The GROMOS Software for (Bio)Molecular Simulation

Volume 2: Algorithms and Formulae for Modelling of Molecular Systems

October 27, 2012

Contents

CHAPTER 1

Introduction

In this volume, the molecular model, the molecular interactions, and the computational procedures used in GROMOS are described. Chap. 2 discusses the basic choices in the definition of a molecular model in a general way. The application of these choices within GROMOS are outlined in Chap. 3. Chap. 4 discusses various aspects of the spatial boundary conditions implemented in GROMOS. Chaps. 5, 6 and 7 deal with the description of molecular interactions divided into covalent interactions (Chap. 5), nonbonded van der Waals interactions (Chap. 6) and nonbonded electrostatic interactions (Chap. 7). Chaps. 8 and 9 deal with geometrical boundary conditions, which may be either imposed on the system in a soft way, using restraints (Chap. 8) or enforced in a hard way, defined as constraints (Chap. 9). The next three chapters involve searching and sampling of conformational space using energy minimization (Chap. 10), molecular dynamics (Chap. 11), or stochastic dynamics (Chap. 12). Chap. 13 deals with the calculation of free energies and Chap. 14 with path-integral simulations. QM/MM simulations are described in Chap. 15. In Chap. 16 the implementation of replica exchange methods in GROMOS is discussed. In Chap. 17 all contributions to the molecular forces and the virial due to the interactions described in Chaps. 5 - 7 are collected and Chap. 18 forms an appendix with mathematical details on the conversion between different types of force constants.

CHAPTER 2

Basic choices in the definition of a molecular model

2.1. Introduction

The four *basic choices*[1,2] involved in the definition of a molecular model are (Fig. 2.1) :

A. Choice of the *degrees of freedom*[3] (Sec. 2.2) :
What are the variables of the model, *i.e.* those uniquely defining a system configuration in this model. Most commonly, this choice is equivalent to that of the elementary particles, *i.e.* smallest entities, considered by the model, *i.e.* the model resolution. The selected degrees of freedom are treated explicitly by the model. In contrast, degrees of freedom internal to the selected elementary particles are handled implicitly.

B. Choice of the description of the *interaction*[3] (Sec. 2.3) :
What is the Hamiltonian operator for quantum-mechanical degrees of freedom or the Hamiltonian function for classical degrees of freedom describing the interaction and kinetic energy associated with the selected degrees of freedom. This interaction must incorporate the mean effect of the implicit degrees of freedom on the explicit ones.

C. Choice of method for the *configuration generation*[4,5] (Sec. 2.4) :
What method will be used to generate configurations of the system *i.e.* values for the variables defining the degrees of freedom of the model, generally based on information from the corresponding interaction. The selection of a method, e.g., MC, MD or SD, affects the properties of the resulting set of configurations in terms of structural, thermodynamic and dynamic properties.

D. Choice of the *boundary conditions*[6] (Sec. 2.5) :
What are the global restrictions imposed on the system during the configuration generation. These restrictions may represent the interface of the internal degrees of freedom of the system to the outside world and are typically of a spatial/geometric, thermodynamic or alchemical nature.

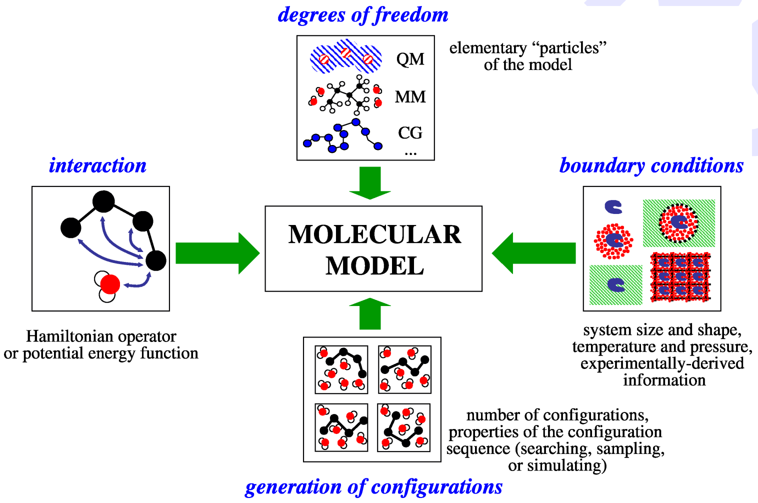

FIGURE 2.1. Four basic choices involved in the definition of a molecular model

Two important considerations typically precede the definition of a molecular model. First, most molecular models result in equations that are too complex or too numerous for an analytical solution. They must thus be solved numerically on a computer. Because the computing power is limited, one is always compelled to strike a balance between the required accuracy of the process or property of interest, the quality of the models and the size of the relevant, accessible configurational space of the system that is to be probed in a tractable amount of computer time.

2.2. Choice of degrees of freedom

The possible physical choices of degrees of freedom in the context of the modelling of molecular systems, from the highest to the lowest resolution, can be broadly classified as**3** (Fig. 2.2) :

A. Nuclei and electrons of atoms
B. Nuclei and valence electrons of atoms
C. Nuclei and (valence) electrons of solute atoms
D. Atoms
E. United-atoms
F. Solute atoms
G. Beads representing atom groups
H. Rigid bodies representing molecules
I. Local properties of finite volume elements

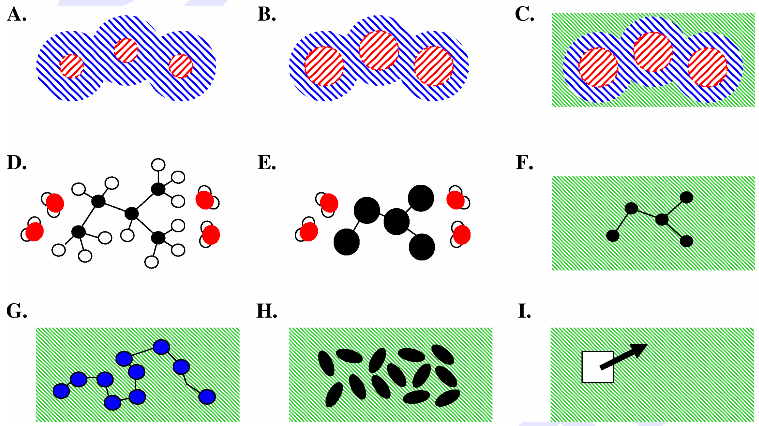

FIGURE 2.2. Broad classification of the possible levels of resolution or choice of the degrees of freedom in the modelling of molecular systems.

Each lower level of resolution turns a set of degrees of freedom that were handled explicitly at the previous level into implicit degrees of freedom at the next level. Models at the levels of resolution of Pts. A-C are commonly referred to as *Quantum Mechanical* (QM) models. Models at the levels of resolution of Pts. D-F are commonly referred to as *Atomistic* (AA for All Atoms) models. Models at the level of resolution of Pt. G are commonly referred to as *Coarse Grained* (CG) models. A number of *hybrid methods* attempt to merge different levels of resolution for different parts of a molecular system.

The selected level of resolution determines what should be referred to as the *configuration* or microstate of a corresponding system of $\mathcal{N}_d$ degrees of freedom. At the levels of resolution of Pts. A-C, a *quantum-mechanical* description is the most appropriate. In this case, a configuration of a system of particles is defined in the position representation by a wavefunction $\Psi \doteq \Psi(\boldsymbol{r})$, where $\boldsymbol{r}$ is the $\mathcal{N}_d$-dimensional vector characterizing the $\mathcal{N}_d = 3\mathcal{N}_a$ positional degrees of freedom associated with the $\mathcal{N}_a$ quantum-mechanical particles considered. At the levels of resolution of Pts. D-G, a *classical* description is the most appropriate. In this case, a configuration of a system of particles is defined by two $\mathcal{N}_d$-dimensional vectors $\boldsymbol{q}$ and $\boldsymbol{p_q}$, containing $\mathcal{N}_d = 3\mathcal{N}_a$ generalized coordinates and momenta associated with the $\mathcal{N}_a$ classical particles considered, respectively. In the special case where a Cartesian coordinate system is adopted, $\boldsymbol{q}$ is written as $\boldsymbol{r}$ and contains the components of the Cartesian particle coordinates as triplets, while $\boldsymbol{p_q}$ is written $\boldsymbol{p}$ and contains the components of the Cartesian particle momenta as triplets. This situation may be noted

$$\boldsymbol{r}^N \doteq \{\mathbf{r}_i \mid i = 1..\mathcal{N}_a\} \quad \text{and} \quad \boldsymbol{p}^N \doteq \{\mathbf{p}_i \mid i = 1..\mathcal{N}_a\} \;, \tag{2.1}$$

where $\mathbf{r}_i$ and $\mathbf{p}_i$ are the three-dimensional Cartesian coordinate and momentum vectors of a particle i. For simplicity and unless otherwise specified, a Cartesian coordinate description will be used systematically throughout this manual. At the level of resolution of Pt. H, there are twelve degrees of freedom per rigid molecule (position, orientation, translational velocity, rotational velocity). Finally, at the level of resolution of Pt. I, a *continuous-material* description is adopted and the number of degrees of freedom per volume element depends on the (scalar or vector) properties that are considered.

In many cases, the choice of the degrees of freedom included in a model is not physical, *i.e.* matching exactly the real degrees of freedom of a molecular system. Important examples include :

J. The *Car Parrinello* (CP) approach[7] :
Here the electronic degrees of freedom in the form of the coefficients of the electronic wavefunction are treated as classical degrees of freedom and evolved in time simultaneously with the classically treated nuclei.

K. The *Charge On Spring* (COS) approach[8] :
Here, the instantaneous dipoles associated with electronic polarization effects are accounted for by means of artificial point charges connected to the atoms by means of harmonic springs.

L. The *Path Integral* (PI) approach[9] :
Here, a system of particles treated at the classical level is used to emulate the statistical-mechanical properties of a corresponding quantum-mechanical system a single quantum-mechanical particle being assigned to a given number of the classical particles, connected by harmonic springs in a ring topology.

M. Various *extended-system* methods :
Here, one or a few additional degrees of freedom are added to those of the physical system and evolved simultaneously, *e.g.* thermostat or barostat variables,[6,10] force-field parameters to be refined,[11] atomic charges to account for polarization effects[8] as in the fluctuating charge model, Hamiltonian coupling parameters [cf λ-dynamics].

N. Ensemble propagation approaches[4,12] :
Here the degrees of freedom do not correspond to those of a single molecular system, but to the coefficients of an ensemble probability distribution of such systems (as projected in a given basis set).[12]

The choice of the degrees of freedom or particles that are handled explicitly in the model automatically implies a definition of the implicit degrees of freedom, the internal degrees of freedom of the selected particles that are not considered explicitly in the model.

The possibilities in the domain of application of the GROMOS package in terms of degrees of freedom are discussed in Sec. 3.2.

2.3. Choice of the description of the interaction

When the degrees of freedom of the model are to be treated according to the laws of quantum mechanics, the interaction associated with the selected degrees of freedom is described by a *Hamiltonian operator* $\hat{\mathcal{H}}$. Assuming that this operator is constant *i.e.* not explicitly dependent on any external parameter or on time, and that all elementary, subatomic, particles are explicitly included in the model, the expectation value $\langle\hat{\mathcal{H}}\rangle$ of the Hamiltonian operator in terms of the wavefunction $\Psi \doteq \Psi(\boldsymbol{r})$ representing a system configuration represents the total, kinetic plus potential, energy of the system in this configuration, *i.e.*

$$\langle\hat{\mathcal{H}}\rangle[\Psi] \doteq \langle\Psi \mid \hat{\mathcal{H}} \mid \Psi\rangle \ , \tag{2.2}$$

where the square brackets indicate a functional dependence and it is assumed that Ψ is normalized. The Hamiltonian operator is itself the sum of a *kinetic energy operator* $\hat{\mathcal{K}}$ and a *potential energy operator* $\hat{\mathcal{V}}$, *i.e.*

$$\hat{\mathcal{H}} \doteq \hat{\mathcal{K}} + \hat{\mathcal{V}} \tag{2.3}$$

with

$$\hat{\mathcal{K}} \doteq -\sum_{i=1}^{\mathcal{N}_a} \frac{\hbar^2}{2m_i}\nabla_i^2 \ , \tag{2.4}$$

where m_i is the mass of particle i and $\hbar \doteq (2\pi)^{-1}h$, h being Planck's constant. The potential energy operator accounts for the proper interaction between the different quantum-mechanical particles. When the quantum-mechanical degrees of freedom included in the model do not encompass all particles in the system, the interaction may still tentatively be formulated using an effective Hamiltonian operator that encompasses the mean effect of the implicit degrees of freedom on the explicit ones, e.g. as in the use of pseudo-potentials accounting for the implicit core electrons or of effective solvation terms for the implicit solvent in the potential energy operator.

When the degrees of freedom of the model are to be treated according to the laws of classical mechanics, the interaction associated with the selected degrees of freedom is described by a *Hamiltonian function* $\mathcal{H}$. Assuming that this function is constant, i.e. not explicitly dependent on any external parameter or on time, that all elementary particles or atoms are explicitly included in the model, and using a Cartesian coordinate system, the Hamiltonian function can be written $\mathcal{H}(\boldsymbol{r}^N, \boldsymbol{p}^N)$, and its value is equal to the total, kinetic plus potential, energy of the system in the configuration $(\boldsymbol{r}^N, \boldsymbol{p}^N)$. The Hamiltonian function is the sum of a *kinetic energy function* $\mathcal{K}$ that solely depends on the Cartesian momenta of the particles and a *potential energy function* $\mathcal{V}$ that solely depends on the Cartesian coordinates of the particles, *i.e.*

$$\mathcal{H}(\boldsymbol{r}^N, \boldsymbol{p}^N) \doteq \mathcal{K}(\boldsymbol{p}^N) + \mathcal{V}(\boldsymbol{r}^N) \tag{2.5}$$

with

$$\mathcal{K}(\boldsymbol{p}^N) \doteq \sum_{i=1}^{\mathcal{N}_a} \frac{1}{2m_i} \mathbf{p}_i^2 \ . \tag{2.6}$$

When a generalized coordinate system is employed, the kinetic energy is in the general case a function of both coordinates and momenta. The potential energy function accounts for the proper interaction between the different classical particles. When the classical degrees of freedom included in the model do not encompass all particles in the system, the interaction may still tentatively be formulated using an effective Hamiltonian function that encompasses the mean effect of the implicit degrees of freedom on the explicit ones. According to classical statistical mechanics, the ensemble properties of the reduced system will be identical to those of a corresponding all-particle system if the potential energy function involved in the effective Hamiltonian is defined as a *potential of mean force*, *i.e.* a potential, the derivative of which is equal to the force on the explicit degrees of freedom ensemble-averaged over the compatible values of the implicit degrees of freedom. In this situation, the notation $\overline{\mathcal{V}}$ (potential of mean force) instead of $\mathcal{V}$ (all-particle potential) will sometimes be used when the distinction is important.

Finally, when the degrees of freedom of the model are to be treated according to a continuous-material description, the interaction is typically described in terms of *gradients* of specific intensive properties and *conservation equations* for specific extensive properties.

When modelling aims at emulating the physical behaviour of a molecular system as closely as possible, the approximations employed to define the interaction should be as representative as possible for the real physical situation. In some cases, however, alterations are performed, typically in the form of extra unphysical and possibly time-dependent terms included into the potential energy function to serve specific purposes, *e.g.* enhanced probing of the configurational space, inclusion of experimental data as a boundary condition. The possibilities in the domain of application of the GROMOS package in terms of interactions are discussed in Sec. 3.3.

2.4. Choice of method for configuration generation

The possible choices of methods to generate configurations of a molecular system can be broadly classified as:**[5]**

A. *Search methods*:
   Generation of a set of relevant configurations within the configurational space of the system, typically configurations representing local energy minima, without any further requirement on the distribution or time sequence of these configurations.

B. *Sampling methods*:
   Generation of a set of system configurations obeying a well-defined probability distribution in terms of energy, a configurational ensemble.

C. *Simulation methods*:
   Generation of a time sequence of system configurations obeying a particular probability distribution both governed by the chosen equation of motion.

Both sampling and simulation methods permit the evaluation of *thermodynamic properties* from the generated configurational ensemble. Only simulation methods permit the evaluation of *dynamical properties* from the generated trajectory.

At the quantum-mechanical level assuming that the Hamiltonian operator is constant, the simulation of a molecular system involves the integration of the *time-dependent Schrödinger equation*

$$\hat{\mathcal{H}}\Psi(\boldsymbol{r},t) = i\hbar\frac{\partial}{\partial t}\Psi(\boldsymbol{r},t) \ , \tag{2.7}$$

where $\Psi(\boldsymbol{r},t)$ represents the instantaneous state of the system at time t. Integrating this equation numerically forward in time on a computer to simulate quantum-dynamical behaviour of a molecular system is called *quantum molecular dynamics* (QMD) simulation. Two simplifying approximations are often made in QMD : (*i*) A molecule is built up from just two types of particles, nuclei with negligible size and irrelevant internal structure and electrons, the motion of which can be separated using the Born-Oppenheimer approximation. Thus, a complex molecular system can be described as a system of point masses moving in an effective potential field; (*ii*) Some particles, e.g. the nuclei, obey the laws of classical mechanics. This is a reasonable assumption at room temperature and for all but the lightest atoms. For the latter the path-integral formalism can be used to obtain the quantum equilibrium distribution using classical equations of motion.

At the classical level, using a Cartesian coordinate system, the simulation of a molecular system involves the integration of the *Newtonian equations of motion*

$$\dot{\boldsymbol{p}}^N(t) = -\nabla\mathcal{V}(\boldsymbol{r}^N(t)) \quad \text{and} \quad \dot{\boldsymbol{r}}^N(t) = \underline{\mathbf{m}}^{-1}\boldsymbol{p}^N(t) \tag{2.8}$$

where $\underline{\mathbf{m}}$ is a $\mathcal{N}_d$-dimentional (diagonal) mass matrix of a system of $\mathcal{N}_a$ particles containing the mass of the atoms by triplets along its diagonal and $(\boldsymbol{r}^N(t), \boldsymbol{p}^N(t))$ represents the instantaneous configuration of the system at time t. These equations represent a particular case of the more general Hamiltonian equations of motion in the special case of a Cartesian coordinate system. Two relevant quantities entering in Eq. 2.8 are the $\mathcal{N}_d$-dimensional force and velocity vectors, *i.e.*

$$\boldsymbol{v}^N(t) \doteq \underline{\mathbf{m}}^{-1}\boldsymbol{p}^N(t) = \dot{\boldsymbol{r}}^N(t) \tag{2.9}$$

and

$$\boldsymbol{f}^N(t) \doteq -\nabla\mathcal{V}(\boldsymbol{r}^N(t)) = \dot{\boldsymbol{p}}^N(t) \ , \tag{2.10}$$

with the notation

$$\boldsymbol{v}^N \doteq \{\mathbf{v}_i \mid i = 1..\mathcal{N}_a\} \quad \text{and} \quad \boldsymbol{f}^N \doteq \{\mathbf{f}_i \mid i = 1..\mathcal{N}_a\} \ . \tag{2.11}$$

where $\mathbf{v}_i$ and $\mathbf{f}_i$ are the three-dimensional Cartesian velocity and force vectors of a particle i. In terms of these quantities, the first equation in Eq. 2.8 may equivalently be written

$$\underline{\mathbf{m}}\dot{\boldsymbol{v}}^N(t) = \boldsymbol{f}^N(t) \ . \tag{2.12}$$

Integrating Eq. 2.8 numerically forward in time on a computer to simulate classical-dynamical behaviour of a molecular system is called classical *molecular dynamics* (MD) simulation.

An alternative to *Newtonian equations of motion* are the *Langevin equations of motion*

$$\underline{\mathbf{m}}\dot{\boldsymbol{v}}^N(t) = \overline{\boldsymbol{f}}^N(t) + \boldsymbol{f}^{st,N}(t) - \underline{\mathbf{m}}\underline{\gamma}\boldsymbol{v}^N(t) \ , \tag{2.13}$$

where

$$\overline{\boldsymbol{f}}(t) \doteq -\nabla\overline{\mathcal{V}}(\boldsymbol{r}(t)) \ . \tag{2.14}$$

The mean force $\overline{\boldsymbol{f}}$ is the negative gradient of the potential of mean force $\overline{\mathcal{V}}$, *i.e.* it includes the forces between the explicit particles but also accounts in an effective fashion for the average forces exerted by the implicit particles.

The *stochastic force* is denoted by $\boldsymbol{f}^{st}(t)$ and the *frictional force* is proportional to the velocities $\mathbf{v}_i$ with the proportionality factor $\underline{\mathbf{m}}\underline{\gamma}$, in which $\underline{\gamma}$ is a diagonal matrix containing the atomic friction coefficients γ_i. A stochastic force introduces energy into the system, and a frictional force removes energy from it. The condition of zero energy loss on average will relate the two forces. If the stochastic force ${\boldsymbol{f}^{st}}_i$ obeys a Gaussian probability distribution with zero mean, if it is not correlated with prior velocities or forces, and if the friction coefficient γ_i is independent of time, this condition reads

$$< ({\boldsymbol{f}^{st}}_i)^2 > \quad = \quad 6 m_i \gamma_i \, k_B \, T_{ref} \tag{2.15}$$

where a time average is denoted by <...>, k_B is Boltzmann's constant, and T_{ref} is the reference temperature of the system. Numerical integration of the stochastic equations of motion is called *stochastic dynamics (*SD*) simulation*. Different approximations to this purpose are *Stochastic Dynamics* (SD), *high viscosity* SD *or Brownian Dynamics* (BD) or *Diffusive Particle Dynamics* (DPD).

In a *path-integral simulation* (PI) classical MD simulation using a modified interaction function $\mathcal{V}(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N)$ is used to generate an ensemble of configurations that corresponds to the quantum-mechanical partition function for the molecular system.

Finally, when the degrees of freedom of the model are to be treated according to a continuous-material description, the equations of motion are typically formulated in terms of *transport equations* that relate the flow of specific extensive properties to the gradients of specific intensive properties.

A molecular system can be coupled to external quantities in different ways (see also Sec. 2.5):

1. Inclusion of an extra term $\mathcal{V}^{(res)}(\boldsymbol{r})$, a *penalty function*, in the interaction function of the system that restrains the motion of the particles such that the simulated value approaches the prescribed value of the given external quantity.
2. The prescribed value of the external quantity can be imposed as a *constraint* on to the system, such that it is satisfied by every configuration.
3. A first-order equation can be added to the particle equations of motion that drives the simulated value of the external quantity towards the prescribed value: the so-called *weak-coupling method*.

The latter two methods imply modification of the classical equations of motion of the molecular system.

We note that the total energy of the system will only be conserved in an MD simulation if the potential energy $\mathcal{V}$ is only a function of the current configuration. When applying weak coupling or a penalty function with time-averaged parameter values, this condition is not fulfilled.

The possibilities in the domain of application of the GROMOS package in terms of configuration generation are discussed in Sec. 3.4.

When performing *energy minimization*, the potential energy $\mathcal{V}$ molecular system is minimized using the negative gradient of this function.

2.5. Choice of the boundary conditions

The term *boundary condition* refers to global restrictions imposed on the system during the configuration generation. A restriction may be *hard*, a constraint affecting all individual configurations, or *soft*, a constraint on the average value of an observable over the generated configurations or a restraint biasing this average value towards a specified target value. One may distinguish the following types of boundary conditions:

A. *Spatial* boundary conditions
These concern the nature of the confinement or periodicity applied to the molecular system. Typical examples are: *Vacuum Boundary Conditions* (VBC), a set of molecules in vacuum, *Fixed Boundary Conditions* (FBC), a system confined to a fixed region of space by means of a wall, and *Periodic Boundary Conditions* (PBC), a periodic system defined by the replication of a computational box of space-filling shape.

B. *Geometric* boundary conditions
These concern the possible presence of constraints or restraints on specific internal coordinates within the system. Typical examples are the use of rigid bond lengths, the use of experiment-based restraints affecting specific internal coordinates or averages thereof, the use of artificial constraints or restraints on a specific internal coordinate, *e.g.* to avoid structural distortions during the equilibration phase of a simulation, or to bias the generation of configurations towards specific regions of the configurational space.

C. *Thermodynamic* boundary conditions
These concern the thermodynamic state point associated with the generated configurational ensemble. Typical examples are a constant *number of particles vs* a constant *chemical potential*, a constant *volume vs* a constant *pressure*, a constant *energy* or enthalpy *vs* a constant *temperature*. In specific cases, other variables may be required for the definition of the thermodynamic state point.

The possibilities in the domain of application of the GROMOS package in terms of boundary conditions are discussed in Sec. 3.5.

CHAPTER 3

Scope of the GROMOS package

3.1. Introduction

The GROMOS package for molecular simulation currently consists of three subpackages that are tightly linked together but can largely be compiled and used as separate entities.

a. GROMOS++ is a library of supporting programs for pre- and post-MD tasks. It mainly consists of programs to facilitate the setup of simulations and of programs to analyze trajectories that are the results of such simulations.
b. PROMD is a program which can perform energy minimizations, molecular dynamics simulations or stochastic dynamics. It is written in FORTRAN.
c. MD++ is an alternative energy minimizer and simulator, written in *C++* in an object oriented manner. Although intrinsically both MD-engines are based on the same algorithms and offer similar possibilities, they differ in the implementation of selected features.

This chapter discusses the options offered by all three subpackages regarding the four basic aspects of molecular simulation.

3.2. Choice of the degrees of freedom

GROMOS can only consider degrees of freedom behaving according to the laws of classical mechanics. These are usually atoms or united atoms with two exceptions : (*i*) beads assumed to behave classically in a coarse-grained representation of molecular systems; (*ii*) beads connected by harmonic springs in a ring topology as a path-integral representation of a quantum-mechanical system. GROMOS is primarily intended for simulations of condensed-phase systems, solutions, pure liquids and crystals, with an explicit representation of the solvent molecules in the case of solutions. However, the simulation of systems in the gas phase and of solutions with an implicit-solvent representation are also possible. Since GROMOS was originally developed for atomic degrees of freedom, particles are generally named atoms, although they actually may be groups of atoms or path-integral beads.

3.3. Choice of the description of the interaction

In classical simulations, the Hamiltonian of a molecular system has the form

$$\mathcal{H}(\boldsymbol{p}^N, \boldsymbol{r}^N) \quad = \quad \mathcal{K}(\boldsymbol{p}^N) + \mathcal{V}(\boldsymbol{r}^N) \tag{3.1}$$

The first term is the *kinetic energy* term

$$\mathcal{K}(\boldsymbol{p}^N) = \sum_{i=1}^{N} \frac{{\mathbf{p}_i}^2}{2m_i} = \sum_{i=1}^{N} \tfrac{1}{2} m_i {\mathbf{v}_i}^2 \ , \tag{3.2}$$

which is independent of the particle coordinates $\boldsymbol{r}^N$ in the absence of geometric positional constraints. If constraints are imposed, the components of the momenta $\mathbf{p}_i$ or the velocities $\mathbf{v}_i$ along the constrained degrees of freedom must be zero. The second term is the *potential energy* term or interaction function, which describes the interaction energy in terms of particle coordinates $\mathbf{r}$

$$\mathcal{V}(\boldsymbol{r}; \mathbf{s}) \quad \equiv \quad \mathcal{V}(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N; \mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_M). \tag{3.3}$$

Here $\boldsymbol{r}^N$ denotes the $3\mathcal{N}_a$-dimensional Cartesian coordinate vector of the system (periodic copies of the atoms within the reference box if PBC is applied). Generally, such a potential energy function $\mathcal{V}(\boldsymbol{r}^N; \mathbf{s})$ depends on a number (M) of parameters, also called force-field parameters, here indicated by $\mathbf{s} \equiv (\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_M)$.

In practice, the interaction function $\mathcal{V}(\boldsymbol{r}^N)$ consists of a sum of terms which represent different types of interactions. We distinguish two types of interactions or forces:

1. The standard *physical* atomic *interaction* function $\mathcal{V}^{(phy)}(\boldsymbol{r}^N, \mathbf{s})$, e.g. the GROMOS force field.
2. The *non-physical* atomic *interaction* function terms $\mathcal{V}^{(spe)}(\boldsymbol{r}^N)$, which are included to serve a special purpose, e.g. atomic position restraining, distance restraining, dihedral-angle restraining, etc.

So, we have

$$\mathcal{V}(\boldsymbol{r};\mathbf{s}) \;=\; \mathcal{V}^{(phy)}(\boldsymbol{r}^N;\mathbf{s}) + \mathcal{V}^{(spe)}(\boldsymbol{r}^N) \tag{3.4}$$

The physical potential energy term $\mathcal{V}^{(phy)}$ is further divided into a term $\mathcal{V}^{(cov)}$ corresponding to *covalent interactions* and a term $\mathcal{V}^{(nbd)}$ corresponding to *non-bonded interactions*. When a path-integral calculation is performed, an additional term $\mathcal{V}^{(pir)}$ is introduced to account for *path-integral ring interactions*, *i.e.* the harmonic interaction between the successive beads in the rings representing each atom. This results in

$$\mathcal{V}^{(phy)}(\boldsymbol{r}^N;\mathbf{s}) \doteq \mathcal{V}^{(cov)}(\boldsymbol{r}^N;\mathbf{s}) + \mathcal{V}^{(nbd)}(\boldsymbol{r}^N;\mathbf{s}) + \mathcal{V}^{(pir)}(\boldsymbol{r}^N;\mathbf{s}) \ . \tag{3.5}$$

The covalent term is further partitioned as a sum of contributions from *bond stretching*, *bond-angle bending*, *improper dihedral-angle bending* and *proper dihedral-angle torsion* interactions, namely

$$\mathcal{V}^{(cov)}(\boldsymbol{r}^N;\mathbf{s}) \doteq \mathcal{V}^{(b)}(\boldsymbol{r}^N;\mathbf{s}) + \mathcal{V}^{(\theta)}(\boldsymbol{r}^N;\mathbf{s}) + \mathcal{V}^{(\xi)}(\boldsymbol{r}^N;\mathbf{s}) + \mathcal{V}^{(\varphi)}(\boldsymbol{r}^N;\mathbf{s}) \ . \tag{3.6}$$

The non-bonded term is further partitioned as a sum of contributions from *van der Waals* and *electrostatic* interactions, namely

$$\mathcal{V}^{(nbd)}(\boldsymbol{r}^N;\mathbf{s}) \doteq \mathcal{V}^{(vdw)}(\boldsymbol{r}^N;\mathbf{s}) + \mathcal{V}^{(ele)}(\boldsymbol{r}^N;\mathbf{s}) \ . \tag{3.7}$$

The different terms involved in Eqs. 3.5-3.7 are described in Chaps. 5, 6 and 7. The force $\mathbf{f}_i$ on particle i due to a particular interaction term is given by the relation

$$\mathbf{f}_i \;=\; -\frac{\partial}{\partial \mathbf{r}_i}\,\mathcal{V}(\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_N), \tag{3.8}$$

which can also be used to obtain the interaction energy term (apart from a constant) that corresponds to a given force $\mathbf{f}_i$. We note that a MD trajectory only depends on the forces on the atoms, not on the energies.

The derivatives of the potential energy terms with respect to atomic coordinates, box or coupling parameters are provided in Chap. 17, while the parameters involved in these terms are included as part of the GROMOS force-field description in Vol. 3.

3.3.1. Charge groups, searching neighbours. The bulk of the computer time required by a simulation time step is used for calculating the non-bonded interactions, that is, for finding the nearest neighbour atoms and subsequently evaluating the van der Waals and electrostatic interaction terms for the obtained atom pairs. Therefore, various schemes for performing this task as efficiently as possible are available.[13]

Since the non-bonded interaction between atoms decreases with the distance between them, only interactions between atoms closer to each other than a certain cut-off distance R_c are generally taken into account in simulations.

When the (partial) atomic charges of a group of atoms add up to exactly zero, the leading term of the electrostatic interaction between two such groups of atoms is of dipolar ($1/r^3$) character. The sum of the 1/r monopole contributions of the various atom pairs to the group-group interaction will be zero. Therefore, the range of the electrostatic interaction can be considerably reduced when atoms are assembled in so-called charge groups, which have a zero net charge, and for which the electrostatic interaction with other (groups of) atoms is either calculated for all atoms of the charge group or for none.

The GROMOS force fields make use of this *concept of charge groups*. The atoms that belong to a charge group are chosen such that their partial atomic charges add up to zero. For groups of atoms with a total charge of $+e$ or $-e$, like the sidechain atoms of Arg or Asp, the partial atomic charges of the charge group may add up to $+e$ or $-e$. When a cut-off radius is used, one can choose to base this cutoff on the atomic positions (AT) or to use a charge group based cutoff. The *position of a charge group* is defined differently for a charge group belonging to the "solute" part of the molecular topology and one in the "solvent" part of the molecular topology.

- The position of a "solute" charge group is taken to be its centre of geometry:

$$R_{cg} = \sum_{i=1}^{N_{cg}} \mathbf{r}_i / N_{cg} \tag{3.9}$$

where the number of atoms belonging to the charge group is denoted by N_{cg}.
- The position of a "solvent" charge group is taken to be the position of the first atom of a solvent molecule. A "solvent" molecule may only contain one charge group.

Therefore, in the GROMOS non-bonded interaction subroutines the cut-off radius R_c, denoted by RCUTP, is used to select nearest-neighbour charge groups.

The simplest way to find the neighbouring charge groups of a charge group, that is, the charge groups that lie within R_c, is to scan all possible charge group pairs in the system. For a system consisting of N_{CG} charge groups, the number of pairs amounts to $1/2{N_{CG}}^2$, which makes the computer time required for finding the neighbours in this way proportional to ${N_{CG}}^2$. Once the neighbours have been found, the time required for calculating the non-bonded interaction is proportional to N_{CG}. We note that non-bonded interactions within a charge group may need to be calculated, when the charge group contains many atoms.

3.3.2. Twin-range method for long-range interactions. In order to evaluate the non-bonded interaction (Eq. 3.7) with sufficient accuracy, a long cut-off radius R_{cl} has to be used; for molecular systems a value of at least 1.4 nm seems necessary. But such a range is very expensive if pair interactions are evaluated; the number of neighbour atoms within 1.4 nm will exceed 300. Therefore, in GROMOS the non-bonded interaction can be evaluated using a *twin-range method*.[14] Secondly, the electrostatic interactions beyond the long-range cutoff R_{cl} can be approximated by a Poisson-Boltzmann generalized *reaction field* term.

The non-bonded interactions in Eq. 3.7 are evaluated at every simulation step using the charge group pair list that is generated with a short range cut-off radius R_{cp} (=RCUTP). The longer range non-bonded interactions, that is, those between charge groups at a distance longer than R_{cp} and smaller than R_{cl} (=RCUTL), are evaluated less frequently, viz. only at every n-th (=NSNB) simulation step when also the pair list is updated. They are kept unchanged between these updates. In this way the long-range non-bonded forces can be approximately taken into account, without increasing the computing effort significantly, at the expense of neglecting the fluctuation of the forces beyond R_{cp} during n simulation steps.

The long-range interaction, which is calculated for charge group pairs at distances between R_{cp} and R_{cl}, is evaluated by using the electrostatic term in Eq. 3.7 and the normal van der Waals parameters in the Lennard-Jones term. It is assumed that no excluded neighbours, no third-neighbour or 1,4-interactions and no intra-charge-group interactions exist at these distances. So R_{cp} must not be chosen too small.

The evaluation of the nonbonded interactions in GROMOS relies on the application of the twin-range method.[15–17,1] The GROMOS implementation (both PROMD and MD++) of this approach includes an increased amount of flexibility, and relies on the definition of: (1) a short-range pairlist distance R_{cp}; (2) a corresponding cutoff distance $\tilde{R}_{cp} \leq R_{cp}$ (optional); (3) a lower bound for the intermediate-range pairlist distance R_s; (4) a corresponding cutoff distance $\tilde{R}_{cs} \leq R_s$ (optional); (5) an upper bound for the intermediate-range pairlist distance R_{cl}; (6) a corresponding cutoff distance $\tilde{R}_{cl} \leq R_{cl}$ (optional); (7) a short-range pairlist update frequency N_s; (8) an intermediate-range pairlist update frequency N_l. Short-range interactions are computed every time step based on a short-range pairlist containing pairs in the distance range $[0; R_{cp}]$, or a filtered subset of this list corresponding to pairs currently (i.e. at the given timestep) in the distance range $[0; \tilde{R}_{cp}]$.

The short-range pairlist is reevaluated every N_s time steps. It can be generated either on the basis of distances between charge groups (groups of covalently linked atoms defined in the system topology) or of distances between individual atoms. In the former case, the filtering (based on the distance $\tilde{R}_{cp}$) may be based either on distances between charge groups or on distances between atoms. In the latter case, only atom-based filtering is possible. Intermediate range interactions are computed every N_l time steps based on all pairs in the distance range $[R_s; R_{cl}]$, or a filtered subset of these pairs in the distance range $[\tilde{R}_{cs}; \tilde{R}_{cl}]$ at the time of the evaluation of these interactions. Only an atom-based filtering is possible here, and it is only meaningful when the initial set of pairs is generated on the basis of distances between charge groups. The energy, forces, and virial contributions associated with intermediate-range interactions are assumed constant between two updates (i.e., during N_l steps).

The evaluated interaction includes Lennard-Jones and electrostatic components. The latter component may include a reaction-field contribution or the real-space contribution to a lattice-sum method. Note that the real-space contribution to a lattice-sum method may only be computed within the short-range contribution to the interaction.

The pairlist construction may be performed in four different ways: (1) using the standard double-loop algorithm included in the GROMOS96 program (merely extended to include the possibility of an atom-based cutoff and of filtering); (2) using an optimized version that improves processor cache usage; (3) using a grid-based pairlist algorithm[18] (PROMD only); (4) using a slight variation of the above grid-based algorithm, which permits easier parallelization and avoids periodicity corrections during the interaction evaluation (MD++ only).

3.3.3. Grid-based pairlist construction. A grid-based pairlist algorithm is implemented in PROMD to allow for a fast construction of cutoff-based nonbonded pairlists in molecular simulations under periodic boundary conditions based on an arbitrary box shape (rectangular, truncated-octahedral, or triclinic).[18] The key features of this algorithm are: (1) the use of a one-dimensional mask array (to determine which grid cells contain interacting atoms) that incorporates the effect of periodicity, and (2) the grouping of adjacent interacting cells of the mask array into stripes, which permits the handling of empty cells with a very low computational overhead. Testing of the algorithm on water systems of different sizes (containing about 2000 to 11,000 molecules) has shown that the method: (1) is about an order of magnitude more efficient compared to a standard (double-loop) algorithm, (2) achieves quasi-linear scaling in the number of atoms, (3) is weakly sensitive in terms of efficiency to the chosen number of grid cells.

MD++ includes a slightly modified version of this grid-based pairlist algorithm following ideas similar to those of a published pairlist algorithm.[19] In an effort to reduce the number of nearest image determinations during the pairlist generation (or filtering) and the nonbonded force calculation, the system gets extended on all sides before the pairlist construction. The additional atom or charge-group positions are obtained by simple shifts of the original positions by the box vectors. To allow for more efficient (distributed- memory) parallelization and to save time, the central computational box is divided into N layers. Each of the P parallel processes only has to extend over N/P layers. After every extension, the atom pairs consisting of one atom within the layer and a second atom from one of the above (not extended) layers are added to the respective pairlist (using a one-dimensional mask array). Filtering or energy and force evaluation can then be carried out right away (without nearest image determinations owing to the preshifted atomic positions), or at a later stage with the information on the shift vectors encoded into the pairlist, thus enabling a fast reconstruction of the shifted positions.

3.4. Choice of the method for the configuration generation

The GROMOS programs PROMD and MD may be used to perform energy minimizations, molecular dynamics simulations or stochastic dynamics simulations. Details of the algorithms that are implemented and considerations to be kept in mind when setting up the calculations are given in Chaps. 10, 11 and 12 respectively.

3.5. Choice of the boundary conditions

GROMOS offers a wide range of algorithms to apply boundary conditions to an energy minimization, molecular or stochastic dynamics simulation. Spatial boundary conditions are limited to vacuum boundary conditions, fixed boundary conditions or periodic boundary conditions in various shapes. Details for spatial boundary conditions are described in Chap. 4. Geometric boundary conditions may be applied in the form of restraints through the use of special force-field terms (Chap. 8) or as constraints through a direct adaptation of the equations of motion (Chap. 9). GROMOS currently cannot perform simulations in a grand-canonical ensemble, *i.e.* with constant chemical potential rather than constant number of particles.