GROMACS Workshop

@ Virtual Winter School on Computational Chemistry
https://winterschool.cc/program/day-4/gromacs-2026

Q&A for Lecture introduction to MD

- why for atom of florine it created an imaginary atoms?
    - Not sure I understand the question. Was a dummy atom created or are you referring to virtual sites?
- which force field is used for metal complexes?
    - There are specific forcefileds for metal ions, but they are very constrained. In the biosimulation forcefield like AMBER or CHARMM all ions by default will have the most spherical arrangement of the ligands (octahedral).
- How do ML-based force fields work? For instance, do you do other force fields and “learn” behavior? I guess I am unsure what is learned in ML-based force fields.
    - There are different ways of doing this, but the most common approach is perhaps to train the force field from quantum mechanical simulations to get more detailed interactions to a lower computational cost. A similar approach can also be used for larger, coarse-grained systems that are trained on atomistic MD simulations.
    - Usually ML forecfields are made to reproduce results of ab initio quantum chemistry. They are also orders slower in simulation speed than classical forcefields but still orders faster than doing pure DFT dynamics.
- Are the force field parameters from existing experimental results?
    - Usually not, but they are tested to reproduce experimental/theoretical results, e.g. macroscopic properties such as densities, heat of vaporization.
- Why FF has to be transferable. Could you expand the idea?
    - The idea of the forcefield is to speed-up simulations. If it is not transferable then for each new system you would need to re-parametrize the forcefield, which will take a lot of time and there will be no speed-up anymore.
- Is dispersion included in these non-bonded long-range interactions?
    - Dispersion will be modeled by Lennard-Jones interactions and electrostatics.

- I’m curious to hear your opinion on the best parameterisation method. Some tools use heat of vaporisation as a generalisable approach, while others use more specific use cases. For example, fine tuning parameters to replicate observables specific to their model. I suppose my question is, do you think there is a single set of parameters that can correctly represent a molecule for classical simulations? Or do you think a number of different parameter sets for the same molecule is appropriate to accommodate the specific research question? I’ve struggled with this as many papers use different parameters for the same molecules, with different research questions. For this question, assume I’m talking about the same forcefield being used.
    - Nowadays parametrization is done by ab initio methods like MP2 or DFT, sometimes with further tuning towards desired experimental properties.
- Are the 26 cell images explicitly represented in the molecular dynamics code, or is there only one cell with a PBC? Why do we need to highlight the actual number of neighboring cells?
    - There’s only one cell with PBC in the simulation. However, if you think about your system, you’re simulating several copies of your molecules extending infinitely in all directions. It might be helpful to visualize in this way when thinking about the size of the box, such that the molecule doesn’t interact with itself (or its “neighbor”).
- Can u use curved cells to better fit a geometry-non-trivial boundary? Hence leading to harder implementation of boundary conditions.
    - It’s difficult to pack curved surfaces in 3D, but you can use cells such as dodecahedra.
- How does the box shape affect the results of a study? Is it safe to say that a more spherical box shape leads to more reliable results?
    - It depends. The closer you get to a spherical cell, the lower the computational cost since the volume is lower (fewer particles, usually solvent, to simulate). You must be able to pack the box in 3D to use PBC, so that puts some restriction on the shapes you can use. Then, it depends on your system. If you have a membrane, it’s better to use something rectangular, while for a small protein in water, it’s better to use something more “spherical”.
- there's any possibility to have BioExcel school available for african student or why not doing a version for african student?
    - BioExcel is an EU project, so we can currently only offer workshops within the EuroHPC member countries. Attendees will need to be located in the EU to attend.
- How should I choose the time constant value for thermostats and barostats? How should I choose the time constant value for thermostats and barostats?

    - It depends on the algorithm. There are usually good default values to select, but the best is to read up on the specific algorithm you're using.
- If we remove non-bonding interactions, isn't there a risk for clashes in the system? Or are we doing this statically?
    - Yes, atoms that are not bonded to each other might then overlap. Bonded atoms will still be separated by the harmonic potential modeling the bond. Also, sometimes a small portion of L-J are always kept to prevent such behavior

Q&A for Tutorial Introduction to MD · link

- can gromacs handle leaving the crystal waters in there?
    - Yes, you can leave the water molecules (that you had in the pdb file) when you process pdb2gmx. Then attention has to pay if you want to have a separate topology for the crystal waters, and solvent waters
    - Are there additional steps to take while solvating the system when the crystal waters were left in the structure?
        - Not in the solvation, but when adding ions it is good to have the crystal water in a different index group and not in the SOL group to avoid to replace the crystal water with an ions
    -
- Is the topology file is the same as the configuration file or so?
    - Not really, the topology file describes the atom connectivity. But to the same connectivity you can have different configurations. Note in force field based simulation you can not break bonds. The information of the positions (structure/configuration) of the atom you have in the gro file (starting configuration). The simulations will sample different configurations and same in function of time in the output file (trajectory)
- Regarding charge group numbers, if they are no longer used in simulations, can I just leave the space empty or should I type any character? Is topology file read by gromacs column by column (as PBD files) or can I write as many spaces as I want between values?
    - I guess you mean if you write the topology manually. In that case you still have to follow the syntax of the topology. The order of the entry should match the entries, but the space is between entry is not defined as in the old pdb file format

- how do would I put crystal waters in a separate index group than the SOL group?
    - To create group in index.ndx you can use gmx make_ndx tool

- Is the -conc flag in gmx genion based on the box volume or the volume of the molecule you wish to replace with ions? (i.e water molecules).
    - It is based on the number of molecules and convert it in volume
- what fluctuations are okay? Do big ones indicate that a bigger water box is needed
    - *The magnitude of the fluctuations depends completely on the system size. The fluctuations of any intensive property, like the pressure and temperature, scale as 1/sqrt(nr. atoms). And the pressure fluctuates due to collisions and vibrations of atoms, so its relative fluctuation is high. E.i >400 bar is high, but could be ok for a very small system.*

- In npt you let I am guessing volume as the one free variable . But if you keep the simulation in a box, how do you allow the volume to fluctuate?
    - We use barostat to couple the system to the desidered average pressure (corresponding fluctuations) - see lecture for details - the coupling occur every X ps, then each X ps the position of the atom is changed to correct for the pressure deviation. or *Pressure can be controlled by varying the volume of the system and scaling the position of the molecules*

Q&A for Tutorial: Solvation free energy · link

- Could you please tell us why we changed the integrator here to a stochastic one?
    - Since we have a system without solvent, it's troublesome to use the thermostat without the molecule accumulating kinetic energy. Therefore, we need to use Langevin dynamics instead of regular MD to model solvent implicitly.
- so, just to confirm, lambda state = 1 in this case is the 0, state 2 is the 0.2 .. etc?
    - Almost, but the lambdas should be 0-indexed, not 1-indexed.
- Can this method (BAR) be used to calculate entropy?
    - Not really. There are better methods for that.
- Is it a Delta G or a Delta H ?
    - Delta G
- Does it matter to run all simulations for the same time? would it still work if we for example only ran run the one with the bigger error longer?
    - A better approach is often to add more lambda points. The simulations need have the same lengths to work with the BAR implementation.

Q&A for Tutorial AWH basic · link

- Why are you using cutoff for vdwtype, not ewald?
    - In force field, PME / RF are used to describe long range non-bonded interactions, LJ and electrostatic potential are used to describe short range interaction up to the cut-off. How to describe the long-range interaction is intrinsically connected to the selection of the force field setting. In the AWH basic tutorial OPLS-AA force field is used.
- Thank you very much for the nice tutorial. I was wondering if you think this method could be used to estimate an adsorption energy on a metal slab, for example? Or construct a dissociation curve from the metal lab surface with an adsorbent?
    - It might be possible, but the absorption estimation requires quantum mechanics. However, it might be possible to combine AWH and QM/MM for this.

Q&A for QM/MM demonstration

- you mentioned the advantages of using cp2k, but is it possible to use another QM software? Is it possible to use ML potentials?
    - Yes, some ML potentials are available in CP2K. Gromacs has its own ML potentials you can try. The Gromacs team is working on a new unified QM/MM interface as well which will become available in the future.
- Is basis set automatically selected?
    - Yes.
- If there is an MM environment, would this be electrostatic embedding, then? Also, how are the bonds across the boundaries treated here? And are theere capping atoms across the boudries as well?
    - Yes, it’s electrostatic embedding. Periodic boundary conditions are taken into account.
- Can you define QM region dynamically? For example, to have all molecules within 5nm of your chosen group be treated as QM
    - No, this would break energy and forces continuity over the trajectory
- The availability of basis set and functional is limited. Is there a plan for future to implement more basis set and functional?
    - Yes, it’s limited, but by purpose.
- Since you mentioned the Kohn-Sham approach in the powerpoint, are you unable to do wavefunction-based or multi configurational approaches?
    - CP2K is mostly a DFT code, so hybrid potentials are slower but still possible. Multi-configurational approaches are currently not possible.

- how we can describe a part from the protein residues in the active site and the synthesis compounds
    -
- sbatch ?? what command is that
    - It's a command you can use to submit compute jobs to a scheduling queue on a supercomputer (if it uses the SLURM queueing system). Your job is then placed in a queue and assigned to the compute resources you asked for in a fair way for all users on the system.
- so the protein is the MM part and the ligand is the QM part if i understand
    - In most cases, yes
- when you say that you equilibrate at the MM level, it means that you have find some parameters for the chromophore, right ?
    - MM equilibration usually means NVT and/or NPT equilibration (as we did in the intro to MD tutorial). This is to make the MM system stable and reach equilibrium.
- Can this technique only perform QM/MM simulations, or can it also perform single-point or geometry optimization QM/MM calculations if needed?
    - Every type of simulation, maybe with exception of FEP, could be performed with QMMM in GROMACS
- Can GROMACS simulate charged systems without issues? even if the excess charge is in the QM region?
    - Simulation-wise, yes. The problem lies in the physical representation of the system, e.g. if you have a system with different dielectric regions (such as a membrane), you will force ions to move into the membrane, which shouldn't happen. You can compensate for extra charges in the QM-region by adding ions in the MM region.
- If we use CP2K input file , does this implies that we can use any functional or basis set available in cp2k ?
    - Yes if you want to use other DFT functional and/or basis set, please use your own CP2K input file.
- can we also take some amino acids as a part of the QM part or not ??
    - Yes. It's not advisable to break the peptide bonds at the QM/MM interface though. Instead, cut around the Ca or Cb atoms.