Protein Preparation in AMBER

Based on the original work by Dr. Adrián Durán-Vargas

Introduction to AMBER

AMBER (Assisted Model Building with Energy Refinement) is a widely used suite of tools designed for molecular dynamics (MD) simulations of biomolecular systems. The package provides utilities for preparing molecular systems, performing simulations, and analyzing the resulting trajectories. Accurate preparation of input systems is essential to ensure that MD simulations reflect biologically meaningful behavior.

Building Systems for Molecular Dynamics

Before starting any MD simulation, the system must be well defined. This involves several preparatory steps:

1. **Selection of the Molecular Model**
   Choose an appropriate protein structure, ensuring good resolution and completeness. Determine whether additional components such as ligands, metal ions, or cofactors are required.
2. **Definition of the Solvent Environment**
   Select a water model compatible with your force field, such as TIP3P or OPC. Configure the system's ionic strength to match physiological or experimental conditions.
3. **Generation of Topological Parameters**
   Use a suitable force field, such as ff19SB for proteins, to assign parameters like charges, bonds, angles, and atom types.

Together, these steps ensure the system is physically and chemically realistic before launching simulations.

System Preparation with pdbfixer, pdb4amber, and tleap

1. Initial Structure Preparation

Start with a PDB structure file. This file should be reviewed and cleaned. Common actions include removing crystallographic water molecules and confirming the presence and proper naming of ligands or cofactors.

Using pdbfixer for Protonation

pdbfixer adds missing hydrogen atoms to the structure based on the desired pH level. This is essential because AMBER requires all hydrogen atoms to be explicitly present for force field assignment.

Example:

```
pdbfixer structure.pdb --ph 7.4 --output=protonated_structure.pdb
```

Explanation:

- Adds hydrogens to standard residues at physiological pH (7.4).
- Prepares the structure for downstream parameterization and validation.

Adjustment of Protonation States for Key Residues

The protonation states of specific residues affect simulation behavior and must be explicitly defined:

- **Cysteine (CYS)**
  Typically exists in its neutral thiol form (–SH). If involved in disulfide bonds (CYS–CYS), no changes are needed. Deprotonated thiolate forms ($–S^-$) are rare and typically only occur under high pH conditions.
- **Histidine (HIS)**
  The imidazole side chain of histidine has a pKa of approximately 6.0. Its protonation depends on the local environment. AMBER defines three protonation states:
  - HID: Proton on the delta nitrogen (Nδ)
  - HIE: Proton on the epsilon nitrogen (Nε)
  - HIP: Both nitrogens protonated, resulting in a +1 charge

  At neutral pH, histidine is often neutral (usually HIE). However, protonation may need to be adjusted manually if the residue is part of a catalytic triad or involved in metal coordination or hydrogen bonding.

Using pdb4amber for Structure Correction

pdb4amber processes and sanitizes the PDB file for AMBER compatibility. It fixes non-standard residue names, removes non-ATOM entries, and checks for issues in the connectivity.

Example:

```
pdb4amber -i protonated_structure.pdb -o modified_structure.pdb
```

Optional flags can be used to rename residues, correct chain IDs, and strip additional non-standard atoms or alternate conformations.

---

2. Parameterization with tleap

tLeap is the core tool used to build AMBER input files. It generates the topology (.parm7) and coordinate (.rst7) files required to run simulations.

Example tleap Script

```
source leaprc.protein.ff19SB
source leaprc.water.opc

mol = loadpdb modified_structure.pdb

solvateoct mol OPCBOX 10.0
charge mol
addIons mol Na+ 0
check mol

saveamberparm mol system.parm7 system.rst7
quit
```

Explanation of the Script Steps:

- source leaprc.protein.ff19SB: Loads the protein force field (ff19SB).
- source leaprc.water.opc: Loads the OPC water model.
- loadpdb: Reads the cleaned and protonated PDB file into tleap.
- solvateoct: Surrounds the solute with an octahedral water box. The argument 10.0 sets a 10 angstrom buffer between the solute and the box edges.
- charge: Prints the net charge of the system, which should be neutralized.
- addIons: Randomly adds sodium and chloride ions to neutralize the system and mimic ionic strength.
- check: Verifies that all atoms and bonds are defined and no parameters are missing.
- saveamberparm: Outputs the topology and coordinate files needed for MD.

---

3. Post-processing with ParmEd

ParmEd is a tool for editing and analyzing AMBER parameter and coordinate files. It is useful for preparing PDB snapshots and for performing hydrogen mass repartitioning (HMR), which allows for longer MD time steps.

Example ParmEd Script

```
parm system.parm7
loadRestrt system.rst7

writePDB solvated_system.pdb

strip :WAT,Cl-
writeCoordinates unsolvated_system.pdb

Hmassrepartition
outparm system_parmed.parm7
quit
```

Explanation:

- Loads the system topology and coordinates.
- Saves a solvated PDB snapshot of the system.
- Strips out water and chloride ions to produce an unsolvated structure.
- Performs hydrogen mass repartitioning, which redistributes mass from heavy atoms to bonded hydrogens, allowing a 4 fs time step instead of 2 fs in MD simulations.
- Saves a new, modified topology file (system_parmed.parm7).

---

4. System Verification

Before running simulations, the final system should be visualized with software such as **UCSF Chimera**, **VMD**, or **PyMOL**. Confirm that:

- The protein and any ligands or cofactors are correctly positioned.
- Solvent and ions are distributed around the solute.
- No steric clashes or unnatural geometries are present.
- Protonation and terminal modifications are consistent with expected chemistry.

Classical Molecular Dynamics Workflow

Once the system is properly prepared, the simulation proceeds using either pmemd (optimized for speed) or sander. A complete simulation typically involves:

1. **Energy Minimization**
   Removes unfavorable contacts and relaxes the system geometry.
2. **Equilibration**
   Gradually adjusts temperature and pressure while maintaining system integrity.
3. **Production Run**
   Performs the actual MD simulation, generating trajectories for structural and dynamic analysis.

Each step requires specific input files that define simulation parameters such as temperature, timestep, restraints, and cutoff distances. These files are available in AMBER tutorials or can be customized as needed.