Force fields, thermo- and barostats

Berk Hess
What is a force field?

A force field usually consists of three parts:

- a set of functional forms
- parameters for the functional forms that, usually, depend on the atom types
- a set of building blocks, e.g. for amino acid residues
Often one can split a force field into two main contributions:

Bonded interactions work only within molecules:

\[ V_b(r_i, \ldots, r_N) = \sum_{m \in \text{molecules}} \left( \sum_{i,j \in \text{bonds/pairs}(m)} V_{ij}(r_i, r_j) + \sum_{i,j,k \in \text{angles}(m)} V_{ijk}(r_i, r_j, r_k) + \sum_{i,j,k,l \in \text{dihedrals}(m)} V_{ijkl}(r_i, r_j, r_k, r_l) \right) \]

Non-bonded interactions work between and within molecules:

\[ V_{nb}(r_i, \ldots, r_N) = \sum_{i < j} V_{ij}(\| r_i - r_j \|) = \sum_{i < j} V_{ij}(r_{ij}) \]
Non-bonded interactions

Non-bonded interactions in GROMACS are pair-additive and centro-symmetric:

Electrostatic interactions
- Coulomb \( V_{ij} = \frac{q_i q_j}{4\pi \varepsilon_0 \varepsilon} \frac{1}{r_{ij}} \)
- tabulated \( V_{ij} = \frac{q_i q_j}{4\pi \varepsilon_0 \varepsilon} f_C(r_{ij}) \)

Van der Waals interactions
- Lennard Jones \( V_{ij} = C_{ij}^{(12)} \frac{1}{r_{ij}^{12}} - C_{ij}^{(6)} \frac{1}{r_{ij}^6} \)
- tabulated \( V_{ij} = C_{ij}^{(12)} f_R(r_{ij}) - C_{ij}^{(6)} f_D(r_{ij}) \)
- Buckingham \( V_{ij} = A_{ij} \exp(-B_{ij}r_{ij}) - \frac{C_{ij}}{r_{ij}^6} \)
**Bonded interactions**

Bonds (two-body):

- harmonic bond stretching: \( V_b(r_{ij}) = \frac{1}{2}k_{ij}^b (r_{ij} - b_{ij})^2 \)
- 4th power bond bond stretching: \( V_b(r_{ij}) = \frac{1}{4}k_{ij}^b (r_{ij}^2 - b_{ij}^2)^2 \)
- ...

Angles (three-body):

- harmonic angle potential: \( V_a(\theta_{ijk}) = \frac{1}{2}k_{ijk}^\theta (\theta_{ijk} - \theta_{ijk}^0)^2 \)
- cosine based potential: \( V_a(\theta_{ijk}) = \frac{1}{2}k_{ijk}^\theta \left( \cos(\theta_{ijk}) - \cos(\theta_{ijk}^0) \right)^2 \)
- ...

GROMACS workshop 2007
**Constraints**

In GROMACS bonds are usually replaced by constraints, i.e. the distances between bonded atoms are fixed to their equilibrium values.

Constraints are actually a more faithful representation of reality, since the dynamics of bonds is of quantum nature.

Removing the bond vibrations with constraints allows one to increase the timestep from 0.5 or 1 fs to 2 or 4 fs.
Improper dihedrals (four-body), for planarity or stereochemistry:
  • harmonic

Improper dihedrals (four-body):
  • single cosine: \( V_d(\phi_{ijkl}) = k_\phi (1 + \cos(n\phi - \phi_s)) \)
  • Ryckaert-Bellemans: \( V_{rb}(\phi_{ijkl}) = \sum_{n=0}^{5} C_n \cos^n(\psi) \)
    • ...

1-4 interactions (two-body):
  • (scaled) Lennard-Jones interactions
  • (scaled) Coulomb interactions

1-4 interactions go together with dihedrals, more on that later
Dihedrals

Proper dihedral

Also 1-4 interactions between atoms $i$ and $j$ contribute to the torsional potential around $\phi$!

Improper dihedral
The choice of force field is one of the most important factors in an MD study.

A good force field requires years of work by many people.
All parameters for all molecules should be consistent!

The first question is which of the force field have parameters for the molecules present in your system.

When some parameters are missing:
- search the literature
- if no success, you need to do force field parametrization :-(

GROMACS workshop 2007
### Biomolecular force fields in GROMACS

<table>
<thead>
<tr>
<th>GROMOS</th>
<th>OPLS-AA</th>
<th>AMBER</th>
</tr>
</thead>
<tbody>
<tr>
<td>(6 versions)</td>
<td>all-atom</td>
<td>all-atom</td>
</tr>
<tr>
<td>united atom</td>
<td>proteins</td>
<td>proteins</td>
</tr>
<tr>
<td>aliphatic CH$_n$</td>
<td>hydrocarbons</td>
<td>DNA/RNA</td>
</tr>
<tr>
<td>proteins</td>
<td>hydrocarbons</td>
<td>many liquids</td>
</tr>
<tr>
<td>hydrocarbons</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

An AMBER port for Gromacs is available at:
http://folding.stanford.edu/ffamber/
Charges and Lennard-Jones parameters:

Very tricky!
Try to do it the same way as the force field has been parametrized

Bonded terms:

Relatively easy, from quantum calculations But beware of the 1-4 terms for dihedrals!
Torsional potential setup

The original idea behind this is:
the dihedral potentials give the 2-3 dependencies only
the 1-4 pair potentials give the 1-4 dependencies only

Dihedrals:
- 9: $C_1$-$C_2$-$C_3$-$C_4$, $H_{21}$-$C_2$-$C_3$-$C_4$, $H_{22}$-$C_2$-$C_3$-$C_4$, ...
- 1 (GROMOS96): only $C_1$-$C_2$-$C_3$-$C_4$

1-4 LJ/Coulomb pair interactions:
- 9: $C_1$-$C_2$-$C_3$-$C_4$, $H_{21}$-$C_2$-$C_3$-$C_4$, $H_{22}$-$C_2$-$C_3$-$C_4$, ...

GROMACS workshop 2007
**Torsional potential parametrization**

What to do when parameters for a certain dihedral are not present in the force field?

1) Scan the dihedral in steps of e.g. 15 degrees and optimize the structures using quantum calculations (e.g. Gaussian)

2) Determine the force-field energies for the optimized quantum structures using `mdrun -rerun`, including all interactions, except for the dihedral to parametrize

3) Determine the difference between the quantum and force-field potential energies and fit one of the dihedral functional forms to this
Your topology is contained in the .top file

The .top file “includes” the force field .itp files, which are stored in the directory $GMXDIR/shared/top

For proteins your molecular topology, i.e. the set of bonded interactions for your molecule, can be generated from the force field building blocks with the program pdb2gmx
Topology file

; Include forcefield parameters
#include "ffG53a6.itp"

[ moleculetype ]
; Name      nrexcl
Methanol        3

... molecule definition stuff, next slide ...

; Include water topology
#include "spc.itp"

[ system ]
; Name
Methanol in water

[ molecules ]
; Compound       #mols
Methanol        1
SOL             640
Molecule definition

[ moleculetype ]
; Name nrexcl
Methanol 3

[ atoms ]
; nr type resnr residue atom cgnr charge mass
 1 CH3 1 SER CB 1 0.266 15.035
 2 OA 1 SER OG 1 -0.674 15.9994
 3 H 1 SER HG 1 0.408 1.008

[ bonds ]
; ai aj funct c0 c1
 1 2 2 gb_18 gb_18
 2 3 2 gb_1 gb_1

[ angles ]
; ai aj ak funct c0 c1
 1 2 3 2 ga_12 ga_12
Force field files

ffG53a6.itp:

```c
#define _FF_GROMOS96
#define _FF_GROMOS53A6

[ defaults ]
; nbfunc comb-rule gen-pairs fudgeLJ fudgeQQ
  1 1 no 1.0 1.0

#include "ffG53a6nb.itp"
#include "ffG53a6bon.itp"
```

ffoplsaa.itp:

```c
#define _FF_OPLS
#define _FF_OPLSAA

[ defaults ]
; nbfunc comb-rule gen-pairs fudgeLJ fudgeQQ
  1 3 yes 0.5 0.5

#include "ffoplsaanb.itp"
#include "ffoplsaabon.itp"
```
**OPLS-AA files**

**ffoplsaanb.itp:**

```plaintext
[ atomtypes ]
; full atom descriptions are available in ffoplsaa.atp

; name    bond_type  mass  charge  ptype  sigma  epsilon
  opls_001  C    6  12.01100  0.500  A  3.75000e-01  4.39320e-01 ; SIG
  opls_002  O    8  15.99940 -0.500  A  2.96000e-01  8.78640e-01 ; SIG
  opls_003  N    7  14.00670 -0.570  A  3.25000e-01  7.11280e-01 ; SIG
...
```

**ffoplsaabon.itp:**

```plaintext
[ bondtypes ]
; i    j    func  b0      kb
  OW   HW  1  0.09572  502080.0 ; For TIP4P Water - wlj 1/98
  OW   LP  1  0.01750  753120.0 ; -idem-
  C*   HC  1  0.10800  284512.0 ;
...
```

```plaintext
[ angletypes ]
; i    j    k    func  th0   cth
  HW   OW   HW  1  109.500  627.600 ; For TIP4P Water - wi 1/98
  HW   OW   LP  1   54.750   418.400 ; For TIP4P Water - wi 1/98
```

GROMACS workshop 2007
GROMOS96 files

ffG53a6nb.itp:

[ atomtypes ]
; name at.num mass charge ptype c6 c12 1
  O 8 0.000 0.000 A 0.0022619536 1e-06
  OM 8 0.000 0.000 A 0.0022619536 7.4149321e-07
  OA 8 0.000 0.000 A 0.0022619536 1.505529e-06
...

[ nonbond_params ]
; i j func c6 c12
  OM O 1 0.0022619536 8.611e-07
  OA O 1 0.0022619536 1.38651e-06
  OA OM 1 0.0022619536 2.258907e-06
...

[ pairtypes ]
; i j func c6 c12
  O O 1 0.0022619536 7.4149321e-07
  OM O 1 0.0022619536 7.4149321e-07
...

GROMACS workshop 2007
GROMOS96 files

ffG53a6bon.itp:

; ICB(H)[N] CB[N] B0[N]
;
#define gb_1 0.1000 1.5700e+07
; H - OA 750
;
#define gb_2 0.1000 1.8700e+07
; H - N (all) 895
...

; ICT(H)[N] CT[N] (T0[N])
;
#define ga_1 90.00 380.00
; NR(heme) - FE - C 90
;
#define ga_2 90.00 420.00
; NR(heme) - FE - NR(heme) 100
...
Making topologies

Proteins:
  - use `pdb2gmx`

Small molecules:
  - write a topology by hand, but be very careful about the details of the force field used

(hetero-)polymers:
  - write building blocks (an .rtp file)
  - then use `pdb2gmx`
Thermostats and barostats
Ensembles

“Plain” molecular dynamics is Hamiltonian dynamics and this conserves energy, leading to an constant NVE or microcanonical ensemble.

However, in most cases a constant NVT (canonical) or NPT ensemble will be a more suitable representation of the real system.

To achieve this, the equation of motions need to be modified to include a thermostat and/or barostat.

There are two types of thermostats:

- global thermostats
- local thermostats
The Berendsen thermostat

A global thermostat, weak coupling to a bath:

\[ \frac{dT}{dt} = \frac{T_0 - T}{\tau} \]

To achieve this first order decay of temperature deviations, all velocities are scaled with the factor:

\[ \lambda = \left[ 1 + \frac{\Delta t}{\tau_T} \left\{ \frac{T_0}{T(t - \frac{\Delta t}{2})} - 1 \right\} \right]^{1/2} \]

\( \tau_T \) is set with the parameter \texttt{tau_t} in the mdp file

Different \( T_0 \) and \( \tau_T \) can be used for subgroups of the system.
A global thermostat, using the extended-ensemble approach.

An additional degree of freedom $\xi$ is added to the Hamiltonian:

$$\frac{d^2 r_i}{dt^2} = \frac{F_i}{m_i} - \xi \frac{dr_i}{dt}, \quad \frac{d\xi}{dt} = \frac{1}{Q} (T - T_0)$$

The dynamics for the extended Hamiltonian is canonical, but leads to an NVT ensemble for the original system.

In the mdp file not the mass $Q$ is set, but the period of the oscillations.
Global thermostats

Berendsen thermostat
advantage: little influence on kinetics
disadvantage: does not generate an ensemble (but error is small)

Nosé-Hoover thermostat
advantage: produces an NVT ensemble
disadvantage: can induce oscillations in the dynamics

Temperature jump from 300 to 330 K for 1000 $H_2O$

![Temperature graph](image)
Ergodicity

Global thermostats only work well when all degrees of freedom in the system are coupled strong enough to one another.

Consider the following system:

<table>
<thead>
<tr>
<th></th>
<th>octane</th>
<th>water</th>
</tr>
</thead>
</table>

Here the octane and water slabs are very weakly coupled. Using one global thermostat can lead to a temperature difference between the octane and water slabs.

A solution is: using separate thermostats for octane and water

But what about the 3 degrees of freedom coupling the octane and water slabs?

GROMACS removes the center of mass motion for octane and water.
Local thermostats

One can also use Stochastic Dynamics (SD):

\[ m_i \frac{d^2 r_i}{dt^2} = -m_i \xi \frac{dr_i}{dt} + F_i(r) + \eta_i \]

advantages: NVT ensemble, always ergodic, no oscillations
disadvantage: all velocity correlations decay with \( \exp(\xi t) \)

In the overdamped limit SD reduces to Brownian Dynamics:

\[ \frac{dr_i}{dt} = \frac{1}{\gamma_i} F_i(r) + \eta_i \]

BD is not useful for dynamics,
but can be very useful for equilibrating bad starting structures.
Barostats

Berendsen barostat, or weak coupling:

\[
\frac{dP}{dt} = \frac{P_0 - P}{\tau_p}
\]

The box/coordinate scaling matrix \( \mu \) is given by

\[
\mu_{ij} = \delta_{ij} - \frac{\Delta t}{3 \tau_p} \beta_{ij} \{ P_{0ij} - P_{ij}(t) \}
\]

Parrinello-Rahman barostat:

\[
\frac{db^2}{dt^2} = VW^{-1} b'^{-1} (P - P_0)
\]

\[
\frac{d^2 r_i}{dt^2} = \frac{F_i}{m_i} - M \frac{dr_i}{dt}, \quad M = b^{-1} \left[ b \frac{db'}{dt} + \frac{db}{dt} b' \right] b'^{-1}
\]
Calculating non-bonded interactions
Neighbor searching

In principle for a periodic system of $N$ atoms we would need to determine $N(N + 1)/2$ pair interactions, plus the interactions will all periodic images.

When interactions are shorted ranged we can use cut-off’s.

For efficiency we used a fixed neighbor list for $n_s$ steps:

- construct neighbor list for all charge groups pairs within $r_c$
- $n_s$ times:
  - evaluate $V$ and $F$ for all pairs in the list
  - integrate the equations of motion
- construct neighbor list for all charge groups pairs within $r_c$
- ...

GROMACS workshop 2007
Charge groups

We use charge groups because of two reasons:

- Neighbor searching efficiency
  Example: For a pair of 3-site water molecules we need 9 atom pair distances, but only 1 charge group distance

- Avoiding electrostatic artifacts due to cut-off’s
  Example: For water without charge groups partial charges enter the cut-off, with charge groups dipoles enter the cut-off
Cut-off, charge groups and diffusion
Cut-off, charge groups and diffusion
Electrostatics with cut-off

Cutting off interactions abruptly causes artefacts:

- In water: over-orientation of dipoles: anti-parallel before the cut-off, parallel beyond the cut-off
- Like charged ions will accumulate just beyond the cut-off

Some artefacts get worse with increasing cut-off distance!!!

This is because Coulomb goes as $1/r$, but the surface of the cut-off sphere goes as $r^2$
Reaction-field electrostatics

\[ \varepsilon_{rf} = \infty \implies V = f \frac{q_i q_j}{\varepsilon_r} \left[ \frac{1}{r_{ij}} + \frac{1}{2} \frac{r_{ij}^2}{r_c^3} - \frac{3}{2} \frac{1}{r_c} \right] \]

Advantage: much less cut-off artefacts

Disadvantage: \( V = 0 \) beyond \( r_c \)
Particle Mesh Ewald (PME) electrostatics

Split the electrostatics into a short- and long-range part
Do the short-range part with particle-particle interactions
Solve the long-range part on a grid using FFT’s

![Graph showing the potential vs distance](image)

Advantages: no cut-off artefacts, infinite interaction range
Disadvantages: computationally more expensive, periodicity (but this is due to the boundary conditions)

Can be used with systems with net charge:
a uniform background charge is applied
**Lennard-Jones potential**

![](image)

LJ decays fast: \(-r^{-6}\)

but the dispersion contributes up to relatively long distances:

\[
V_{disp} = \int_{r_c}^{\infty} \rho_n \langle C^{(6)} \rangle 4\pi r^2 r^{-6} \, dr = -\rho_n \langle C^{(6)} \rangle \frac{4\pi}{3} r_c^{-3}
\]
## Some useful box shapes

[Image of box shapes]

<table>
<thead>
<tr>
<th>box type</th>
<th>image distance</th>
<th>box volume</th>
<th>box vectors</th>
<th>box vector angles</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>cubic</td>
<td>(d)</td>
<td>(d^3)</td>
<td>(d)</td>
<td>0</td>
</tr>
<tr>
<td>rhombic dodecahedron</td>
<td>(d)</td>
<td>(\frac{1}{2}\sqrt{2} d^3)</td>
<td>0</td>
<td>(\frac{1}{2} d)</td>
</tr>
<tr>
<td>truncated octahedron</td>
<td>(d)</td>
<td>(\frac{4}{9}\sqrt{3} d^3)</td>
<td>(\frac{1}{3} d)</td>
<td>(\frac{2}{3} \sqrt{2} d)</td>
</tr>
</tbody>
</table>
Neighbor search with a grid

grid search in two dimensions, the arrows are the box vectors
Neighbor search/interaction setting

Neighbor search is C-code ⇒ slow
Force loops are SSE/Altivec/... assembly ⇒ fast

For efficiency you want to use nstlist = 10

To avoid electrostatic artifacts use PME

<table>
<thead>
<tr>
<th>force field</th>
<th>neighb. list</th>
<th>Elec. cut-off</th>
<th>PME grid</th>
<th>VdW type</th>
<th>VdW cut-off</th>
<th>disp. corr.</th>
</tr>
</thead>
<tbody>
<tr>
<td>GROMOS</td>
<td>1.0</td>
<td>1.0</td>
<td>0.135</td>
<td>shift</td>
<td>0.9</td>
<td>yes</td>
</tr>
<tr>
<td>united atom</td>
<td></td>
<td></td>
<td></td>
<td>cut-off</td>
<td>1.4</td>
<td>no</td>
</tr>
<tr>
<td>OPLS-AA</td>
<td>0.9</td>
<td>0.9</td>
<td>0.125</td>
<td>shift</td>
<td>0.8</td>
<td>yes</td>
</tr>
<tr>
<td>all atom</td>
<td></td>
<td></td>
<td></td>
<td>cut-off</td>
<td>1.4</td>
<td>no</td>
</tr>
</tbody>
</table>