CHL Lab
Rosalind J. Xu 18’ Fall 2016

Cyanylation on protein from PDB code — using cyanylator_protein.py

Good for: setting up a simulation from a raw PDB file. The output can be further processed via CaM-gmx-md-proc.txt, which goes through equilibration - md run - output processing.

1. Cyanylation: in Chimera.
    1. use cyanylator_protein.py. Edit the first few lines of file: enter folder name, protein name, and sites to be cyanylated one by one. Then open .py file in Chimera. Generate $**protein-$site.pdb**.
    2. To introduce a new fragment/probe group: this can be either done in Chimera, using code similar to cyanylator.py. Or, you can explore the “build” functionality of Pymol, which allows you to draw out the new fragment. There is a command-line counterpart as well.
2. Structural editing: in Pymol.
    1. zoom in to added MSCN. If extra bonds were added by Pymol (because of close interatomic distance), use *unbond* command to remove bonds.
    2. Set torsion angle of CA-CB-SD-CE to 180, using *set_dihedral* command. If steric clashes exists after CA-CB-SD-CE dihedral is set to 180, increase or decrease the dihedral by the least amount until all non-bonded interatomic distances are less than bond lengths. Check to ensure that the dihedral has converged to one of the most stable configurations at the end of equilibration.
    3. Zoom out. If protein is capped, go to wizard-mutagenesis. Select the N-terminal residue; change N-end to “acetyl”, or other capping groups; apply. Similarly for C-term.
    4. If protein has missing residues: if < 5 residues missing from termini of protein, simply use the “*Build*” functionality in Pymol to fill them in. If > 5 residues missing, or residues missing from more structured regions of protein (such as helixes/sheets), consider using programs such as Modeller.
    5. save protein: $**protein-$site-edited.pdb.**
3. .pdb file trimming: in text editor/command line (awk)
    1. change “atom type” of MSCN from “HETATM” to “ATOM”. Adjust name of each atom to that in aminoacids.rtp in the forcefield folder. Make sure “residue type” of MSCN is “Protein” in residuetype.dat in the the top folder (the top folder contains the forcefield folder).
    2. If protein is capped, go to C-term. Move “TER” record to end of chain. Compare names of each atom in the added cap and the capped residue to the terminal entries in aminoacids.rtp. Similarly for N-term.
    3. If using Amber forcefields: if protein contains Lys, His or Cys, check atoms

in residue against aminoacids.rtp**.** Change “residue type” to either LYS or LYN; HID, HIE or HIP; CYN or CYX. Check the naming for the other residues and ions, too. For other forcefields, check documentation and make similar changes.

4. Compare names of hydrogen atoms in each residue to aminoacids.rtp. In case of mismatch, modify switchHydrogenNames.awk to switch in correct names. Similarly for mismatches in other atom types. Please never use the -ignh option: you might miss the protonation states of residues / mistake one residue for another if you do so!
5. Edit (if the forcefield you are using has a different naming scheme for the hydrogens) and run switchHydrogenNames.awk. Generate $**protein-$site-trimmed.pdb.**