Few concepts in protein MD

MD of a protein molecule

- Start with a structure file
  - If a structure is not available, create one, using modelling methods
  - Structure file contains atomic coordinates of all atoms in 3D space, which are present in the molecule
  - Structures determined by NMR method will contain hydrogen atoms present in the molecule
  - **Structures arrived at via X-ray diffraction methods will not have hydrogens**
- If hydrogens are absent, add the missing hydrogens
  - The **pdb2gmx** command will do this for you for many proteins
  - If the protein contains ligand atoms or modified amino acids or posttranslational modifications
  - In such situations, use an external site of choice to add hydrogens to obtain the starting pdb
- Many a times, few loop residues will be missing in the downloaded PDB file
  - This is moreso in case of larger receptor molecules
  - If electron density is not seen, the atom will not be incorporated in structure determinations
  - **Hence inspect the downloaded structure carefully before proceeding further.**
  - **If needed, add the missing atoms or residues and you may need to do an energy minimization**
- Large systems with many atoms are computationally intensive
  - A 150-200 residue protein is on the upper limit computation capability for a personal PC that we encounter usually
  - When interested in protein motion longer than ns timescale, write out structures at larger intervals
  - Run the simulation over longer time
- **Large molecular structures, will need a larger box size and more water molecules for solvation. This will increase the computation time substantially**
- If ligand molecules are present, they need to be deleted from the PDB and processed
  - Choose an external site as desired by the force field of choice
  - Create the gro files for protein and ligand separately and later on, cat them manually, keeping the Gromacs syntax