Segger – Probabilistic Modeling (ProMod)

Last updated: Oct. 29, 2015 (Segger v1.9.2, Chimera Version 1.11.0)

Overview

In this tutorial, we will use ProMod to build a probabilistic model derived from multiple modeling and flexible fitting results. How the models were arrived at is not the focus here. Modeling can include loop modeling (as has been covered in a previous tutorial), and flexible fitting (e.g. with MDFF, Direx, or FlexEM).

The main idea behind probabilistic models is that at medium to low resolution, the final model is not strongly constrained by the density, and many results are possible. Ways to visualize multiple results include to look at it as an ensemble, or to cluster them and find representative models. Here, we look at multiple results as defining a probabilistic function for the position of the αC in each residue.

The probability function is commonly defined using a standard (or normal, or Gaussian) distribution. Such a distribution is defined using a mean value μ, and the standard deviation around that value σ. With these two parameters specified for a given residue, we can calculate, given the priors (the atomic structure and any paramters the modeling methods used), and the observed data (the cryoEM map), the probability of finding that αC in a given area of space.

So, to calculate the probabilistic model, we need as input a number of samples, and we simply calculate the mean value and standard deviations around that mean for each αC atom.

For this tutorial, download the following file from the following link:
https://www.dropbox.com/s/msu77fq9dj78wqc/segger_tutorial_promod.zip

1. Opening the modeling results and density

- First, open the density maps of a single P22 portal protein, in the file P22_c12_portal_1p.mrc
- Then, open all the pdb models from the downloaded file.
- After varying the color, transparency, and threshold on the density file, you should see something like the image below, on the left.

2. Creating the probabilistic model

- In the Segment Dialog, open the Shortcuts panel and press the **ProMod** button to the right of 'Other tools:'.
- The method will use only atomic models that are visible in the main window, i.e. in the Model Panel the models that have a checkmark in the S (Shown) column.
- Press the **Find Average Model** button on the ProMod dialog.
    - You should see "found: portal_hub_fitted_1p_7.pdb" appear next to the button.
- Then, press the **Calculate** button in the ProMod dialog.
    - This will calculate the standard deviations of each αC atom around the position of the same αC atom in the 'average model'. The standard deviation is placed in each atom of the residue in the B-factor column.

The 10 resulting models that will be combined into a probabilistic model.

The probabilistic model. The 'average model' is shown, with color coding and thickness corresponding to deviation around the mean: blue/less thick is lower deviation, and red/thicker is higher deviation.

3. Displaying the probabilistic model

- We can visualize the probabilistic model using the 'average model', and the standard deviation at each residue.
- The average model is the model that is the closest resulting model to a fictional, *true* average model.
    - In the *true* average model, the position of each atom is the average of all the same atoms from the set of resulting models.
    - This true average model is fictional, because it does not have good structural properties – very likely the bond lengths, angles, dihedrals would be very bad, and there would likely be a lot of steric clashes.
- Start by showing only the 'average model' found as before, and hiding all the other resulting models.
    - The average model was reported to be portal_hub_fitted_1p_7.pdb.
- In Tools -> Depiction, select Render by Attribute.
- In that dialog:
    - Select Attributes of 'residues'
    - Models: portal_hub_fitted_1p_7.pdb
    - Attribute: b-factor
    - Press **Apply**
    - Then, select the 'Worms' tab
    - Press Apply again
    - You can vary the thresholds a bit under both tabs and press Apply again, which can highlight the parts with lower standard deviations as well.

4. Conclusions

In the image above, the probabilistic model is shown. Not surprisingly, the parts of the protein with the highest deviations are the flexible loop that was added using the loop modeling procedure. Also the density is not well resolved in this area, so it doesn't constrain the model very tightly. On the other hand, some parts of the density are very alpha-helical and constrain the model more tightly, leading to lower deviations. The latter could also be interpreted as there being a higher probability of finding that residue in a smaller area, given the priors and the observed data.