MapQ Tutorial

Last updated: Aug. 27, 2021
MapQ v1.8.0
Chimera Version 1.13+

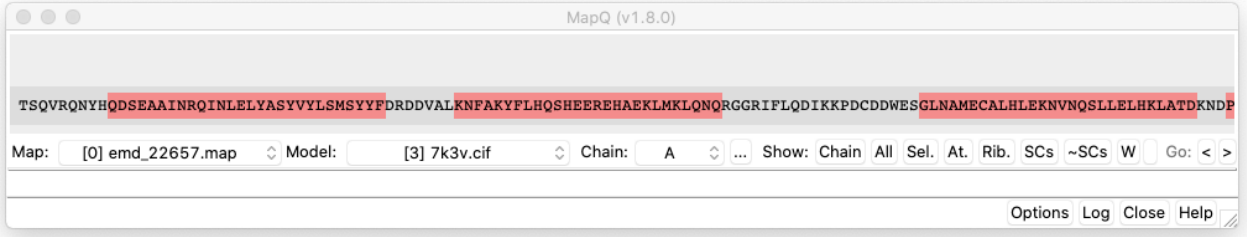
Greg Pintilie
gregp@slac.stanford.edu

Steps

1. Download and open the map EMD:22657, and the model, PDB:7k3v.

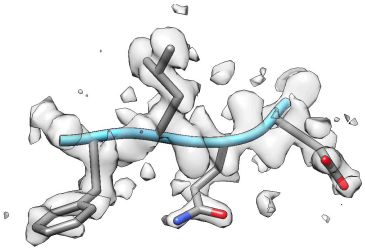
2. Open the MapQ dialog:
   - From Segger -> Options Panel -> MapQ (press the Options button at the bottom of the Segger dialog).
   - From the Chimera window, Tools -> Volume Data -> MapQ.
   - If neither of these are present there, install the latest version from www.github.com/gregdp/mapq

3. In the MapQ dialog, select the map and model in the Map: and Model: fields:

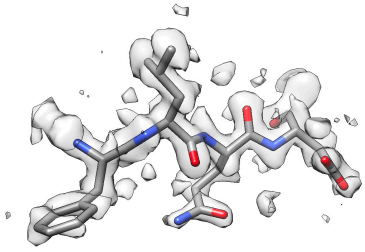


- Chain A is selected by default, and its sequence will be shown in the dialog.
  - The sequence is made up of amino acid residues (e.g. Tyrosine, T, Serine, S, Glutamine, Q, etc.)
  - Residues that are in 'loops' are shown with gray background.
  - Residues that are in 'helices' are shown with reddish background.
- The 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.

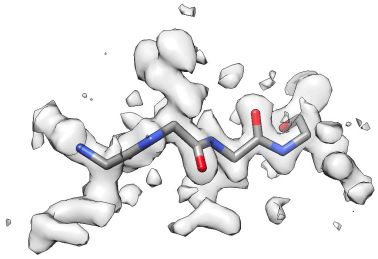
- With each part selected, click '**At.**', '**Rib.**', '**SCs**', '**~SCs**', '**W**' to show/switch between seeing Atoms, Ribbon, Side Chains, No Side Chains, Wire-bonds. This will show the protein in different ways, as illustrated below.



As Ribbon, with Side Chains

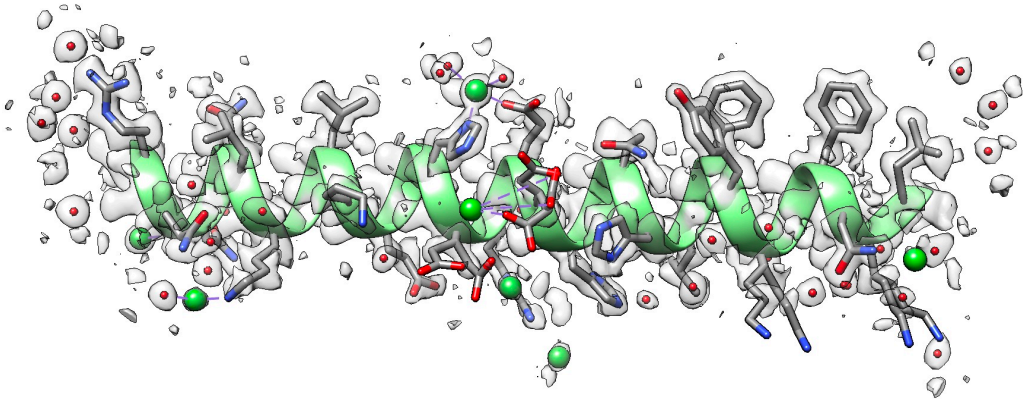


As 'Atoms', with Side Chains.



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

- Selecting the entire second helix:

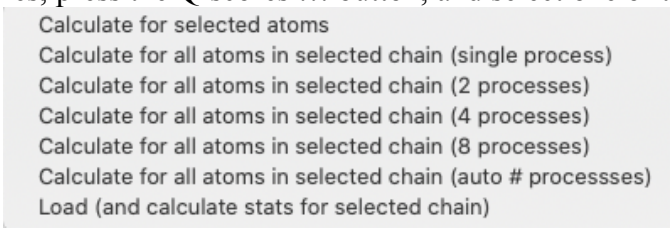


- **Options**
  - The "Options" button at the bottom of the dialog will show another line in the dialog:

  - Resolution – this value is used to compare Q-scores to expected Q-scores at a given resolution.
  - 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.
  - Gaps – gaps are shown in the sequence with .
  - Extract – when selecting part of the sequence, it will be shown by itself along with masked density
  - Mesh – also adds a mesh when creating a masked map
  - Ligands – show ligands near selected sequence
  - Keep – keeps previously shown selection, otherwise it is hidden
  - +Map – also keeps previous masked map
  - H – show hydrogen atoms (if SC are shown), otherwise they are hidden

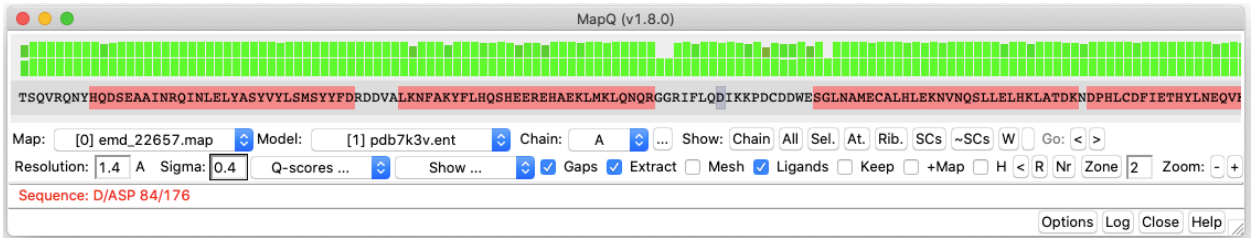
- < - when using Keep, this reverts the last selection
- R – redoes the last selection
- Nr – show residues near selection
- Zone – masks map with current selection
- Zoom: '-' and '+' decrease the sequence text size, '<' and '>' go to the beginning and end of sequence, respectively.

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

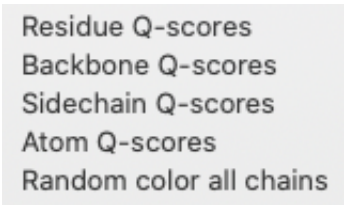


  - Press the second option for single processor (will take about 5min), or one of the multi-processing options which will be faster.

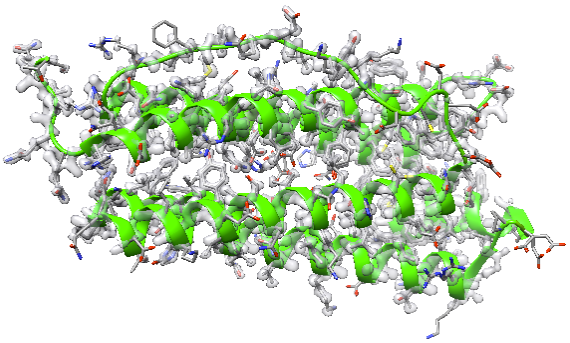
- After calculating Q-scores using the **Calc** button (or loading them with the **Load** option), you will see the area above the sequence show a bar graph of Q-scores for each residue.
  - On the bottom are backbone Q-scores, on the top are side chain Q-scores:



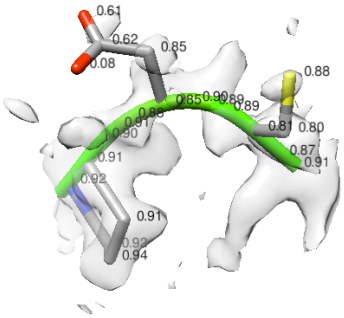
  - The resolution entered in the "Resolution:" field is used to calculate what the height of the bars should be, i.e. the expected Q-scores for a residue should be.
  - To visualize Q-scores on the model, press "Show…" button and select one of the options:



    - **Residue** – colors ribbon by entire residue Q-score
    - **Backbone** – colors the model ribbon by backbone score (green is high, red is lower Q-score).
    - **Sidechain** – colors ribbon by side chain score
    - **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.)

4. Per-residue Q-scores

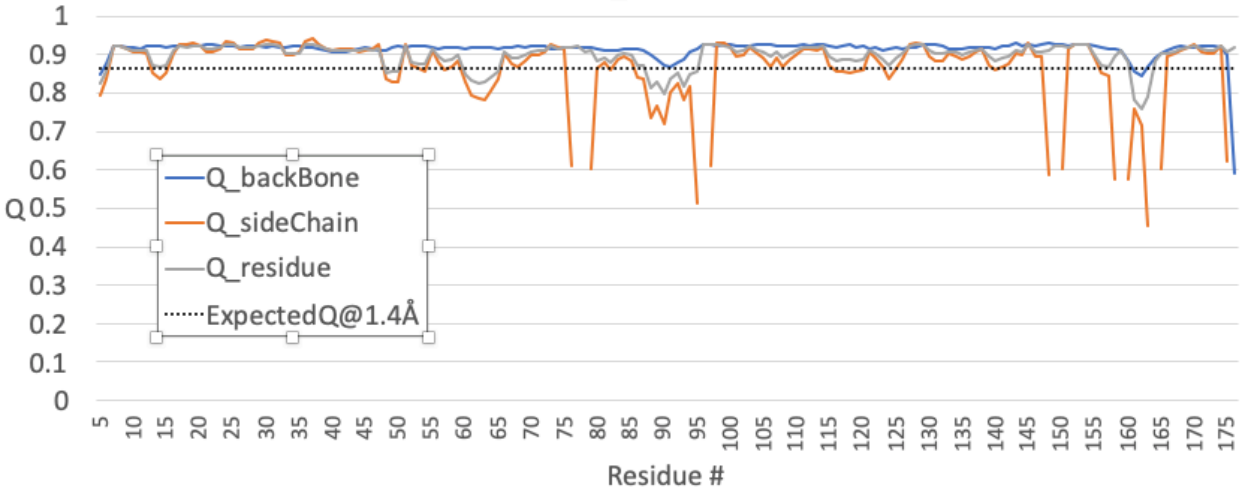
- After Q-scores are calculated you will see two new files in the folder where your model is saved:
  - [model_name]_Q_[map_name].pdb
    - This file stores Q-scores in the occupancy column
    - Q-scores are loaded from this file when you press the **Load** button, if it exists
  - [model_name]_Q_[map_name]_[chain].txt
    - 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 average Q-score at a resolution of xxÅ is shown, where xx is the value entered in the Resolution:[_] field in the MapQ dialog. This value is calculated from the equations in the plots below, where Q-scores are correlated to the reported resolutions of maps in the EMDB.
- 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 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



- The expected Q-score is calculated for a given resolution (entered in the Resolution field in the MapQ dialog) using the correlation formulas described in the following publications.
  - For maps at resolutions of 1.5Å and lower with sigma=0.6:
    - https://www.nature.com/articles/s41592-020-0731-1
  - For maps at resolutions up to 1.0Å, with sigma=0.4:
    - https://www.nature.com/articles/s41422-020-00432-2