$

where the summation over *i* runs over the *N* ESP sampling points, while that over *A* runs over the *M* atoms of the molecule (and, in the simplest force field models also to the non-bonding interaction sites). Eq. (32) must be solved iteratively with respect to the partial charges, which must maximize the agreement with the QM-ESP. The location of the sampling points is crucial both for the quality of the derived charges and for the robustness of the fitting procedure. As a general rule, these points should be located in those regions of space where the most relevant electrostatic inter-molecular activity occurs. For this purpose, in most of the fitting schemes, the QM-ESP is sampled only outside the van der Waals surface of the molecule. This is defined as the union of atomic spheres the radius of which corresponds to the distance where inter-atomic repulsion starts exerting its effects. Usually, the points used in the ESP procedure are located on multiple van der Waals-like surfaces obtained by gradually increasing the atomic radii. The region of space enclosed by the van der Waals volume is excluded since it is not relevant for inter-molecular forces and because the high electron density in proximity of nuclei would dominate the fitting procedure introducing either noise or a strong dependence upon the grid resolution where the fitting is done. If molecules were rigid bodies, ESP-derived charges would be the best possible charge model in additive force fields. Unfortunately, they suffer in general of two severe drawbacks. The first limitation is that the derived charges strongly depend upon the molecular conformation chosen to calculate the QM-ESP. It is important to underline that the electronic density, and, by consequence, also the best fitting charges vary when the molecule changes conformation. Therefore, simulative methods that explore the molecular conformational space should in principle use conformation dependent partial charges (73). Within fixed partial charge models this is obviously not allowed, although this issue might be faced by polarizable force fields (discussed below). The second limitation concerns the fitting procedure, in which the most buried interaction sites are also the less robustly determined. This is a consequence of the nature of the fitted function (the ESP), which is highly sensitive to the more exposed atoms. This results in very large fluctuations of the partial charges in the buried interaction sites during the fitting procedure. In the end, buried non-polar atoms might have a large partial charge magnitude, with a detrimental effect on the quality of the intra-molecular description and wrong conformational energies. To face this issue, Bayly and coworkers developed a scheme named RESP (restrained ESP) (74-75). The basic idea of the method is to restrain the charges during the fitting to the arbitrary value of zero by adding hyperbolic penalties to the cost function (Eq. (32)) in a stepwise procedure. The rationale for the use of restraints is to damp the oscillations on the buried sites during the fitting, whereas the hyperbolic functional form ensures a limited impact of the restraints onto the more polar atoms (since the partial charges of these latter are greater in magnitude than those of non-polar atoms). The RESP procedure has been employed to derive amino acid charges within the Amber series of additive force fields starting from the so-called ff94 release (41). Since the method is intended to derive charges that are mainly used in methods such as MD and to limit the charge dependency on conformation, chemical equivalencies are imposed as an additional restraint to be satisfied during the fitting procedure. In addition, it should be mentioned that the RESP procedure is flexible enough to allow charge derivation from multiple QM-ESP belonging to different conformations of the same molecule simultaneously. In this variant, each QM-ESP is weighted during the fitting procedure according to the Boltzmann weight of the corresponding conformation (76).

The last class of charge determination methodology consists of so-called *empirical charges*. This method, adopted for instance by CHARMM, can also relies upon QM calculations, but is meant to find a proper balance among solute-solute, solute-solvent (usually water) and solvent-solvent interactions. To this aim, charges are fitted to reproduce both interaction energies and minimum energy distances calculated between one or more water molecules (possibly considered in an ensemble of orientations) and each polar site of the molecules included in given datasets. In the variant adopted in the OPLS-AA and GROMOS force fields, charges are derived from experimental thermodynamic data. More specifically, they are derived along with van der Waals parameters by fitting the experimental heat of vaporization and densities of a variety of pure liquids having similar moieties to those found in proteins (see later). An advantage of this approach over the others is that many-body effects are automatically taken into account by the parameterization, even though in an average and unspecific manner (effective potential).

QM calculations used to fit partial charges are usually performed in vacuum. Generalizing the evidence of what happens with bulk water, the empirical dipole of which is larger than the corresponding gas-phase value (2.35 D compared to 1.86 D), a widespread workaround to reproduce bulk properties by means of gas-phase derived partial charges, is to enhance the QM multipole moments before applying the charge fitting procedure. Within the RESP procedure, this is done simply by deriving the QM-ESP at a level of theory (HF/6-31G(d)) that is known to overestimate the polarizability of molecules, so that the derived charges will consistently inherit a desired enhancement in magnitude (74-75). In contrast, in the CHARMM fitting procedure the solute-solvent interaction energies are scaled by a factor of 1.16 (further possible

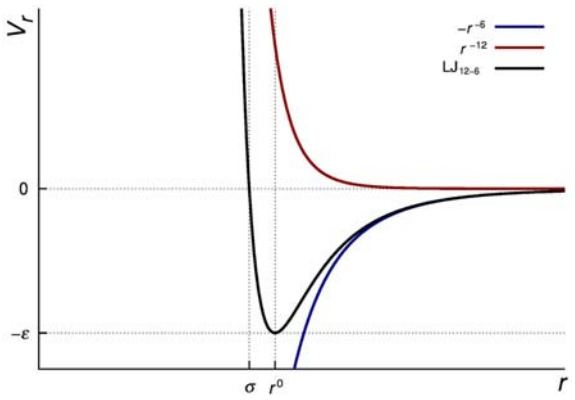

Figure 5. Lennard-Jones, "12-6", potential plotted together with the r^{-12} and r^{-6} components.

adjustments are allowed) (42). Clearly, the effective potential approach may be considered progressively less satisfactory moving from the bulk of a component to its interface, where polarization effects can only be properly modeled by using polarizable force fields (see later).

It is important to note that the point charge model is not the only way to represent electrostatic interactions. An alternative are the so called distributed multipole models where multipoles up to a certain order are assigned to the molecule and are not necessarily located at the nuclear positions. As an example, atom centered multipoles are employed in the MM2, MM3 and AMOEBA force fields (47-48, 77).

4.1.2.2. Van der Waals term

The van der Waals term consists of an attractive and a repulsive term representing the *dispersive* and the *exchange* interactions, respectively. The *dispersive interactions* (also known as London forces) are due to the instantaneous correlation of mutually induced multipoles arising from fluctuations in electronic clouds between non-bonded approaching atoms. As an example, they are responsible for deviations from the ideal gas behavior for rare gases and, more generally, for any attractive interaction between non-polar molecules or groups of atoms. Involving polarization between electronic clouds, dispersive interactions are intrinsically non-additive in nature. However, they are generally safely modeled in force fields by using a r^{-6} distance dependent function. The other contribution is due to exchange interactions, which are repulsive and can be rationalized in terms of the Pauli's exclusion principle. This principle states that two electrons cannot be in the same quantum state, and in particular have the same spin state and position. Exchange interactions are effective at shorter distances compared with dispersive interactions, and mainly for reasons of computational convenience they are usually modeled in force field by means of a r^{-12} distance dependent function, even though QM calculations would rather point at an exponential behavior. Taken as a whole, the van der Waals interactions described here give rise to the Lennard-Jones 12-6 function:

$$V_{ij}^{\mathrm{LJ}} = 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \qquad 33)$$

where r_{ij} is the separation distance, ε_{ij} is the well depth (not to be confused with the dielectric constant of the electrostatic term), and σ_{ij} is the collision diameter (the separation distance corresponding to a null van der Waals potential) between the considered interaction sites (see Figure 5). Given the well depth and the collision diameter, it is possible to derive the minimum energy distance r^0_{ij} as the point where the derivative of Eq. (33) with respect to the separation distance is zero ($r^0_{ij} = 2^{1/6}\sigma_{ij}$). Thus, the Lennard-Jones potential can be alternatively expressed in terms of the minimum energy distance:

$$V_{ij}^{\mathrm{LJ}} = \epsilon_{ij}\left[\left(\frac{r^0_{ij}}{r_{ij}}\right)^{12} - 2\left(\frac{r^0_{ij}}{r_{ij}}\right)^{6}\right] \qquad 34)$$

The determination of the parameters needed to define the Lennard-Jones potential for all the atom types of a force field may be a cumbersome procedure. Therefore, the parameters needed to evaluate the cross interactions ($i \neq j$) are often derived from those available from pure atomic species ($i = j$) using simple combination (or *mixing*) rules. Within the Lorentz-Berthelot mixing rules, adopted by the Amber and CHARMM force fields, the collision diameter and the well depth are calculated by means of the arithmetic and geometric mean of the pure species, respectively:

$$\sigma_{ij} = \frac{\sigma_{ii} + \sigma_{jj}}{2} \qquad 35)$$

$$\epsilon_{ij} = \sqrt{\epsilon_{ii} \cdot \epsilon_{jj}} \qquad 36)$$

As a consequence, between the minimum energy distance and the van der Waals radii of the considered atoms holds the following relationship:

$$r^0_{ij} = \frac{r^0_{ii} + r^0_{jj}}{2} = r_i^{\mathrm{vdW}} + r_j^{\mathrm{vdW}} \qquad 37)$$

It must be noted that the previous mixing rule is not the only one adopted. For instance, both OPLS-AA and GROMOS employ the geometric mean to determine the collision diameter. The parameterization of the van der Waals terms strongly depends upon the purpose of the force field. If it is intended to reproduce gas-phase properties, the parameters are usually derived from crystallographic data. Conversely, in the case of condensed-phase force fields, van der Waals parameters are derived by reproducing experimental thermodynamic data. This strategy, first introduced by Jorgensen to develop a revised version of Amber-UA which led to the so called Amber-OPLS force field, had so strong an impact on the community that it has been adopted by all the condensed-phase force fields with only minor variations (55, 57-58, 68). In particular, collision diameters and well depths were empirically obtained by matching densities and enthalpy of

vaporization (calculated by means of Monte Carlo simulations) with the experimental observables for a series of organic liquids (hydrocarbons, alcohols, amides, thioesters, etc.) representing the functional groups (or atom types) found in neutral amino acids (58). Since the corresponding thermodynamic properties are clearly not available for the functional groups representing charged amino acids, experimental heats of hydration and QM-derived water interaction energies for ammonia and carboxylate derivatives were rather employed to parameterize those (58). This procedure directly leads to the derivation of an effective potential where the many-body effects may be considered accounted for in an average way.

Since non-bonded interactions only apply to atoms that are not directly involved in bonding relationships, atoms belonging to a 1-2 and 1-3 topological relationship (see Figure 2) are excluded from the non-bonded evaluation. In contrast, 1-4 atoms usually experience a scaled non-bonded interaction, in addition to the usual dihedral terms. In the OPLS-AA force field the 1-4 non-bonded interactions are scaled of a factor of 1/2, whereas in the Amber force field a differential scaling is adopted: 1/2 and 1/1.2 is employed for the van der Waals and the electrostatic interactions, respectively. No scaling factor is adopted in the CHARMM force field, whereas a system specific approach is used within the GROMOS force field. The van der Waals scaling is an attempt to correct the too steep repulsive term of the Lennard-Jones 12-6 potential compared to the more appropriate exponential form (proportional to $\exp(-r/r_0)$), which would be particularly appropriate for 1-4 atoms. Besides, the motivation for the electrostatic scaling lies in the fact that for 1-4 bonded atoms it is reasonable to expect some charge redistribution along the dihedral leading to a decrease in magnitude of the interactions. As a matter of fact, this latter is another example of accounting for polarization effects in additive force fields.

4.2. Force fields revisions and extensions

As it may be inferred from the previous discussion, force fields are not invariant entities, as they rather change over the years either for *revisions* or *extensions*. Usually, general force fields undergo revisions whenever an extension covering a larger atom type data set is needed. Indeed, gas-phase force fields are extremely optimized against available experimental data, limitations and pitfalls are known and well documented, therefore revisions over the same data set are in general not needed. Assessing the quality for specific force fields, such as those designed to model peptides, is not a trivial task. Moreover, the parameterization strategy of fitting around energy minima may be suboptimal when the system explores conformations which are far from those minima. As a matter of fact, to properly validate the performance of condensed-phase force fields an extensive configurational space sampling is required to allow a statistical comparison against experimental data (63). It is therefore reasonable that as soon as a significant increase in computational power occurs, allowing longer simulation and more extensive sampling, limitations in the available force fields are recognized and revisions invoked (78). Rather than a re-derivation of all the parameters in the light of the discovered deficiencies, the strategy to re-parameterize only dihedral terms exploiting their corrective meaning is more often preferred. Accordingly, by taking advantage of the increased computational power, the OPLS-AA amino acids dihedrals were completely re-derived at higher level of QM theory leading to the OPLS-AA/L force field (43, 79). Conversely, in the case of Amber ff94, deficiencies were mostly identified in the φ/ψ backbone dihedral parameters that were completely refitted giving rise to the ff99 and later to the ff99SB releases (41, 80-81). The recent ff99SB-ILDN is the result of a further refinement of the latter revision where the efforts were mostly focused towards corrections of the χ_1 side chain dihedrals for some critical amino acids (78). In parallel to these ff94 variants, it has to be mentioned that a new generation of the Amber force field based upon a higher level of QM theory for charges derivation was developed too (ff03) (82). A somewhat different strategy was employed to derive a revision for the CHARMM22 force field. In this case, as already mentioned, an energy correction map based on a grid-based interpolation procedure to better reproduce the QM-PES in the φ/ψ space is used, and for this reason the obtained release is referred to as CHARMM22-CMAP (66).

Aside from revisions, specific peptide force fields are also subjected to extensions in their range of applicability. A comprehensive overview of the attempt to incorporate chemical species other than peptides in the most popular condensed-phase force field may be found in the literature (83). As a short summary, it should be recalled that parameters for nucleic acids were present in the Amber force field since the earlier versions (Amber-AA), even though they have been consistently revised over the years (59). In contrast, CHARMM27 was intentionally developed to provide a set of nucleic acids parameters consistent with the underlying protein force field (84-85). On the other side, lipids have been extensively studied and parameterized to be used with the united atom version of the GROMOS force field, whereas only recently a GROMOS96 lipid dataset was released (86-87). A similar evolution holds for the CHARMM force field, where lipids were supported since earlier versions (CHARMM19), even though only later a thorough re-parameterization was undertaken (CHARMM36) (60, 88). Very recently, a consistent set of lipid parameters to be used together with the Amber force field has been developed too (89-90). In the field of drug design, and in particular to study protein-ligand interactions, the availability of parameters to treat small organic molecules consistently with the peptide counterpart is obviously of pivotal importance. For all the most important condensed-phase force fields an impressive effort has been done in order to cover a remarkably broad range of chemical space. From this perspective, virtually all condensed-phase force fields, such as OPLS-AA, Amber (GAFF), and CHARMM (CGenFF), come with a real general force field companion (91-93).

4.3. Polarizable force fields

Electronic polarization consists in the redistribution of the electron density in response, *reaction*,

to an external electric field. Under the hypothesis of *linear response*, which is usually, but not always, assumed, the effect of such reaction can be schematized as the induction of dipoles at several interaction sites, in the considered molecule, having a magnitude proportional to the local field $\mathbf{E}^{\mathrm{tot}}$:

$$\boldsymbol{\mu}_{\mathrm{ind}} = \boldsymbol{\alpha}\mathbf{E}^{\mathrm{tot}} \tag{38}$$

where $\boldsymbol{\alpha}$ is the electronic polarizability tensor of the molecule under investigation. It is worth recalling that by *linear response* one intends that the reaction field is locally proportional to the permanent field and, by consequence, also to the total field, which is the sum of the two. This model can be formalized by performing an hypothetical Taylor expansion of the induced field as a function of the total field truncated at the first order (94-95). This procedure is completely different from the so-called *multipole expansion*, which is often done in electrostatics. The latter can also be an expansion of an electric field, although it more often involves the potential, but it is done with respect to the inverse distance to a point in space, regardless of the nature of the field, whether it is permanent or induced, or total.

From an inter-molecular standpoint, polarizability is closely related to non-additive and cooperative effects, and it can be modeled in traditional force field only in an average way with effective pair potentials. The development of polarizable force fields is an active area of research and for all the major condensed-phase additive force fields a polarizable extension has been developed too. However, the process to add polarizability to the already existing terms without a reparameterization may become problematic, and for this reason more recent force fields (such as AMOEBA) include the polarization term consistently since the beginning of their development (77, 96-97).

In the following, we present four models that incorporate electronic polarizability in classical force fields: (*i*) the Continuum Dielectric model (*ii*) the Point Polarizable Dipoles, (*iii*) the Drude oscillator model, and (*vi*) the Fluctuating Charge model. In any of these approaches, polarizable terms are schematized as a supplementary polarization contribution added to the familiar bonded and non-bonded terms: $V_{\mathrm{FF}} = V_{\mathrm{b}} + V_{\mathrm{nb}} + V_{\mathrm{pol}}$. In this description, the potential energy accounting for the permanent electrostatic interactions is included in the V_{nb} term.

4.3.1. Continuum dielectric model

The Continuum Dielectric (CD) model is probably the simplest approach to polarizability, although not the most widely used. It assumes there is a linear response of the molecule to the local electric field, that the induced dipoles are *point-like* and, unlike other approaches, that they are spread over the matter rather than being localized at given sites. A *point-like* dipole is such that the absolute value of its positive and negative charges, which are equal in magnitude, tends to infinity while their distance tends to zero in a way that the product of the two, i.e. the dipole moment, stays constant. The CD model borrows its formalism from the traditional electrostatics of continuum media, which provides expressions for calculating the electrostatic energy of polarizable systems (98). In this description, the polarization energy can be singled out from the total electrostatic energy as follows:

$$\begin{aligned} V_{\mathrm{elect}} &= \frac{1}{2}\int_{\mathbb{R}^3} \rho^{\mathrm{perm}}\phi^{\mathrm{tot}}\,\mathrm{d}\mathbf{x} = \\ &= \frac{1}{2}\int_{\mathbb{R}^3} \rho^{\mathrm{perm}}\phi^{\mathrm{perm}}\,\mathrm{d}\mathbf{x} + \frac{1}{2}\int_{\mathbb{R}^3} \rho^{\mathrm{perm}}\phi^{\mathrm{pol}}\,\mathrm{d}\mathbf{x} = \\ &= V_{\mathrm{perm}} + V_{\mathrm{pol}} \end{aligned} \tag{39}$$

This expression decomposes electrostatic energy in two contributions (here identified by the superscripts *perm* and *pol*, respectively), that due to the permanent charge distribution (which is the one considered in the traditional force fields), and that due to polarization effects, also called *reaction field energy* since Φ^{pol} is the potential generated by the polarization charge in reaction to the local electric field (98). By using Green's lemma the polarization term can be rewritten as:

$$V_{\mathrm{pol}} = \frac{1}{2}\int_{\mathbb{R}^3} \rho^{\mathrm{pol}}\phi^{\mathrm{perm}}\,\mathrm{d}\mathbf{x} \tag{40}$$

A suitable manipulation of Eq. (40), performed by addition and subtraction of the total electric potential $\Phi^{\mathrm{tot}} = \Phi^{\mathrm{perm}} + \Phi^{\mathrm{pol}}$, allows a partitioning which leads to a deeper physical interpretation of this term:

$$\begin{aligned} V_{\mathrm{pol}} &= \frac{1}{2}\int_{\mathbb{R}^3} \rho^{\mathrm{pol}}\left(\phi^{\mathrm{perm}} + \phi^{\mathrm{tot}} - \phi^{\mathrm{tot}}\right)\,\mathrm{d}\mathbf{x} = \\ &= \int_{\mathbb{R}^3} \rho^{\mathrm{pol}}\phi^{\mathrm{perm}}\,\mathrm{d}\mathbf{x} - \frac{1}{2}\int_{\mathbb{R}^3} \rho^{\mathrm{pol}}\phi^{\mathrm{tot}}\,\mathrm{d}\mathbf{x} + \\ &\quad + \frac{1}{2}\int_{\mathbb{R}^3} \rho^{\mathrm{pol}}\phi^{\mathrm{pol}}\,\mathrm{d}\mathbf{x} \end{aligned} \tag{41}$$

The last three terms of Eq. (41) represent the interaction of the polarization charge with the permanent field (V_{stat}), the always unfavorable energy contribution required to distort the electron distribution to create the polarization (V_{dist}), and the mutual interaction between the polarization charges ($V_{\mathrm{ind/ind}}$), respectively.

Expressions in Eqs. (39) to (41) are quite widely applicable, and are consistent with the general relationship, $\mathbf{D} = \varepsilon_0\,\mathbf{E} + \mathbf{P}$, between displacement, electric and polarization fields. The continuum description of linear dielectric media specializes the latter relationship so as that a direct proportionality between them holds: $\mathbf{D} = \varepsilon_0\varepsilon_r\mathbf{E}$, $\mathbf{P} = \varepsilon_0(\varepsilon_r - 1)\mathbf{E}$. Here, ε_r is the, possibly space-varying, relative dielectric constant of the medium, accounting for how strongly it polarizes (although the dielectric constant is most often considered as a scalar quantity, in principle it could be a tensor, just like the polarizability tensor, to which the former is closely related).

When molecules in solution are studied, there are CD approaches that keep the traditional atomistic description for the solute and treat only the solvent as a continuum, while other methods model both the solute and the solvent as polarizable media, with different dielectric constants. This formalism is extended to consider also the ionic strength of an aqueous solution in the Poisson-Boltzmann equation, the solution of which allows the derivation of the so-called reaction field (99-101). Alternative to the numerical solution of the Poisson-Boltzmann equation is the set of analytical methods based on the generalized Born formalism (102-104). Fundamental is the choice of the dielectric constant value, that in this kind of approaches is an effective parameter quantifying the entity of the response and accounting for all the rearrangements that are not described explicitly. Its actual value is debated and in some applications also different dielectric values are used to reproduce inhomogeneous polarizabilities (105-106).

The resulting electrostatic energy is a free energy in the sense that the electrostatic contribution of most of the degrees of freedom, namely those giving rise to water molecular polarizability, salt effects and those that are accounted for by the dielectric constant value, are implicitly equilibrium averaged. The most advanced derivation of a force field describing a polarizable solute in an implicit solvent are due to Tan and Luo (107-108).

In the Poisson-Boltzmann approach, the polarization charge which originates Φ^{pol} can be derived, for instance, by using Gauss' law for the **P** field once the equation is solved (for details, see (101)). One advantage of this model with respect to others, such as the PPD and the Drude models, is that it avoids the so-called *polarization catastrophe*. The latter is a phenomenon that can occur when neighboring point dipoles interact with each other. It can be partially corrected by an *ad hoc* damping (109).

4.3.2. Polarizable point dipoles model

The point polarizable dipoles (PPD, also known as induced point dipoles) is so far the most popular method to include polarizability in force fields, being adopted by Amber (ff02 (110-111) and ff02EP (112)), OPLS-AA (OPLS-PFF (113-115)), and by a CHARMM22 extension (PIPF-CHARMM (116)) that is under development. Moreover, the already mentioned AMOEBA force field (77, 96-97), which includes the polarization term since its foundation, also adopts an induced point dipole approach. The PPD model represents the reaction to the local field as a set of inducible dipoles (variable in magnitude and direction) as in equation (38), located in predetermined sites. These induced dipoles generate a field which algebraically adds up to the permanent field. It can be interesting to find a formal correspondence between this model and the expressions given in the continuum dielectric model. This can be done by assuming that the polarization field vector **P**, the physical dimension of which is dipole moment per volume unit, has the following form:

$$\mathbf{P} = \sum_i \boldsymbol{\mu}_i \delta(\mathbf{x} - \mathbf{x}_i) \tag{42}$$

where $\mathbf{x}_i$ are the positions of the interacting polarizable sites. Assuming that the distribution of permanent charges is finite in space, and using Gauss' law for polarization charge, a useful identity can be found:

$$\begin{aligned} 0 &= \frac{1}{2}\int_{\mathbb{R}^3} \boldsymbol{\nabla}\cdot(\mathbf{P}\phi)\,\mathrm{d}\mathbf{x} = \\ &= \frac{1}{2}\int_{\mathbb{R}^3} (\mathbf{P}\boldsymbol{\nabla}\phi + \phi\boldsymbol{\nabla}\cdot\mathbf{P})\,\mathrm{d}\mathbf{x} = \\ &= -\frac{1}{2}\int_{\mathbb{R}^3} \mathbf{P}\mathbf{E}\,\mathrm{d}\mathbf{x} - \frac{1}{2}\int_{\mathbb{R}^3} \rho^{\text{pol}}\phi\,\mathrm{d}\mathbf{x} \end{aligned} \tag{43}$$

This expression holds regardless of the specific nature of the potential Φ provided that the appearing electric field **E** corresponds to the negative of its gradient. By applying Eq. (43) to Φ^{perm}, by substituting the expression in Eq. (42) for **P**, and by means of Eq. (40), one obtains the following expression for the polarization energy term in the PPD model:

$$V_{\text{pol}} = -\frac{1}{2}\int_{\mathbb{R}^3} \mathbf{P}\mathbf{E}^{\text{perm}}\,\mathrm{d}\mathbf{x} = -\frac{1}{2}\sum_i \boldsymbol{\mu}_i \mathbf{E}_i^{\text{perm}} \tag{44}$$

where the subscript i to the electric field $\mathbf{E}_i$ indicates the field at the point $\mathbf{x}_i$.

Moreover, by applying Eq. (43) also to Φ^{tot}, by using again the expression in Eq. (42) for **P**, the energetic partitioning in Eq. (41) can be tailored to the PPD model:

$$\begin{aligned} V_{\text{pol}} &= V_{\text{stat}} + V_{\text{dist}} + V_{\text{ind/ind}} = \\ &= \sum_i \boldsymbol{\mu}_i \left(-\mathbf{E}_i^{\text{perm}} + \frac{1}{2}\mathbf{E}_i^{\text{tot}} - \frac{1}{2}\mathbf{E}_i^{\text{pol}}\right) \end{aligned} \tag{45}$$

It can be useful to recast the three components of V_{pol} as follows. Assuming that, as in the simplest case, the permanent field $\mathbf{E}_i^{\text{perm}}$ is generated only by point fixed charges, V_{stat} can be expressed as:

$$V_{\text{stat}} = \frac{1}{4\pi\epsilon_0}\sum_{i,j\neq i} \boldsymbol{\mu}_i \frac{q_j(\mathbf{r}_j - \mathbf{r}_i)}{r_{ij}^3} \tag{46}$$

where the induced dipole at the i^{th} polarizable site does not interact with the fixed charge at the same site. Similarly, V_{dist} can be written as:

$$\begin{aligned} V_{\text{dist}} &= \frac{1}{2}\sum_i \boldsymbol{\mu}_i \mathbf{E}_i^{\text{tot}} = +\frac{1}{2}\sum_i \boldsymbol{\mu}_i \boldsymbol{\alpha}(\mathbf{x}_i)^{-1}\boldsymbol{\mu}_i = \\ &= \sum_i \frac{\boldsymbol{\mu}_i^2}{2\alpha(\mathbf{x}_i)} \end{aligned} \tag{47}$$

where the last equality holds only if the polarizability tensor is isotropic and diagonal, and $\alpha(\mathbf{x}_i)$ is its diagonal element. Consistently with intuition, the energy needed to create the dipoles grows with the intensity of the dipole moments and decreases with the local polarizability. The set of diagonal elements per interaction site i of the

polarizability tensor are the parameters needed by this model, and can be estimated either experimentally or theoretically. Finally, it is possible to rewrite $V_{\text{ind/ind}}$ according to:

$$V_{\text{ind/ind}} = -\frac{1}{2}\sum_i \boldsymbol{\mu}_i \mathbf{E}_i^{\text{pol}} = \frac{1}{2}\sum_{i,j\neq i} \boldsymbol{\mu}_i \mathbf{T}_{ij} \boldsymbol{\mu}_j \quad 48)$$

where $\mathbf{T}_{ij}$ is the dipole interaction tensor, which relates the electric field generated by a dipole to its dipole moment, and that can be derived from geometrical considerations. As already mentioned, a pathological behavior affecting the point polarizable model consists in a singularity arising in the induced dipoles at a small distance from the interaction sites (polarization catastrophe). This undesirable effect may be avoided either by excluding short range interactions or by attenuating them using a distance dependent switching function (96, 110-111, 113-116).

The last step needed to implement the PPD model consists in determining the set of all induced dipole moments $\{\boldsymbol{\mu}_i\}$. One way to do it is by solving for $\{\boldsymbol{\mu}_i\}$ the following relationship, which can be obtained by instantiating Eq. (38) in the case of point dipoles and by decomposing the total field in its two contributions:

$$\begin{aligned}\boldsymbol{\mu}_i &= \boldsymbol{\alpha}(\mathbf{x}_i)\mathbf{E}_i^{\text{tot}} = \boldsymbol{\alpha}(\mathbf{x}_i)\left(\mathbf{E}_i^{\text{perm}} + \mathbf{E}_i^{\text{pol}}\right) = \\ &= \boldsymbol{\alpha}(\mathbf{x}_i)\left(\mathbf{E}_i^{\text{perm}} - \sum_{j\neq i}\mathbf{T}_{ij}\boldsymbol{\mu}_j\right)\end{aligned} \quad 49)$$

The very same expression can be obtained by suitably recasting Eq. (45), differentiating it with respect to $\{\boldsymbol{\mu}_i\}$ and imposing that the result is null so as to identify the dipole configuration that minimizes the energy. Expression in Eq. (49) is obviously consistent with the PPD model, where the electric field at the i^{th} site due to polarization can be expressed as the sum of the fields generated by all the other induced dipoles, which are expressed through the tensor $\mathbf{T}_{ij}$. Conversely, $\mathbf{E}^{\text{perm}}$ is generated by all permanent field sources, which can be point charges or other permanent multipoles. For instance, the Amber ff02 (110-111) and ff02EP (112) use only permanent atomic and off-site charges, OPLS-PFF (113-115) also employs permanent dipoles, whereas an expansion up to the quadrupole is considered in the AMOEBA (77, 96-97) force field.

Eq. (49) represents a system of linear, non-homogenous, equations relating each induced dipole to all the remaining ones. To solve this system, three methods are available: (*i*) the analytical approach by matrix inversion, (*ii*) a self-consistent iterative optimization approach, and (*iii*) the extended Lagrangian approach, where auxiliary degrees of freedom (in this case the dipole momenta), endowed with fictitious masses and velocities, are introduced in the Lagrangian of the system, and the extended system is evolved together with the solution of the global equations of motion. The reader is redirected to specific reviews for more details (117-118).

4.3.3. Drude oscillator model

The Drude oscillator model is also known as shell model: the former nomenclature is mostly employed in the community of condensed-phase simulations, whereas the latter is preferred in the solid state physics field. The basic principle is the same, and it goes back to the Drude theory of polarizability (1902) which was originally derived to describe London forces. In the context of force fields, the Drude oscillator model has been extensively used to develop polarizable water models, and it has also been employed in a CHARMM27 extension for nucleic acids (119-120). Within this model, each polarizable center is described by a pair of interaction sites connected by a harmonic spring of elastic constant k_i. The core site is located at the nuclear position and it is associated to a partial charge $q_i^C = z_i + q_i^{\text{pol}}$, whereas the Drude particle site consists in a partial charge $q_i^D = -q_i^{\text{pol}}$ so as to preserve the formal atomic charge z_i. Polarizability is thus accounted for by the relative displacement of these sites by mimicking the valence electron density redistribution occurring in atoms subjected to external electric fields. Similarly to the PPD model, the corresponding energy partitioning can be derived from the continuum electrostatics theory. It is interesting to observe that, although this model engenders dipoles, which have the form $\boldsymbol{\mu}_i = q_i^{\text{pol}}(\mathbf{r}_i^C - \mathbf{r}_i^D) = -q_i^{\text{pol}}\,\mathbf{d}_i$, there is no constraint on the fact that they can be considered as *point-like* dipoles. Therefore, the assumption done in Eq. (42) cannot be done. The electrostatic interaction among non polarizable charged centers can rather be evaluated by the classical Coulomb potential as in non polarizable force fields. Moreover, being primarily electronic in nature, van der Waals interaction sites are centered on the Drude particles rather than into the cores.

As per polarization, by applying Eq. (40), we obtain the following expression for the reaction field energy corresponding to the Drude model:

$$V_{\text{pol}} = \frac{1}{8\pi\epsilon_0}\sum_i^N\sum_{j\neq i}^N q_i^{\text{pol}} z_j \left(\frac{1}{|r_i^C - r_j^C|} + - \frac{1}{|\mathbf{d}_i + r_i^C - r_j^C|}\right) \quad 50)$$

Similarly, the two following energy contributions can be derived:

$$V_{\text{stat}} = \frac{1}{4\pi\epsilon_0}\sum_i^N\sum_{j\neq i}^N q_i^{\text{pol}} z_j \left(\frac{1}{|r_i^C - r_j^C|} + - \frac{1}{|\mathbf{d}_i + r_i^C - r_j^C|}\right) \quad 51)$$

$$V_{\text{ind/ind}} = \frac{1}{8\pi\epsilon_0}\sum_i^N\sum_{j\neq i}^N q_i^{\text{pol}} q_j^{\text{pol}} \left(\frac{1}{|r_i^C - r_j^C|} + - \frac{1}{|\mathbf{d}_i + r_i^C - r_j^C|} - \frac{1}{|r_i^C - r_j^C - \mathbf{d}_i|} + - \frac{1}{|r_i^C - r_j^C + \mathbf{d}_i - \mathbf{d}_j|}\right) \quad 52)$$

The remaining term, V_{dist}, could also be calculated by means of the same strategy. However, one should consider that q_i^{pol} and $\mathbf{d}_i$ are not independent parameters. Moreover, as Eqs. (50) and (51) show, interactions between the core and the Drude particle belonging to the same polarizable center are excluded. Indeed, those interactions account for the Drude expression of the V_{dist} term, that, according to this model, can be cast as follows:

$$V_{\text{dist}} = \frac{1}{2}\sum_{i}^{N} k_i \mathbf{d}_i^2 = \frac{1}{2}\sum_{i}^{N} k_i \frac{\boldsymbol{\mu}_i^2}{q_i^{\text{pol}\,2}} \qquad 53)$$

By comparing Eq. (53) to Eq. (47), one can find the equivalent, isotropic and diagonal, polarizability induced by the Drude model:

$$\alpha(\mathbf{x}_i) = \frac{q_i^{\text{pol}\,2}}{k_i} \qquad 54)$$

From Eq. (54) one sees that the partial charge and the force constant of the Drude sites may be derived from polarizability values obtained either experimentally or theoretically. In any case, it must be recognized that q_i^{pol} and k_i are not independent parameters, and their relative value may be chosen by convenience considerations. In the case of the Drude oscillator model, the dipole moments are calculated via self-consistent solution of the system of equations obtained by nullifying the gradient of the energy with respect to $\{\mathrm{d}_i = \left| \boldsymbol{r}_i^C - \boldsymbol{r}_i^D \right| \}$.

The Drude oscillator model is particularly suited to be coupled with the extended Lagrangian approach. In this case, fictitious masses are assigned to the Drude particle sites. By using this approach care must be taken in the choice of spring constants and masses because of their impact on the integration time step of the overall dynamics.

4.3.4. Fluctuating charge model

All the polarizable models considered so far can be reconciled in a general framework for treating polarizability within a classical continuum electrostatics framework. Based on these premises, in previous sections we were able to clearly identify the V_{stat}, $V_{\text{ind/ind}}$, and V_{dist} components of the polarization energy term following a unified procedure for CD, PPD and the Drude oscillator model. Unlike these methods, the Fluctuating Charge (FQ) model is rather based upon a different perspective which, historically, goes back to the development of earlier derivation schemes to compute partial charges. The theory behind the FQ model is grounded in the Sanderson's principle of electronegativity equalization (1951), which states: "When two or more atoms initially different in electronegativity combine chemically, their electronegativities become equalized in the molecule." For this reason, the method is also known as *electronegativity equalization* model. The FQ model was firstly employed in force fields by Rappé and coworkers in the development of their Universal Force Field (UFF) (121-122). More recently, it has also been implemented in OPLS-AA and CHARMM22 extensions (123-125). The approach is based upon the idea that charge can flow within a molecule and find a configuration where electronegativity is balanced. In this context, polarization is intrinsically taken into account. Since there is no compulsory locality in the displacement of the equalizing charges and the distinction in inducing and induced charges is lacking, or at least less evident, the formalism of classical continuum electrostatics is more difficult to apply in this context.

The energy required to introduce a charge q on a neutral atom can be Taylor expanded at the second order (126):

$$V(q) = V^0 + \left(\frac{\partial V}{\partial q}\right)_0 q + \frac{1}{2}\left(\frac{\partial^2 V}{\partial q^2}\right)_0 q^2 \qquad 55)$$

According to this expression, the energy of the unperturbed state is V^0 and the energy required to introduce a formal charge on an atom can be calculated by substituting q in Eq. (55) with the added charge. By taking advantage of the definitions of *ionization potential* (IP = $V(+1) - V(0)$), being the energy required to extract an electron from a neutral atom in the ground state ($\mathrm{A} \rightarrow \mathrm{A}^+ + \mathrm{e}$), and *electron affinity* (EA = $V(0) - V(-1)$), being the energy required to extract an electron from a singly charged negative ion ($\mathrm{A}^- \rightarrow \mathrm{A} + \mathrm{e}$), it is possible to derive useful expressions for the linear and quadratic coefficients of the Taylor expansion:

$$\left(\frac{\partial V}{\partial q}\right)_0 = \frac{\mathrm{IP} + \mathrm{EA}}{2} = \chi^0 \qquad 56)$$

$$\left(\frac{\partial^2 V}{\partial q^2}\right)_0 = \mathrm{IP} - \mathrm{EA} = 2\eta^0 \qquad 57)$$

where χ^0 is the Mulliken absolute electronegativity and η^0 is the absolute hardness, that is the ability of an isolated atom to attract electrons and its resistance upon charge flow, respectively (127-128). Eq. (55) can be rearranged to describe the energy of a group of N atoms belonging to a molecule:

$$V(q) = V_0 + \sum_{i}^{N}(\chi_i^0 q_i + \eta_{ii}^0 q_i^2) + \sum_{i}^{N}\sum_{j\neq i}^{N} 2\eta_{ij}^0(r_{ij}) q_i q_j \qquad 58)$$

In Eq. (58), the generic potential energy term V_0 encompasses all the bonded and van der Waals force field contributions. The off-diagonal hardness coefficients $2\eta_{ij}^0$ can be taken to be proportional to r_{ij}^{-1}, so as that the third and last term of the right hand side of Eq. (58) recovers the standard Coulomb potential. Moreover, since one is here considering the energy required to add a charge q to atoms that are subjected to the electric field generated by the already existing charges, it is possible to relate the second term of Eq. (58) to the V_{dist} energy of the previous models. Eq. (58) is solved by nullifying the gradient of the energy

with respect to the $\{q_i\}$ under the additional constraint of total charge conservation ($\Sigma q_i = q_{tot}$):

$$\frac{\partial V}{\partial \mathbf{q}_i} \equiv \chi_i = \chi_i^0 + 2\eta_{ii}^0 q_i + \sum_{j \neq i}^{N} 2\eta_{ij}^0 (r_{ij}) q_j = 0 \quad 59)$$

At the minimum of this energy, the partial derivatives are null and the electronegativities of the atoms in the molecular environment are equalized (in contrast to what happens when atoms are isolated), from whence the name of electronegativity equalization principle. Remarkably, Parr and coworkers have shown that the Mulliken electronegativity closely approximates the negative of the chemical potential for the ground state of a many-electron system in Density Functional Theory (129). In this context, applying Eq. (59) to derive a polarized set of charges for a molecule is equivalent to minimize the chemical potential of the system, which provides a physically sound interpretation to the Sanderson's principle and to the whole procedure as well. This should not be overrated though, since as in any force field-based model the quality of the results is strongly dependent upon the parameterization. Clearly, the parameters in the FQ model are represented by atomic electronegativity and hardness, which may be derived either from experimental data or, more often, from QM calculations.

It is interesting to note that in principle Eq. (58) could be generalized to consider also an inter-molecular charge flow, and for this reason it is sometimes claimed that FQ is the only polarizable model able to account for charge transfer effects. However, this statement should not be taken too literally since it must be kept in mind that this is still done in a classical force field framework. An evident advantage of the FQ model relies on the fact that, differently from PPD and the Drude oscillator models, the number of interaction sites is the same as in non polarizable force fields, thus the overall computational cost is relatively limited.

From the above discussion, it emerges that the FQ model bears some resemblance with charge derivation procedures, with the important difference that the charges have to be re-derived for *each* configuration sampled by the MD or MC engine. From this standpoint, it is appealing to compare the FQ model with the Gasteiger-Marsili scheme, which has become quite popular, especially in the pharmaceutical community, as a fast method to calculate partial charges. The Gasteiger-Marsili method (also known as Partial Equalization of Orbital Electronegativity) can be thought of as an approximate and empirical solution to Eq. (59), where the orbital electronegativity (defined as the change in energy due to the variation of the occupation of that orbital) of the i^{th} atom in the molecule is expressed as:

$$\chi_i = a_i + b_i q_i + c_i q_i^2 \quad 60)$$

where a_i is the Mulliken electronegativity of the isolated atom (as in Eq, 59), while b_i and c_i are other empirical parameters derived by the experimental ionization potentials and electron affinities (69). Rather than fully equalizing the electronegativities of directly bonded atoms (which led to chemically unacceptable results in earlier analogous implementations), the Gasteiger-Marsili scheme conceives a successively damped charge transfer from the less electronegative atom *i* to the more electronegative atom *j* according to the formula:

$$q_{i \to j}^k = \frac{\chi_j^k - \chi_i^k}{\chi_i^+} \left(\frac{1}{2}\right)^k \quad 61)$$

where χ_i^+ is the orbital electronegativity of the positive ion of the less electronegative species, and the superscript k stands for the iteration step. Eq. (61) is solved under the usual total charge conservation constraint. Because of the damping factor $(1/2)^k$, as the iterations proceed the charge transfer becomes more and more reduced, simulating a resistance in charge flow growing with the increasing electrostatic field (69). The charges derived by the Gasteiger-Marsili method depend on the parameters $\{a_i\}$, $\{b_i\}$, and $\{c_i\}$, and on the connectivity of the atoms in the considered molecule, but they do not depend upon the conformation of the molecule itself. For this reason, these charges are often referred to as *topological charges*. In general, the method is no longer considered accurate enough to be used with current force fields, nor it is considered theoretically sound due to the requirement of a full equalization of electronegativity as demonstrated by Parr and coworkers (129).

5. CONCLUSIONS

In this review, we treat two among the most important bases of molecular simulation, namely Molecular Dynamics and Molecular Mechanics. In the first part, specific attention was devoted to the numerical aspects related to the integration of the equations of motion and to the algorithms used to impose the control variables, such as temperature and pressure, that characterize the most frequently used statistical mechanical ensembles. In the second part, the main elements of the force fields are described in some detail. A specific section is devoted to the polarizable force fields, which represent an often necessary extension of the Molecular Mechanics aimed at improving its accuracy.

6. ACKNOWLEDGEMENTS

The Authors gratefully acknowledge the IIT platform "*Computation*". Funding from the NIGMS, NIH, grant number, 1R01GM093937-01 is also acknowledged.

7. REFERENCES

1. W. F. van Gunsteren, D. Bakowies, R. Baron, I. Chandrasekhar, M. Christen, X. Daura, P. Gee, D. P. Geerke, A. Glättli, P. H. Hünenberger, M. A. Kastenholz, C. Oostenbrink, M. Schenk, D. Trzesniak, N. F. A. van der Vegt and H. B. Yu: Biomolecular Modeling: Goals, Problems, Perspectives. *Angew Chem Int Ed*, 45(25), 4064-4092 (2006)

2. N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller and E. Teller: Equation of State Calculations by Fast Computing Machines. *J Chem Phys*, 21(6), 1087-1092 (1953)

3. B. J. Alder and T. E. Wainwright: Phase Transition for a Hard Sphere System. *J Chem Phys*, 27(5), 1208-1209 (1957)

4. B. J. Alder and T. E. Wainwright: Studies in Molecular Dynamics. I. General Method. *J Chem Phys*, 31(2), 459-466 (1959)

5. M. P. Allen and D. J. Tildesley. Computer Simulation of Liquids. Oxford University Press, New York (1987)

6. D. Frenkel and B. Smit. Understanding Molecular Simulation. Academic Press (1996)

7. L. Verlet: Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys Rev*, 159(1), 98-103 (1967)

8. R. W. Hockney. Potential calculation and some applications. In: *Methods in Computational Physics: Advances in Research and Applications*. Elsevier (1970)

9. W. C. Swope, H. C. Andersen, P. H. Berens and K. R. Wilson: A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. *J Chem Phys*, 76(1), 637-649 (1982)

10. W. B. Streett, D. J. Tildesley and G. Saville: Multiple time-step methods in molecular dynamics. *Mol Phys*, 35(3), 639 - 648 (1978)

11. G. J. Martyna, M. E. Tuckerman, D. J. Tobias and M. L. Klein: Explicit reversible integrators for extended systems dynamics. *Mol Phys*, 87(5), 1117-1157 (1996)

12. 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. *J Comput Phys*, 23(3), 327-341 (1977)

13. H. C. Andersen: Rattle: A "velocity" version of the shake algorithm for molecular dynamics calculations. *J Comput Phys*, 52(1), 24-34 (1983)

14. P. J. Steinbach and B. R. Brooks: New spherical-cutoff methods for long-range forces in macromolecular simulation *J Comput Chem*, 15, 667-683 (1994)

15. L. Greengard and V. Rokhlin: A fast algorithm for particle simulations. *J Comput Phys*, 73, 325-348 (1987)

16. K. E. Schmidt and M. A. Lee: Implementing the fast multipole method in three dimensions. *J Stat Phys*, 63, 1223-1235 (1991)

17. S. W. De Leeuw, J. W. Perram and E. R. Smith: Simulation of electrostatic systems in periodic boundary conditions. I. Lattice sums and dielectric constants. *Proc R Soc London, Ser A*, 373, 27-56 (1980)

18. S. W. De Leeuw, J. W. Perram and E. R. Smith: Simulation of electrostatic systems in periodic boundary conditions. II. Equivalence of boundary conditions. *Proc R Soc London, Ser A*, 373, 56-66 (1980)

19. S. W. De Leeuw, J. W. Perram and E. R. Smith: Simulation of electrostatic systems in periodic boundary conditions. III. Further theory and applications. *Proc R Soc London, Ser A*, 388, 177-193 (1983)

20. J.-P. Hansen: Molecular-dynamics simulations of Coulomb systems in two and three dimensions. In: *Molecular Dynamics Simulations of Statistical Mechanics Systems*. Ed G. C. a. W. G. Hoover. North Holland, Amsterdam (1986)

21. W. H. Press, B. P. Flannery, S. A. Teukolsky and W. T. Vetterling. Numerical Recipes: The Art of Scientific Computing. Cambridge University Press, Cambridge (1986)

22. R. W. Hockney and J. W. Eastwood. Computer Simulations Using Particles. McGraw Hill, New York (1981)

23. T. Darden, D. York and L. Pedersen: Particle mesh Ewald: An N log(N) method for Ewald sums in large systems. *J Chem Phys*, 98, 10089-10092 (1993)

24. M. Deserno and C. Holm: How to mesh up Ewald sums. I. A theoretical and numerical comparison of various particle mesh routines. *J Chem Phys*, 109, 7678-7693 (1998)

25. G. Brancato, N. Rega and V. Barone: Reliable molecular simulations of solute-solvent systems with a minimum number of solvent shells. *J Chem Phys*, 124, 214505-214513 (2006)

26. W. Im, S. Berneche and B. Roux: Generalized solvent boundary potential for computer simulations. *J Chem Phys*, 114(7), 2924-2937 (2001)

27. P. H. Hünenberger: Thermostat algorithms for molecular dynamics simulations. *Adv Polymer Sci*, 173, 105-149 (2005)

28. L. V. Woodcock: Isothermal molecular dynamics calculations for liquid salts. *Chem Phys Lett*, 10(3), 257-261 (1971)

29. H. J. C. Berendsen, J. P. M. Postma, W. F. v. Gunsteren, A. DiNola and J. R. Haak: Molecular dynamics with coupling to an external bath. *J Chem Phys*, 81(8), 3684-3690 (1984)

30. H. C. Andersen: Molecular dynamics simulations at constant pressure and/or temperature. *J Chem Phys*, 72(4), 2384-2393 (1980)

31. S. Nosé: A molecular dynamics method for simulations in the canonical ensemble. *Mol Phys*, 52, 255-268 (1984)

32. S. Nosé: A unified formulation of the constant temperature molecular dynamics methods. *J Chem Phys*, 81(1), 511-519 (1984)

33. W. G. Hoover: Canonical dynamics: Equilibrium phase-space distributions. *Phys Rev A: At Mol Opt Phys*, 31(3), 1695-1697 (1985)

34. G. J. Martyna, M. L. Klein and M. Tuckerman: Nosé-Hoover chains: The canonical ensemble via continuous dynamics. *J Chem Phys*, 97(4), 2635-2643 (1992)

35. T. Schneider and E. Stoll: Molecular-dynamics study of a three-dimensional one-component model for distortive phase transitions. *Phys Rev B: Condens Matter*, 17(3), 1302-1322 (1978)

36. S. C. Harvey, R. K. Z. Tan and T. E. Cheatham: The flying ice cube: Velocity rescaling in molecular dynamics leads to violation of energy equipartition. *J Comput Chem*, 19(7), 726-740 (1998)

37. M. Parrinello and A. Rahman: Crystal Structure and Pair Potentials: A Molecular-Dynamics Study. *Phys Rev Lett*, 45(14), 1196-1199 (1980)

38. M. Parrinello and A. Rahman: Polymorphic transitions in single crystals: A new molecular dynamics method. *J Appl Phys*, 52(12), 7182-7190 (1981)

39. J. D. van der Waals: On the continuity of the gas and liquid state. PhD Thesis, Leiden (The Netherlands) (1873)

40. P. Geerlings, F. De Proft and W. Langenaeker: Conceptual Density Functional Theory. *Chem Rev*, 103, 1793-1873 (2003)

41. W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell and P. A. Kollman. *J Am Chem Soc*, 117, 5179 (1995)

42. A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin and M. Karplus: All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J Phys Chem B*, 102(18), 3586-3616 (1998)

43. W. L. Jorgensen, D. S. Maxwell and J. Tirado-Rives: Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J Am Chem Soc*, 118(45), 11225-11236 (1996)

44. X. Daura, A. E. Mark and W. F. V. Gunsteren: Parametrization of aliphatic CHn united atoms of GROMOS96 force field. *J Comput Chem*, 19(5), 535-547 (1998)

45. N. L. Allinger: Conformational analysis. 130. MM2. A hydrocarbon force field utilizing V1 and V2 torsional terms. *J Am Chem Soc*, 99(25), 8127-8134 (1977)

46. N. L. Allinger, Y. H. Yuh and J. H. Lii: Molecular mechanics. The MM3 force field for hydrocarbons. 1. *J Am Chem Soc*, 111(23), 8551-8566 (1989)

47. J. H. Lii and N. L. Allinger: Molecular mechanics. The MM3 force field for hydrocarbons. 2. Vibrational frequencies and thermodynamics. *J Am Chem Soc*, 111(23), 8566-8575 (1989)

48. J. H. Lii and N. L. Allinger: Molecular mechanics. The MM3 force field for hydrocarbons. 3. The van der Waals' potentials and crystal data for aliphatic and aromatic hydrocarbons. *J Am Chem Soc*, 111(23), 8576-8582 (1989)

49. T. A. Halgren: Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *J Comput Chem*, 17(5-6), 490-519 (1996)

50. T. A. Halgren: Merck molecular force field. II. MMFF94 van der Waals and electrostatic parameters for intermolecular interactions. *J Comput Chem*, 17(5-6), 520-552 (1996)

51. T. A. Halgren: Merck molecular force field. III. Molecular geometries and vibrational frequencies for MMFF94. *J Comput Chem*, 17(5-6), 553-586 (1996)

52. S. Mayo, B. Olafson and W. Goddard: DREIDING: a generic force field for molecular simulations. *J Phys Chem*, 94(26), 8897-8909 (1990)

53. A. C. T. van Duin, S. Dasgupta, F. Lorant and W. A. Goddard: ReaxFF: A Reactive Force Field for Hydrocarbons. *J Phys Chem A*, 105(41), 9396-9409 (2001)

54. F. Jensen and P.-O. Norrby: Transition states from empirical force fields. *Theor Chem Acc*, 109(1), 1-7 (2003)

55. S. J. Weiner, P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Profeta and P. Weiner: A new force field for molecular mechanical simulation of nucleic acids and proteins. *J Am Chem Soc*, 106(3), 765-784 (1984)

56. B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan and M. Karplus: CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J Comput Chem*, 4(2), 187-217 (1983)

57. W. L. Jorgensen and C. J. Swenson: Optimized intermolecular potential functions for amides and peptides. Structure and properties of liquid amides. *J Am Chem Soc*, 107(3), 569-578 (1985)

58. W. L. Jorgensen and J. Tirado-Rives: The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of

cyclic peptides and crambin. *J Am Chem Soc*, 110(6), 1657-1666 (1988)

59. S. J. Weiner, P. A. Kollman, D. T. Nguyen and D. A. Case: An all atom force field for simulations of proteins and nucleic acids. *J Comput Chem*, 7(2), 230-252 (1986)

60. E. Neria, S. Fischer and M. Karplus: Simulation of activation free energies in molecular systems. *J Chem Phys*, 105(5), 1902-1921 (1996)

61. L. Monticelli, S. K. Kandasamy, X. Periole, R. G. Larson, D. P. Tieleman and S.-J. Marrink: The MARTINI Coarse-Grained Force Field: Extension to Proteins. *J Chem Theory Comput*, 4(5), 819-834 (2008)

62. V. Tozzini: Coarse-grained models for proteins. *Curr Opin Struct Biol*, 15(2), 144-150 (2005)

63. J. W. Ponder, D. A. Case and D. Valerie. Force Fields for Protein Simulations. In: *Advances in Protein Chemistry*. Academic Press (2003)

64. P. M. Morse: Diatomic Molecules According to the Wave Mechanics. II. Vibrational Levels. *Phys Rev*, 34(1), 57-64 (1929)

65. V. Pophristic and L. Goodman: Hyperconjugation not steric repulsion leads to the staggered structure of ethane. *Nature*, 411(6837), 565-568 (2001)

66. A. D. Mackerell, M. Feig and C. L. Brooks: Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J Comput Chem*, 25(11), 1400-1415 (2004)

67. A. Rahman and F. H. Stillinger: Molecular Dynamics Study of Liquid Water. *J Chem Phys*, 55(7), 3336-3359 (1971)

68. W. L. Jorgensen: Quantum and statistical mechanical studies of liquids. 10. Transferable intermolecular potential functions for water, alcohols, and ethers. Application to liquid water. *J Am Chem Soc*, 103(2), 335-340 (1981)

69. J. Gasteiger and M. Marsili: Iterative partial equalization of orbital electronegativity--a rapid access to atomic charges. *Tetrahedron*, 36(22), 3219-3228 (1980)

70. T. A. Halgren and R. B. Nachbar: Merck molecular force field. IV. conformational energies and geometries for MMFF94. *J Comput Chem*, 17(5-6), 587-615 (1996)

71. R. S. Mulliken: Electronic Population Analysis on LCAO-MO Molecular Wave Functions. I. *J Chem Phys*, 23(10), 1833-1840 (1955)

72. R. F. W. Bader: Atoms in molecules. *Acc Chem Res*, 18(1), 9-15 (1985)

73. U. Dinur and A. T. Hagler: Geometry-dependent atomic charges: Methodology and application to alkanes, aldehydes, ketones, and amides. *J Comput Chem*, 16(2), 154-170 (1995)

74. C. I. Bayly, P. Cieplak, W. Cornell and P. A. Kollman: A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J Phys Chem*, 97(40), 10269-10280 (1993)

75. W. D. Cornell, P. Cieplak, C. I. Bayly and P. A. Kollmann: Application of RESP charges to calculate conformational energies, hydrogen bond energies, and free energies of solvation. *J Am Chem Soc*, 115(21), 9620-9631 (1993)

76. P. Cieplak, W. D. Cornell, C. Bayly and P. A. Kollman: Application of the multimolecule and multiconformational RESP methodology to biopolymers: Charge derivation for DNA, RNA, and proteins. *J Comput Chem*, 16(11), 1357-1377 (1995)

77. J. W. Ponder, C. Wu, P. Ren, V. S. Pande, J. D. Chodera, M. J. Schnieders, I. Haque, D. L. Mobley, D. S. Lambrecht, R. A. DiStasio, M. Head-Gordon, G. N. I. Clark, M. E. Johnson and T. Head-Gordon: Current Status of the AMOEBA Polarizable Force Field. *J Phys Chem B*, 114(8), 2549-2564 (2010)

78. K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror and D. E. Shaw: Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins: Struct Funct Bioinform*, 78(8), 1950-1958 (2010)

79. G. A. Kaminski, R. A. Friesner, J. Tirado-Rives and W. L. Jorgensen: Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides *J Phys Chem B*, 105(28), 6474-6487 (2001)

80. J. Wang, P. Cieplak and P. A. Kollman: How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J Comput Chem*, 21(12), 1049-1074 (2000)

81. V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg and C. Simmerling: Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Struct Funct Bioinform*, 65(3), 712-725 (2006)

82. Y. Duan, C. Wu, S. Chowdhury, M. C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. Wang and P. Kollman: A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J Comput Chem*, 24(16), 1999-2012 (2003)

83. O. Guvench and A. D. MacKerell. Comparison of Protein Force Fields for Molecular Dynamics Simulations. In: *Molecular Modeling of Proteins*. Humana Press (2008)

84. N. Foloppe, A. D. MacKerell and Jr.: All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. *J Comput Chem*, 21(2), 86-104 (2000)

85. A. D. MacKerell and N. K. Banavali: All-atom empirical force field for nucleic acids: II. Application to molecular dynamics simulations of DNA and RNA in solution. *J Comput Chem*, 21(2), 105-120 (2000)

86. A. R. van Buuren, S. J. Marrink and H. J. C. Berendsen: A molecular dynamics study of the decane/water interface. *J Phys Chem*, 97(36), 9206-9212 (1993)

87. L. D. Schuler, X. Daura and W. F. van Gunsteren: An improved GROMOS96 force field for aliphatic hydrocarbons in the condensed phase. *J Comput Chem*, 22(11), 1205-1218 (2001)

88. J. B. Klauda, R. M. Venable, J. A. Freites, J. W. O'Connor, D. J. Tobias, C. Mondragon-Ramirez, I. Vorobyov, A. D. MacKerell and R. W. Pastor: Update of the CHARMM All-Atom Additive Force Field for Lipids: Validation on Six Lipid Types. *J Phys Chem B*, 114(23), 7830-7843 (2010)

89. A. A. Skjevik, B. D. Madej, R. C. Walker and K. Teigen: LIPID11: A Modular Framework for Lipid Simulations Using Amber. *J Phys Chem B*, 116(36), 11124-11136 (2012)

90. C. Dickson, L. Rosso, R. Betz, R. Walker and I. Gould: GAFFlipid: a General Amber Force Field for the accurate molecular dynamics simulation of phospholipid. *Soft Matter*, 8(37), 9617-9627 (2012)

91. M. L. P. Price, D. Ostrovsky and W. L. Jorgensen: Gas-phase and liquid-state properties of esters, nitriles, and nitro compounds with the OPLS-AA force field. *J Comput Chem*, 22(13), 1340-1352 (2001)

92. J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case: Development and testing of a general amber force field. *J Comput Chem*, 25(9), 1157-1174 (2004)

93. K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov and A. D. Mackerell: CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J Comput Chem*, 31(4), 671-690 (2010)

94. S. Giordano and W. Rocchia: Shape-dependent effects of dielectrically nonlinear inclusions in heterogeneous media. *J Appl Phys*, 98(10), 104101-104110 (2005)

95. S. Giordano and W. Rocchia: Predicting the dielectric nonlinearity of anisotropic composite materials via tensorial analysis. *J Phys: Condens Matter*, 18, 10585-10599 (2006)

96. P. Ren and J. W. Ponder: Consistent treatment of inter- and intramolecular polarization in molecular mechanics calculations. *J Comput Chem*, 23(16), 1497-1506 (2002)

97. P. Ren, C. Wu and J. W. Ponder: Polarizable Atomic Multipole-Based Molecular Mechanics for Organic Molecules. *J Chem Theory Comput*, 7(10), 3143-3161 (2011)

98. J. D. Jackson: Classical Electrodynamics. John Wiley & Sons, Inc. (1998)

99. M. Gilson and B. Honig: Calculation of the total electrostatic energy of a macromolecular system: solvation energies, binding energies, and conformational analysis. *Proteins: Struct Funct Bioinform*, 4, 7-18 (1988)

100. B. Honig and A. Nicholls: Classical electrostatics in biology and chemistry. *Science*, 268, 1144-1149 (1995)

101. W. Rocchia, E. Alexov and B. Honig: Extending the Applicability of the Nonlinear Poisson-Boltzmann Equation: Multiple Dielectric Constants and Multivalent Ions. *J Phys Chem B*, 105(28), 6507-6514 (2001)

102. A. Onufriev. Continuum Electrostatics Solvent Modeling with the Generalized Born Model. In: *Modeling Solvent Environments*. Wiley-VCH Verlag GmbH & Co. KGaA (2010)

103. A. Onufriev, D. A. Case and D. Bashford: Effective Born radii in the generalized Born approximation: the importance of being perfect. *J Comput Chem*, 23(14), 1297–1304 (2002)

104. W. C. Still, A. Tempczyk, R. C. Hawley and T. Hendrickson: Semianalytical treatment of solvation for molecular mechanics and dynamics. *J Am Chem Soc*, 112(16), 6127-6129 (1990)

105. C. N. Schutz and A. Warshel: What are the dielectric "constants" of proteins and how to validate electrostatic models? *Proteins: Struct Funct Bioinform*, 44(4), 400-417 (2001)

106. L. Wang, Z. Zhang, W. Rocchia and E. Alexov: Using DelPhi capabilities to mimic protein's conformational reorganization with amino acid specific dielectric constants. *Commun Comput Phys*, 13, 13-30 (2013)

107. T.-H. Tan and R. Luo: Continuum treatment of electronic polarization effect. *J Chem Phys*, 126(9), 094103-094108 (2007)

108. T.-H. Tan, C. Tan and R. Luo: Continuum polarizable force field within the Poisson-Boltzmann framework. *J Phys Chem*, 112(25), 7675–7688 (2008)

109. B. Thole: Molecular polarizabilities calculated with a modified dipole interaction. *Chem Phys*, 59(3), 341-350 (1981)

110. P. Cieplak, J. Caldwell and P. Kollman: Molecular mechanical models for organic and biological systems going beyond the atom centered two body additive approximation: aqueous solution free energies of methanol and N-methyl acetamide, nucleic acid base, and amide hydrogen bonding and chloroform/water partition coefficients of the nucleic acid bases. *J Comput Chem*, 22(10), 1048-1057 (2001)

111. Z.-X. Wang, W. Zhang, C. Wu, H. Lei, P. Cieplak and Y. Duan: Strike a balance: Optimization of backbone torsion parameters of AMBER polarizable force field for simulations of proteins and peptides. *J Comput Chem*, 27(6), 781-790 (2006)

112. R. W. Dixon and P. A. Kollman: Advancing beyond the atom-centered model in additive and nonadditive molecular mechanics. *J Comput Chem*, 18(13), 1632-1646 (1997)

113. G. A. Kaminski, H. A. Stern, B. J. Berne, R. A. Friesner, Y. X. Cao, R. B. Murphy, R. Zhou and T. A. Halgren: Development of a polarizable force field for proteins via ab initio quantum chemistry: First generation model and gas phase tests. *J Comput Chem*, 23(16), 1515-1531 (2002)

114. G. A. Kaminski, H. A. Stern, B. J. Berne and R. A. Friesner: Development of an Accurate and Robust Polarizable Molecular Mechanics Force Field from ab Initio Quantum Chemistry. *J Phys Chem A*, 108(4), 621-627 (2003)

115. J. R. Maple, Y. Cao, W. Damm, T. A. Halgren, G. A. Kaminski, L. Y. Zhang and R. A. Friesner: A Polarizable Force Field and Continuum Solvation Methodology for Modeling of Protein-Ligand Interactions. *J Chem Theory Comput*, 1(4), 694-715 (2005)

116. W. Xie, J. Pu, A. D. MacKerell and J. Gao: Development of a Polarizable Intermolecular Potential Function (PIPF) for Liquid Amides and Alkanes. *J Chem Theory Comput*, 3(6), 1878-1889 (2007)

117. S. W. Rick and S. J. Stuart: Potentials and Algorithms for Incorporating Polarizability in Computer Simulations. *Reviews in Computational Chemistry*, 18, 89-146 (2002)

118. P. Cieplak, F.-Y. Dupradeau, Y. Duan and J. Wang: Polarization effects in molecular mechanical force fields. *J Phys: Condens Matter*, 21(33), 333102 (2009)

119. G. Lamoureux, J. A. D. MacKerell and B. Roux: A simple polarizable model of water based on classical Drude oscillators. *J Chem Phys*, 119(10), 5185-5197 (2003)

120. V. M. Anisimov, G. Lamoureux, I. V. Vorobyov, N. Huang, B. t. Roux and A. D. MacKerell: Determination of Electrostatic Parameters for a Polarizable Force Field Based on the Classical Drude Oscillator. *J Chem Theory Comput*, 1(1), 153-168 (2004)

121. A. Rappe and W. Goddard: Charge equilibration for molecular dynamics simulations. *J Phys Chem*, 95(8), 3358-3363 (1991)

122. A. K. Rappe, C. J. Casewit, K. S. Colwell, W. A. Goddard and W. M. Skiff: UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *J Am Chem Soc*, 114(25), 10024-10035 (1992)

123. J. L. Banks, G. A. Kaminski, R. Zhou, D. T. Mainz, B. J. Berne and R. A. Friesner: Parametrizing a polarizable force field from ab initio data. I. The fluctuating point charge model. *J Chem Phys*, 110(2), 741-754 (1999)

124. S. Patel and C. L. Brooks: CHARMM fluctuating charge force field for proteins: I parameterization and application to bulk organic liquid simulations. *J Comput Chem*, 25(1), 1-16 (2004)

125. S. Patel, A. D. Mackerell and C. L. Brooks: CHARMM fluctuating charge force field for proteins: II Protein/solvent properties from molecular dynamics simulations using a nonadditive electrostatic model. *J Comput Chem*, 25(12), 1504-1514 (2004)

126. R. P. Iczkowski and J. L. Margrave: Electronegativity. *J Am Chem Soc*, 83(17), 3547-3551 (1961)

127. R. S. Mulliken: A New Electroaffinity Scale; Together with Data on Valence States and on Valence Ionization Potentials and Electron Affinities. *J Chem Phys*, 2(11), 782-793 (1934)

128. R. G. Parr and R. G. Pearson: Absolute hardness: companion parameter to absolute electronegativity. *J Am Chem Soc*, 105(26), 7512-7516 (1983)

129. R. G. Parr, R. A. Donnelly, M. Levy and W. E. Palke: Electronegativity: The density functional viewpoint. *J Chem Phys*, 68(8), 3801-3807 (1978)

Key Words: Molecular mechanics, Molecular Dynamics, Force Fields, Polarizable Force Fields, Configurational Space, Atomistic Simulation, Review

Send correspondence to: Walter Rocchia, Drug Discovery and Development, Istituto Italiano di Tecnologia, via Morego 30, 16163 Genova, Italy, Tel:39-01071781552, Fax:39-01071781228, E-mail: walter.rocchia@iit.it