**Pipeline**

1. Protein-docking modeling software to get substrate-protein conformations at each key step
2. QM/MM MD simulation using amber at each of the 6 steps
3. Find free energy at each step

**List of parameters to consider for Amber**

1. Protein force field
   1. *GAFF* (general Amber force field): simpler, possibly [1]
      1. *ff14SB* for TIP3P water [2]
   2. May need to find more protein-specific force field [1]
      1. CHARMM 22 [3]
2. Choice of simulation program
   1. *Sander*: parallel processing (may not be possible with M1) [1]
      1. Might be the only way to access the QM/MM capabilities
   2. *Pmemd*: optimizes single-processor performance [1]
   3. *Nmode*: old, for <3000 atoms [1]
3. Water model
   1. Explicit water model [1]
      1. TIP3P [3]
      2. TIP4P
      3. TIP4P-Ew
      4. TIP5P
      5. SPC/E
      6. POL3
      7. OPC
         1. Shown to be more accurate for proteins [2]
   2. Implicit water model
      1. Replace discrete water molecule by “virtual water”
      2. Infinite continuum with some of the dielectric and hydrophobic properties of water [1]
      3. Many advantages of explicit model [1]
4. Boundary conditions
   1. Stochastic [3]
5. Ways to enhance conformational sampling
   1. Locally enhanced sampling (*LES*) [1]
      1. Reduces internal barrier heights
      2. Replaces portion of interest with multiple copies (mean-field approach)
      3. Prepared using *addles* module
      4. Useful for global optimization problems
         1. E.g., structure refinement, protein loop conformation optimization
      5. 3-5 copies suggested
         1. Single residues to improve protein sidechain rotamer sampling
         2. 3-4 residues to sample alternate backbone conformations
   2. Replica exchange molecular dynamics (*REMD*) [1]
      1. Simulated annealing by increasing the temperature of the system – provides more kinetic energy to overcome barriers to transitions
      2. Multiple copies are simulated at different temperatures
   3. Umbrella sampling [3]
6. QM/MM parameters
   1. Define QM and MM regions
      1. QM: substrate hydrocarbon framework, metal-pyrophosphate cluster PP-(Mg2+)3 (+ mutated side chain atoms)
      2. MM: remaining enzyme-solvent system
   2. “Treatment”
      1. DFT [3]
   3. Functional
      1. M06-2X [3]

**Amber Workflow**

1. Startup Amber environment
   1. conda activate AmberTools22
2. Prepare PDB files
   1. Choice of PDB file [4]
      1. 3p5r, chain A
      2. 60 N-terminal residues are truncated to give active protein
      3. Components
         1. C- terminal domain: S553-V862
         2. N-terminal domain: M107-I135 and S349-Q552
            1. DDIPRLSANYHGDL important in catalysis
         3. Insertion domain: S136-Y348
   2. Fix PDB file: [5]
      1. Non-standard residues
         1. This is a more-complicated fix since our protein has metal atoms and a substrate, but I’m not sure what to do about it
      2. Metals [6]
         1. Mg2+ is a divalent ion but not very charged 🡪 use Li/Merz 12-6 ions for TIP3P water
            1. *Lib file*: atomic\_ions.lib

Topologies for monoatomic ions

* + - * 1. *Frcmod file*: frcmod.ions234lm\_126\_tip3p

Li/Merz ion parameters for +2 to +4 ions in TIP3 water (12-6 normal usage)

* + 1. Experimental methods [4]
       1. No disulfide bonds
       2. 2-fluorogeranylgeranyl diphosphate (FGP) substrate bound in active site
       3. Crystallized in the presence of Mg2+
       4. 107 residues removed from N-terminus:
          1. Add back 30 residues (cleaving only 77 residues which is within the 60-79 you can cleave to remain active)

maqlsfnaal kmnalgnkai hdptncraks erqmmwvcsr sgrtrvkmsr gsggpgpvvm mssstgtskv vsetsstivd diprlsanyh gdlwhhnviq tletpfress

* + - 1. Hexahistidine tag added to C terminus
         1. Removed
    1. Solvent molecules or crystallization buffer
       1. Water molecules were removed using PyMOL
       2. Substrate (FGP) was removed
       3. No other solvents or phosphates seen in the PDB file
    2. Missing electron density
       1. Use PyMOL to fill gaps
       2. Use Modeller ModLoop to relax loops
          1. Modeller deletes the metal ions 🡪 these were added back manually using the original positions

Since only the added residues were relaxed and these are far from the active site, I assumed it was ok to keep the original Mg2+ ions

* + 1. Check for disulfide bonds
       1. None identified
    2. Check protonation states
       1. Does taxadiene synthase have any non-standard protonation states? None mentioned in the paper
       2. Histidine protonation predicted using the H++ program
          1. Standard parameters, pH 7.0
          2. Total charge: -19

1. *tleap.in, pdb.in:* Use LEaP to build protein system in explicit solvent
   1. *prmtop*: make topology
   2. *inpcrd*: make coordinate files
2. Relax the system [7]
   1. Relax the surrounding water and ions
      1. *1min.in*
         1. Minimize the added water and ions
      2. *2mdheat.in*
         1. Heat up the system from 100K to 298K over 1ns under constant volume
      3. *3md*.*in*
         1. Relax the system at constant pressure
      4. *4md.in*
         1. Lower the restraints on the system by running the MD over 1ns at constant pressure and 298K with a smaller restraint of 10 kcal/mol•Å2
   2. Relax the backbone
      1. *5min.in*
         1. Minimize the system with restraints just on the backbone of the molecule (Ca, N, C atoms)
      2. *6md.in*
         1. Relax the system over 1ns at a constant pressure with backbone restrains set to 10 kcal/mol•Å2
      3. *7md.in*
         1. Reduce the backbone restrain force by relaxing for 1ns at a constant pressure with restraints on the backbone lowered to 1 kcal/mol•Å2
      4. *8md.in*
         1. Continue to reduce the backbone restrain force by relaxing for 1ns at a constant pressure with restraints on the backbone lowered to 0.1 kcal/mol•Å2
   3. Relax the system with no restraints
      1. *9md.in*
         1. 1ns of simulation time at constant pressure

**References**

[1] D. A. Case *et al.*, "The Amber biolmolecular simulation programs," *Journal of Computational Chemistry,* vol. 26, no. 16, pp. 1668-1688, 2005, doi: 10.1002/jcc.20290.

[2] "Protein Force Fields." The Amber Project. <https://ambermd.org/AmberModels_proteins.php> (accessed May 15, 2023).

[3] T. Ansbacher, Y. Freud, and D. T. Major, "Slow-starter enzymes: Role of active-site architecture in the catalytic control of taxadiene by taxadiene synthase," *Biochemistry,* vol. 56, no. 26, pp. 3773-3779, 2018, doi: /10.1021/acs.biochem.8b00452.

[4] M. Köksal, Y. Jin, R. M. Coates, R. Croteau, and D. W. Christianson, "Taxadiene synthase structure and evolution of modular architecture in terpene biosynthesis," *Nature,* vol. 469, pp. 115-120, 2011, doi: <https://doi.org/10.1038/nature09628>.

[5] A. Held and M. Nagan. "Building protein systems in explicit solvent." <https://ambermd.org/tutorials/basic/tutorial7/index.php> (accessed May 4, 2023).

[6] "Ion Parameters." <https://ambermd.org/AmberModels_ions.php> (accessed May 10, 2023).

[7] "Relaxation of Explicit Water Systems." The Amber Project. <https://ambermd.org/tutorials/basic/tutorial13/index.php> (accessed.