DOI: 10.1051/sfn/201112009

Force fields and molecular dynamics simulations

M.A. González

Institut Laue Langevin, 6 rue Jules Horowitz, 38042 Grenoble Cedex 9, France

This paper reviews the basic concepts needed to understand Molecular Dynamics simulations and will hopefully serve as an introductory guide for the non-expert into this exciting topic.

Abstract. The objective of this review is to serve as an introductory guide for the non-expert to the exciting field of Molecular Dynamics (MD). MD simulations generate a phase space trajectory by integrating the classical equations of motion for a system of N particles. Here I review the basic concepts needed to understand the technique, what are the key elements to perform a simulation and which is the information that can be extracted from it. I will start defining what is a force field, which are the terms composing a classical force field, how the parameters of the potential are optimized, and which are the more popular force fields currently employed and the lines of research to improve them. Then the Molecular Dynamics technique will be introduced, including a general overview of the main algorithms employed to integrate the equations of motion, compute the long-range forces, work on different thermodynamic ensembles, or reduce the computational time. Finally the main properties that can be computed from a MD trajectory are briefly introduced.

1. INTRODUCTION

The use of computers to explore the properties of condensed matter goes back to the decade of 1950's and the first Monte Carlo (MC) and Molecular Dynamics (MD) simulations of model liquids performed by Metropolis et al. [1] and Alder and Wainwright [2], respectively. Since then, continuous progress in hardware and software has led to a rapid growth on this field and at present MD simulations are applied in a large variety of scientific areas. Furthermore their use is no longer reserved to experts and many experimentalists use computer simulations nowadays as a tool to analyze or interpret their measurements whenever they are too complex to be described by simple analytical models. This is particularly true in the case of neutron scattering, as in this case the experimental observables correlate directly with the properties obtained in MD simulations and the spatial and time scales that can be measured match very well those amenable to computation. Such features, together with the availability of user-friendly and reliable software to perform this kind of calculation (e.g. Charmm [3], NAMD [4], Amber [5], Gromacs [6], Gromos [7], DL_POLY [8], etc.) and to visualize and analyse their output (VMD [9], gOpenMol [10], nMoldyn [11], ...), enable that today MD simulations are routinely used by many neutron users.

Even if the available software makes quite easy for a novice user to perform a MD simulation without needing lengthy training, it is advisable to acquire a minimum knowledge on the principles of the technique, as well as on the meaning of the terms and choices that one may encounter when using any MD program. The goal of the present chapter is precisely to give to the reader such basic knowledge. But if we want to study a particular system by means of a MD simulation, the requirement of a good program to perform the computation is a necessary but not sufficient condition. In addition we need a good model to represent the interatomic forces acting between the atoms composing the

This is an Open Access article distributed under the terms of the Creative Commons Attribution-Noncommercial License 3.0, which permits unrestricted use, distribution, and reproduction in any noncommercial medium, provided the original work is properly cited.

170 Collection SFN

system. Ideally this can be done from first principles, solving the electronic structure for a particular configuration of the nuclei, and then calculating the resulting forces on each atom [12]. Since the pioneering work of Car and Parrinello [13] the development of *ab initio* MD (AIMD) simulations has grown steadily and today the use of the density functional theory (DFT) [14] allows to treat systems of a reasonable size (several hundreds of atoms) and to achieve time scales of the order of hundreds of ps, so this is the preferred solution to deal with many problems of interest. However quite often the spatial and/or time-scales needed are prohibitively expensive for such *ab initio* methods. In such cases we are obliged to use a higher level of approximation and make recourse to empirical force field (FF) based methods. They allow to simulate systems containing hundreds of thousands of atoms during times of several nanoseconds or even microseconds [15, 16]. On the other hand the quality of a force field needs to be assessed experimentally. And here again neutrons play a particularly important role because while many different experimental results can be used to validate some of the FF parameters, the complementarity mentioned above makes neutron scattering a particularly useful tool in this validation procedure.

Therefore in the first part of this short review I will introduce briefly what is a FF, which are the main terms entering into the description of a standard FF, and which are the current lines of research in order to develop more accurate interatomic potentials. A historical account of the development of FFs in connection with molecular mechanics is given by Gavezzotti [17]. He also reviews the problems to derive the terms entering into the FF directly from the electronic density obtained from quantum mechanical (QM) calculations and the inherent difficulty in assigning a physical meaning to these terms. General reviews about empirical force fields can be found in references [18–20] and a comprehensive compilation of available FFs is supplied in [21]. In the second part I will review the principles of MD simulations and the main concepts and algorithms employed will be introduced. Useful reviews have been written recently by Allen [22], Sutmann [23–25], and Tuckerman and Martyna [26]. And for a more comprehensive overview the interested reader is referred to the excellent textbooks of Allen and Tildesley [27], Haile [28], Rapaport [29] or Frenkel and Smit [30].

2. FORCE FIELDS

2.1 Definition

A force field is a mathematical expression describing the dependence of the energy of a system on the coordinates of its particles. It consists of an analytical form of the interatomic potential energy, $U(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)$, and a set of parameters entering into this form. The parameters are typically obtained either from *ab initio* or semi-empirical quantum mechanical calculations or by fitting to experimental data such as neutron, X-ray and electron diffraction, NMR, infrared, Raman and neutron spectroscopy, etc. Molecules are simply defined as a set of atoms that is held together by simple elastic (harmonic) forces and the FF replaces the true potential with a simplified model valid in the region being simulated. Ideally it must be simple enough to be evaluated quickly, but sufficiently detailed to reproduce the properties of interest of the system studied. There are many force fields available in the literature, having different degrees of complexity, and oriented to treat different kinds of systems. However a typical expression for a FF may look like this:

$$U = \sum_{\text{bonds}} \frac{1}{2} k_b (r - r_0)^2 + \sum_{\text{angles}} \frac{1}{2} k_a (\theta - \theta_0)^2 + \sum_{\text{torsions}} \frac{V_n}{2} [1 + \cos(n\phi - \delta)]$$

$$+ \sum_{\text{improper}} V_{imp} + \sum_{\text{LL}} 4\epsilon_{ij} \left(\frac{\sigma_{ij}^{12}}{r_{ij}^{12}} - \frac{\sigma_{ij}^6}{r_{ij}^6} \right) + \sum_{\text{elec}} \frac{q_i q_j}{r_{ij}}, \tag{2.1}$$

where the first four terms refer to intramolecular or local contributions to the total energy (bond stretching, angle bending, and dihedral and improper torsions), and the last two terms serve to describe the repulsive and Van der Waals interactions (in this case by means of a 12-6 Lennard-Jones potential) and the Coulombic interactions.

2.2 Intramolecular terms

As shown in equation (2.1), bond stretching is very often represented with a simple harmonic function that controls the length of covalent bonds. Reasonable values for r_0 can be obtained from X-ray diffraction experiments, while the spring constant may be estimated from infrared or Raman spectra. The harmonic potential is a poor approximation for bond displacements larger than 10% from the equilibrium value. Additionally the use of the harmonic function implies that the bond cannot be broken, so no chemical processes can be studied. This is one of the main limitations of FF based MD simulations compared to *ab initio* MD. Occasionally some other functional forms (in particular, the Morse potential) are employed to improve the accuracy. However as those forms are more expensive in terms of computing time and under most circumstances the harmonic approximation is reasonably good, most of the existing potentials use the simpler harmonic function.

Angle bending is also usually represented by a harmonic potential, although in some cases a trigonometric potential is preferred:

$$U_{\text{bending}} = \frac{1}{2} k_a (\cos \theta - \cos \theta_0)^2. \tag{2.2}$$

Some times other terms are added to optimize the fitting to vibrational spectra. The most common addition consists of using the Urey-Bradley potential [31]:

$$U_{\rm UB} = \sum_{\rm angles} \frac{1}{2} k_{\rm UB} (s - s_0)^2, \tag{2.3}$$

where s is the distance between the two external atoms forming the angle.

In any molecule containing more than four atoms in a row, we need to include also a dihedral or torsional term. While angle bending, and in particular bond stretching, are high frequency motions that often are not relevant for the study of the properties of interest and can be replaced by a rigid approximation (see section 3.8), torsional motions are typically hundreds of times less stiff than bond stretching motions and they are necessary to ensure the correct degree of rigidity of the molecule and to reproduce the major conformational changes due to rotations about bonds. Therefore they play a crucial role in determining the local structure of a macromolecule or the relative stability of different molecular conformations. As an example, the definition of a dihedral angle in the 1,2-dichloroethane molecule and the corresponding torsional potential are shown in figures 1 and 2.

Torsional energy is usually represented by a cosine function such as the one used in equation (2.1), where ϕ is the torsional angle, δ is the phase, n defines the number of minima or maxima between 0 and 2π , and V_n determines the height of the potential barrier. The combination of two or more terms with different n allows to construct a dihedral potential with minima having different depths. But alternative representations for the dihedral potential can be found in the literature. For example the OPLS potential uses the following expression [32]:

$$U_{\text{tors}} = \sum_{\text{torsions}} k_0 + \frac{k_1}{2} (1 + \cos\phi) + \frac{k_2}{2} (1 - \cos 2\phi) + \frac{k_3}{2} (1 + \cos 3\phi). \tag{2.4}$$

The torsional parameters are usually derived from *ab initio* calculations and then refined using experimental data such as molecular geometries or vibrational spectra.

Finally, an additional term is needed to ensure the planarity of some particular groups, such as sp² hybridized carbons in carbonyl groups or in aromatic rings. This is because the normal torsion terms

172

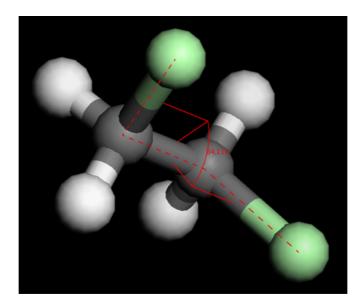


Figure 1. Definition of the dihedral angle in 1,2-dichloroethane. The figure shows the gauche isomer.

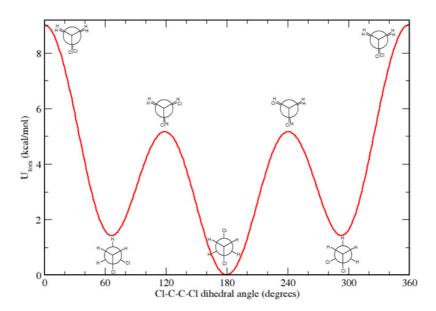


Figure 2. Torsional potential of 1,2-dichloroethane showing the gauche ($\approx 60^{\circ}$ and 300°) and trans ($\approx 180^{\circ}$) conformational isomers and the saddle points corresponding to the eclipsed conformations [33].

described above are not sufficient to maintain planarity, so this extra component describes the positive contribution to the energy of those out-of-plane motions. Typical expressions for the improper torsion term are:

$$U_{\text{imp}} = \sum_{\text{impropers}} \frac{k_{\text{imp}}}{2} [1 + \cos(2\omega - \pi)], \text{ or}$$

$$U_{\rm imp} = \sum_{\rm impropers} \frac{k_{\rm imp}}{2} (\omega - \omega_0)^2, \tag{2.5}$$

where ω is the improper angle corresponding to the deviation from planarity.

2.3 Intermolecular terms

Van der Waals interactions between two atoms arise from the balance between repulsive and attractive forces. Repulsion is due to the overlap of the electron clouds of both atoms, while the interactions between induced dipoles result in an attractive component that varies as r^{-6} . The 12-6 Lennard-Jones (LJ) potential is very often used to represent these interactions, but it is not uncommon to find some "softer" terms to describe the repulsive part, in particular the exponential function, which is used in the Buckingham potential [34] and is preferred over the LJ function in studies of organic molecular crystals [35, 36]. Other functions used occasionally are the 9-6 LJ potential used in the COMPASS FF [37] or even a buffered 14-7 LJ term, as employed in MMFF [38]. Van der Waals forces act between any pair of atoms belonging to different molecules, but they also intervene between atoms belonging to the same molecule that are sufficiently separated, as described later. It is possible to define a set of parameters (e.g. σ_{ij} and ϵ_{ij}) for each different pair of atoms, but for convenience most force fields give individual atomic parameters (i.e. σ_i and ϵ_i), together with some rules to combine them. For example, for the LJ potential the corresponding depth of the well for the interaction between two atoms i and j is given by the geometric mean, $\epsilon_{ij} = (\epsilon_i \epsilon_j)^{1/2}$, while the value at which the potential becomes zero can be given by the geometric, $\sigma_{ij} = (\sigma_i \sigma_j)^{1/2}$, or the arithmetic mean, $\sigma_{ij} = \frac{1}{2}(\sigma_i + \sigma_j)$, depending on the FF chosen.

The final term in equation (2.1) serves to describe the electrostatic interactions. While the molecular electronic density can be obtained with a high accuracy by means of high-level quantum-mechanical calculations, the problem of reducing such density to a manageable description to be used in a MD simulation is not trivial. The usual choice is to assign a partial atomic charge to each nucleus and use Coulomb's law to compute their contribution to the total energy. The partial charges can be derived from a fit to experimental thermodynamic data, but this approach is only practical for small molecules [18]. There are also some fast methods to determine the charge distribution which are based on the atom electronegativities [39-41]. However the most common way to obtain reliable partial charges consists of performing an ab initio calculation and then deriving them from the quantum mechanical potential. Unfortunately they cannot be derived unambiguously because atomic charges are not experimental observables, so many different methods have been developed to determine them and they do not always produce the same distribution of partial charges [42, 43]. An important point to note is that in condensed phases there are polarization effects that cannot be fully described using fixed partial charges. Although polarizable force fields exist (section 2.4), generally these non-pairwise additive effects are just taken into account in an average way by using effective charges. As a consequence the resulting dipole moment of a molecular model is usually larger than the experimental dipole moment corresponding to the isolated molecule. For example, most water models employ charge distributions giving $\mu = 2.3$ -2.6 D instead of the experimental value for the vapour of 1.85 D [44]. Another point to note is that electrostatic interactions are long-ranged, so they require a particular treatment when truncating them in our computation of the forces. The techniques to deal with long-range interactions will be presented in section 3.5. A final remark is that for computational efficiency and practical reasons, partial charges are normally assigned only to the atomic sites. But nothing precludes placing them outside the atom positions. Again water gives us a good example, as we can find water models having only 3 sites (with partial charges assigned to the O and H atoms), but also models with 4, 5 and even 6 sites [44].

In most cases both the Van der Waals and electrostatic intramolecular interactions between atoms separated by more than three bonds (>1-4) are treated in the same way as if they were intermolecular, while interactions between 1-2 and 1-3 pairs are excluded. This is to avoid numerical problems, as the potential can become strongly repulsive or attractive due to the small distances involved, and also

because the interactions between those atoms are considered to be already correctly described by the intramolecular terms. The treatment of 1-4 interactions is less standard, but often they are partially excluded by introducing a scaling factor. However the scaling can be different from one force field to another. This has to be taken into account if parameters coming from different force fields are mixed. Furthermore it has to be considered that it is the particular combination of the dihedral potential with the 1-4 Van der Waals and electrostatic interactions that determines the rotational barriers, so in general it is not advised to modify them or to combine parameters coming from different sources.

2.4 Special terms

The contributions to the energy described above appear in almost all force fields, but in certain cases we can encounter some other additional terms.

Thus it is possible to find cross terms describing the coupling between stretching, bending, and torsion. They bring corrections to the intramolecular energy and allow to reproduce better the observed spectra of intramolecular vibrations. Examples of such cross-terms are:

$$U_{\text{bond-bond}} = \frac{1}{2} k_{bb} (r - r_0) (r' - r'_0), \tag{2.6}$$

$$U_{\text{bond-bend}} = \frac{1}{2} k_{ba} [(r - r_0) + (r' - r'_0)](\theta - \theta_0). \tag{2.7}$$

Hydrogen bonding is usually reproduced by an adequate choice of van der Waals parameters and partial charges, but some times an extra term is included to improve the accuracy of the H-bonding energy. This H-bonding potential may adopt different forms, but a general expression often used is [17]:

$$U_{\rm HB} = \left(\frac{A}{r_{AD}^{12}} - \frac{B}{r_{AD}^{10}}\right) \left(\cos^2 \theta_{AHD}\right),\tag{2.8}$$

where r_{AD} is the distance between the donor and the acceptor, and θ_{AHD} is the angle between the donor, the hydrogen, and the acceptor.

Finally the most important addition to the standard potential given by eqn. (2.1) is the inclusion of polarization effects explicitly [45]. As explained above, local electric fields emerging in condensed phases will induce the appearance of dipoles. This is a non-pairwise additive interaction, so the use of effective charges to describe the average polarization is only a partial solution to this problem and furthermore reduces the transferability of the FF. For example, the partial charges used to describe a molecule in the liquid state will not be adequate to describe the same molecule in the gas phase. Or the same amino-acid should have different charge distributions in different proteins due to the different electrostatic environment. Furthermore the polarizability will affect the solvation energy of ions in a nonpolar environment, the directionality and energetics of H-bonds, the interactions between cations and aromatic molecules, etc. For this reason, most modern potentials include explicitly the polarization. In this case, for each atom or molecule the surrounding particles will induce a charge redistribution that needs to be modelled. There are three principal methods to do this [46, 47]:

- 1. Fluctuating charges: The charges are allowed to fluctuate according to the environment, so the charge flows between atoms until their instantaneous electronegativities are equal [48].
- 2. Shell models (Drude particle): Electronic polarization is incorporated by representing the atom as the sum of a charged core and a charged shell connected by a harmonic spring whose force constant is inversely proportional to the atomic polarizability. The relative displacement of both charges depends on the electrostatic field created by the environment [49].
- 3. Induced point dipoles: An induced point dipole μ_i is associated to each polarizable atom:

$$\mu_i = \alpha_i (\mathbf{E}_i^q + \mathbf{E}_i^{\text{ind}}), \tag{2.9}$$

where α_i is the isotropic polarizability of the atom, \mathbf{E}_i^q is the electrostatic field created on the atom i by the permanent charges, and $\mathbf{E}_i^{\mathrm{ind}}$ is the field created by the induced dipoles of the rest of the atoms in the system.

The polarization energy is:

$$U_{\text{pol}} = -\frac{1}{2} \sum_{i} \mu_i \mathbf{E}_i. \tag{2.10}$$

As the induced dipoles depend both on the static charges and the remaining induced dipoles, they have to be determined self-consistently by solving the previous equations iteratively until the values for all the induced dipoles converge [50].

Incorporating polarization effects into the potential describing the intermolecular interactions is an important step forward. However a force field is a global entity, so the construction of new polarizable force fields requires also a reoptimization of the Van der Waals and intramolecular parameters. This represents a huge task, so current polarizable force fields have not yet achieved a fully mature state [47]. This could explain the unconvincing results obtained when comparing the ability of polarizable and non polarizable potentials to describe the structure of liquid water [51, 52]. Nevertheless recent comparisons on biological systems indicate that they are able to provide a better agreement with experiment than non-additive force fields [47] and they provide a more physical representation of intermolecular interactions, increasing their transferability. Thus most of the new force fields currently being developed include polarization effects in one way or another.

2.5 Popular force fields

The first force fields appeared in the 1960's, with the development of the molecular mechanics (MM) method and their primary goal was to predict molecular structures, vibrational spectra and enthalpies of isolated molecules [17]. They were therefore mainly oriented to treat small organic molecules, but some of them have survived and continue to be developed and used today. The best example is given by the MM potentials developed by Allinger's group [53]: MM2 [54], MM3 [55] and MM4 [56]. The first of these force fields was originally created to study hydrocarbons, but they were later extended to be able to deal with many different types of organic or functionalized molecules (alcohols, ethers, sulphides, amides, etc.).

Since then the scope of research has moved to deal with much more complex systems, leading to the development of more widely applicable force fields. Good examples are the Dreiding [57] and Universal (UFF) [58] force fields, that contain parameters for all the atoms in the periodic table. Other very popular force fields are CHARMM [59], AMBER [60], GROMOS [61], OPLS [32], and COMPASS [37]. All of them are quite general, but the first three are often employed in simulations of biomolecules, while OPLS and COMPASS were originally developed to simulate condensed matter. It must also be noted that many of these force fields are continuously evolving and different versions are available (e.g. CHARMM19, CHARMM22, CHARMM27 [62]; GROMOS96, GROMOS45A3, GROMOS53A5, GROMOS53A6 [63]; AMBER91, AMBER94, AMBER96, AMBER99, AMBER02 [64]; etc.).

The force fields using an energy expression similar to equation (2.1) are sometimes called first-generation or class I force fields. Second-generation or class II force field include the cross terms mentioned in section 2.4. COMPASS, UFF, MM2, MM3 and MM4 belong to this second class. Other popular class II force fields are CFF (consistent force field) [65] and MMFF (Merck molecular force field) [38].

During the 1990s the first general polarizable force fields appeared. The PIPF (polarizable intermolecular potential function) [66], DRF90 [67] and AMOEBA [68] force fields are good examples of such developments. In addition, some of the general force fields mentioned above have also developed polarizable versions. Thus polarization has been introduced into CHARMM using either fluctuating

charges [69] or the shell model [70], and AMBER [71], OPLS [72] and GROMOS [73] have also been extended to include polarization. A more comprehensive list of polarizable force fields is given in reference [47].

In addition to those general force fields, there is a vast number of specific potentials developed to describe just a particular system or a class of compounds. Water merits a particular mention in this respect, as due to its importance a large number of water models has been proposed since the first MC simulation of Barker and Watts [74]. Many of those models are listed in reference [44], together with a critical review of their strengths and weaknesses. And recently Vega et al. performed an extensive comparison of the most popular rigid non-polarizable water potentials (TIP3P, TIP4P, TIP5P, SPC and SPC/E), finding that a modified version of the second one, called TIP4P/2005, provides the best description of 9 out of the 10 experimental properties investigated [75].

It is even more difficult to compare the performance of the existing general force fields, as the result will depend strongly on the system and properties simulated. Some comparisons can be found in the literature, in particular concerning the accuracy of the CHARMM, AMBER and OPLS force fields in the context of biomolecular simulations, but the results are not unambiguous. Thus Price and Brooks found that the three provided remarkably similar results concerning the structure and dynamics of three different proteins [76]. On the other hand Yeh and Hummer found significant differences in the configurations of two peptides studied with CHARMM and AMBER, as well as in the endto-end diffusion coefficient [77]. Aliev and Courtier-Murias also found a strong dependence of the structure of small open chain peptides on the force field used [78]. A similar result was found in simulations of insulin using CHARMM, AMBER, OPLS, and GROMOS, as different structural trends are favoured depending on the force field chosen [79]. One of the most detailed evaluation of existing potentials has been performed by Paton and Goodman. They evaluated the interaction energies of 22 molecular complexes of small and medium size and 143 nuclear acid bases and amino-acids using seven different force fields (MM2, MM3, AMBER, OPLS*, OPLSAA, MMFF94, and MMFF94s) and benchmarked them against the energies obtained from high-level ab initio calculations. Their results show that all the potentials tested, and specially OPLSAA and MMFF94s, describe quite accurately electrostatic and van der Waals interactions, but that the magnitude of hydrogen bonding interactions are severely underestimated [80]. The main conclusion is that each force field has its particular strengths and weaknesses related to the data and procedure employed in its parametrization, so the final choice depends on the particular problem being considered. However, as stated by Jorgensen and Tirado-Rives, "all leading force fields now provide reasonable results for a wide range of properties of isolated molecules, pure liquids, and aqueous solutions" [81].

2.6 Parametrizing a force field

The ultimate goal of a force field is to describe in classical terms all the quantum mechanical facts, partitioning the total electronic energy into well separated atom-atom contributions, such as Coulombic, polarization, dispersion, and repulsive energies. Unfortunately, even disposing of very accurate QM calculations, it is impossible to fully separate the intricate electronic effects (see reference [17]). This implies that we are obliged to employ significant physical approximations in order to describe in a tractable manner the intermolecular interactions, which limits their accuracy. Therefore they are called empirical potentials or empirical force fields, and depending on the procedure followed to develop them and the set of input data used to optimize their parameters, we will obtain different force fields applicable to different systems or problems. This has led T. Halgren to say that "force field development is still as much a matter of art as of science" [82].

From the above considerations, it should be clear that the development of a force field is not a trivial task. A thorough account of the methodology needed to parametrize a general FF is given in the series of papers describing the Merck Molecular Force Field, MMFF94 [38, 83–86], and serves to exemplify

the complexity of this task. Here we will just mention the four basic steps needed to develop a new force field:

- 1. Select the function to model the system energy: Are we going to employ cross-terms to describe intramolecular interactions? Shall we include polarization explicitly or is an additive potential enough? What form are we going to use to represent van der Waals interactions (12-6 LJ, exponential or other)?
- 2. Choose the data set required to determine all the parameters needed in the previously selected function: There is a large number of experimental data that can be employed, such as equilibrium bond lengths obtained from x-ray or neutron diffraction, force constants from vibrational spectroscopy, heats of sublimation or vaporization, densities, etc. Nevertheless in many cases experimental data are scarce or unreliable, so nowadays *ab initio* calculations constitute the main source of input data.
- 3. Optimize the parameters: There is a large number of parameters, so very often they are refined in stages. Additionally most of them are not decoupled, so a change in one value may affect some other parameter. Therefore the optimization becomes an iterative procedure. An alternative approach consists in using a least-squares fitting to determine the whole set of parameters providing the best agreement with the input information. Further details about both approaches are given in [18, 82] and references therein. Finally, a new and attractive approach is the force-matching method, consisting in directly fitting the potential energy surface derived from QM calculations [87, 88].
- 4. Validate the final set of parameters by computing the properties of systems not employed in the parametrization procedure.

This is an extremely hard and time-consuming job that is performed only by specialized groups, so the end user normally only needs to select an appropriate FF from the existing literature. Nevertheless if the parameters for one group or molecule are missing, he may be obliged to "find" the necessary values. This requires an iterative procedure, as intermolecular parameters will affect also the resulting intramolecular geometries, vibrations and conformational energies [89]. The starting point consists of borrowing a set of parameters from similar atoms or groups. Then partial atomic charges can be assigned using the electrostatic potential derived from high level QM calculations and employing fitting methods such as CHELPG [90] and RESP [91], although sometimes faster approaches using semiempirical methods also provide good results [92]. Note also that non-polarizable models may need to account for the lack of explicit inclusion of non-pairwise additive effects by using effective charges enhanced by 10-20% with respect to those obtained from the QM calculation. It is also important to use the same level of OM theory employed in developing the existing set of parameters. And if the molecule can adopt different conformations it is advised to perform the QM calculations for several of them. Once a set of charges has been found, we need to optimize the van der Waals parameters. This is the most difficult step of the whole procedure. Due to the limited accuracy of QM calculations when dealing with dispersion interactions, normally they are adjusted in order to reproduce experimental data such as heats of vaporization, densities, isocompressibilities, heat capacities and equilibrium intermolecular geometries. Unfortunately in most cases the optimized parameters are underdetermined due to the small quantity of available information compared to the large number of values to optimize, implying that the parameters obtained can be strongly correlated [19]. Once a reasonable set of nonbonded parameters is obtained, intramolecular terms are adjusted to ensure that the molecular geometry and the normal modes of vibration are well described. This is followed by a simulation of the condensed phase to test if the experimental properties of the crystal or liquid are reproduced with the desired accuracy. If not, the process is repeated iteratively to achieve convergence [89].

2.7 Limits, advantages and beyond empirical force fields

Compared to AIMD, the use of empirical force field based simulations has some intrinsic limitations. In particular, they cannot provide any information about the electronic structure and they are unable

to handle reactions such as the breaking and formation of bonds, electron excitations, charge transfers, etc. Furthermore their predictive power is limited, as they can be applied only to systems containing the functional groups included in the development of the FF and their accuracy depends on the quality of its parametrization. Their transferability is also limited, so great care is needed when applying a potential to simulate a system under conditions that differ strongly from the conditions used in the parametrization (e.g. different solvent, pressure, temperature, etc.). Finally we must not forget that any FF is based on numerous approximations and derived from different types of data. This is the reason why they are called empirical and thus any simulation should be validated by an appropriate comparison with experimental results

On the other hand, force field based simulations are several orders of magnitude faster than AIMD simulations, due to the simplicity of the force calculation using an expression like equation (2.1) compared to the huge problem of solving the electronic structure to obtain the interatomic forces. As a consequence we can handle much larger systems and explore much longer simulation times. Thus today it is possible to perform simulations on systems containing millions of atoms and to reach times of the order of microseconds, as shown by recent examples in the literature [93–96]. A less obvious advantage is that it is possible to modify easily the energy expression to bias the calculation. In this way it is possible to determine which are the precise interactions that are particularly relevant to explain a specific property or to examine how the observed behaviour depends on the strength of each of the terms appearing in equation (2.1) or in particular atomic contributions to it.

There are several specialized groups working actively in developing new methods to overcome some of the limitations of empirical force fields. For example, reactive force fields have appeared allowing chemical reactivity to be treated [97–99]. Another increasingly popular approach consists of combining quantum mechanical and molecular mechanical potentials (QM/MM) [100–102]. While QM/MM hybrid methods solve reasonably most of the deficiencies of classical potentials in the region treated explicitly using QM, those difficulties remain in the MM part of the calculation. Therefore a new approach that treats the entire system equally has been proposed: the X-POL force field. X-POL is a quantum mechanical model in which bonded interactions are fully represented by an electronic structure theory augmented with some empirical torsional terms and nonbonded interactions are modelled by an iterative, combined quantum mechanical and molecular mechanics method, in which partial charges are derived from the molecular wave functions of individual fragments [103]. It is an empirical force field, but based on a quantum mechanical formalism, that allows to treat many-body polarization and charge delocalization effects and can be applied to large systems which are out of reach of current AIMD simulations [104].

3. MOLECULAR DYNAMICS

3.1 Principle

An MD simulation is a technique to produce a dynamical trajectory for a system composed of N particles by integrating Newton's equations of motion. We need a set of initial conditions (positions and velocities of each particle), a good model to represent the forces acting between the particles (either from electronic structure calculations or using the empirical force fields presented in the previous section), and to define the boundary conditions to be employed. Then we need to solve the classical equation of motion:

$$m_i \frac{d^2 \mathbf{r}_i}{dt^2} = \mathbf{f}_i = -\frac{\partial}{\partial \mathbf{r}_i} U(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N),$$
 (3.1)

where $U(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)$ is the potential energy depending on the coordinates of the N particles. This is a system of N coupled second order non linear differential equations that cannot be solved exactly, so equation 3.1 has to be solved numerically step by step using an appropriate integration algorithm. In the following sections we will review briefly each of these ingredients.

3.2 Initial conditions

We need to know the initial positions and velocities of each particle in the system. In the case of a crystal the positions will be typically available in the form of a crystallographic file and we can construct a supercell combining several unit cells. For a disordered system, the positions can be generated randomly or we can create an ordered structure and then melt it. The velocity of each particle is attributed randomly from a Maxwellian distribution centered on the desired temperature and then they are adjusted in order to zero the angular momentum and the center of mass velocity of the total system.

3.3 Evaluation of forces

For the pairwise potentials used in most cases, the force calculation is simple. For example, for a Lennard-Jones potential the force exerted by the atom j on the atom i is:

$$\mathbf{f}_{ij} = -\frac{\partial}{\partial \mathbf{r}_{ij}} \left\{ 4\epsilon \left[\left(\frac{\sigma}{r_{ij}} \right)^{12} - \left(\frac{\sigma}{r_{ij}} \right)^{6} \right] \right\} = \frac{48\epsilon}{\sigma^{2}} \left[\left(\frac{\sigma}{r_{ij}} \right)^{14} - \frac{1}{2} \left(\frac{\sigma}{r_{ij}} \right)^{8} \right] \mathbf{r}_{ij}, \tag{3.2}$$

and

$$\mathbf{f}_i = \sum_{j \neq i} \mathbf{f}_{ij},\tag{3.3}$$

and

$$\mathbf{f}_{ji} = -\mathbf{f}_{ij}.\tag{3.4}$$

For the intramolecular potentials the derivation is more involved, but one can obtain also analytical expressions for the forces (see [27]). Finally if the potential is given as a table of values of U vs r, then the forces have to be determined by numerical differentiation.

3.4 Boundary conditions

We may simulate our system of N particles in isolation, i.e. surrounded by vacuum. But in most cases we are interested on the bulk properties of a liquid or solid system, so we need to impose some boundary conditions. We could use rigid walls, but then the surface effects would blur the real bulk physics. This is because the fraction of atoms near the walls is proportional to $N^{-1/3}$. While this fraction is negligible in a macroscopic sample, in a typical simulation it can go from 0.06 for $N = 10^6$ atoms to 0.49 for $N = 10^3$. So unless we are really interested in those effects (for example if we study a confined liquid), we need to use periodic boundary conditions (PBC). Therefore the simulation box is surrounded by an infinite number of replicas of itself, as shown in figure 3. Only the N atoms inside the main cell are considered explicitly, but as soon as one of the atoms leaves the cell, an image particle enters from the opposite side to replace it.

It must be noted that in this case the real system being simulated consists of the primitive box and all its replicas. This is not a problem when simulating a crystal, but in the case of disordered systems we are introducing an artificial periodicity. If the size of the simulation box is sufficiently large, the effects are normally not important, but we need to be careful when considering any property that depends on long-range correlations. For example, the only allowed fluctuations are those having wavelengths that are compatible with the box size. Thus the longest wavelength permitted in the system will be $\lambda = L$, where L is the length of the simulation box, and it will be problematic to simulate any situation where longer wavelength fluctuations are important, such as phase transitions. Even in solid simulations, the strain field generated by any inhomogeneity will be truncated and modified at the boundaries. And when simulating a macromolecule in solution, the latter could interact with its own images if the box is too

180

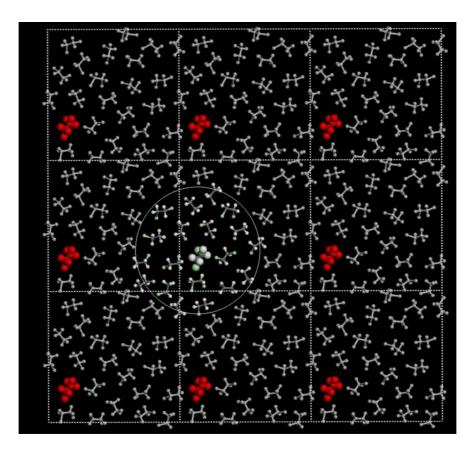


Figure 3. Schematic representation of periodic boundary conditions in 2D. The figure shows a selected molecule in the primary cell together with 8 replicas. The coloured molecules represent those molecules which are at a distance from the reference molecule smaller than the cutoff radius. Typically the cutoff is applied to full molecules (for example using the distance between their center of masses as reference) or neutral groups, in order to minimize truncation effects.

small, so for example Souza and Ornstein recommend to use a minimum layer of $10\,\text{Å}$ of water when simulating DNA [105].

PBC are used together with the minimum image convention, meaning that only the interactions with the nearest image are considered. However this is not the best solution for two reasons. First because then the time needed to compute the forces will scale as N^2 . But if we have a large system where the interatomic potential has decayed to a negligible value for distances much smaller than L/2, there is no need to compute all the interactions. Second because the potential is not constant on the surface of a cube around a given particle. Therefore the usual approach consists of employing a spherical truncation such that only the interactions between particles separated by a distance smaller than R_c are taken into account:

$$U_{\text{trunc}}(r) = \begin{cases} U^{LJ}(r), r \le R_c, \\ 0, r > R_c \end{cases}$$
(3.5)

The cutoff radius R_c cannot be larger than L/2 because otherwise some interactions between a particle and more than two images would occur, but it has to be large enough so that the potential can be safely neglected beyond R_c . In the case of the Lennard-Jones potential one can use $R_c \ge 2.5\sigma$. Even if the value of U(r) for $r > R_c$ is small, the contribution of the excluded interactions to the total energy

is not negligible because the number of neighbours increases asymptotically as r^{d-1} , where d is the dimensionality of the system. Nevertheless it is possible to estimate the correction to the potential energy by assuming that beyond R_c the atoms are distributed isotropically, so $g(r > R_c) = 1$ and

$$U_{\text{tot}} = \sum_{i < j, r_{ij} < R_c} U(r_{ij}) + \frac{N\rho}{2} \int_{R_c}^{\infty} 4\pi r^2 U(r) dr.$$
 (3.6)

In the case of the Lennard-Jones potential, this tail correction is:

$$U_{\text{tail}} = \frac{8}{3}\pi\rho\epsilon\sigma^{3} \left[\frac{1}{3} \left(\frac{\sigma}{R_{c}} \right)^{9} - \left(\frac{\sigma}{R_{c}} \right)^{3} \right]. \tag{3.7}$$

While the value of the LJ potential at R_c is $U(R_c) = -0.016\epsilon$, the tail correction for a typical liquid density $\rho\sigma^3 = 1$ amounts to $U_{\text{tail}} = -0.535\epsilon$, which is about 10% of the total potential energy, clearly showing that this tail correction is not negligible. An equivalent correction can be applied to the calculation of the pressure [27].

The simple truncation scheme presents some problems concerning the energy conservation due to the discontinuity in the potential at R_c . This can be avoided using a truncated and shifted potential:

$$U_{\text{trunc}}(r) = \begin{cases} U(r) - U(R_c), & r \le R_c, \\ 0, & r > R_c \end{cases}$$
 (3.8)

More complex shifting schemes that avoid also discontinuities in the forces can be applied [106]. Another possibility is to multiply the potential by a smoothing function that brings the value of U(r) to zero at R_c .

There is not an ideal truncation method, but the differences between them are less significant for larger cutoff radii, so R_c should be sufficiently large [107]. It must also be noted that this truncation should be applied only to short-ranged interactions. Long-range forces (typically electrostatic interactions) should never be truncated. As an example, Yonetani has shown that large artefacts appear in water simulations when truncating the Coulombic forces and that surprisingly those artefacts can be larger when using a cutoff $R_c = 18$ than with $R_c = 9$ Å [108]. Therefore in the case of long-range interactions some of the methods described in the following section must be applied. Finally it must be noted that in most cases the choice of the truncation method is not important, but that in some particular conditions such as close to a critical point the results may differ significantly depending on the method chosen [109].

3.5 Long-range interactions

The electrostatic potential given in equation (2.1) does not decay with distance rapidly enough to allow us to use safely the simple truncation scheme. A simple alternative is to use the reaction field method, consisting in considering all the molecules beyond a certain cutoff as a dielectric continuum. For a given atom, all the interactions with the particles inside the cutoff sphere are explicitly considered and the charge distribution inside that sphere polarizes the dielectric continuum, so this one produces an additional electrical field into the cavity. This reaction field is given by the equation:

$$E_{\rm rf} = \frac{1}{4\pi\epsilon_0} \frac{2(\epsilon_{\rm rf} - 1)}{2\epsilon_{\rm rf} + 1} \frac{1}{R_c^3} \sum_{i \in P} \mathbf{l}_i, \tag{3.9}$$

where ϵ_0 is the dielectric constant of the vacuum, $\epsilon_{\rm rf}$ is the dielectric constant of the system, R_c is the cutoff radius which normally coincides with the cutoff employed for the Lennard-Jones interactions, and \mathbf{l}_j is the dipole moment of molecule j. The previous equation requires to know the dielectric constant of the system being modelled, which usually is not known. Fortunately the factor $2(\epsilon_{\rm rf}-1)/(2\epsilon_{\rm rf}+1)$ goes rapidly to one as $\epsilon_{\rm rf}$ increases, so the results are quite independent of the exact value used and

we can reasonably approximate ϵ_{rf} by the dielectric constant of the real system. Another point to take into account is that whenever a molecule j is inside the cavity around i there will be an orientational correlation between both. When the molecule j leaves the cavity this correlation is lost, so if j goes back into the cavity it will enter with a random orientation that on average will correspond to a situation of higher energy. As a consequence the discontinuous jumps in energy occurring when a molecule leaves or enters the cavity do not cancel exactly and the system will heat up. Additionally some spurious features may appear in the radial distribution function around $r \approx R_c$. In order to solve this problem a tapering function should be introduced [110]. The reaction field method was applied for the first time by Barker and Watts to simulate liquid water [111] and later generalized to deal also with ions [112]. The expressions needed to calculate the electric properties of the system when using this method have also been derived [113]. The main criticism that can be made of the reaction field method is that considering the system as a continuum for distances beyond 9–15 Å (which are typical cutoff values) is a very crude approximation. Nevertheless some studies comparing this method with the reference method using the Ewald sums (see below) conclude that in general they produce comparable results [114–116].

The more rigourous way of treating the electrostatic interactions in a periodic system consists of taking into account the contribution to the potential energy from all the replicated cells. Then we have:

$$U_{\text{coul}} = \sum_{i,j,\mathbf{R}} \frac{q_i q_j}{|\mathbf{r}_{ij} + \mathbf{R}|},\tag{3.10}$$

where **R** represents the vector connecting the primitive cell to each of its replicas. But this sum is only conditionally convergent, so it cannot used to compute the electrostatic potential. The solution was proposed by Ewald [117] and consists basically of adding to each point charge a charge distribution of opposite sign, such that the electrostatic potential caused by the screened charges becomes short-ranged and can then be treated in real space. A convenient choice is to use a Gaussian distribution:

$$\rho = -q_i \left(\frac{\alpha}{\pi}\right)^{3/2} \exp(-\alpha r^2),\tag{3.11}$$

and the resulting contribution in real space is:

$$U_{\text{coul,r}} = \frac{1}{2} \sum_{ij} \frac{q_i q_j \text{erfc}(\sqrt{\alpha} r_{ij})}{r_{ij}}.$$
 (3.12)

Here α is a parameter chosen to make this potential sufficiently small at the chosen cutoff to be safely truncated. Now in order to recover the original system we need to add another compensating Gaussian distribution of opposite sign to that of equation (3.11), which can be treated in reciprocal space to give:

$$U_{\text{coul},k} = \frac{1}{2V} \sum_{\mathbf{k} \neq \mathbf{0}} \frac{4\pi}{k^2} |\rho(\mathbf{k})|^2 \exp(-k^2/4\alpha), \tag{3.13}$$

with

$$\rho(\mathbf{k}) = \sum_{i} q_i \exp(i\mathbf{k}\mathbf{r}_i). \tag{3.14}$$

And finally we have to correct the self-interaction emerging from the unphysical interaction between the point charge and its compensating cloud:

$$U_{\text{coul,self}} = \left(\frac{\alpha}{\pi}\right)^{1/2} \sum_{i} q_i^2. \tag{3.15}$$

The sum in reciprocal space has to be performed up to a sufficiently large k vector, $k_{\text{max}} = \frac{2\pi}{L} n_{\text{max}}$, which depends on the chosen value for α . The convergence of the Ewald sum depends then on three parameters: the cutoff R_c , α and n_{max} . R_c is normally taken as the same cutoff employed for the Lennard-Jones potential for computational convenience. This imposes a lower limit for α , as in order to get a relative error in the real space sum smaller than 4×10^{-5} we need to set $\alpha \ge 3.2/R_c$.

Larger values of α will reduce the error, but at the cost of slowing down the convergence of the reciprocal sum. To get the same relative error of 4×10^{-5} in the latter, we need to extend the sum up to $k_{\text{max}} \approx 6.2\alpha$. Thus in a cubic system and using $R_c = L/2$ we obtain $n_{\text{max}} = 7$. Those values can be further refined to optimize the computing time without loosing accuracy [118].

Ewald summation is the most correct method to treat Coulombic interactions, even if in the case of simulations of disordered systems is sometimes criticized because it imposes an artificial periodicity in a system that is not inherently periodic. Another inconvenience is that it is computationally expensive. An optimized Ewald sum scales with the number of atoms as $O(N^{3/2})$, while for large systems the computation of the short-range potential can scale as O(N). There are several variants of the Ewald sums allowing scalings of $O(N \log N)$. They are called particle-mesh methods because they exploit the fact that the Poisson equation can be solved much more efficiently if the charges are distributed in a mesh. The most popular ones are the Particle-Particle / Particle-Mesh (PPPM) [119], the Particle Mesh Ewald (PME) [120], and the Smooth Particle Mesh Ewald (SPME) [121] methods.

3.6 Integration algorithms

We already said that equation (3.1) can only be solved numerically. So we need to discretize the trajectory and use an integrator to advance over small time steps:

$$\mathbf{r}_i(t_0) \to \mathbf{r}_i(t_0 + \Delta t) \to \mathbf{r}_i(t_0 + 2\Delta t) \to \cdots \mathbf{r}_i(t_0 + n\Delta t).$$
 (3.16)

The desirable properties of such an integrator are [27, 30]:

- Minimal need to compute the forces: This is the most time consuming step in the simulation, so any algorithm requiring more than one cycle of evaluation of the forces per time step will not be efficient.
- Good stability if large time steps Δt are used.
- Good accuracy.
- Good conservation of energy and momentum.
- Time-reversibility.
- Conservation of the phase space volume (in this case they are called symplectic integrators).
 The immediately apparent solution would be to use a simple Taylor expansion, so:

$$\mathbf{r}_i(t_0 + \Delta t) = \mathbf{r}_i(t_0) + \frac{\mathbf{d}\mathbf{r}_i(t_0)}{\mathbf{d}t}\Delta t + \frac{1}{2}\frac{\mathbf{d}^2\mathbf{r}_i(t_0)}{\mathbf{d}t^2}\Delta t^2 + O(\Delta t^3)$$
(3.17)

But this algorithm is unstable and inaccurate. A better solution was proposed by Verlet. If we sum the Taylor expansions for $+\Delta t$ and $-\Delta t$, the terms in Δt , Δt^3 , etc. cancel and we obtain:

$$\mathbf{r}_{i}(t_{0} + \Delta t) = -\mathbf{r}_{i}(t_{0} - \Delta t) + 2\mathbf{r}_{i}(t_{0}) + \mathbf{a}_{i}(t_{0})\Delta t^{2} + O(\Delta t^{4}). \tag{3.18}$$

The velocities are not used in the algorithm, but they can be obtained as:

$$\mathbf{v}_i(t_0) = \frac{1}{2\Delta t} \left[\mathbf{r}_i(t_0 + \Delta t) - \mathbf{r}_i(t_0 - \Delta t) \right]. \tag{3.19}$$

The intrinsic error of the algorithm or local truncation error is therefore of $O(\Delta t^4)$. Two equivalent algorithms producing the same trajectory as the Verlet integrator are the leap-frog algorithm:

$$\mathbf{r}_{i}(t_{0} + \Delta t) = \mathbf{r}_{i}(t_{0}) + \mathbf{v}_{i} \left(t_{0} + \frac{\Delta t}{2}\right) \Delta t, \tag{3.20}$$

$$\mathbf{v}_i \left(t_0 + \frac{\Delta t}{2} \right) = \mathbf{v}_i \left(t_0 - \frac{\Delta t}{2} \right) + \mathbf{a}_i(t_0) \Delta t. \tag{3.21}$$

And the velocity-Verlet algorithm:

$$\mathbf{r}_i(t_0 + \Delta t) = \mathbf{r}_i(t_0) + \mathbf{v}_i(t_0)\Delta t + \frac{1}{2}\mathbf{a}_i(t_0)\Delta t^2,$$
(3.22)

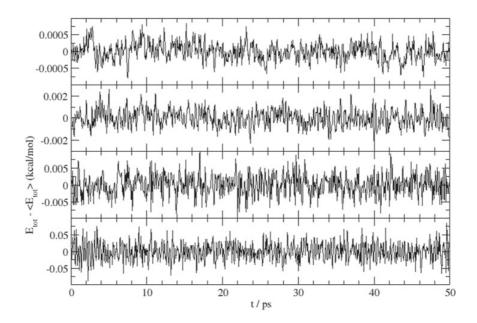


Figure 4. Energy fluctuations observed in a 50 ps simulation of water using a flexible model and different time steps. From top to bottom the four frames correspond to simulations performed with timesteps $\Delta t = 0.2, 0.5, 1$ and 2 fs, respectively. Simulations with $\Delta t = 3$ fs are already unstable. Note the different scales on the y axis.

$$\mathbf{v}_{i}(t_{0} + \Delta t) = \mathbf{v}_{i}(t_{0}) + \frac{1}{2} \left[\mathbf{a}_{i}(t_{0}) + \mathbf{a}_{i}(t_{0} + \Delta t) \right] \Delta t.$$
 (3.23)

Those three algorithms are very simple, but nevertheless they are efficient, stable and reasonably accurate. Furthermore they are also time-reversible and symplectic, so they constitute a good choice as integrators for an MD simulation.

The other type of algorithms sometimes used to integrate the equations of motion are of the predictor-corrector type. The iteration starts with a prediction of the new positions and all the required derivatives (\mathbf{v} , \mathbf{a} , $\dot{\mathbf{a}}$, etc.). Then the correct forces at $t+\Delta t$ are evaluated using the new positions. And finally the difference between the predicted and corrected forces is used to correct the predicted values. For small time steps this algorithm is more accurate than the Verlet integrator and their equivalents, but it is neither time reversible nor symplectic, so energy conservation in long simulations may be worse, in particular if a long time step is used. An extensive discussion on the merits of each algorithm is given in [122].

We also need to choose a correct time step, which requires a compromise. A small time step will improve the accuracy of the numerical solution of equation (3.1), but at the cost of requiring a much larger number of time steps to obtain a trajectory of the required length. On the other hand, a too large value for Δt will cause large fluctuations or drifts in the energy, and can even make the simulation unstable. In any case Δt should not be larger than the mean time between collisions, so for example a typical time step to simulate a Lennard-Jones fluid such as argon close to the triple point is 5 fs. This can also be taken as a usual value when simulating rigid molecules. In the case of flexible molecules, Δt needs to be small enough to model correctly the fastest intramolecular vibrational motions, so usually $\Delta t = 0.5$ –1 fs in simulations of systems presenting C-H stretches. A good way to see if the time step employed is reasonable is to check if the total energy is conserved in a microcanonical simulation (see next section). Figure 4 shows an example of the influence of the time step on the energy conservation in a standard simulation. Even if the local truncation error of the algorithm and the round-off errors of

the computer make the total energy fluctuate, these fluctuations must be small. There are two standard criteria based on energy fluctuations to ensure the stability of the integration results [123]:

$$|\Delta E| = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{E_i - E_0}{E_0} \right| \le 0.003,$$
 (3.24)

where E_0 is the initial total energy of the system and N is the number of time steps, or

$$\Delta E^{\text{rms}} = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^{N} (E_i - \langle E_i \rangle)^2}}{\sqrt{\frac{1}{N} \sum_{i=1}^{N} (K_i - \langle K \rangle)^2}} \le 0.01,$$
(3.25)

where K is the kinetic energy. However very often less strict criteria can be employed [123].

A final remark is that we cannot expect to obtain an exact solution to the equations of motion for a very long time. This is due to the extreme sensitivity of the trajectory to the initial conditions, which results in a so called Lyapunov instability [30]. Fortunately we are not interested in obtaining exact trajectories, but in computing statistical properties, and we can assume that the trajectory generated by our algorithm is representative of a true trajectory, even if we do not know which one, and use it to obtain the desired properties.

3.7 Thermodynamic ensembles

The integration of equation (3.1) keeps constant the number of particles N, the volume of the simulation cell V, and the total energy of the system, E. This implies that we will generate a trajectory in the microcanonical or NVE ensemble. But integration errors, force fluctuations and inconsistencies in the forces usually generated by the cutoff may cause slow drifts in the total energy. Additionally, the total energy is constant, but not the kinetic and potential energy contributions, so a system that is not in equilibrium will go to it while the temperature changes. It is therefore desirable to have some way of controlling the temperature. Furthermore we may prefer to work at constant temperature or pressure in order to be able to compare the simulation results with experiment. This can be done using one of the available thermostats and barostats. A thorough review of the various thermostats proposed in the literature is given in [124].

- Constant temperature simulations: The temperature of the system is directly connected with its kinetic energy, so we can change T by modifying the velocities of the atoms. The simplest scheme then to control the temperature is the so called 'velocity scaling'. It consists on scaling all the velocities by a factor $\sqrt{T_B/T(t)}$, where T_B is the desired temperature and T(t) is the instantaneous temperature of the system before the scaling. This scaling can be done regularly, every N_{scale} steps, or whenever T(t) goes out of some specified limits. However this method does not allow to sample the true canonical or NVT ensemble and the scaling affects the dynamics of the system in an unpredictable way.

A more physical way of controlling the temperature is to use the thermostat proposed by Berendsen [125], consisting in a weak coupling of the system to a heat bath. This is done by modifying the equation of motion using:

$$\mathbf{a}_i = \frac{\mathbf{F}_i}{m_i} + \frac{1}{2\tau_T} \left(\frac{T_B}{T(t)} - 1 \right) \mathbf{v}_i. \tag{3.26}$$

The additional term acts as a frictional force and the coupling constant τ_T determines the strength of the coupling. This method allows to achieve rapidly the desired temperature and it is very flexible. In the limit of infinitely weak coupling, $\tau_T \to \infty$, we recover the NVE ensemble, while if $\tau_T = \Delta t$ the method is equivalent to the velocity scaling. However it does not allow to sample the correct canonical ensemble and it can affect the system dynamics when small values of τ_T

are used. Occasionally some strange artefacts due to the transfer of energy from high frequency to low-frequency degrees of freedom have also been observed, producing the so called "flying ice cube effect" [124].

The Andersen thermostat also uses the coupling to a heat bath, but in this case the coupling is represented by stochastic forces that act occasionally on randomly selected particles and the coupling strength determines the frequency of the collisions. At each collision the particle gets a new velocity selected from a Maxwell-Boltzmann distribution corresponding to the bath temperature, T_B [126]. Contrary to the previous methods, the Andersen thermostat samples the canonical ensemble. But the dynamics of the system are found to depend on the collision frequency [30], so only the thermodynamic and static properties can be used safely.

Finally the most rigourous method to control the temperature is to use the Nosé-Hoover thermostat. In this case the heat bath becomes an integral part of the system by adding an artificial variable with an associated effective mass, so we have an extended lagrangian [127–129]. This method samples the correct canonical ensemble and it does not seem to produce severe effects on the dynamics, even if some care is always needed [30]. The main inconvenience is that T(t) goes to T_B in an oscillatory way and that if the effective mass of the artificial variable is too large the coupling will be too loose, causing a poor temperature control and requiring very long times to attain the canonical distribution. It has also been found that in some exceptional cases the Nosé-Hoover dynamics is not ergodic [124] and a solution has been proposed consisting of coupling the thermostat to another thermostat or to a chain of thermostats [130].

In practice it is recommended to equilibrate the system using the simple velocity scaling or Berendsen's thermostat with a small coupling constant in order to attain quickly the desired temperature. The production runs can be performed in the NVE ensemble if the total energy and the temperature do not experience large drifts. Otherwise the Nosé-Hoover thermostat can be employed to keep the temperature constant or even Berendsen's thermostat with a large coupling constant ($\tau_T > 0.5 \, \mathrm{ps}$).

Constant pressure simulations: The approaches to control the pressure are similar to those employed for the temperature. In this case the instantaneous virial takes the role of the temperature and the pressure is controlled by scaling the system volume. It has also to be noted that the volume of the cell can be varied uniformly in the three spatial directions, or that both the shape and the volume of the cell can be changed. This is for example needed when studying crystal phase transformations [131]. The interested reader can consult the technical details concerning different constant pressure algorithms in references [126, 132–134].

3.8 Speeding up the simulations

There are several methods to increase the efficiency of an MD simulation. Here we discuss briefly some of the standard ways usually employed to save CPU time:

Neighbour and cell lists: In a system composed of N atoms, for each atom i we need to compute the distances with the remaining N-1 atoms, and then evaluate the total force on i arising from all the atoms inside a sphere of radius R_c . The first step requires $\frac{1}{2}N(N-1)$ operations, but if R_c is much smaller than the size of the simulation cell this number can be strongly reduced using neighbour and cell lists. In the first case, for each atom we keep a list of all the atoms that are at a distance smaller than $R_c + \Delta_{\text{list}}$, where Δ_{list} is a value chosen by the user (typically $\approx 1-2 \,\text{Å}$). So at each time step, the search for the atoms situated at a distance $< R_c$ of a given atom is done only over the atoms included in its neighbour list. The list has to be updated regularly to take into account the atoms that enter or leave the sphere, or even better one can keep a record of the atomic displacements and update it whenever one of them has moved more than Δ_{list} . For very large systems, where $L >> R_c$, it is more convenient to use the cell list or linked-list method. It consist of dividing the simulation box in several subcells of size $L' \geq R_c$ and then each particle of the system is assigned to one of the

subcells. For a given atom, only the atoms located in the same subcell or the neighbouring subcells can be at a distance $< R_c$, so the search is limited to those subcells. Both methods can be combined and an optimized MD code may achieve a scaling with the number of atoms of O(N) if there are no long-range forces.

- Intramolecular constraints: Another way to increase the efficiency of the simulation is to use a large time step. But as we saw previously, Δt must be always much smaller than the fastest motion in the system. In the case of a molecular system this limit will be given by the intramolecular vibrations (mainly stretching and to a lesser extent angle bending), so if we freeze these high frequency motions by applying distance constraints to the bond lengths and angles, then we can use a larger time step in the simulation. There are several methods to apply these constraints, but possibly the most used is the SHAKE method proposed by Ryckaert, Ciccotti and Berendsen [135]. It is up to the user to decide which motions can be constrained or not. As a rule of thumb fixing the bond distances does not affect much the observed dynamics, but constraining also the bond angles may entail some risks. As an example van Gunsteren and Karplus showed that the introduction of bond length constraints in a protein model did not perturb the structure and dynamics, but constraining the bond angles altered severely the dynamics of the protein [136]. Tironi et al. have also pointed out that treating the intramolecular high-frequency motions as rigid is probably a more correct approximation to the quantum dynamics than describing them with a harmonic classical potential, recommending the use of rigid models for simulations of molecular liquids or solvents [137].
 - In the case of small molecules, such as water, methane, benzene, etc., it is even possible to go a step further and define the full molecule as a rigid body. In this case the motion is decomposed into a center of mass translation and a rotation about the center of mass. The first one is treated in the standard way, while the rotations are described using Hamilton's quaternions [27]. The use of a rigid molecule may result in saving computer time by a factor of 5–10, as Δt can be increased while maintaining the same energy conservation.
- **Multiple time steps:** Another approach consists of maintaining a flexible molecule, but using different rates to integrate different degrees of freedom. Thus high frequency motions, such as bond stretching, are evaluated at each time step, while the low frequency motions are evaluated more rarely (typically every $5-10 \, \Delta t$) and the same values for the calculated low-frequency forces are used until the next update. Normally two different time scales are employed, but nothing forbids using more than two. In any case, the evaluation of the physical properties should be done only at the larger time step. Time reversible algorithms to perform multiple time scale MD simulations exist [138, 139]. They are particularly appropriate to simulate proteins or other macromolecules showing a broad spectrum of relaxation times.
- United atom models: A final strategy that allows to save a significant amount of computing time is to employ a united atom model, where several atoms are joined into a single interaction site. The most typical case consists in replacing CH₃ or CH₂ groups by a single site. In this way, not only the fastest degrees of freedom are removed, but additionally the number of atoms composing the system is largely reduced. However some atomic details are inevitably lost, so this method can be considered already as a kind of coarse grained simulation [140].

3.9 Running the simulation

We can consider that there are four different steps in a MD simulation:

1. Setting up the system: This stage includes most of the aspects mentioned in the preceding sections. We have to create a reasonable configuration of the system we are interested in, choose a valid force field for this system, select a cut-off radius and the method to treat the electrostatic interactions if the system contains partial charges, choose an integration algorithm (if the MD code we use allows it) and a correct time step, and select the working ensemble (NVE, NVT or NPT). At this point we have to be sure that the size of the system is enough to explore the physics of interest. For example, if we

- are interested in studying structural inhomogeneities of a size of ξ_a , the box size must be at least $2\xi_a$. Similarly, if we want to investigate the sound propagation on a liquid, L must be sufficiently large to allow us to access the low Q values at which the spectrum shows a well resolved Rayleigh-Brillouin triplet [141].
- 2. Equilibration: Normally the initial configuration will not be representative of the conditions we want to explore. For example, if it comes from a previous MD simulation we may want to simulate the system at a different temperature. If the initial positions were determined randomly we may have some atoms that are too close or a lack of short-range correlations which is not real. On the other hand, if we start from an ordered configuration and we are interested in the liquid, we need to melt the system and guarantee that the long-range order has completely disappeared. This step usually requires the use of a NPT simulation in order to allow the system to achieve the equilibrium density corresponding to the desired pressure and temperature. However the choice of the thermostat and barostat is not critical, as this part of the simulation will not be used to compute any property. It is important to ensure that we have achieved an equilibrated state, so during this phase we should follow the pressure, the density and the different energy components, which at equilibrium should fluctuate around some average value without showing any drift. If we have started from a lattice configuration and we are melting it, we can also check the evolution of a translational and/or rotational order parameter [27]. Finally we can verify that some standard properties such as the radial distribution function or the mean square displacement are independent of the segment of the trajectory employed to compute them, but we have to be aware that this will be the case only when the segment is long enough so that the ergodic hypothesis holds (see below).
- 3. **Production:** Once we have attained an equilibrium situation at the desired temperature and pressure, we can perform the production run. It may be convenient to work in the NVE ensemble to avoid any influence from the thermostat or the barostat, as we saw before. Nevertheless in very long simulations the drift in temperature originated mainly from truncation effects can be embarrassing, so we may need to apply any of the thermostats described above. The pressure should oscillate around the average value achieved during the equilibration, so normally there is no need to use any barostat. Thus we can work with a fixed cell and avoid the inconvenience of dealing with volume fluctuations during the analysis of the trajectory. A clear exception here is in the case where we are interested in simulating a phase transition [131, 142]. We also have to decide how long to run the simulation. We need to run it for a time much longer than the relaxation time of the property of interest. But the simulation time should also be sufficiently long to ensure the validity of the ergodic hypothesis, i.e. that the time average of a given property coincides with the ensemble average [30]. While this condition is true in most cases (such as when simulating a standard liquid or crystal), some systems such as glasses are not ergodic, so the results will depend on the history of the system. Similarly, if we study a large biomolecule that can adopt different conformations, the simulation time should be enough to allow the macromolecule to explore all the possible configurations. Otherwise the results obtained (e.g. atomic displacements) may correspond to one of the global conformations of the macromolecule, but not to the real situation.
- 4. **Analysis:** Finally the simulated trajectory must be analysed to extract the desired properties. We have access to the atomic positions, velocities and even forces as a function of time, so any statistical mechanical property that can be expressed in terms of those variables can be computed. In the following section we present some examples of the different properties that can be computed in a MD simulation, while a more detailed account can be found in this collection [143].

3.10 Data analysis and error estimation

We will distinguish three kind of properties that can be computed from a standard MD simulation: thermodynamic, structural and dynamical properties.

1. **Thermodynamic properties:** The temperature and the pressure are directly related to the kinetic energy and to the virial, respectively:

$$T = \frac{2E_k}{3(N - N_c)k_B} = \frac{\sum_{i}^{N} m_i v_i^2}{3(N - N_c)k_B},$$
(3.27)

where N_c is the number of constraints.

$$P = \frac{1}{V} \left[Nk_B T - \frac{1}{3k_B T} \sum_{i}^{N} \sum_{j>i} \mathbf{r}_{ij} \mathbf{F}_{ij} \right].$$
 (3.28)

Both are normally calculated at each time step and used to control the simulation by means of any of the thermostats or barostats that we showed previously. Two other quantities that are easily computed and can be compared to experimental values in order to validate the simulation are the density and the potential energy. The latter can be compared with the vaporization or sublimation enthalpies for a liquid or a crystal, respectively.

The heat capacity at constant volume could be calculated from the thermodynamic expression:

$$C_v = \left(\frac{\partial U}{\partial T}\right)_V,\tag{3.29}$$

but this requires a series of simulations at different temperatures. Alternatively it can be obtained from a single simulation from the fluctuations in the total energy, as:

$$C_v = \frac{\langle \delta E^2 \rangle_{NVT}}{k_B T^2}. (3.30)$$

Here it must be noted that the previous equation can be applied only to a simulation performed in the canonical ensemble (see section 3.7). Otherwise the correct expressions for the microcanonical [144, 145] or Berendsen's thermostat [146] should be used. Similar expressions exist for the adiabatic compressibility κ_s , and the thermal pressure coefficient, γ_v , while the remaining response functions can be derived from C_v , κ_s , and γ_v using classical thermodynamic relations [28].

Entropic properties cannot be computed directly, as they are not defined as time averages over a phase-space trajectory. Usually they are obtained from MC simulations, but there are also some specialized methods that can be applied to MD simulations in order to compute the entropy or the free energy [28, 147].

2. **Structure:** The most obvious structural quantity to compute is the radial distribution function, which describes how the atoms are radially packed around each other on average. It can be defined as the ratio between the local density on a sphere shell of thickness δr at a distance r from the chosen atom and the average density. For an isotropic monatomic system [27, 28]:

$$g(r) = \frac{1}{N\rho} \left\langle \sum_{i}^{N} \sum_{j \neq i} \delta(r - r_{ij}) \right\rangle. \tag{3.31}$$

In the case of a polyatomic system we can compute all the n partial correlation functions between different elements α and β in the same way, and then obtain the total pair-correlation function as [148]:

$$G(r) = \sum_{\alpha,\beta}^{n} c_{\alpha} c_{\beta} \bar{b_{\alpha}} \bar{b_{\beta}} [g_{\alpha\beta}(r) - 1], \qquad (3.32)$$

where c_{α} and c_{β} are the number concentration and \bar{b}_{α} and \bar{b}_{β} the weight for each element, so using the neutron scattering lengths or the x-ray form factors in \bar{b} we can compare with experimental results.

In the case of molecular liquids, the structural information provided by the radial distribution functions is limited even when we calculate the full set of partials, as some information is lost

190 Collection SFN

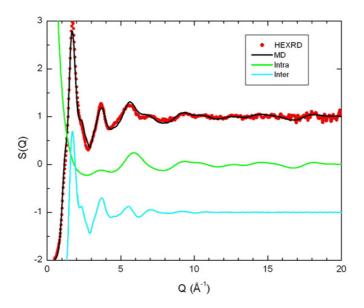


Figure 5. Comparison of experimental (x-ray diffraction) and simulated structure factors for liquid 1-ethyl-3-methylimidazolium.

due to the averaging over the angular part of the atomic pair distribution. We can avoid this by computing the spatial distribution function [10, 149].

The static structure factor can be obtained as the Fourier transform of G(r) or directly by means of the Debye equation:

$$S(Q) = \sum_{i,j}^{N} b_i b_j \frac{\sin(Qr_{ij})}{Qr_{ij}}$$
(3.33)

This expression can also be employed to calculate the diffraction pattern of a powder crystal, but it does not take into account the infinite periodicity of the crystal lattice, so the resolution of the calculated pattern will be quite poor.

Figure 5 shows an example comparing the simulated structure of a room temperature ionic liquid, 1-ethyl-3-methylimidazolium bromide, with the experimental data obtained from high-energy X-ray diffraction measurements [150]. The partial radial distribution functions involving the bromide anion are shown in figure 6, and the spatial distribution of those anions around the imidazolium cation is shown in figure 7. While the experimental data serve to validate the model and the simulation, the detailed information presented in those figures can only be obtained from the simulation.

3. **Dynamical properties:** One of the simplest properties to compute from an MD trajectory is the atomic mean square displacement:

$$u^{2}(t) = \frac{1}{N} \left\langle \sum_{i} |\mathbf{r}_{i}(t_{0} + t) - \mathbf{r}_{i}(t_{0})|^{2} \right\rangle, \tag{3.34}$$

where the brackets indicate an average over different time origins, t_0 . In the long time limit the slope of $u^2(t)$ gives the self-diffusion coefficient, D:

$$D = \lim_{t \to \infty} \frac{u^2(t)}{6t},\tag{3.35}$$

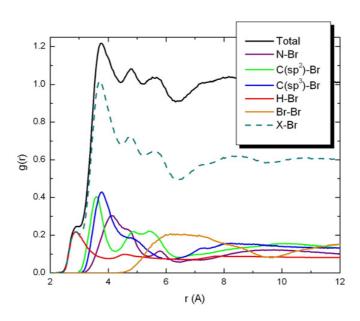


Figure 6. Total radial distribution function for liquid 1-ethyl-3-methylimidazolium and partial r.d.f.s containing the Br anion. Each partial has been weighted by the factor $c_{\alpha}c_{\beta}\bar{b}_{\alpha}\bar{b}_{\beta}$ in order to show its contribution to the total signal measurable experimentally.

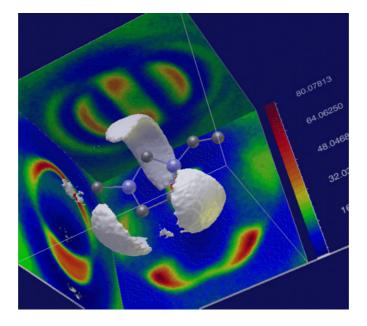


Figure 7. Spatial radial distribution showing the distribution of Br anions around the imidazolium cation for liquid 1-ethyl-3-methylimidazolium.

which can be compared with the diffusion coefficient derived from quasielastic neutron scattering or NMR measurements. It should be noted that in many cases (supercooled or viscous liquids, macromolecules, etc.) the diffusive regime is achieved only after a long time and is preceded by a subdiffusive regime where $u^2(t) \propto t^{\alpha}$, with $\alpha < 1$, so it is recommended to check that the diffusive regime has been well attained before comparing with any experimental data.

It is also possible to compute all kinds of correlation functions as:

$$C(t) = \langle \mathbf{A}_i(t_0 + t)\mathbf{A}_i(t_0) \rangle, \tag{3.36}$$

where **A** refers to any variable and the brackets indicate again an average over different time origins. For example, setting $\mathbf{A} = \mathbf{v}$ we obtain the velocity autocorrelation function and its Fourier transform will give the vibrational density of states, $G(\omega)$, which can be compared to that measured using inelastic neutron scattering spectroscopy. Other autocorrelation functions can be connected with different measurable transport coefficients such as bulk and shear viscosities or thermal conductivities via Green-Kubo relationships [27].

Orientational correlation functions can be calculated as:

$$C_l(t) = \langle P_l[\cos \theta(t)] \rangle, \tag{3.37}$$

where P_l indicates the Legendre polynomial of order l and $\theta(t)$ is the angle swept by a reference vector.

It is also possible to use the MD simulations to compute the self- and distinct Van Hove functions [151]:

$$G_s(\mathbf{r},t) = \frac{1}{N} \left\langle \sum_{i}^{N} \delta[\mathbf{r} - \mathbf{r}_i(t_0 + t) - \mathbf{r}_i(t_0)] \right\rangle, \tag{3.38}$$

$$G_d(\mathbf{r},t) = \frac{1}{\rho N} \left\langle \sum_{i}^{N} \sum_{j \neq i}^{N} \delta[\mathbf{r} - \mathbf{r}_i(t_0 + t) - \mathbf{r}_j(t_0)] \right\rangle.$$
(3.39)

These are the Fourier transforms in space and time of the incoherent and coherent dynamic structure factors, $S(Q, \omega)$, but they cannot be obtained by any experimental technique. In order to compare to experiment it is easier to calculate the intermediate scattering functions:

$$I_s(\mathbf{Q}, t) = \frac{1}{N} \sum_{i=1}^{N} \left\langle \exp[-i\mathbf{Q}\mathbf{r}_i(t_0)] \exp[i\mathbf{Q}\mathbf{r}_i(t_0 + t)] \right\rangle, \tag{3.40}$$

$$I(\mathbf{Q},t) = \frac{1}{N} \sum_{i}^{N} \sum_{j}^{N} \left\langle \exp[-i\mathbf{Q}\mathbf{r}_{i}(t_{0})] \exp[i\mathbf{Q}\mathbf{r}_{j}(t_{0}+t)] \right\rangle.$$
(3.41)

They can be compared directly with the intermediate function measured on a neutron spin-echo instrument, or Fourier transformed to give the analogue of $S(Q, \omega)$ measured on a neutron triple-axis, time-of-flight or backscattering spectrometer.

4. **Errors:** As in the analysis of an experimental result, any value obtained from a simulation should be computed with its associated error. For any property calculated as the average of the value of a given observable during the simulation, we could estimate the error as the variance of the set of data. However, contrary to most experimental situations, the set of data obtained from the simulation does not represent independent measurements. The error corresponding to averaging over a series of correlated data can be calculated using several techniques, such as the block method or the integration of the autocovariance [30, 152–154]. In the case of correlation functions, a simple estimate for the error has been proposed by Zwanzig and Ailawadi [155]:

$$R(t) = \frac{C(t)}{C(0)} \pm \sqrt{\frac{2\tau}{T}} \left[1 - \frac{C(t)}{C(0)} \right],\tag{3.42}$$

where $\frac{C(t)}{C(0)}$ is the normalized correlation function, τ is the relaxation time of this correlation function, and T is the total simulation time. Although this estimation is obtained assuming that C(t) is a random Gaussian variable, it has been shown that this is also a good order-of-magnitude estimate for

the uncertainty of the correlation function of a non-Gaussian velocity distribution [156]. If we are computing a self-correlation function we can compute R(t) for each of the N atoms and then average over them, so this error will be divided by an additional factor \sqrt{N} .

3.11 Molecular dynamics codes and utilities

There is a large number of good software to perform MD simulations available in the public domain. Some general, reliable and popular programs worth mentioning are DL_POLY [157], NAMD [158], LAMMPS [159], Gromacs [160], TINKER [161], or GULP [162], but many others can be found in the web. Popular commercial codes include GROMOS [63], CHARMM [163], AMBER [64] or the Forcite module inside Materials Studio [164]. Some useful software to visualize the simulation configurations or trajectories includes VMD [165], gOpenMol [166], or Rasmol [167], among others. Many of the above cited programs also contain some analysis tools to compute a large range of properties from the MD trajectories. Other specific analysis packages for MD simulations are MD-Tracks [168] and nMoldyn [169], which is specially focused on the computation of neutron scattering functions [143].

4. CONCLUSION

Molecular Dynamics simulations represent a simple and efficient method to explore the structure and dynamics of complex systems. Their accuracy and reliability depends on the quality of the force field employed to model the intra- and intermolecular interactions, so they should always be validated by comparing their output with experimental results. Neutron scattering is one of the best suited techniques to perform this operation, as the measured observables (static and dynamic structure factors) and the sampled time and length scales correspond exactly to the quantities that can be computed from an MD simulation. However the utility of a simulation is not to reproduce a experimental result, but to allow to understand the microscopic origin of the physical properties observed or to predict qualitatively the behaviour expected at conditions that cannot be accessed experimentally. Thus the main output of a simulation is a phase space trajectory and the scientific challenge consists of using the detailed information about the atomic positions, velocities, and forces to obtain relevant information about the physics of the system studied. The developments occurring during the last 50 years have now made available a number of simple and user friendly software packages, together with reliable force fields, so today computer simulations are easily accessible to experimentalists and they are increasingly becoming an essential tool to analyze and interpret complex experiments and/or systems.

Acknowledgements

I would like to thank Bachir Aoun, Eric Pellegrini, and Mark Johnson for their valuable advice and enlightening discussions during the writing of this review.

References

- [1] 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.*, 21:1087–1092, 1953.
- [2] B. J. Alder and T. E. Wainwright. Phase transition for a hard sphere system. *J. Chem. Phys.*, 27:1208–1209, 1957.
- [3] B. R. Brooks et al. Charmm: The biomolecular simulation program. *J. Comput. Chem.*, 30:1545–1615, 2009.
- [4] J. C. Phillips et al. Scalable molecular dynamics with NAMD. J. Comput. Chem., 26:1781–1802, 2005.

- [5] D. A. Case et al. The Amber biomolecular simulation programs. *J. Comput. Chem.*, 26:1668–1688, 2005.
- [6] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen. GROMACS: Fast, flexible, and free. *J. Comput. Chem.*, 26:1701–1718, 2005.
- [7] M. Christen et al. The GROMOS software for biomolecular simulation: Gromos05. *J. Comput. Chem.*, 26:1719–1751, 2005.
- [8] W. Smith, C. W. Yong, and P. M. Rodger. DL_POLY: Application to molecular simulation. *Mol. Sim.*, 28:385–471, 2002.
- [9] W. Humphrey, A. Danke, and K. Schulten. VMD: Visual molecular dynamics. *J. Mol. Graph.*, 14:33–38, 1996.
- [10] D. L. Bergman, L. Laaksonen, and A. Laaksonen. Visualization of solvation structures in liquid mixtures. *J. Mol. Graphics Mod.*, 15:301–306, 1997.
- [11] T. Róg, K. Murzin, K. Hinsen, and G. R. Kneller. nMoldyn: A program package for a neutron scattering oriented analysis of molecular dynamics simulations. *J. Comput. Chem.*, 24:657–667, 2003.
- [12] R. M. Martin. *Electronic structure: Basic theory and practical methods*. Cambridge University Press, 2004.
- [13] R. Car and M. Parrinello. Unified approach for molecular dynamics and density-functional theory. *Phys. Rev. Lett.*, 55:2471–2474, 1985.
- [14] T. Deutsch, Collection SFN, 12: 33–76, 2011.
- [15] T. Hansson, C. Oostenbrink, and W. F. van Gunsteren. Molecular dynamics simulations. *Curr. Opin. Struct. Biol.*, 12:190–196, 2002.
- [16] M. Karplus and J. A. McCammon. Molecular dynamics simulations of biomolecules. *Nature Struct. Biol.*, 9:646–652, 2002.
- [17] A. Gavezzotti. *Molecular Aggregation: Structure analysis and molecular simulation of crystals and liquids.* Oxford University Press, 2007.
- [18] A. R. Leach. *Molecular modelling: Principles and applications*, chapter 3: Empirical force field models: molecular mechanics, pages 196–204. Addison Wesley Longman Limited, 1996.
- [19] A. D. Mackerell. *Computational Methods for Protein Structure Prediction and Modeling*, chapter 2: Empirical force fields, pages 45–69. Springer New York, 2007.
- [20] A. D. Mackerell. Empirical force fields for biological macromolecules: Overview and issues. *J. Comput. Chem.*, 25:1584–1604, 2004.
- [21] M. Jalaie and K. B. Lipkowitz. *Reviews in Computational Chemistry*, volume 14, chapter Published force field parameters for molecular mechanics, molecular dynamics and Monte Carlo simulations, pages 441–486. Wiley-VCH, 2000.
- [22] M. P. Allen. *Computational Soft Matter: From synthetic polymers to proteins*, volume 23 of *NIC series*, chapter Introduction to molecular dynamics simulation, pages 1–28. 2004.
- [23] G. Sutmann. *Quantum simulations of complex many-body systems: From theory to algorithms*, volume 10 of *NIC series*, chapter Classical molecular dynamics, pages 211–254. 2002.
- [24] G. Sutmann. *Computational nanoscience: Do it yourself!*, volume 31 of *NIC series*, chapter Molecular dynamics Vision and reality, pages 159–194. 2006.
- [25] G. Sutmann. *Multiscale simulation methods in molecular sciences*, volume 42 of *NIC series*, chapter Molecular dynamics Extending the scale from microscopic to mesoscopic, pages 1–49 2009
- [26] M. E. Tuckerman and G. J. Martyna. Understanding modern molecular dynamics: Techniques and applications. *J. Phys. Chem. B*, 104:159–178, 2000.
- [27] M. P. Allen and D. J. Tildesley. Computer simulation of liquids. Oxford University Press, 1987.
- [28] J. M. Haile. Molecular dynamics simulation: Elementary methods. Wiley-Interscience, 1992.
- [29] D. C. Rapaport. The art of molecular dynamics simulation. Cambridge University Press, 2004.
- [30] D. Frenkel and B. Smit. Understanding Molecular Simulation. Academic Press, 2002.

- [31] H. C. Urey and C. A. Bradley Jr. The vibrations of pentatonic tetrahedral molecules. *Phys. Rev.*, 38:1969–1978, 1931.
- [32] 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.*, 118:11225–11236, 1996.
- [33] Y. El Youssoufi, M. Herman, and J. Liévin. The ground electronic state of 1,2-dichloroethane I. Ab initio investigation of the geometrical, vibrational and torsional structure. *Mol. Phys.*, 94:461–472, 1998.
- [34] R. A. Buckingham. The classical equation of state of gaseous helium, neon and argon. *Proc. Royal Soc. London A*, 168:264–283, 1938.
- [35] D. E. Williams. Improved intermolecular force field for molecules containing H, C, N and O atoms, with application to nucleoside and peptide crystals. *J. Comput. Chem.*, 22:1154–1166, 2001.
- [36] G. Filippini and A. Gavezzotti. Empirical intermolecular potentials for organic crystals: The 6-exp approximation revisited. *Acta Cryst. B*, 49:868–880, 1993.
- [37] H. Sun. COMPASS: An ab initio force field optimized for condensed-phase application Overview with details on alkane and benzene compounds. J. Phys. Chem., 102:7338–7364, 1998.
- [38] T. A. Halgren. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comput. Chem.*, 17:490–519, 1996.
- [39] J. Gasteiger and M. Marsili. Iterative partial equalization of orbital electronegativity: Rapid acess to atomic charges. *Tetrahedron*, 36:3219–3288, 1980.
- [40] A. K. Rappe and W. A. Goddard III. Charge equilibration method for molecular dynamics simulations. *J. Phys. Chem.*, 95:3358–3363, 1991.
- [41] M. K. Gilson, H. S. R. Gilson, and M. J. Potter. Fast assignment of accurate partial atomic charges: An electronegativity equalization method that accounts for alternate resonance forms. *J. Chem. Inf. Comput. Sci.*, 43:1982–1997, 2003.
- [42] E. Sigfridsson and U. Ryde. Comparison of methods for deriving atomic charges from the electrostatic potential and moments. *J. Comput. Chem.*, 19:377–395, 1998.
- [43] D. F. Green and B. Tidor. Evaluation of *ab initio* charge determination methods for use in continuum solvation calculations. *J. Phys. Chem. B*, 107:10261–10273, 2003.
- [44] B. Guillot. A reappraisal of what we have learnt during three decades of computer simulations on water. *J. Mol. Liq.*, 101:219–260, 2002.
- [45] A. Warshel, M. Kato, and A. V. Pisliakov. Polarizable force fields: History, test cases and prospects. *J. Chem. Theory Comput.*, 3:2034–2045, 2007.
- [46] T. A. Halgren and W. Damm. Polarizable force fields. Curr. Opin. Struct. Biol., 11:236–242, 2001.
- [47] P. Cieplak, F.-Y. Dupradeau, Y. Duan, and J. Wang. Polarization effects in molecular mechanical force fields. *J. Phys.: Condens. Matter*, 21:333102–21, 2009.
- [48] S. W. Rick, S. J. Stuart, and B. J. Berne. Dynamical fluctuating charge force fields: Application to liquid water. *J. Chem. Phys.*, 101:6141–6156, 1994.
- [49] P. J. Mitchell and D. Fincham. Shell model simulations by adiabatic dynamics. *J. Phys.: Condens. Matter*, 5:1031–1038, 1993.
- [50] P. Ahlström, A. Wallqvist, S. Engström, and B. Jönsson. A molecular-dynamics study of polarizable water. *Mol. Phys.*, 68:563–581, 1989.
- [51] P. Jedlovsky and J. Richardi. Comparison of different water models from ambient to supercritical conditions: a Monte Carlo simulation and molecular Ornstein-Zernike study. *J. Chem. Phys.*, 110:8019–8031, 1999.
- [52] J. M. Sorensen, G. Hura, R. M. Glaeser, and T. Head-Gordon. What can x-ray scattering tell us about the radial distribution functions of water. *J. Chem. Phys.*, 113:9149–9161, 2000.

- [53] N. L. Allinger. Allinger's molecular mechanics research lab (home of mm2/mm3/mm4 force field). http://europa.chem.uga.edu.
- [54] N. L. Allinger. Conformational analysis. 130. MM2. a hydrocarbon force field utilizing V1 and V2 torsional terms. *J. Am. Chem. Soc.*, 99:8127–8134, 1977.
- [55] N. L. Allinger, Y. H. Yuh, and J.-H. Lii. Molecular mechanics. the MM3 force field for hydrocarbons. 1. *J. Am. Chem. Soc.*, 111:8551–8565, 1989.
- [56] N. L. Allinger, K. Chen, and J.-H. Lii. Improved force field (MM4) for saturated hydrocarbons. *J. Comput. Chem.*, 17:642–668, 1996.
- [57] S. L. Mayo, B. D. Olafson, and W. A. Goddard III. Dreiding: A generic force field for molecular simulations. *J. Phys. Chem.*, 94:8897–8909, 1990.
- [58] A. K. Rappe, C. J. Casewit, K. S. Colwell, W. A. Goddard III, and W. M. Skiff. UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *J. Am. Chem. Soc.*, 114:10024–10035, 1992.
- [59] A. D. Mackerell et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem.*, 102:3586–3616, 1998.
- [60] W. D. Cornell et al. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.*, 117:5179–5197, 1995.
- [61] C. Oostenbrink, A. Villa, A. E. Mark, and W. F. van Gunsteren. A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. *J. Comput. Chem.*, 25:1656–1676, 2004.
- [62] A. D. Mackerell. CHARMM FF parameters. http://mackerell.umaryland.edu/CHARMM_ff_params.html.
- [63] W. F. van Gunsteren et al. GROMOS web site. http://www.igc.ethz.ch/GROMOS/index.
- [64] AMBER web site. http://ambermd.org.
- [65] J. R. Maple et al. Derivation of class II force fields. I. Methodology and quantum force field for the alkyl functional group and alkane molecules. *J. Comput. Chem.*, 15:162–182, 1994.
- [66] J. Gao, D. Habibollazadeh, and L. Shao. A polarizable intermolecular potential function for simulation of liquid alcohols. *J. Phys. Chem.*, 99:16460–16467, 1995.
- [67] M. Swart and P. Th. van Duijnen. DRF90: Polarizable force field. Mol. Sim., 32:471–484, 2006.
- [68] J. W. Ponder et al. Current status of the AMOEBA polarizable force field. *J. Phys. Chem. B*, 114:2549–2564, 2010.
- [69] S. Patel, A. D. MacKerell Jr., and C. L. Brooks III. CHARMM fluctuating charge force field for proteins: I. Parameterization and application to bulk organic liquid simulations. *J. Comput. Chem.*, 25:1–15, 2004.
- [70] G. Lamoureux and B. Roux. Modeling induced polarization with classical Drude oscillators: Theory and molecular dynamics simulation algorithm. *J. Chem. Phys.*, 119:3025–3039, 2003.
- [71] 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.*, 27:781–790, 2006.
- [72] W. L. Jorgensen, K. P. Jensen, and A. N. Alexandrova. Polarization effects for hydrogen-bonded complexes of substituted phenols with water and chloride ion. *J. Chem. Theory Comput.*, 3:1987–1992, 2007.
- [73] D. P. Geerke and W. F. van Gunsteren. On the calculation of atomic forces in classical simulation using the charge-on-spring method to explicitly treat electronic polarization. *J. Chem. Theory Comput.*, 3:2128–2137, 2007.
- [74] J. A. Barker and R. O. Watts. Structure of water; a Monte Carlo calculation. *Chem. Phys. Lett.*, 3:144–145, 1969.
- [75] C. Vega, J. L. F. Abascal, M. M. Conde, and J. L. Aragones. What ice can teach us about water interactions: a critical comparison of the performance of different water models. *Faraday Discuss.*, 141:251–276, 2009.

- [76] D. J. Price and C. L. Brooks III. Modern protein force fields behave comparably in molecular dynamics simulations. *J. Comput. Chem.*, 23:1045–1057, 2002.
- [77] I.-C. Yeh and G. Hummer. Peptide loop-closure kinetics from microsecond molecular dynamics simulations in explicit solvent. *J. Am. Chem. Soc*, 124:6563–6568, 2002.
- [78] A. E. Aliev and D. Courtier-Murias. Experimental verification of force fields for molecular dynamics simulations using Gly-Pro-Gly-Gly. *J. Phys. Chem. B*, 114:12358–12375, 2010.
- [79] N. Todorova, F. S. Legge, H. Treutlein, and I. Yarovsky. Systematic comparison of empirical forcefields for molecular dynamics simulation of insulin. *J. Phys. Chem. B*, 112:11137–11146, 2008.
- [80] R. S. Paton and J. M. Goodman. Hydrogen bonding and π-stacking: How reliable are force fields? A critical evaluation of force field descriptions of nonbonded interactions. J. Chem. Inf. Model, 49:944–955, 2009.
- [81] W. L. Jorgensen and J. Tirado-Rives. Potential energy functions for atomic-level simulations of water and organic and biomolecular systems. *Proc. Natl. Acad. Sci. USA*, 102:6665–6670, 2005.
- [82] T. A. Halgren. Potential energy functions. Curr. Opin. Struct. Biol., 5:205–210, 1995.
- [83] T. A. Halgren. Merck molecular force field. II. MMFF94 van der Waals and electrostatic parameters for molecular interactions. *J. Comput. Chem.*, 17:520–552, 1996.
- [84] T. A. Halgren. Merck molecular force field. III. Molecular geometries and vibrational frequencies for MMFF94. *J. Comput. Chem.*, 17:553–586, 1996.
- [85] T. A. Halgren and R. B. Nachbar. Merck molecular force field. IV. Conformational energies and geometries for MMFF94. *J. Comput. Chem.*, 17:587–615, 1996.
- [86] T. A. Halgren. Merck molecular force field. V. Extension of MMFF94 using experimental data, additional computational data, and empirical rules. *J. Comput. Chem.*, 17:616–641, 1996.
- [87] F. Ercolessi and J. B. Adams. Interatomic potentials from first principle calculations: The force-matching method. *Europhys. Lett.*, 26:583–588, 1994.
- [88] B. Waldher, J. Kuta, S. Chen, N. Henson, and A. E. Clark. Forcefit: A code to fit classical force fields to quantum mechanical potential energy surfaces. *J. Comput. Chem.*, 31:2307–2316, 2010.
- [89] I. J. Chen, D. Yin, and A. D. MacKerell Jr. Combined ab initio/empirical approach for optimization of Lennard-Jones parameters for polar-neutral compounds. J. Comput. Chem., 23:199–213, 2002.
- [90] C. M. Breneman and K. B. Wibert. Determining atom-centered monopoles from molecular electrostatic potentials. The need for high sampling density in formamide conformational analysis. *J. Comput. Chem.*, 11:361–373, 1990.
- [91] C. I. Bayly, P. Cieplak, W. D. 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.*, 97:10269–10280, 1993.
- [92] A. Jakalian, D. B. Jack, and C. I. Bayly. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J. Comput. Chem.*, 23:1623–1641, 2002.
- [93] K. Kadau, T. C. Germann, and P. S. Lomdahl. Molecular dynamics comes of age: 320 billion atom simulation on BlueGene/l. *Int. J. Modern Phys. C*, 17:1755–1761, 2006.
- [94] P. L. Freddolino, A. S. Arkhipov, S. B. Larson, A. McPherson, and K. Schulten. Molecular dynamics simulations of the complete satellite tobacco mosaic virus. *Structure*, 14:437–449, 2006.
- [95] J. L. Klepeis, K. Lindorff-Larsen, R. O. Dror, and D. E. Shaw. Long-timescale molecular dynamics simulations of protein structure and function. *Curr. Opin. Str. Biol.*, 19:120–127, 2009.
- [96] R. Schulz, B. Lindner, L. Petridis, and J. C. Smith. Scaling of multimillion-atom biological molecular dynamics simulation on a petascale supercomputer. *J. Chem. Theory Comput.*, 5:2798–2808, 2009.
- [97] A. C. T. van Duin, S. Dasgupta, F. Lorant, and W. A. Goddard III. ReaxFF: A reactive force field for hydrocarbons. *J. Phys. Chem. A*, 105:9396–9409, 2001.

- [98] S. Braun-Sand, M. Strajbl, and A. Warshel. Studies of proton translocations in biological systems: Simulating proton transport in carbonic anhydrase by EVB-based models. *Biophys. J.*, 87:2221–2239, 2004.
- [99] D. W. M. Hofmann, L. Kuleshova, and B. D'Aguanno. A new reactive potential for the molecular dynamics simulation of liquid water. *Chem. Phys. Lett.*, 448:138–143, 2007.
- [100] A. Warshel and M. Levitt. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozome. *J. Mol. Biol.*, 103:227–249, 1976.
- [101] H. Lin and D. G. Truhlar. QM/MM: What have we learned, where are we, and where do we go from here? *Theor. Chem. Acc.*, 117:185–199, 2007.
- [102] S. C. L. Kamerlin, M. Haranczyk, and A. Warshel. Progress in *ab initio* QM/MM free-energy simulations of electrostatic energies in proteins: Accelerated QM/MM studies of pK_a, redox reactions and solvation free energies. *J. Phys. Chem. B*, 113:1253–1272, 2009.
- [103] W. Xie and J. Gao. Design of a next generation force field: The X-POL potential. *J. Chem. Theory Comput.*, 3:1890–1900, 2007.
- [104] W. Xie, M. Orozco, D. G. Truhlar, and J. Gao. X-pol potential: An electronic structure-based force field for molecular dynamics simulation of a solvated protein in water. *J. Chem. Theory Comput.*, 5:459–467, 2009.
- [105] O. N. de Souza and R. L. Ornstein. Effect of periodic box size on aqueous molecular dynamics simulation of a DNA dodecamer with particle-mesh Ewald method. *Biophys. J.*, 72:2395–2397, 1997.
- [106] J. Li. *Handbook of materials modeling*, chapter 2.8: Basic molecular dynamics, pages 565–588. Springer, 2005.
- [107] 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.
- [108] Y. Yonetani. Liquid water simulation: A critical examination of cutoff length. *J. Chem. Phys.*, 124:204501–11, 2006.
- [109] B. Smit. Phase diagrams of Lennard-Jones fluids. J. Chem. Phys., 96:8639–8640, 1992.
- [110] D. J. Adams, E. M. Adams, and G. J. Hills. The computer simulation of polar liquids. *Mol. Phys.*, 38:387–400, 1979.
- [111] J. A. Barker and R. O. Watts. Monte Carlo studies of the dielectric properties of water-like models. *Mol. Phys.*, 26:789–792, 1973.
- [112] I. G. Tironi, R. Sperb, P. E. Smith, and W. F. van Gunsteren. A generalized reaction field method for molecular dynamics simulations. *J. Chem. Phys.*, 102:5451–5459, 1995.
- [113] M. Neumann, O. Steinhauser, and G. S. Pawley. Consistent calculation of the static and frequency-dependent dielectric constant in computer simulations. *Mol. Phys.*, 52:97–113, 1984.
- [114] L. Perera, U. Essmann, and M. L. Berkowitz. Effect of the treatment of long-range forces on the dynamics of ions in aqueous solutions. *J. Chem. Phys.*, 102:450–456, 1995.
- [115] A. Gil-Villegas, S. C. McGrother, and G. Jackson. Reaction-field and Ewald summation methods in Monte Carlo simulations of dipolar liquid crystals. *Mol. Phys.*, 92:723–734, 1997.
- [116] P. H. Hünenberger and W. F. van Gunsteren. Alternative schemes for the inclusion of a reaction-field correction into molecular dynamics simulations: Influence on the simulated energetic, structural, and dielectric properties of liquid water. *J. Chem. Phys.*, 108:6117–6134, 1998.
- [117] P. Ewald. Die berechnung optischer und elektrostatischer gitterpotentiale. *Ann. Phys.*, 369:253–287, 1921.
- [118] W. Smith, T. R. Forester, and I. T. Todorov. *The DL_POLY 2 user manual, version 2.19*, chapter 3.3.5: Choosing Ewald sum variables, page 90. 2008.
- [119] R. W. Hockney and J. W. Eastwood. *Computer simulations using particles*. McGraw-Hill, 1981.

- [120] T. A. 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.
- [121] U. Essmann, L. Perera, M. L. Berkowitz, T. A. Darden, H. Lee, and L. Pedersen. A smooth particle mesh Ewald method. *J. Chem. Phys.*, 103:8577–8593, 1995.
- [122] H. J. C. Berendsen and W. F. van Gunsteren. *Molecular dynamics simulation of statistical mechanical systems*, volume Proceedings of the 97th Int. Enrico Fermi school of physics, chapter Practical algorithms for dynamics simulations, pages 43–65. North-Holland, Amsterdam, 1986.
- [123] M. Watanabe and M. Karplus. Dynamics of molecules with internal degrees of freedom by multiple time-step methods. *J. Chem. Phys.*, 99:8063–8074, 1993.
- [124] P. H. Hünenberger. Thermostat algorithms for molecular dynamics simulations. *Adv. Polym. Sci.*, 173:105–149, 2005.
- [125] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, 81:3684–3690, 1984.
- [126] H. C. Andersen. Molecular dynamics at constant pressure and/or temperature. *J. Chem. Phys.*, 72:2384–2393, 1980.
- [127] S. Nosé. A unified formulation of the constant temperature molecular dynamics method. *J. Chem. Phys.*, 81:511–519, 1984.
- [128] S. Nosé. A molecular dynamics method for simulation in the canonical ensemble. *Mol. Phys.*, 52:255–268, 1984.
- [129] W. G. Hoover. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A*, 31:1695–1697, 1985.
- [130] G. J. Martyna, M. L. Klein, and M. E. Tuckerman. Nosé-Hoover chains: The canonical ensemble via continuous dynamics. *J. Chem. Phys.*, 97:2635–2645, 1992.
- [131] M. Parrinello and A. Rahman. Crystal structure and pair potentials: A molecular-dynamics study. *Phys. Rev. Lett.*, 45:1196–1199, 1980.
- [132] S. Nosé and M. L. Klein. Constant pressure molecular dynamics for molecular systems. *Mol. Phys.*, 50:1055–1076, 1983.
- [133] G. J. Martyna, D. J. Tobias, and M. L. Klein. Constant pressure molecular dynamics algorithms. *J. Chem. Phys.*, 101:4177–4189, 1994.
- [134] S. E. Feller, Y. H. Zhang, R. W. Pastor, and B. R. Brooks. Constant pressure molecular dynamics simulation: The Langevin piston method. *J. Chem. Phys.*, 103:4613–4621, 1995.
- [135] J. P. Ryckaert, G. Cicotti, 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.*, 23:327–341, 1977.
- [136] W. F. van Gunsteren and M. Karplus. Effect of constraints on the dynamics of macromolecules. *Macromolecules*, 15:1528–1544, 1982.
- [137] I. G. Tironi, R. M. Brunne, and W. F. van Gunsteren. On the relative merits of flexible versus rigid models for use in computer simulations of molecular liquids. *Chem. Phys. Letters*, 250:19–24, 1996.
- [138] M. E. Tuckerman, B. J. Berne, and G. J. Martyna. Reversible multiple time scale molecular dynamics. *J. Chem. Phys.*, 97:1990–2001, 1992.
- [139] G. J. Martyna, M. E. Tuckerman, D. J. Tobias, and M. L. Klein. Explicit reversible integrators for extended system dynamics. *Mol. Phys.*, 87:1117–1157, 1996.
- [140] S. O. Nielsen, C. F. Lopez, G. Srinivas, and M. L. Klein. Coarse grain models and the computer simulation of soft materials. *J. Phys.: Condens. Matter*, 16:R481–R512, 2004.
- [141] E. Enciso, N. G. Almarza, P. Dominguez, M. A. Gonzalez, and F. J. Bermejo. Dynamic structure factor of a helium-neon dense gas mixture: Crossover from hydrodynamics to the microscopic regime. *Phys. Rev. Lett.*, 74:4233–4236, 1995.
- [142] R. Martonak, A. Laio, and M. Parrinello. Predicting crystal structures: The Parrinello-Rahman method revisited. *Phys. Rev. Lett.*, 90:75503–4, 2003.

- [143] V. Calandrini, E. Pellegrini, K. Hinsen, and G. R. Kneller, *Collection SFN*, 12: 201–232, 2011.
- [144] J. L. Lebowitz, J. K. Percus, and L. Verlet. Ensemble dependence of fluctuations with application to machine computation. *Phys. Rev.*, 153:250–254, 1967.
- [145] P. S. Y. Cheung. On the calculation of specific heats, thermal pressure coefficients, and compressibilities in molecular dynamics simulations. *Mol. Phys.*, 33:519–526, 1977.
- [146] T. Morishita. Fluctuation formulas in molecular-dynamics simulations with the weak coupling bath. *J. Chem. Phys.*, 113:2976–2982, 2000.
- [147] C. D. Christ, A. E. Mark, and W. F. Van Gunsteren. Basic ingredients of free energy calculations: A review. *J. Comput. Chem.*, 31:1569–1582, 2010.
- [148] H. E. Fischer, A. C. Barnes, and P. S. Salmon. Neutron and x-ray diffraction studies of liquids and glasses. *Rep. Prog. Phys.*, 69:233–299, 2006.
- [149] I. M. Svishchev and P. G. Kusalik. Structure in liquid water: A study of spatial distribution functions. *J. Chem. Phys.*, 99:3049–3058, 1993.
- [150] B. Aoun, A. Goldbach, S. Kohara, J.-F. Wax, M. A. Gonzalez, and M.-L. Saboungi. Structure of a prototypic ionic liquid: Ethyl-methylimidazolium bromide. *J. Phys. Chem. B*, 114:12623– 12628, 2010.
- [151] L. van Hove. Correlations in space and time and Born approximation scattering in systems of interacting particles. *Phys. Rev.*, 95:249–262, 1954.
- [152] M. Bishop and S. Frinks. Error analysis in computer simulations. *J. Chem. Phys.*, 87:3675–3676, 1987.
- [153] H. Flyvbjerg and H. G. Petersen. Error estimates on averages of correlated data. *J. Chem. Phys.*, 91:461–466, 1989.
- [154] G. L. Deitrick, L. E. Scriven, and H. T. Davis. A new method of error analysis for molecular simulations. *Comput. Phys. Comm.*, 62:327–335, 1991.
- [155] R. Zwanzig and N. K. Ailawadi. Statistical error due to finite time averaging in computer experiments. *Phys. Rev.*, 182:280–283, 1969.
- [156] I. Bitsanis, M. Tirrell, and H. T. Davis. Statistics of correlation functions from molecular dynamics. *Phys. Rev. A*, 36:958–961, 1987.
- [157] DL_POLY web site. http://www.cse.scitech.ac.uk/ccg/software/DL_POLY.
- [158] NAMD web site. http://www.ks.uiuc.edu/Research/namd.
- [159] LAMMPS web site. http://lammps.sandia.gov.
- [160] Gromacs web site. http://www.gromacs.org.
- [161] TINKER web site. http://dasher.wustl.edu/tinker.
- [162] GULP web site. http://projects.ivec.org/gulp.
- [163] CHARMM web site. http://www.charmm.org.
- [164] Materials Studio web site. http://accelrys.com/products/materials-studio.
- [165] VMD web site. http://www.ks.uiuc.edu/Research/vmd.
- [166] gOpenMol web site. http://www.csc.fi/english/pages/g0penMol.
- [167] Rasmol web site. http://www.openrasmol.org.
- [168] MD-Tracks web site. http://molmod.ugent.be/code/wiki/MD-Tracks.
- [169] nMoldyn web site. http://dirac.cnrs-orleans.fr/plone/software/nmoldyn.