

Luis Vollmers | Dr. Maria Reif | Prof. Martin Zacharias

Molecular Dynamics (SS2025)

Exercise 5

The Significance of Accurate Long-Range Interactions

Report Tasks

Task 1 is optional. We still recommend at least reading through Section 1, as it contains all the necessary background information for the other tasks. In Section 2, you can limit yourself to either A- or B-DNA. Of course, you can also do both if you are interested.

1. (Optional) From the SPC/E water simulation, compare the first two peak positions of the RDF for all six different potential-modifiers with the experimental values. (Section 1)
2. Include RDF plots for all the methods with a qualitative and quantitative comparison to figure 3b. (Section 2)
3. Include time resolved plots and averages of the RMSDs for all the different methods with a brief discussion. (Section 2)
4. Analyse all results regarding quality and performance and try to determine a best-practice method. Include the timings in this discussion. (Section 2)

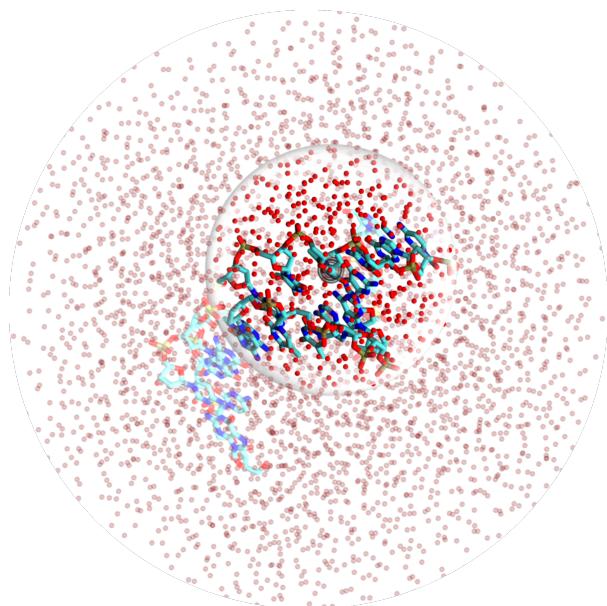


Figure 1: Simplified visualisation of the cut-off radius. Every opaque particle falls within the (enlarged) central particle's cut-off, signified by the *force field bubble*. Nonbonded interactions with (transparent) particles outside this bubble would be omitted in a corresponding MD simulation.

Water Structure Under the Influence of Different Electrostatic Interaction Schemes

The intermolecular forces, resulting from the Coulomb potential V_C and Lennard-Jones potential V_{LJ} , converge to 0 for infinite interatomic distances. Evidently, the finite size of the simulation box and the periodic image forbids the accurate treatment of these long-range interactions. Additionally, the computational time depends largely on the number of atom pairs, for which nonbonded interactions have to be calculated. The introduction of a potential-modifier that usually uses a cut-off distance r_{cut} to truncate nonbonded interactions remedies these issues. On the other hand, the physical exactness must be maintained as well as possible, leading to various algorithms that mingle with this trade-off. This tutorial aims to explore some of these potential-modifiers, analyse them for their strengths and weaknesses, and crown one solution as best practice.

The algorithms modify the nonbonded potential and are discussed in the lecture. This exercise investigates the potential shift, the plain cut-off, and the Particle-Mesh-Ewald method. The two former methods rely on cut-off truncation to simplify the calculation of electrostatic interactions. For these methods, a metric must be defined that tells GROMACS when to shift the potential and truncate the nonbonded interactions. The PME method, on the other hand, makes use of lattice summation (implemented in terms of Fast Fourier Transforms) to simplify the calculation of electrostatic interactions. Its use requires the specification of several MDP options beyond this tutorial's scope. The prepared MDP files can be downloaded from Moodle. As usual, the system has to be set up using GROMACS' tools. The edges of the cubic box should be 4 nm and the SPC/E water model is supposed to be used with the amber99SB-ILDN force field. The MDP files for the equilibrations can be downloaded (EM, NVT, and NPT). The system can be equilibrated once, followed by the pre-production and production run for each long-range electrostatics treatment. In this tutorial, six different modifiers are tested: PME, potential-shift with 8 Å and 12 Å cut-off, and the truncation method with 8 Å, 12 Å and 18 Å cut-off. The time benchmarks should be extracted from the LOG file to be included in the report. Usually, there are two different run time measurements, the wall time and the core time. The wall time measures the real-world time that elapses during the MD run, and the core time refers to the time spent by all processors to execute every command. The latter is more important in this task and should be compared for the different electrostatic interaction schemes, preferably in a table with a brief discussion.

Furthermore, to measure the quality of the simulation, the radial distribution function (RDF) is supposed to be calculated for each method, and a short introduction shall be given for a profound understanding of the GROMACS implementation.

The RDF between particle group A and B ($g_{AB}(r)$) describes the structure of (isotropic) liquids and liquid mixtures. Basically, it works by counting the particles of type B at a certain distance from a central particle of type A. The abundance of B as a function of the distance reveals information about their intermolecular interactions and could be used to calculate several thermodynamic properties. To calculate $g_{AB}(r)$ the average particle density of particle type B for a distance r , $\langle \rho_B(r) \rangle$, is divided by the bulk density $\langle \rho_B \rangle_{\text{bulk}}$,

$$g_{AB}(r) = \frac{\langle \rho_B(r) \rangle}{\langle \rho_B \rangle_{\text{bulk}}} . \quad (1)$$

The bulk density for B is usually calculated with respect to the complete simulation box, hence dividing N_B by the box volume V_{Box} . For a continuous distance, the average particle density B around A can be written in terms of the Kronecker-delta $\delta(\dots)$ of the particle distance r_{ij} and

the surface area of the sphere,

$$\langle \rho_B(r) \rangle = \frac{1}{N_A} \sum_{i \in A}^{N_A} \sum_{j \in B}^{N_B} \frac{\delta(r_{ij}, r)}{4\pi r^2} , \quad (2)$$

where $\delta(r_{ij}, r) = 1$ if $r_{ij} = r$ and $\delta(r_{ij}, r) = 0$ otherwise. The expression becomes slightly more complicated for the non-continuous case, used by computational methods. Firstly, spherical shells with a thickness of Δr are considered instead of a sphere's surface, which effectively turns $g_{AB}(r)$ into a histogram depending on n bins $g_{AB}(n)$ of bin width Δr . Let r_n and r_{n+1} be the lower and upper border of the n -th bin. Then, this histogram function can be calculated by augmenting equation (2) with the Iverson bracket notation and the formula for a spherical shell,

$$\langle \rho_B(n) \rangle = \frac{1}{N_A} \sum_{i \in A}^{N_A} \sum_{j \in B}^{N_B} \frac{[r_n < r_{ij} < r_{n+1}]}{\frac{4}{3}\pi(r_{n+1}^3 - r_n^3)} . \quad (3)$$

This expression can be further simplified by assuming that each shell is very thin, hence $\Delta r \ll r_n$. Given GROMACS' default bin width of 2 pm, this assumption is satisfied, resulting in the final expression of the discretised radial distribution function,

$$g_{AB}(n) = \frac{1}{N_A \langle \rho_B \rangle_{\text{bulk}}} \sum_{i \in A}^{N_A} \sum_{j \in B}^{N_B} \frac{[r_n < r_{ij} < r_{n+1}]}{4\pi r_n^2 \Delta r} . \quad (4)$$

Notice that A and B can also be of the same particle type, which is useful for investigating the structure of pure liquids like pure water, especially since the RDF can be obtained from scattering experiments as well as from MD simulations. A direct comparison of the plot appearance can hint towards the force field quality of the water model and the methods as well.

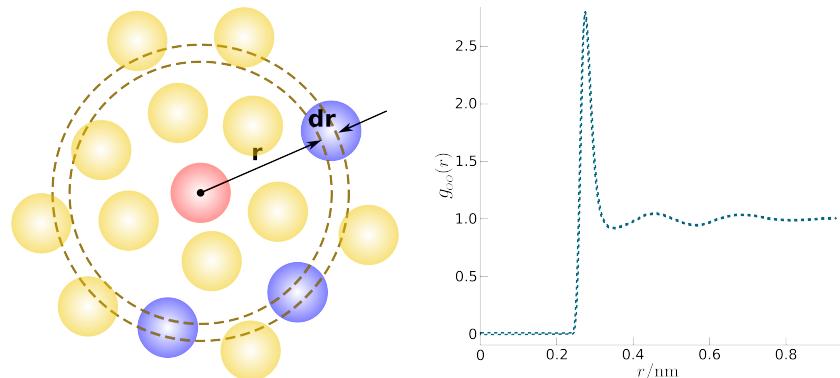


Figure 2: Left: Schematic on how the RDF is calculated.¹ Right: The RDF of water oxygens of pure SPC water calculated from Simulation.²

A quantitative comparison can be conducted utilising RDF peak positions. In figure 2, the peak locations correspond to solvation shells of water molecules. Therefore, the first peak corresponds to the location of the first, very distinct solvation layer. The second peak corresponds to the second, etc. The simulation quality can be accessed by comparison with the experimental peak positions (not heights). Therefore, the report should include the timings

¹https://en.wikipedia.org/wiki/Radial_distribution_function

²Berendsen, Postma, van Gunsteren, Hermans, Interaction models for water in relation to protein hydration. **Intermolecular Forces.** ed. D. Reidel Publishing Company Dordrecht, 1981, 331–342

and the r -values of the first two RDF peaks for each method and a quantitative comparison. Experimental values for comparing r_1 and r_2 were measured by Skinner et al.³. They found r_1 at 2.79983 Å and r_2 at 4.5149 Å for a temperature of 300 K. Notice that the RDF is calculated for the water oxygen atoms, not the whole molecules. For completing this task `gmx rdf` is highly recommended. Make sure to read through the GROMACS command line reference since this command is quite versatile. As a hint, you will need to use the `-ref` and `-sel` flags and specify the atom names given in the GRO file.

A-DNA and B-DNA Simulations with Different Electrostatic Interaction Schemes

After the investigation of pure water, the effects of different methods to calculate electrostatic interactions are probed for complex bio-molecules. Herein, A-DNA and B-DNA are simulated, both stable conformations of Watson-Crick-base-paired DNA. However, B-DNA has been proven to be more stable. Furthermore, DNA is a highly charged molecule with many anionic oxygen atoms in the phosphate backbone. Thus, it is expected that the treatment of long-range electrostatic interactions has a vast impact on DNA stability and structure.

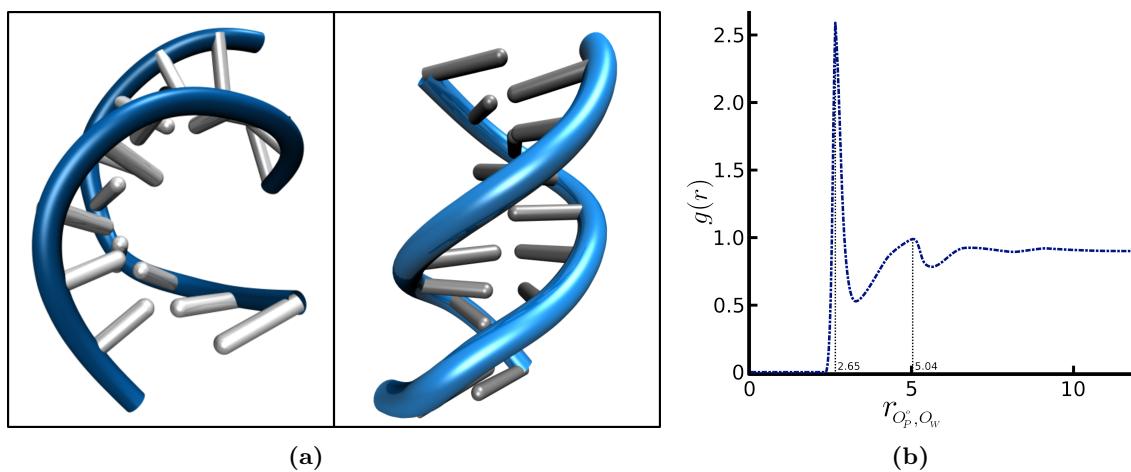


Figure 3: a) A-DNA and B-DNA conformations visualised by VMD. B-DNA exhibits the classical helix shape. b) RDF calculated by Jaiswal et al.¹ between the uncharged phosphate oxygen atoms and the water oxygen atoms. The peak positions are indicated on the x-axis.

The files necessary for the simulations can be downloaded from Moodle. The setup consists of six electrostatic interaction schemes (same as in the prior section). Two PDB files, pictured in figure 3a, contain the coordinates of the DNA molecules. The usual GROMACS programs are used to generate a solvated simulation box from the PDB files.

```
gmx pdb2gmx [...] # Use AMBER99SB-ILDN and SPC/E water; also specify flag: -ignh
gmx solvate [...] # solvates the simulation box; Use cubic box with edge length 5 nm
gmx grompp [...] # just to generate tpr file for gmx genion
gmx genion [...] # replaces water with ions; just neutralise the DNA
```

³Skinner, Benmore, Neufeind, Parise, The structure of water around the compressibility minimum, **J. Chem. Phys.** 141, 214507, 2014

¹Jaiswal, Srivastava, Pandey, Bandyopadhyay, Microscopic picture of water-ethylene glycol interaction near a model DNA by computer simulation., 2018, **PLoS ONE** 13(11): e0206359

After the system generation, the same MDP files as in the first section should be used to equilibrate the system. The electrostatic interaction schemes become important only for the pre-production and production run, from which the data is collected for analysis.

The information of interest consists of the RMSD, the RDF, and the timings of the production run for each method since the calculation of nonbonded interactions is the major bottleneck of the performance. The timings are readily collected from the LOG file. In order to extract the other information from the trajectories, they need to be augmented first. Augmentation in the sense that the DNA molecules should be centered in the simulation box since they can be broken across periodic boundaries, which messes up the measurements. The GROMACS tool that can manipulate trajectories is called `gmx trjconv`, which is used in combination with an index file (NDX) to target a non-standard atom group. Herein, the non-standard group consists of one ssDNA strand, which is necessary because centering the DNA dimer can position each DNA strand on opposing periodic boundaries. GROMACS has its own tool to generate NDX files, fittingly called `gmx make_ndx`, which offers a command-line based user interface. Within this prompt, the user has to pick the correct atoms resembling the desired index group, e.g. 'a 1-315', followed by saving the new index group by typing `q` + `enter`. Since the website information for `gmx trjconv` is quite confusing, the correct options for the command are given below.

```
gmx trjconv -f <PROD-TRAJ>.xtc -s <PROD-STRUC>.tpr -center -n <INDEX>.ndx -pbc mol -o <WHOLE-TRAJ>.xtc
```

After execution, `gmx trjconv` prompts the user for which group to center (the ssDNA) and which group to output (the whole system). The RMSD of the trajectory is to be calculated in reference to its first frame, a.k.a. the last frame of the pre-production run. To make sure that the reference frame is not split across periodic boundaries, `gmx trjconv` needs to be called for `<PRE-GRO>.gro` as well.

```
gmx trjconv -f <PRE-GRO>.gro -s <PRE-STRUC>.tpr -center -n <INDEX>.ndx -pbc mol -o <WHOLE-GRO>.gro
```

The resulting trajectory can then be used to calculate each system's RDF and RMSD curves (The RDF can also be calculated with the non-augmented trajectory). Herein, the RDF should be determined between the neutral phosphate oxygens and the water oxygens. Jaiswal et al. calculated such an RDF with TIP3P water for B-DNA, which serves as a ground truth for comparison (see figure 3b for plot appearance and reference peak positions). Thus, the plot appearances should be compared qualitatively, whereas the peak locations should be compared quantitatively. The command `gmx rdf` is recommended to fulfill this task.

Regarding the RMSD, two measurements are of utmost importance: the time resolved RMSD and the average RMSD of the production run. The reference structure that is always needed to calculate an RMSD is the prepared `<WHOLE-GRO>.gro` file and needs to be specified in the `-s` flag of `gmx rms`. The report should include RMSD plots of all the different methods as well as a comparison with figure 4, assessing which cut-off scheme produces realistic RMSD curves. Furthermore, a brief discussion of the average RMSDs for each method should be included. Eventually, all results from the DNA and the water simulations should be used to crown one potential modifier as best practice to calculate electrostatic interactions, with a brief explanation. While timing information is crucial, the quality of the results is to be ranked as more important in this evaluation.

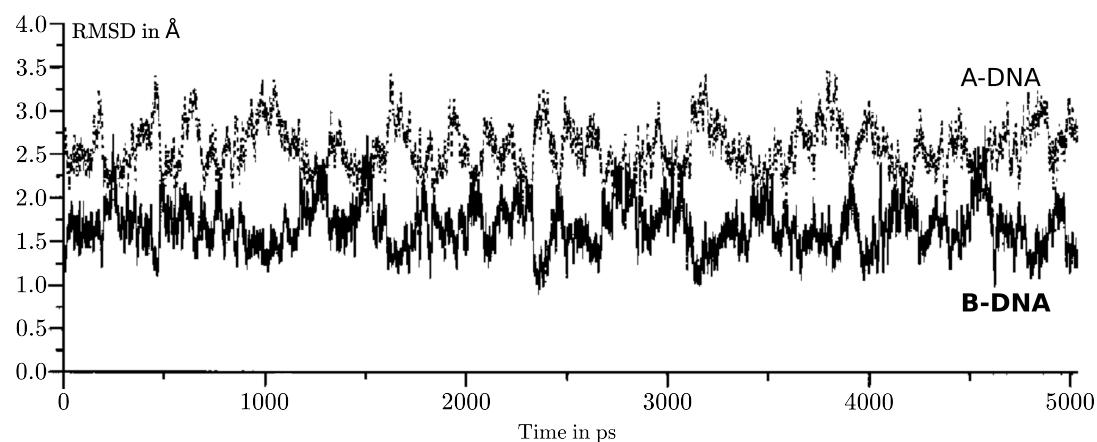


Figure 4: Exemplary time resolved RMSD. Modified from.²B-DNA is slightly more stable, which results in an overall lower RMSD. However, the fluctuations seem comparable to those seen in the visual inspection.

²Norberg, Nilsson. On the truncation of long-range electrostatic interactions in DNA. 2000, **Biophys J.**; 79(3):1537-53.