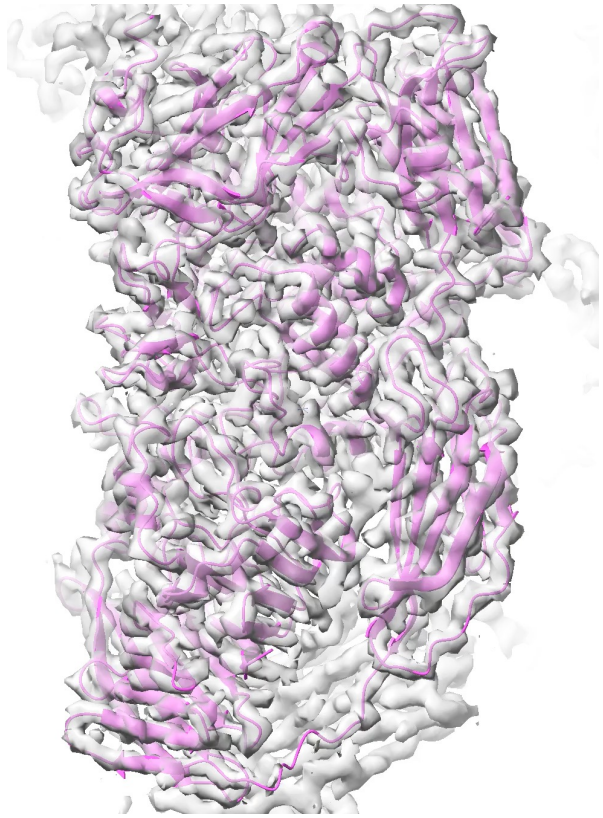
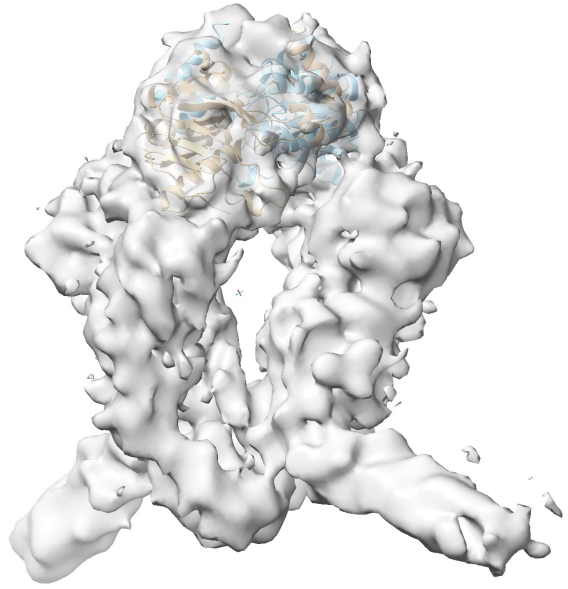
The docking problem in cryo-EM

- We have a map: how can we place an atomic model of a component in that map?
  - scoring problem
    - map correlations?
    - likelihood?
  - search problem: exploring rotations and translations
    - brute-force 6D search?
    - separate rotation and translation search?
  - decision problem
    - how confident can we be in the solution?

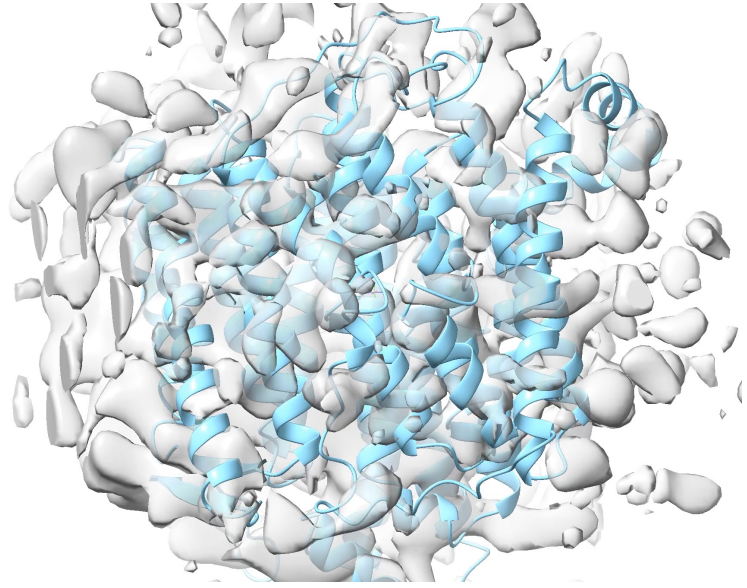
Which docking cases are important?



β-galactosidase
2.2 Å



C-terminal domain of MutS
6.9 Å



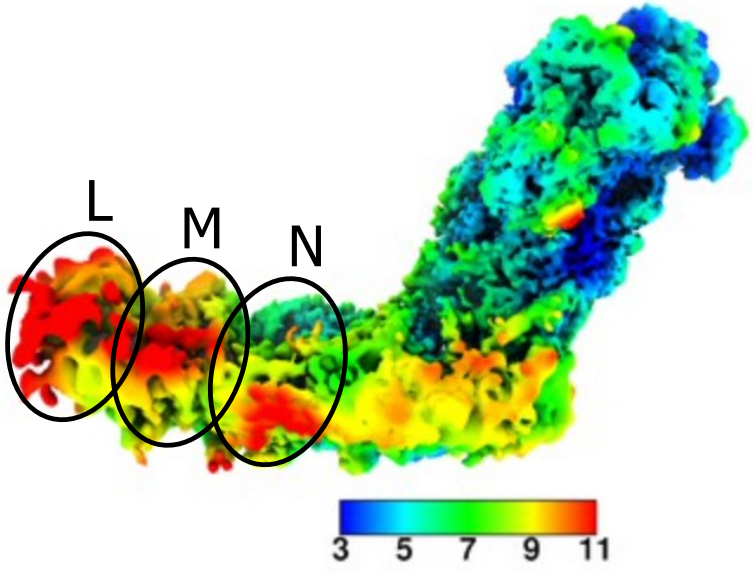
Chain L of *E. coli* complex I
3.8 - 11 Å

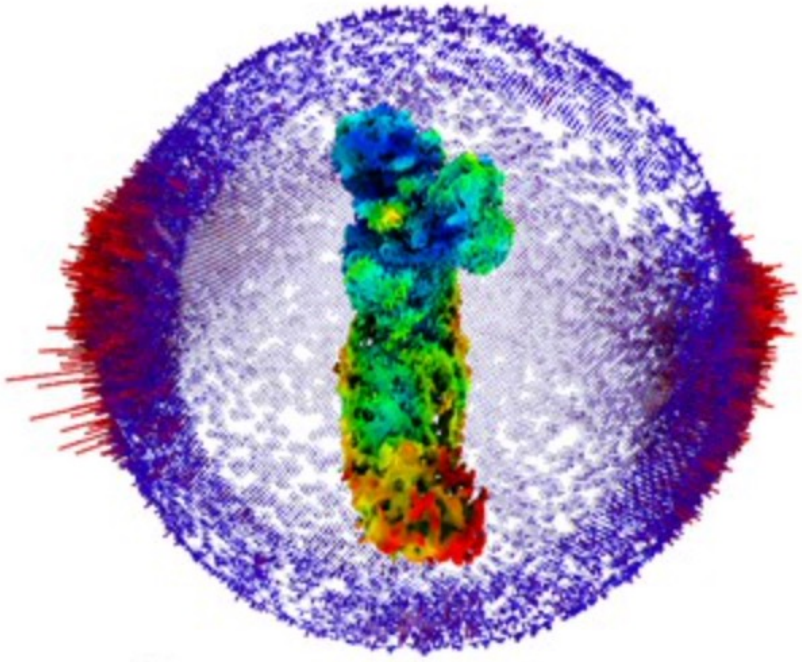
Likelihood: signal and noise in cryo-EM data

- Individual particle images are very noisy
  - use low dose to reduce radiation damage
  - average data from many particles
- Signal reduced by lack of reproducibility of the sample
  - different conformations, errors in particle orientation, radiation damage
- Signal and noise are analysed by comparing half-maps
  - this is used to calibrate the likelihood targets
  - expected signal-to-noise used to optimise search strategy
  - described in
    Read, Millán, McCoy & Terwilliger, *Structural Biology (Acta Cryst D)*

Example: EMDB 12654: PDB 7nyu

- *E. coli* respiratory complex 1 in lipid nanodisc
  - Kolata & Efremov, eLife, 2021
  - resolution ranges from 3.8 to 11 Å





orientation preference

Docking a model to a cryo-EM map

- Break 6D search into two 3D searches for efficiency, as in MR
  - rotation search: uses amplitudes of the Fourier coefficients
    - equivalent to the crystallographic rotation function
  - translation search uses phase information
- Details of strategy adapt to the quality of the data and the model, through the expected log-likelihood-gain (eLLG)

The expected log-likelihood-gain (eLLG)

- Rotation eLLG: same as crystallographic eLLG for space group P1
  - much lower than translation eLLG: rotation is the hard step!
  - rotation LLG and eLLG can be increased by putting the relevant density in a smaller box: inversely proportional to box volume
    - this does require phase information!

Overall docking strategy in *EM_placement*

- Evaluate signal and noise in entire reconstruction
  - will the rotation search probably succeed?
    - YES: run rotation search followed by translation search ← rotation eLLG
    - NO: will rotation search for minimal sub-volume succeed?
      - YES: divide map into sub-volumes, carry on as before
      - NO: do brute-force rotation and translation search ← translation eLLG
- If potential solutions are found, carry out focused docking:
  - cut out volume needed to enclose each solution, do rigid-body refinement
- Implementation and test cases (1.7-8.5Å resolution, 5-50% complete model) described in
  Millán, McCoy, Terwilliger & Read, *Structural Biology (Acta Cryst D)*

Searching in a defined sphere: *emplace_local*

- More sensitive (and much faster) if you know approximately where a molecule should go
- Easiest to run from new ChimeraX plugin
  - see YouTube tutorials by Dorothee Liebschner
    - https://www.youtube.com/c/phenixtutorials
    - Phenix/ChimeraX playlist