MapQ & Q-scores Tutorial

Last updated: Aug. 10, 2022
MapQ v1.9.4 Chimera
Version 1.13+

Greg Pintilie
gregp@slac.stanford.edu

Overview

1. In this tutorial, we will use the MapQ Chimera plugin to first have a look through a map and model, to visually see how backbone and side chain residues fit the density.
2. Then, we will calculate Q-scores and visualize them on the ribbon, to spot areas that may not be resolved or fitted as well as we might expect at the given resolution.
3. Where 'expected-Q' score comes from
4. Finding areas where the model is not optimally fitted to the density
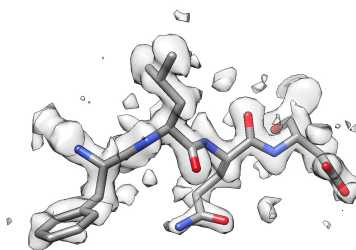
1. Install plugin, visualize map and model

- Download and install the latest version of the plugin
  - www.github.com/gregdp/mapq
  - Please not that on Mac, Chimera versions newer than 1.13 may appear to lag. In this tutorial, we use Chimera v1.13.

- Download and open the map EMD:22657, and the model, PDB:7k3v.
  - https://www.emdataresource.org/EMD-22657
  - https://www.rcsb.org/structure/7k3v

- Open the MapQ dialog:
  - From the Chimera window, Tools -> Volume Data -> MapQ.

- In the MapQ dialog, select the map and model in the Map: and Model: fields:
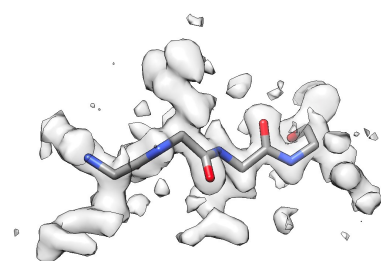 - They may already be selected by default if they were already open.
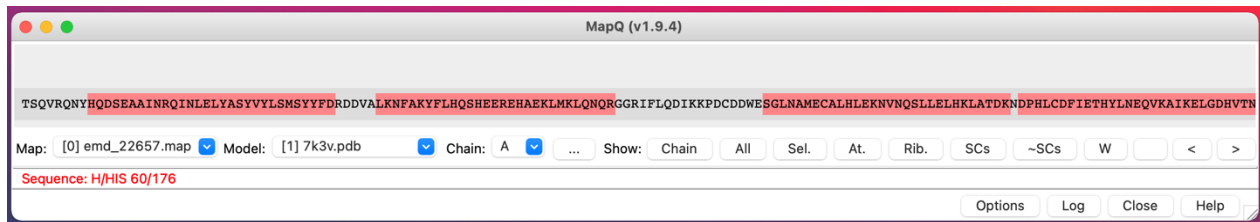
- With Chain A selected, its sequence will be shown in the dialog.
    - The sequence is made up of amino acid residues (e.g. Tyrosine, T, Serine, S, Glutamine, Q, etc.) ○ Residues that are in 'loops' are shown with gray background.
    - Residues that are in 'helices' are shown with reddish background.

- The "…" button can be used to open new maps and models using the usual Chimera file selection dialog.
    - With this option, .cif files are opened using a new parser, which should open faster than with the usual Chimera dialog. However more testing may be needed, please report if anything looks unusual…

- The buttons to the right of 'Show:' are:
    - '**Chain**' shows only the selected chain ○ '**All**' shows all chains ○ '**Sel**.' makes what is selected visible.

- 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.
    - Note that this also selects the corresponding portion of the protein in the main window (i.e. you should see a green outline around it).

- With each part selected, click '**At.**', '**Rib.**', '**SCs**', '**~SCs**', '**W**' to show/switch between seeing Atoms, Ribbon, Side Chains, No Side Chains, Wire-bonds. This will show the protein in different ways, as illustrated below.
    - The blank " " button can be used to hide anything that's currently selected.
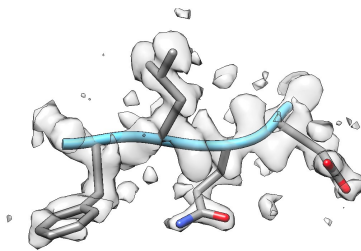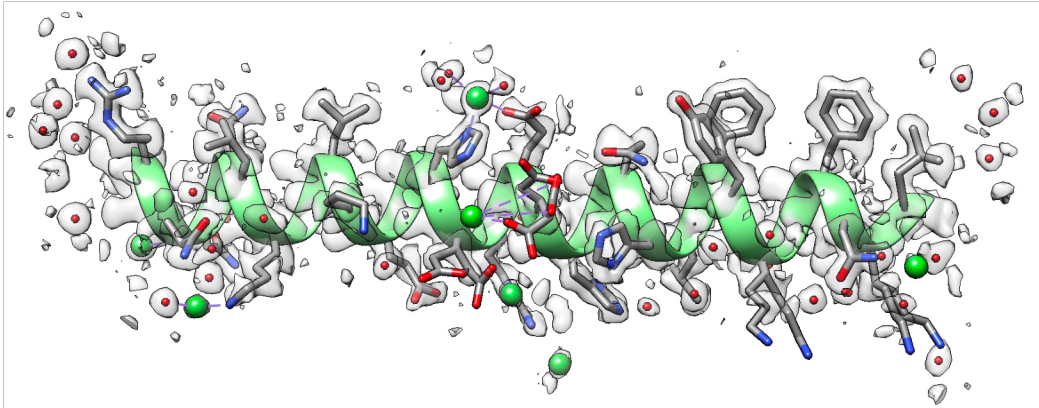

As Ribbon, with Side Chains   *As 'Atoms', with Side Chains.*   *As 'Atoms', without Side Chains, i.e. just the Backbone*

• Selecting the entire second helix:

• **Options**

    o   The "Options" button at the bottom of the dialog will show another line in the dialog:

    o

Resolution: 3   A   Sigma: 0.6   Q-scores ... ⌄   Show ...   ⌄ ✓ Gaps ✓ Extract ☐ Mesh ✓ Ligands ☐ Keep ☐ +Map ☐ H <   Redo   Near   Zoom: -   +

    o   Resolution – this value is used to compare Q-scores to expected Q-scores at a given resolution. o Sigma – by default, 0.6 is used in general, and works well for maps with resolutions of 1.5Å and lower. For higher-resolutions up to 1Å, 0.4 is used.

    o   Gaps – gaps are shown in the sequence with .

    o   Extract – when selecting part of the sequence, it will be shown by itself along with masked density

    o   Mesh – also adds a mesh when creating a masked map o Ligands – show ligands near selected sequence

    o   Keep – keeps previously shown selection, otherwise it is hidden

    o   +Map – also keeps previous masked map

    o   H – show hydrogen atoms (if SC are shown), otherwise they are hidden

    o   < - when using Keep, this reverts the last selection o Redo – Redo the last selection o Near – show residues near selection

    o   Zoom: '-' and '+' decrease the sequence text size, '<' and '>' go to the beginning and end of sequence, respectively.

2. Calculating and visualizing Q-scores

- To calculate Q-scores, press the Q-scores … button, and select one of the options:

> Calc. for selected atoms
> Calc. for all atoms in chain (single process)
> Calc. for all atoms in chain (2 processes)
> Calc. for all atoms in chain (4 processes)
> Calc. for all atoms in chain (8 processes)
> Calc. for all atoms in chain (auto # processses)
> Load (and calculate stats for selected chain)

  - Press the second option for single processor (will take about 1min), or one of the multi-processing options which will be faster.

- After calculating Q-scores using the **Calc** button (or loading them with the **Load** option), 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:
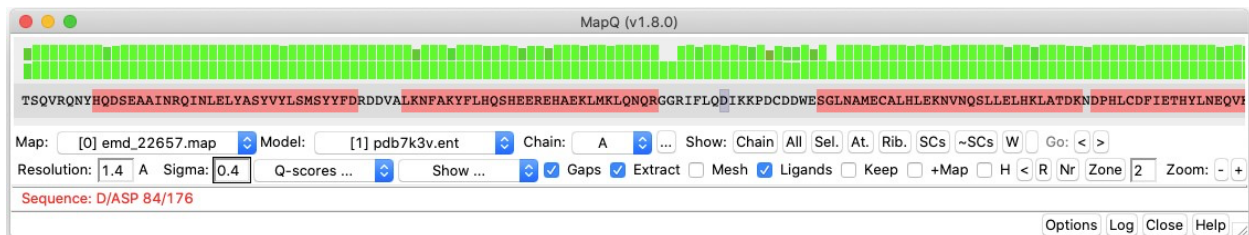
  - The resolution entered in the "Resolution:" field is used to calculate what the height of the bars should be, i.e. the expected Q-scores for a residue should be.
 - To visualize Q-scores on the model, press "Visualize…" button and select one of the options:

> Residue Q-scores
> Backbone Q-scores
> Sidechain Q-scores
> Atom Q-scores
> Random color all chains

 - **Residue** – colors ribbon by entire residue Q-score

- **Backbone** – colors the model ribbon by backbone score (green is high, red is lower Q-score). o **Sidechain** – colors ribbon by side chain score o **Atom** – shows a label with the Q-score next to each atom currently selected.

    When no atoms are selected, all previously shown labels are hidden.

- **Random color all chains** – colors each chain a random color
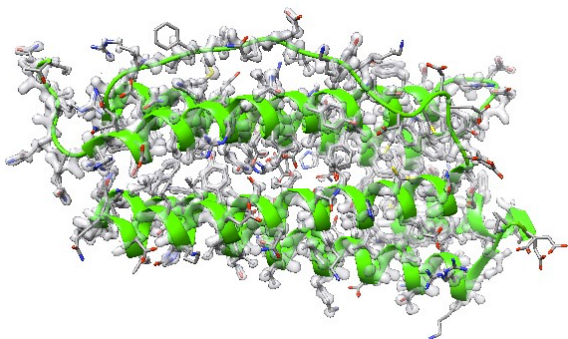
Colored backbone using Residue Q-scores (mostly green because all residues are on average well resolved)
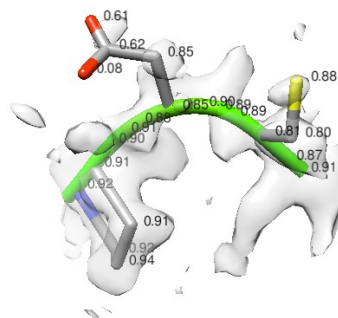

Showing atom Q-scores, displayed as labels next to a few atoms. (Select the atoms, then the Show… Atom Q-scores option.)

- 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 o [model_name]_Q_[map_name]_[chain].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 expected Q-score at a resolution of xxÅ is shown, where xx is the value entered in the Resolution:[_] field in the MapQ dialog. Where this value comes from is explained below.

- 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, Qscores 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 chain will be calculated.

Per-residue Q-scores can be plotted from the .txt file, e.g. using Excel as below.

- The Q-scores are mostly above the expected Q-score at the given resolution (1.4Å, here we used sigma 0.4 instead of the default 0.6 which should be used for 1.5Å and lower resolutions).

- Q-scores for some side chains dip below the expected Q-score line. Such residues should be inspected to see if a better fit to the map can be found, or if the map is simply not too well-resolved there (e.g. side chains may have alternate conformations, or are not resolved due to a more dynamical nature).

- Some per-residue side chain Q-scores are missing (breaks in the line) for residues like GLY which don't have side chain atoms.

Plot of per-residue Q-scores
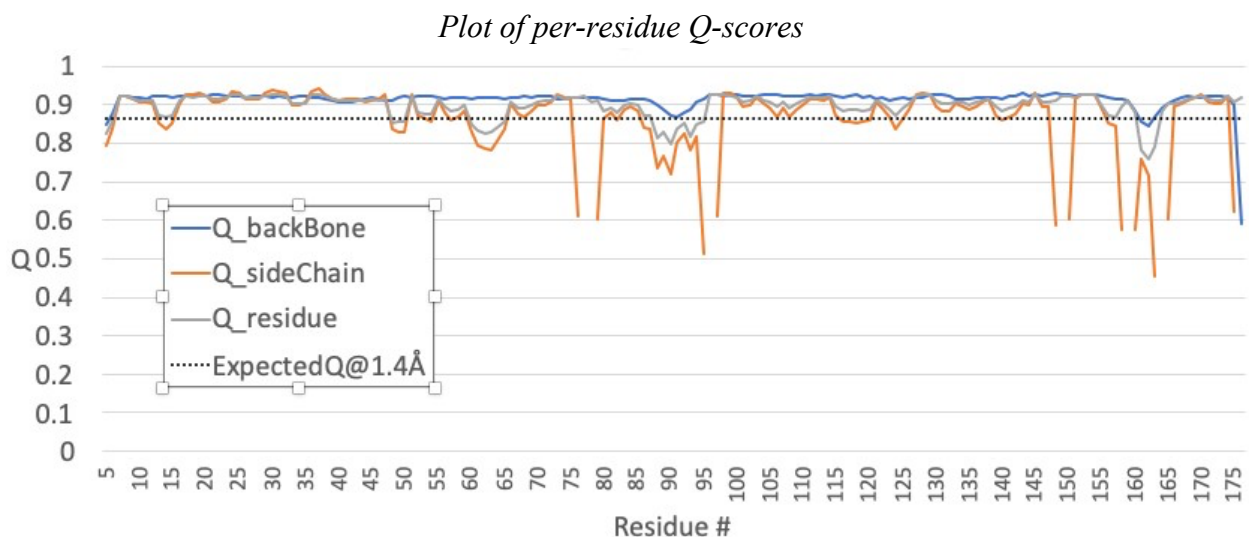
3. Where 'expected Q-score' comes from

- The expected Q-score is calculated for a given resolution (entered in the Resolution field in the MapQ dialog) based on the relationship seen between Q-scores and reported resolution for a number of maps and models in the EMDB, as shown below.
- Two sigmas are possible, and each one gives a slightly different expected Q-score; sigma=0.4 is used for maps up to resolution of 1Å, and sigma=0.6 is used for maps up to resolution 1.5Å. Either sigma is fine to use, but if your map is higher than 1.5Å, sigma=0.4 should be used to maintain the nearly-linear relationship up to 1.0Å resolution.
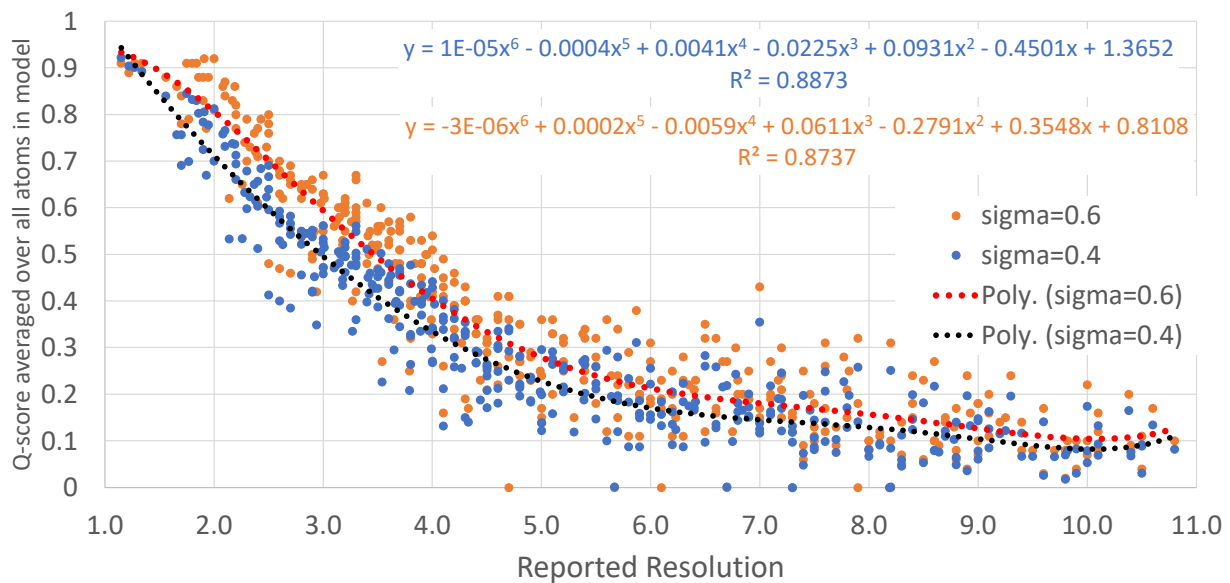
Plot of model Q-score vs reported resolution of map, for ~400 maps and models in the EMDB with resolutions evenly spread between ~1 and 11Å. From the shown fitted curves relating y and x, i.e. average Q-score and reported resolution, we can calculate what is the expected Q-score at a given resolution. Two sigmas are used, 0.4, and 0.6. For 0.4, the expected Q-score increases up to ~1.Å resolution, whereas for 0.6, it starts to taper of beyond 1.5Å as we approach 1.0Å. Hence sigma=0.4 is more appropriate to use for resolutions up to ~1Å.

4. Finding areas where the model is not fitted properly to the map

- Download the following map and model from the EMDB:
    - https://www.emdataresource.org/EMD-22657
    - https://www.rcsb.org/structure/7k3v

- Calculating and visualizing Q-scores as above, we can see a potential problem in the region colored reddish. Upon closer, look, we can see the model may not be fitted optimally to the map.