Day 5 Period 2: Analysis of alchemical binding calculations

- This module is a tour of calculating binding free energies and poses using YANK

Analysis with YANK

- See http://getyank.org/latest/analysis.html#analysis-usage
- You can perform automatic analysis with
  - yank analyze --yaml={Some YAML file which ran with ``yank script``}

```
ccb:[~/YANK]: yank analyze --store=/home/bxie4/restraint_simulation/YANK/TEST_SYSTEMS/bromodomain/3mxf/40flat_output_rep
eat1/experiments
/home/bxie4/local/miniconda3/lib/python3.6/site-packages/openmmtools/multistate/__init__.py:75: UserWarning: Warning: op
enmmtools.multistate API is experimental
  warnings.warn('Warning: openmmtools.multistate API is experimental')
/home/bxie4/local/miniconda3/lib/python3.6/site-packages/yank/multistate/utils.py:253: FutureWarning: Using a non-tuple
sequence for multidimensional indexing is deprecated; use `arr[tuple(seq)]` instead of `arr[seq]`. In the future this wi
ll be interpreted as an array index, `arr[np.array(seq)]`, which will result either in an error or a different result.
  equilibrated_data = cast_data[slc]
Free energy            :    99.786 +- 0.005 kT (59.489 +- 0.003 kcal/mol)
DeltaG solvent         :    99.786 +- 0.005 kT

Enthalpy               :   100.720 +- 0.034 kT (60.045 +- 0.020 kcal/mol)
```

Analysis with YANK

- You can also generate a simulation report in the form of a jupyter notebook
  - yank analyze report --store={experiments} --output={mynotebook.ipynb} {--format ipynb}
  - Let's look at examples
    - https://github.com/choderalab/yank/blob/master/Yank/reports/YANK_Health_Report_Example.ipynb
    - https://github.com/daveminh/Chem456/blob/master/static_files/tutorials/bromodomains-YANK/3mxf.ipynb

Why use YANK for binding poses?

- Molecular docking scores are based on potential energy and don't include entropy
- The free energy of a pose depends on both
  - the enthalpy (proxy is potential energy) and
  - the entropy
- Molecular dynamics simulations of the bound pose are often trapped in a single minimum
- YANK is
  - likely to sample multiple bound conformations, due to Hamiltonian replica exchange
  - accounts for entropy

How to predict a binding pose

- In a molecular simulation, every structure is different
- Every structure has the same statistical weight
- To predict a pose, we need to group together similar structures: clustering
- For pose prediction, I implemented a method based on
  - aligning every protein structure according to alpha carbons
  - calculating a symmetry-corrected RMSD matrix of the ligand atoms
  - hierarchical clustering according to the symmetry-corrected RMSD
  - selecting a representative based on the medoid - the point closest to all other points - of each cluster
  - ranking poses based on the population of each cluster

Example pose prediction

- To extract the bound state from YANK results and align the protein to the first frame: https://github.com/daveminh/Chem456/blob/master/static_files/tutorials/bromodomains-YANK/getBoundState.ipynb
- To predict the ligand binding pose: https://github.com/daveminh/Chem456/blob/master/static_files/tutorials/bromodomains-YANK/ligandPosePrediction.ipynb
- Results are in poses.dcd