MapQ Tutorial

Last Updated: Aug. 25, 2019

Greg Pintilie
gregp@slac.stanford.edu

Overview

- MapQ is a UCSF Chimera plugin that allows calculation and visualization of Q-scores, which measure resolvability of atomic features in cryoEM maps using a fitted model.
- This tutorial covers the following topics:
    1. Get the software (UCSF Chimera and the MapQ plugin)
    2. Getting an example map and model
        - Feel free to use your own map and model here
    3. Use the MapQ interface to calculate and visualize Q-scores for amino acid residues and solvent atoms.
    4. Calculate and plot per-residue Q-scores
- The plugin has been tested on MacOS (10.13.16), Linux (Ubuntu 14.04), and Windows 10 with Chimera 1.12, 1.13 and 1.14.

1. Getting the Software

- Download Chimera for your system from
    - https://www.cgl.ucsf.edu/chimera/download.html
    - Either production or daily release should work.
- Download and install the MapQ plugin
    - https://github.com/gregdp/mapq
    - Follow the Install instructions at the above link.

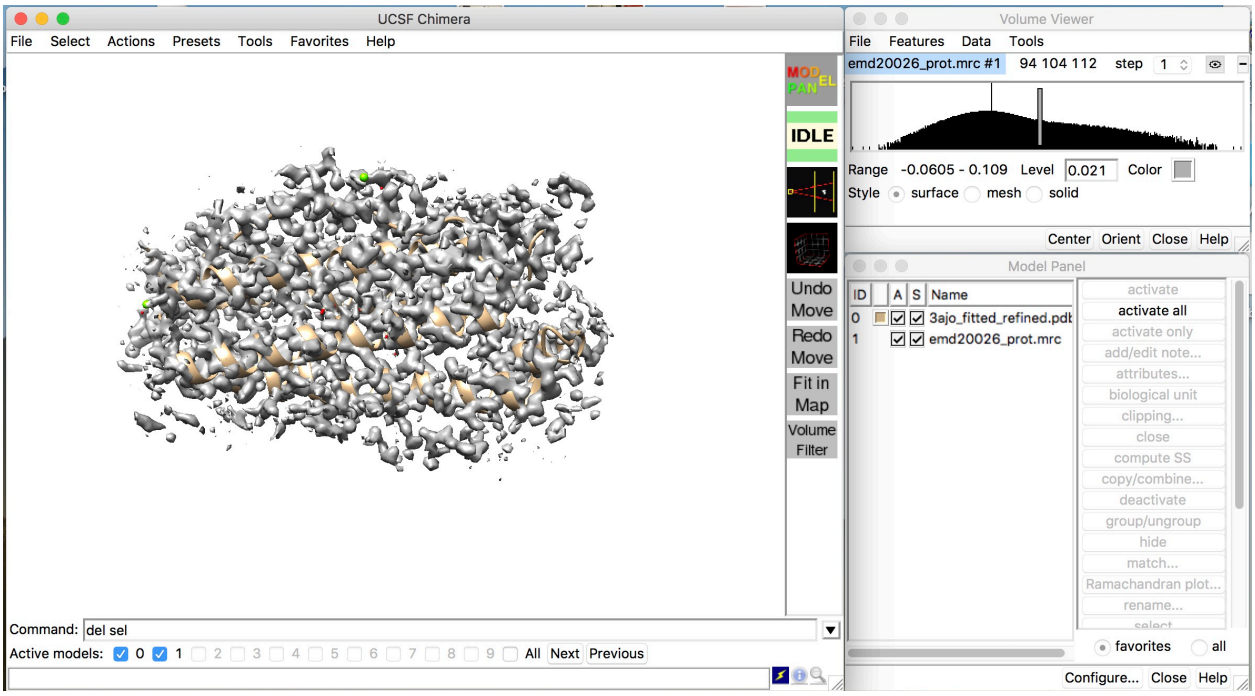
2. Loading the map and model

- Download the map and model from the following link:
    - https://github.com/gregdp/mapq/blob/master/data/QScore_Apoferritin_Tutorial.zip
    - Once downloaded, unzip the file. It should contain two files, a map and a model:
        - Emd20026_prot.mrc – this is a portion of the entire Apoferritin map representing a single protein and solvent atoms around it.
        - 3ajo_fitted_refined.pdb – this is the X-ray model PDB:3ajo which was first rigidly fitted to the cryoEM map, and then refined using Phenix.real_space_refine

- Open Chimera
- Load the map:

    - In Chimera, from File menu, select Open, navigate and select the map (emd20026_prot.mrc), click Open
- Load the model:
    - In Chimera, from File menu, select Open, navigate and select the model (3ajo_fitted_refined.pdb), click Open

- You should see something like this in the Chimera window, and two dialogs should pop up, 'Model Panel, and 'Volume Viewer'
    - If these dialogs go away, they can be opened again from the top menu:
        - Tools -> General Controls -> Model Panel
        - Tools -> Volume Data -> Volume Viewer

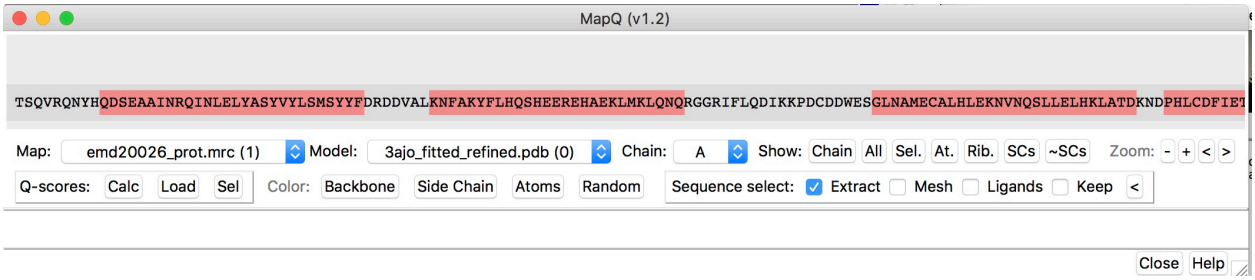


- Volume Viewer
    - This shows the Cryo-EM maps that are open (emd_20026_prot.map).
    - For efficiency, the 'step' may be "2" initially, however this shows less detail. Select 1 to see the full detail.
    - The map is shown in the main window (titled UCSF Chimera) as an 'Iso-surface', i.e. a surface that has the same map value at all the points on it. The range and histogram of the values in the map are shown in the Volume Viewer dialog.
    - Drag the bar on the histogram from left to right (with the mouse left button), and see what happens to the surface shown in the main window.
    - You can click the 'eye' icon to show/hide the map in the main window (useful when dealing with multiple maps).

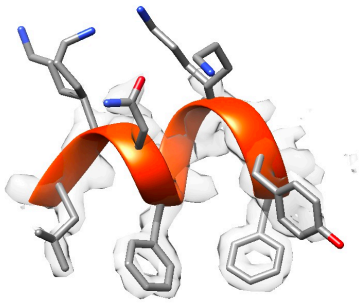
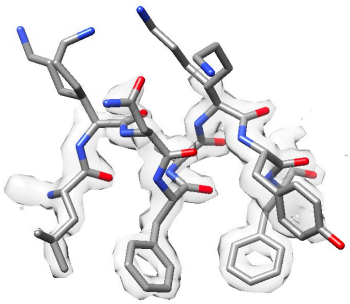
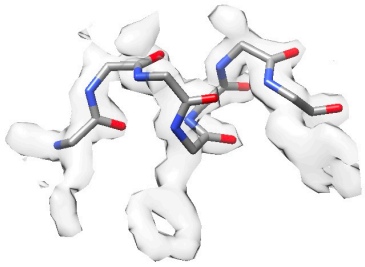
3. Using MapQ to calculate and visualize Q-scores

- For this part, we will use the 'MapQ' tools/plugin

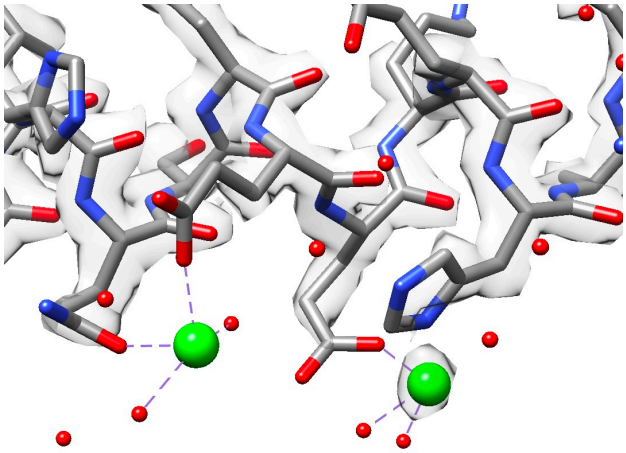
- Open ModelZ dialog:
  - Tools menu -> Volume Data -> MapQ
- The ModelZ Dialog looks like this:



- A map, model, and chain are already selected.
  - If not, or to select a new map/model/chain, click on the respective fields just under the sequence.
- The sequence for the selected chain is shown
  - The sequence is made up of amino acid residues (e.g. Tyrosine, T, Serine, S, Glutamine, Q, etc.)
- Residues that are in 'loops' are shown with gray background.
- Residues that are in 'helices' are shown with reddish background.
- Select an entire loop/helix (Ctrl+Click+Drag on the sequence).
  - This selects just that part of the protein, and,
  - The map around the selection will be extracted and shown.
- With each part selected, click '**At.**', '**Rib.**', '**SCs**', '**~SCs**' to show/switch between seeing Atoms, Ribbon, Side Chains, No-Chains. This will show the protein in different ways, as illustrated below.

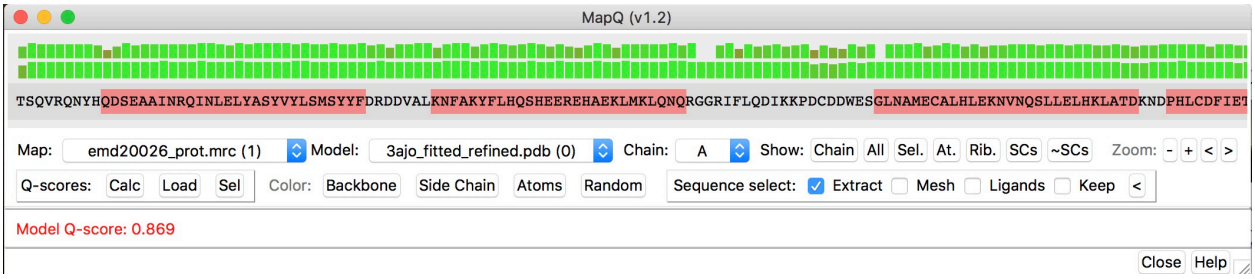
		
As Ribbon, with Side Chains	As 'Atoms', with Side Chains.	As 'Atoms', without Side Chains, i.e. just the Backbone

- Make sure '**Ligands**' is checked in the MapQ dialog (to the right of 'Sequence Select:'). Then, select the entire second helix. The selection and extracted map will now include nearby 'ligand' atoms, which in this case are water and Mg atoms as below:



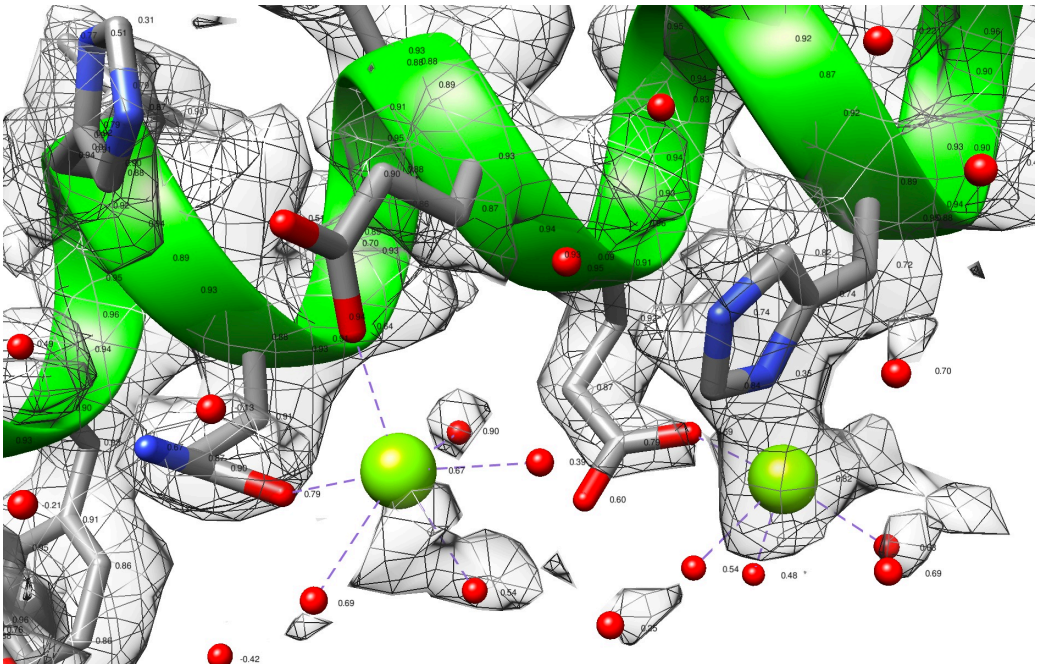
- Note how at lower thresholds you can see contours around some of the Mg and water (O) atoms. These atoms were placed in the X-ray model, and are also resolved in the cryoEM map.

- To calculate Q-scores, you can press either:
  - "**Calc**" button
    - This will calculate Q-scores for all atoms in the selected model and chain. Select 'All' in the Chain field to calculate for all atoms in the model.
    - For this protein, it will take about 4 min to calculate the Q-scores for each atom. You can see the progress and ETA at the bottom of the main Chimera window. To cancel, you can press the X button.
  - "**Load**" button
    - Once the Q-scores are calculated once, a new file will appear in the folder of the map, which is the same as the original model, but will have Q-scores stored in the B-factor column.
    - Pressing this button will look for this file and load it instead of calculating Q-scores again.
  - "**Sel**" button
    - This will calculate Q-scores only for atoms that are selected (with Ctrl+Left Click – press the up arrow to expand selection to residue, secondary structure, or entire chain).
    - It typically takes just a few seconds to calculate the Q-score for each atom, so this is faster than calculating Q-scores for an entire model.
- After calculating Q-scores using the **Calc** button (or loading them with the **Load** button), you will see the area above the sequence show a bar graph of Q-scores for each residue.
  - On the bottom are backbone Q-scores, on the top are side chain Q-scores:



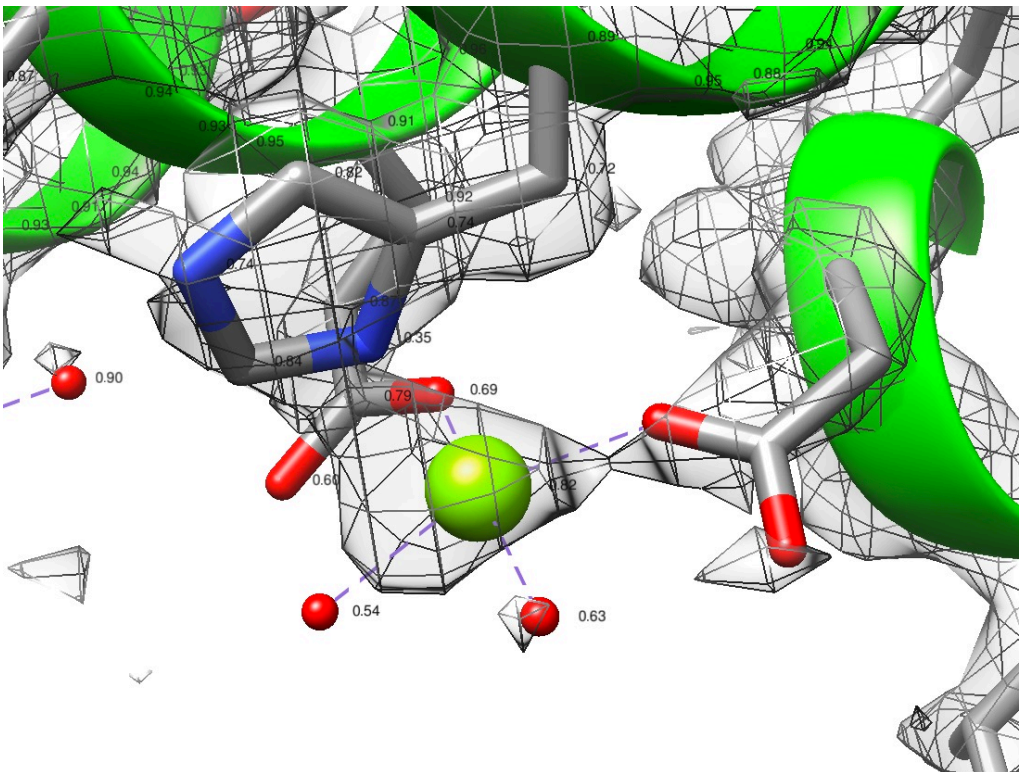
- Because this map is very good, you will see most residues in this case have good Q-scores – the bars for these are green.
    - A couple have lower (red). These tend to be residues which are more dynamic (e.g. Asp, Glu), or portions of loops.

- To visualize Q-scores on the model, you can press a button to the right of the Color: label:
    - **Backbone** – colors the model ribbon by backbone score (green is high, red is lower Q-score).
    - **Side Chain** – colors ribbon by side chain score.
    - **Atoms** – shows a label with the Q-score next to each atom, as below
    - **Random** – colors each chain a random color



- If you select the '**Keep'** checkbox, when selecting a new part of the sequence, the previous selection is also included in what is shown.
    - So for example, you can extract/show two parts of the model at the same time.
    - Press the '<' button to undo the last selection (helpful if the new selection was not what you were looking for).

- Using this mode, we can show, for example, the GLU 27 residue which is close to the Mg on the right, while keeping the previous selection:



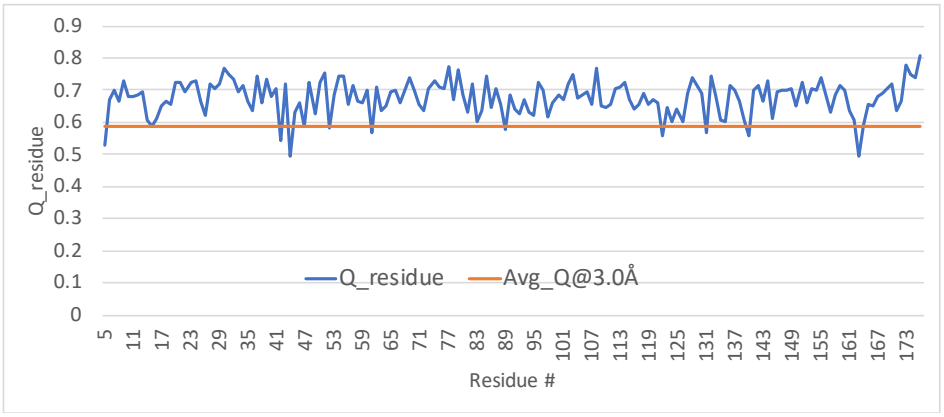
4. Per-residue Q-scores

- After Q-scores are calculated you will see two new files in the folder where your model is saved:
  - [model_name]_Q_[map_name].pdb
    - This file stores Q-scores in the occupancy column
    - Q-scores are loaded from this file when you press the **Load** button, if it exists
  - [model_name]_Q_[map_name].txt
    - This is a text file which saves per-residue Q-scores
    - You can load it in Excel or another graphing program to create per-residue Q-score plots as shown below.
    - At each residue, the average Q score over all atoms in the residue are in the Q_residue column. Average Q-scores for backbone and side chain atoms are in the Q_backBone and Q_sideChain columns.
    - There are a few columns in which Q-scores from nearby residues are averaged (with 1,2,3 and 5 nearby residues). This makes the plot a bit nicer if the Q-scores fluctuate a lot.
    - In the last column in each section, the average Q-score at a resolution of 3.0Å is shown. This column can be recalculated with the resolution of

your map using the following formulas, which are obtained from the plots shown below.

- Protein: avgQ = -0.1794 * RES + 1.1244
- RNA: avgQ = -0.1574 * RES + 1.0673

- Important notes:
  - these files are created only if you loaded your model from a file on disk; if you loaded it with Fetch or copied it from another model somehow, you will not see these files
  - the files will only have Q-scores for the chain selected; if 'All' is selected, Q-scores for all chains will be calculated and included in the files above
  - if you want Q-scores for all chains in the .txt file, make sure to select 'All' next to chains before hitting the **Calc** button. Otherwise, only the Q-scores for the selected chains will be entered.



Per-residue Q-scores example