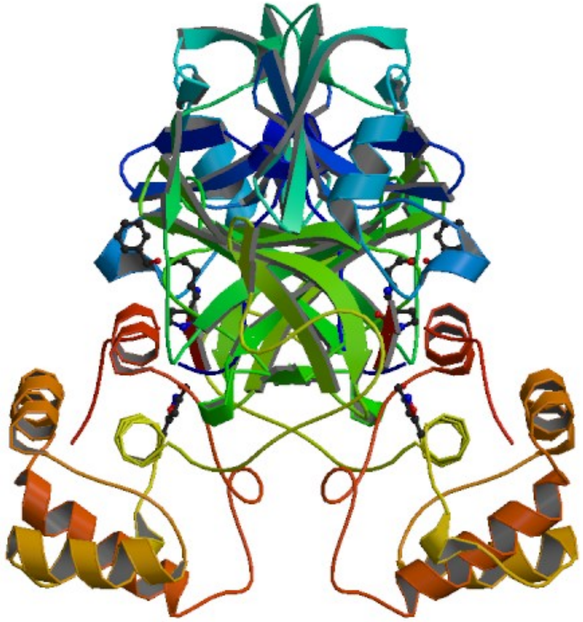
3/26/2020 Week 10 Module 1 Demonstration: Free Energy Calculations with YANK

Scientific Rationale

What is known about 3cl^{pro}?

- There is no proven cure for COVID-19
- The 3C-like protease (3cl^{pro}), also known as the main protease (M^{pro}),
  - is a key part of the life cycle of the COVID-19 virus
  - cleaves polyproteins to allow the virus to assemble
- 3cl-pro is a possible target for structure-based drug design
  - Proteases with similar function are established drug targets in other viruses, like HIV
  - A high-resolution (2.16 Å) structure in complex with an inhibitor has been solved and deposited into the protein data bank as 6LU7 [1]
  - The enzyme is inhibited by several compounds with a micromolar IC_{50} [1]



Previous modeling of 3cl^{pro}

- Molecular dynamics simulations that suggest high flexibility [2]
- Molecular docking suggests that
  - HIV protease inhibitors including lopinavir can be repurposed against the COVID-19 enzyme [3, 4]
    - there are clinical trials of using Kaletra (lopinavir/ritonivir) against COVID-19
  - Natural products may inhibit the enzyme [3]
- Binding free energy calculations
  - can be used to refine docking hits to identify the most promising repurposing candidates
  - may inform dosing recommendations in clinical trials
  - can model fluctuations of the enzyme in the presence of different known inhibitors
- To start this demonstration, I used AutoDock Vina to dock the FDA approved drugs database and known inhibitors [1] against 3cl-pro. Then I prepared the some systems for YANK.

Running binding free energy calculations with YANK

Installing YANK

- If you want to run YANK on Bridges, you need to install it in your directory
- For detailed instructions, see http://getyank.org/latest/installation.html
- In short, once you have anaconda/miniconda installed, you can type
  - conda config --add channels omnia --add channels conda-forge
  - conda create --name yank yank
  - conda update --all

Receptor setup

- The YANK workflow requests a PDB file for the receptor
- You have already done similar steps
- I prepared the receptor file by
  - submitting the structure 6LU7 to the PDB2PQR server
  - removing the ligand and water from the main result, 6LU7.pqr
    - results of these first steps are here: 3cl-pro/YANK/receptor
  - copying 6LU7.pqr to 6LU7.pdb to have a file name accepted by the YANK workflow

Parameterizing the ligand

- Why do we need ligand parameters?
- I am assuming that you docked the ligand using AutoDock Vina or AutoDock 4
- Parameters from docking programs
  - are not carefully assigned, e.g. charges have minimal dependence on environment
  - are not compatible with biomolecular force fields like AMBER

What are some ways to parameterize ligands?

- The generalized AMBER force field (GAFF) has long been the standard publicly available AMBER-compatible force field for small molecules [5]
- The Open Force Field Initiative (https://openforcefield.org)
  - is a nonprofit consortium that has partnered with the pharmaceutical industry to develop high-quality force fields that are compatible with biomolecular force fields
  - smirnoff99Frosst is comparable to the generalized AMBER force field [6]
  - The new Open Force Field 1.0 “Parsley” demonstrates improved comparison with quantum calculations and condensed phase properties
- I have demonstrations for parametrizing ligands with both GAFF and Parsley
  - In this presentation I will focus on GAFF, as the workflow is simpler
  - If you wish, your team may choose to set up your systems with Parsley

Converting AutoDock Vina results to mol2

- The docking results didn't really have a logical naming scheme. I decided to rename them based the ZINC identifier. I did this in a short jupyter notebook: 3cl-pro/ADVina/analyze/renameDocked.ipynb
- The pdbqt formatted files from AutoDock aren't readily accepted by YANK. I converted the files into mol2 format using a short jupyter notebook based on OpenBabel: 3cl-pro/YANK_GAFF/ligands/0-build/convert_ADVina_to_mol2.ipynb
- In the OpenBabel command,
  - the option `-l 1' means that the mol2 file will have one binding pose
  - the option `-p 7' means that hydrogen atoms will be added based on pH 7

Running YANK

- YANK parameters are described in YAML scripts, e.g. 3cl-pro/YANK_GAFF/yaml/MPro_ZINC000001542916.yaml
- Running `yank script -y MPro_ZINC000001542916.yaml' will execute the code, but you shouldn't do it on the login node on Bridges
- On Bridges you can enter `python submit_YANK.py --type shared --yaml MPro_ZINC000001542916.yaml', which will create a script and submit it to the queue.

Caveat: YANK is still under development and may not work

- See a github issue that I raised for problems with both pipelines
- If a system doesn't work, for the purposes of this class
  - just try another system
  - I want you to
    - get familiar with the ideas and programs of binding free energy calculations
    - not beat your head against a wall trying to solve research-level problems

References on M^{Pro}

- [1] Jin, Z.; Du, X.; Xu, Y.; Deng, Y.; Liu, M.; Zhao, Y.; Zhang, B.; Li, X.; Zhang, L.; Peng, C.; Duan, Y.; Yu, J.; Wang, L.; Yang, K.; Liu, F.; Jiang, R.; Yang, X.; You, T.; Liu, X.; Yang, X.; Bai, F.; Liu, H.; Liu, X.; Guddat, L. W.; Xu, W.; Xiao, G.; Qin, C.; Shi, Z.; Jiang, H.; Rao, Z.; Yang, H. Structure of Mpro from COVID-19 Virus and Discovery of Its Inhibitors; preprint; Biochemistry, 2020. https://doi.org/10.1101/2020.02.26.964882.
- [2] Bzówka, M.; Mitusińska, K.; Raczyńska, A.; Samol, A.; Tuszyński, J.; Góra, A. Molecular Dynamics Simulations Indicate the COVID-19 Mpro Is Not a Viable Target for Small-Molecule Inhibitors Design; preprint; Molecular Biology, 2020. https://doi.org/10.1101/2020.02.27.968008.
- [3] Khaerunnisa, S.; Kurniawan, H.; Awaluddin, R.; Suhartati, S.; Soetjipto, S. Potential Inhibitor of COVID-19 Main Protease (Mpro) From Several Medicinal Plant Compounds by Molecular Docking Study; preprint; MEDICINE & PHARMACOLOGY, 2020. https://doi.org/10.20944/preprints202003.0226.v1.
- [4] Contini, A. Virtual Screening of an FDA Approved Drugs Database on Two COVID-19 Coronavirus Proteins; preprint; 2020. https://doi.org/10.26434/chemrxiv.11847381.

References on Ligand Parameters

- [5] Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. Journal of Computational Chemistry 2004, 25 (9), 1157–1174. https://doi.org/10.1002/jcc.20035.
- [6] Mobley, D. L.; Bannan, C. C.; Rizzi, A.; Bayly, C. I.; Chodera, J. D.; Lim, V. T.; Lim, N. M.; Beauchamp, K. A.; Slochower, D. R.; Shirts, M. R.; Gilson, M. K.; Eastman, P. K. Escaping Atom Types in Force Fields Using Direct Chemical Perception. J. Chem. Theory Comput. 2018, 14 (11), 6076–6092. https://doi.org/10.1021/acs.jctc.8b00640.

Additional Resources

- Resource for alchemical binding free energy calculations (http://www.alchemistry.org/wiki/Main_Page)
- Thermodynamic cycle in YANK (http://getyank.org/latest/theory.html)

Installing “Parsley”

- Parsley parameters
  - can be applied using a Python package
  - are designed for use with OpenMM

- To install the openforcefield software and force fields
  - conda activate openmm
  - conda install -c omnia openforcefield openforcefields

- To install a program that outputs OpenMM to AMBER format for YANK
  - conda install parmed

  - conda install --yes -c conda-forge -c omnia openmmforcefields

Parsley parameterization

- After completing the installation of “Parsley”, I wrote and executed the jupyter notebook in 3cl-pro/YANK/ligands/0-build/buildSystems.ipynb
- This creates AMBER files of the ligand in solvent
  - prmtop - describes the forces between atoms
  - inpcrd - coordinates of the atoms
- The positions in the inpcrd file look like the pdbqt and sdf input files, except that there are additional hydrogen atoms

```
(openmm) Minh-IIT-MBP2018:[~/Documents/GitHub/Chem456/static_files/tutorials/3cl
-pro/YANK/ligands]: ls
ParselyParameters.ipynb ZINC000001714738.sdf    ZINC000003951740.prmtop
ZINC000001542916.inpcrd ZINC000002015152.inpcrd ZINC000003951740.sdf
ZINC000001542916.prmtop ZINC000002015152.prmtop ZINC000013985228.inpcrd
ZINC000001542916.sdf    ZINC000002015152.sdf    ZINC000013985228.prmtop
ZINC000001714738.inpcrd ZINC000003951740.inpcrd ZINC000013985228.sdf
ZINC000001714738.prmtop ZINC000003951740.mol2
```

Using Parsley inputs in YANK

- The previous script prepares the ligand and complex in solvent as AMBER prmtop and inpcrd files
- There is a different YAML input to use these prepared files in YANK: 3cl-pro/YANK/yaml/MPro_ZINC000013985228.yaml