Ab initio model building with Buccaneer

Buccaneer is a program for *ab initio* model building. It automatically builds protein chains into cryoEM density maps using sequence information alone and doesn't require a homology model. Only two inputs are required: **(1) 3D map** (MRC format) and **(2) amino acid sequence** expected of the complex (typically fasta format but others supported). Buccaneer works very well with high resolution maps <=3Å but models can still be generated from maps up to ~4Å. This tutorial will show you how to use it via the CCP-EM GUI.

In this example we will use the *E. coli* β-galactosidase map from the RELION 3.0 tutorial. To save time we have segmented the map in Chimera such that only one chain is present. You will use:

Map:	run_class001_flip_seg_prep.mrc (2.9Å resolution)
Sequence:	5a1a.fasta.txt

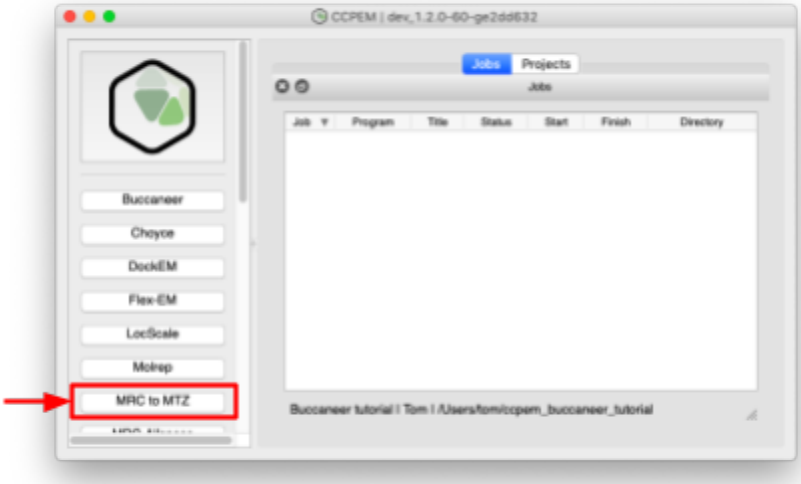
You can find these in the Buccaneer tutorial data.

Part 1) Optimise global map sharpening: MRC to MTZ

Before running Buccaneer, it's recommended to optimise the map sharpening. During the reconstruction process it is not possible to determine the absolute sharpening value of the map, in particular if your target is a multidomain protein. Thus it is recommended to apply an array of blurring and sharpening values and visually inspect your map prior to any model building.

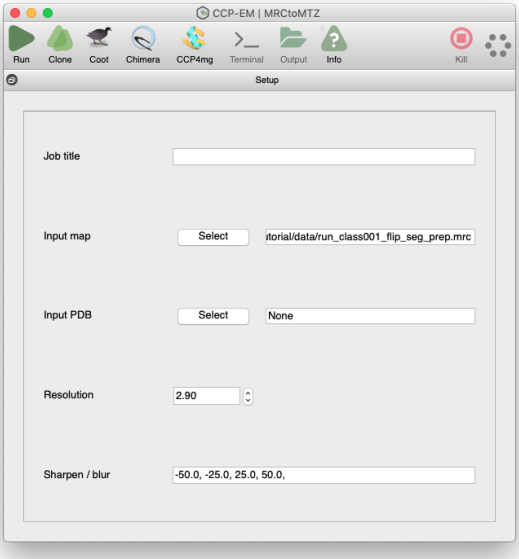
1) Open ccpem and create a new project if you have not done so already.

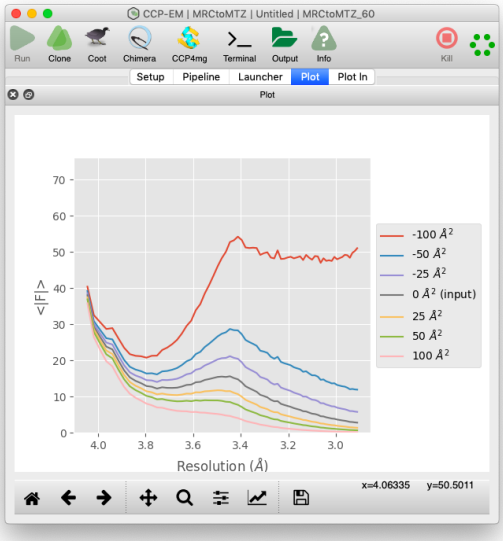
2) Click the 'MRC to MTZ' button to launch a new task.



2) Enter a job title, the input map and resolution and set the Sharpen / blur values (try -100, -50, -25, 25, 50, 100). N.B. negative values sharpen the map, positive values blur it (hover the mouse over the input box to see the tooltip).

You may also find it useful to add a job title. Then press “Run”.





3) Map inspection: power spectrum

When the job is finished the power spectra for all of the blurred and sharpened maps are displayed in the ‘Plot’ and ‘Plot ln’ tabs.

What is the effect of blurring and sharpening? What do you expect the power spectrum to look like?

We will discuss this in more detail during the workshop, but we expect to see a drop off towards higher resolution (on the right). We also expect to see a bump - this is due to commonly recurring feature spacings, e.g. between neighbouring chains, or turns of a helix. Subjectively, the -25 Å² and 0 Å² lines look good to me.

4) Map inspection: visual in Coot

Press the **‘Coot’** button in the toolbar. This opens all of the sharpened and blurred maps, but in this workshop we will hide those and use interactive sharpening in Coot.

Click ‘Display Manager’ in the toolbar to toggle the display of the different maps.

Delete all the maps except `'starting_map.mtz Fout0 Pout0'` (this is the original unsharpened map). Make sure this map is also selected for Scroll. You can then use the mouse scroll wheel or “+” and “-” keys to set the map contour level.

When there is no model, Coot's initial viewing position is at the corner of the map box where there is usually no density visible (unless you drop to a very low contour level, when you will see background noise with no obvious structure). To find the density for the molecule, you will need to move around the map. To move without an atomic model loaded, hold **'ctrl'**, **hold left click and drag**.

You can move around manually to find some of the map density, or you can find it by going directly to the middle of the box using Coot's Cryo-EM module.

Calculate → Modules → Cryo-EM (this creates a new Cryo-EM menu)
Cryo-EM → Go to box middle

Once you have found some map density, set the contour level so you can see the protein backbone clearly. Then, you can explore the effect of map sharpening and blurring interactively using the following tool:

Calculate → Map Sharpening/Blurring → Move the slider

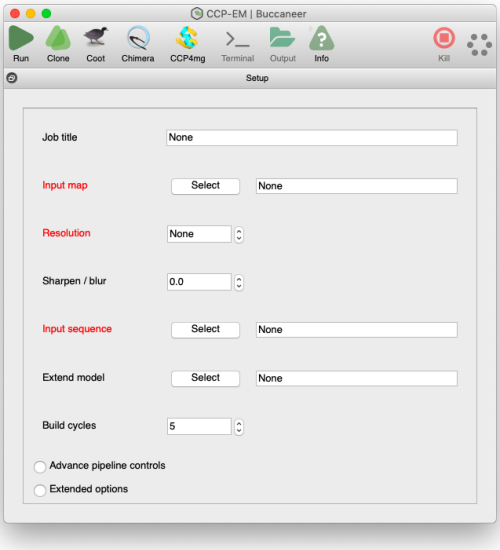
What is the effect of blurring and sharpening? You might also want to adjust the contour level after changing the sharpening to get a feel for the features that can be seen in the map. Which value gives the most interpretable map and what features are desirable?

(Note: Coot's interactive sharpening tool only works for maps in MTZ format. Coot also oversamples MTZ maps so they appear smoother. If you open an MRC format map, interactive sharpening will not work, and the map will look less smooth.)

Part 2) *Ab initio* model building: Buccaneer

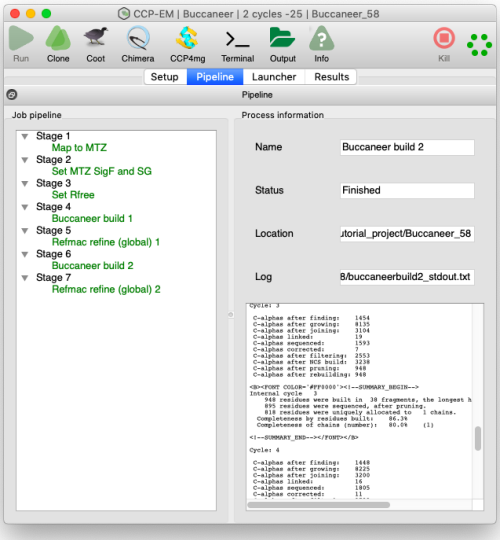
1) From the main GUI launch the Buccaneer task window.

2) Enter the following parameters and hit run:
   Input map: run_class001_flip_seg_prep.mrc
   Resolution: 2.9
   Input sequence: 5a1a.fasta.txt
   Sharpen / blur: <value decided above>
   Build cycles: 2
   Advanced pipeline controls -> 1st Buccaneer cycles: 2
   Advanced pipeline controls -> Nth Buccaneer cycles: 2
   Advanced pipeline controls -> CPU: 2



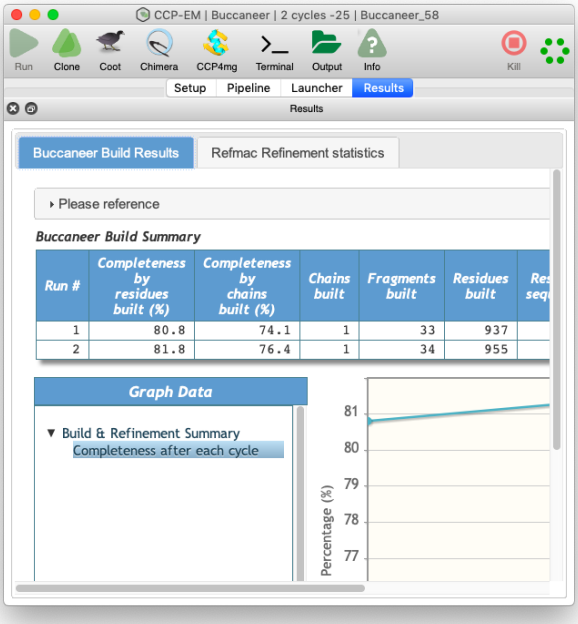
5 build cycles is the default value, 2 is used here to save time. We have also reduced the number of internal iterations to 2. The calculation takes ~30 mins (on MacBook Air) for this dataset.
If you are unsure which sharpen / blur value to use try -25. (If you have a lot of CPUs use the 'Clone' button to generate another task with the same parameters. You can then run multiple jobs in parallel to try different sharpening or resolution values.)
For reference, the original sequence file can be found from the downloads link in PDB entry: https://www.ebi.ac.uk/pdbe/entry/pdb/5a1a Chains B,C,D were manually removed in a text editor.

3) Monitoring job progress in 'Pipeline' tab



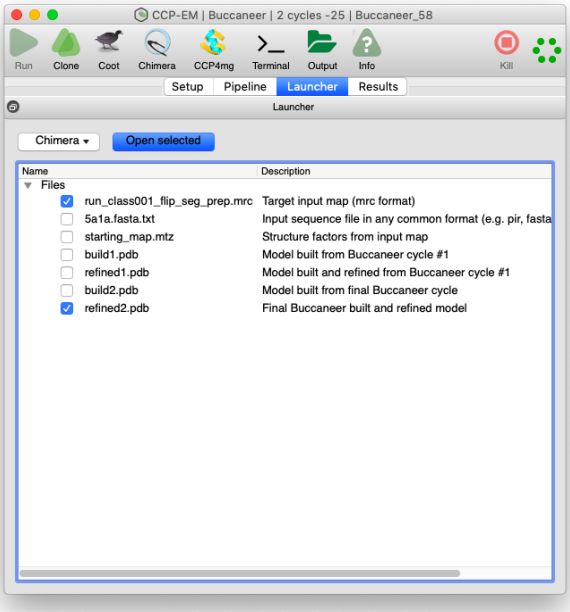
Whilst the job is running the green hexagon in the top right will rotate.
You can see the progression through the various stages in the pipeline on the left.
Green is finished, blue running, grey yet to start, red failed.
Clicking on the stage will display the log file for that stage. Double clicking on the log file window will launch your preferred text editor with the log file.

4) Results inspection. Click the 'Results' and the 'Launcher' to see the output from Buccaneer.



When the job is complete (hexagon solid green) the 'Results' tab shows the number of residues built.
You can see the completeness of the model in the 'Buccaneer build results' tab and the 'Refmac refinement statistics' tab shows an overview of the refinement statistics after initial automatic refinement in Refmac.
'Completeness by residue' is the proportion of residues built which were docked into the sequence rather than being left unsequenced. Buccaneer tends to be optimistic in its sequence docking, so this number is often overestimated.
'Completeness by chain' is the proportion of the expected sequence which was built. This does however depend on Buccaneer having correctly guessed the number of chains present in the map - if the number of chains is correct the completeness by chain is the more accurate measure of success.

The 'Launcher' tab allows you to select which files you are interested in viewing for visual inspection.
Use the post refinement models, i.e. 'refined<cycle>.pdb'.
N.B. Model building is stochastic so you may find models from intermediate steps are better than the final model if you run multiple cycles.



If you ran multiple cycles, which build cycle worked best for you? If you ran parallel runs with different input parameters (e.g. sharpening, resolution cutoff, etc) what effect did these have?

`5) Validating the results in Coot. Select the map and coordinates in the 'Launcher' tab and press the 'Coot' button.`

In Coot you use the 'Display Manager' to select the maps and model. Use the sequence viewer (Draw → Sequence View) to see the sequence of the built fragments. Clicking on the residue will orientate the view.

Do the fragments look correct? Can you see any errors in the structure, e.g. sequencing errors or register errors, and if so can you manually correct them in Coot? Are there incorrectly built sequences that should be removed? At the end of the fragments, can you manually build additional residues? If you have time you can 'walk' through the structure by pressing space and this moves the focus by one residue at a time. You can then optimise each residue's fit to density and stereochemistry (see Coot tutorial for further details).

Cheat hint: load the additional file 5a1a_chaina_molrep.pdb to show a complete model.
This was generated by docking chain A from the PDB entry into the map via Molrep, so it's not a perfect fit but shows a fully sequenced model which you can compare with the ab initio model generated by Buccaneer.

When you are ready you can save the edited coordinates (File → Save Coordinates) and run Refmac (see Refmac tutorial).

Buccaneer references:
Cowtan, K. The Buccaneer software for automated model building. *Acta Cryst* D 62, 1002-101, 2006
Cowtan, K. Fitting molecular fragments into electron density. *Acta Cryst* D64, 83-89, 2008.

CCP-EM reference:
Burnley, T., Palmer, C.M. & Winn, M. Recent developments in the CCP-EM software suite. *Acta Cryst* D73, 469-47, 2017.

Contact:
Do please report any issues or bugs. It's much appreciated and helps us make the software better: **ccpem@stfc.ac.uk**