Secondary Structure Element annotation with Haruspex

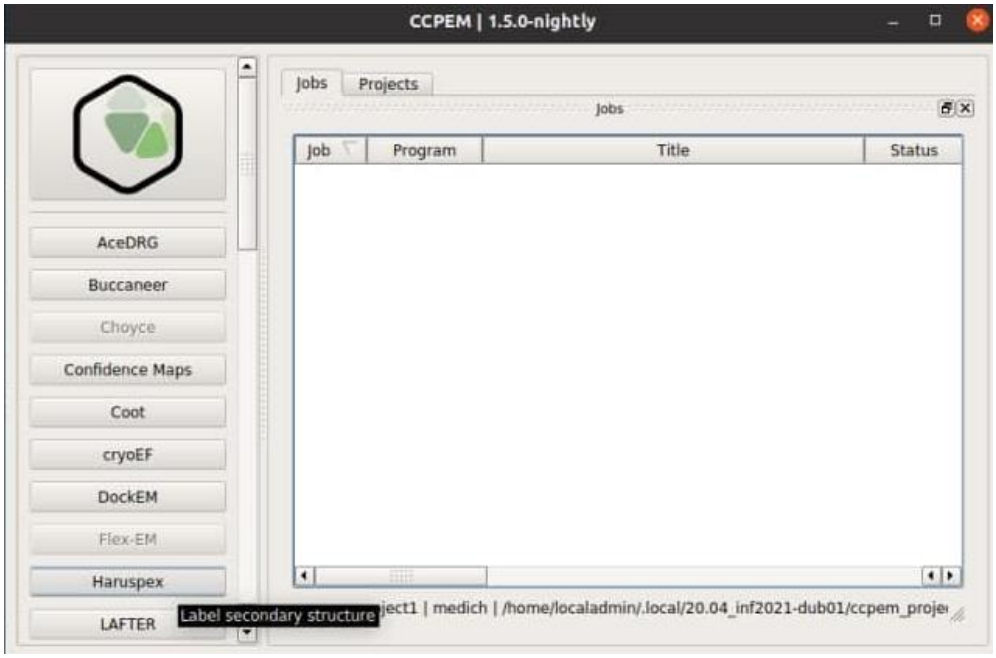
Haruspex is a programm used to automatically annotate densitiy maps from Cryo-EM via convolutional neural networks (CNNs). It was trained on a large set of data of proteins to identify α-helices, β-sheets, and oligonucleotides like DNA or RNA. Haruspex can be used prior model building or for validation of constructed models afterwards.

Haruspex is part of the CCP-EM suite. However, it requires some additional steps, which are covered in the workshop. Check out the additional handout on required installations.
The official Haruspex installation from CCP-EM is found here:
https://www.ccpem.ac.uk/user_help/haruspex_instructions.php

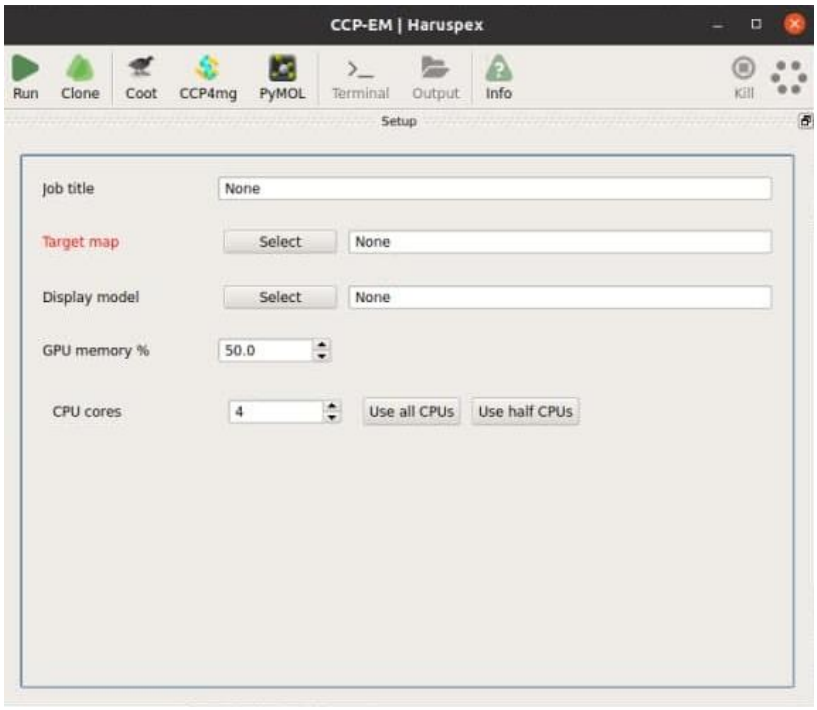
Example of running Haruspex on haemoglobin

Get the files of the map and the structure either from the workshop folder or download them directly from the data bases:
https://www.ebi.ac.uk/emdb/entry/3488
https://www.rcsb.org/structure/5NI1

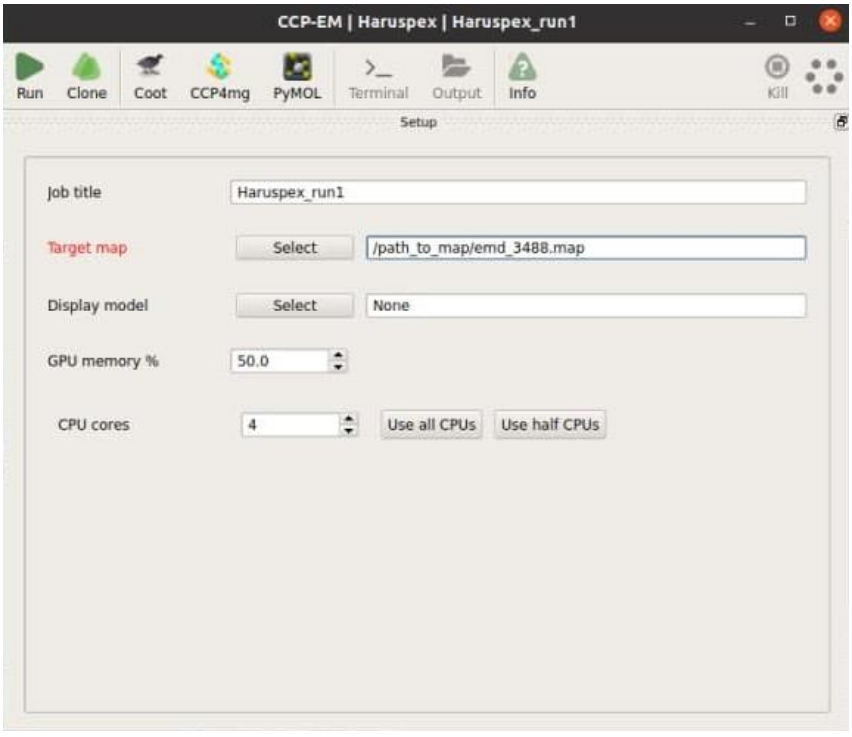
Move the files to a directory of your choice. Afterwards, launch CCP-EM, which will lead you to the following screen. Remember, to source the `setup_ccpem.sh` before you launch the GUI via `ccpem`. You may get some small windows with warnings during the start of ccpem, which occur due to missing software. For this tutorial you should be able to ignore those without affecting the following steps.



If you installed Haruspex correctly, the button “Haruspex” should appear in the list of software tools on the left. Click on it.

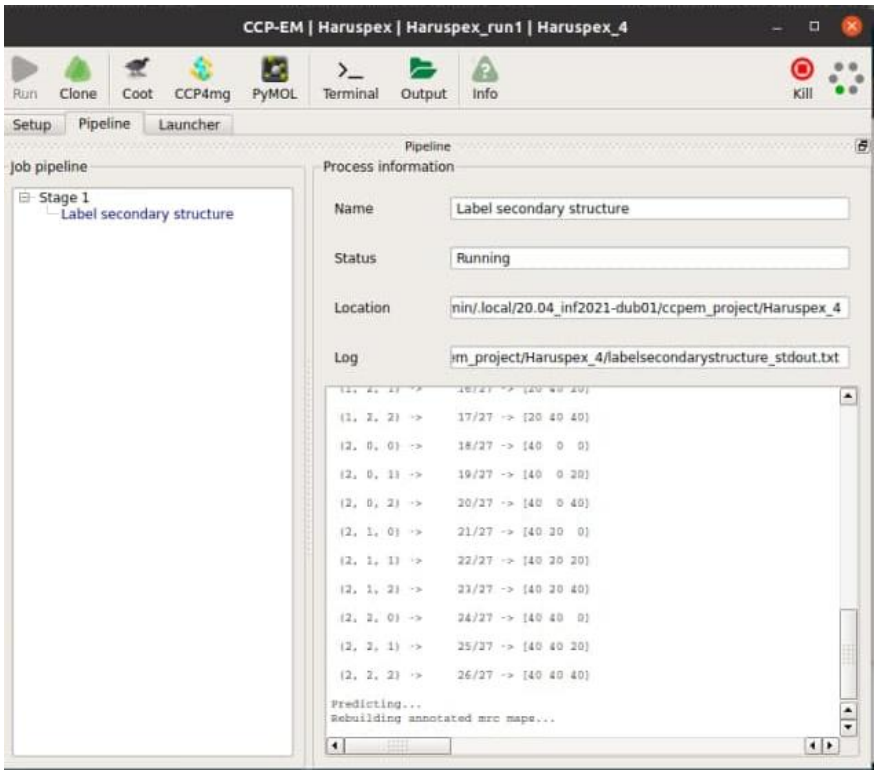


This opens the Haruspex window. The top bar shows the “run” and “clone” button, as well as a collection of additional tools you can use to view the results and some additional buttons to view the terminal or the output folder. The right displays a “kill” button to terminate a running process and a circle indication the state of the current run.
In the center you can enter the input for Haruspex. You can name the job, specify the target map and a structure model, and allocate the GPU and CPU memory to run this job. Only the target map is required, all other fields are optional.

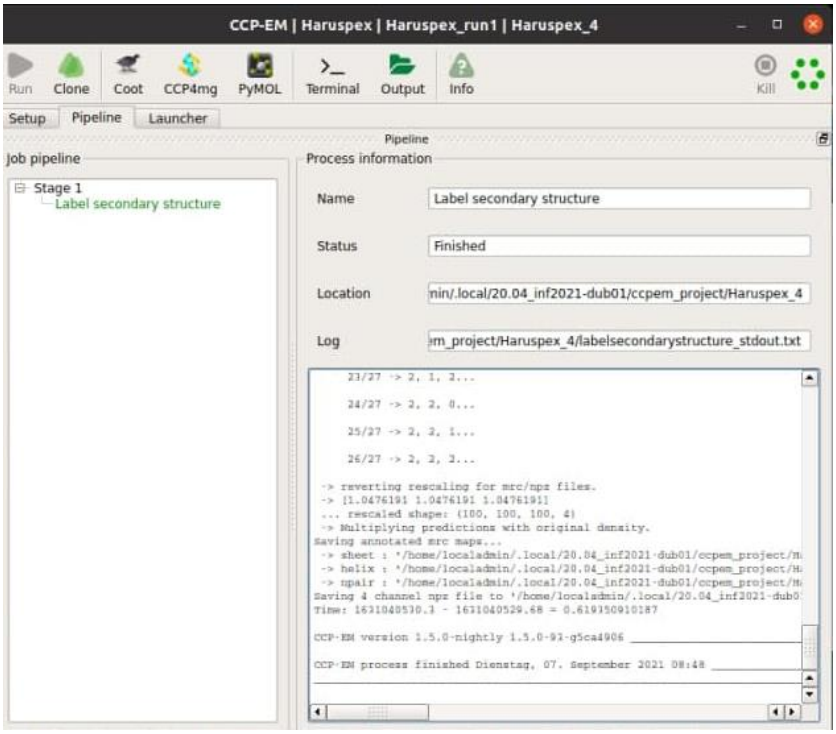


Give the job a title and use the “Select” buttonto choose a file or type in the path to a density map from electron microscopy by yourself. You can also specify the path to a .pdb file related to this map,

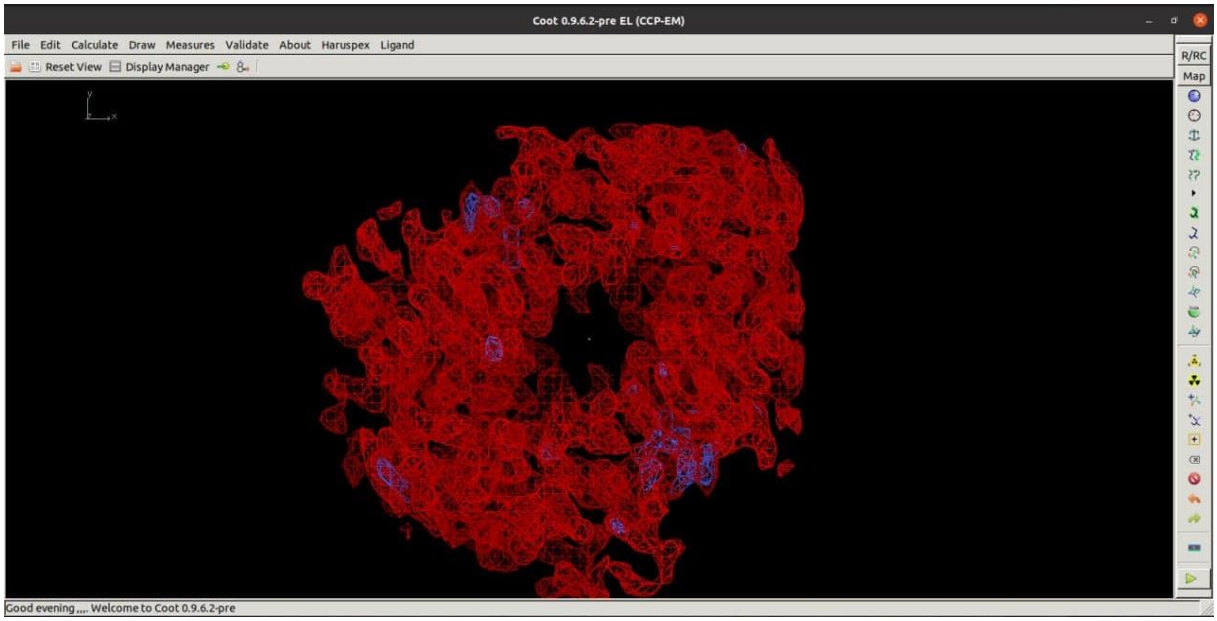
but this is optional. Specifying a structure has no influnce on the annotaions made by haruspex. It is for display purpose only. Before you launch the job, check the memory specifications. It is recommended to use not the maximum capacity, since this could crash your computer. If your computer is not too old, Haruspex will require only a few minutes. The required time depends also on the size of the map you want to annotate. After setting up Haruspex you can click on “Run” at the top left corner.



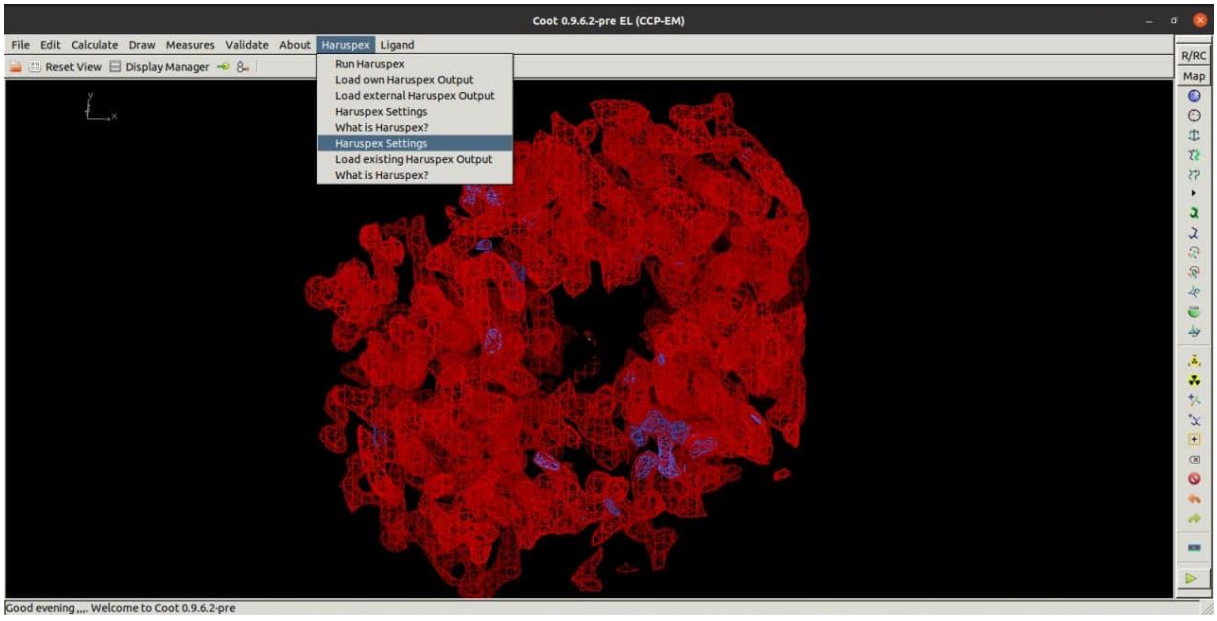
Your window should now show the Pipeline tab instead of the Setup. The green circles at the top right corner indicate that the process is running. You get also feedback in the included terminal. Wait until the process has finished.



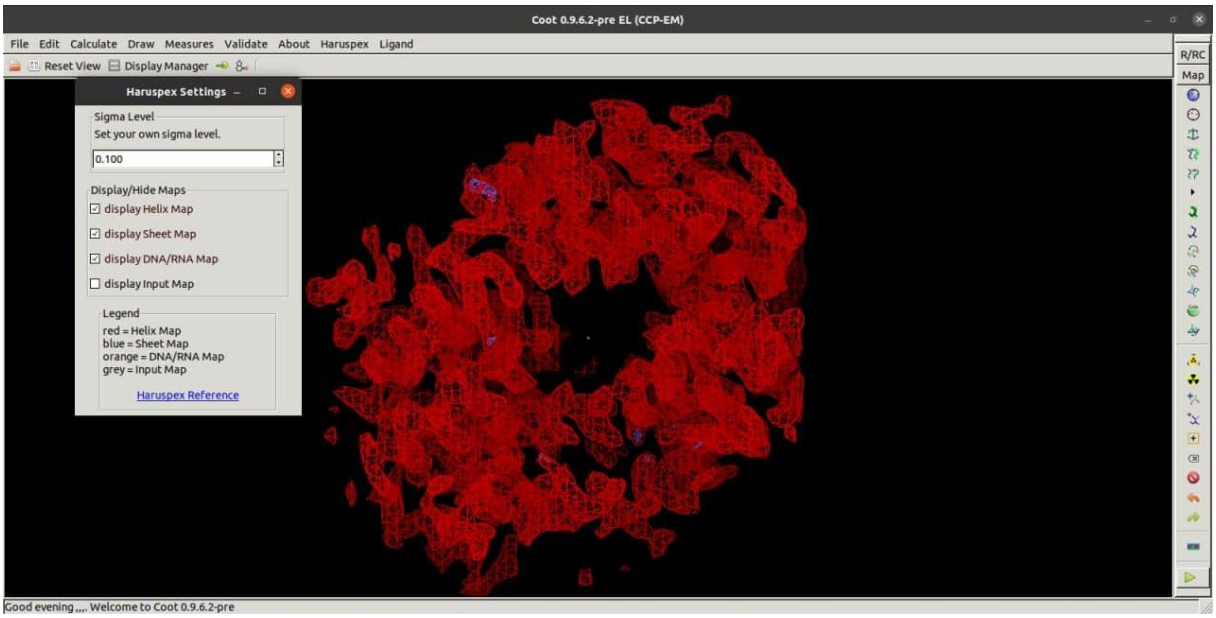
The process has finished, if the ring of circles is complety green and the bar next to “Status” shows “Finished”. Now you can click one of the tools in the top bar. Note, that depending on your installation you may have a different display here. For the next steps we are clicking the button of Coot. If this button is not displayed, you have to make sure, that Coot is installed und functional. Clicking the button will launch Coot.



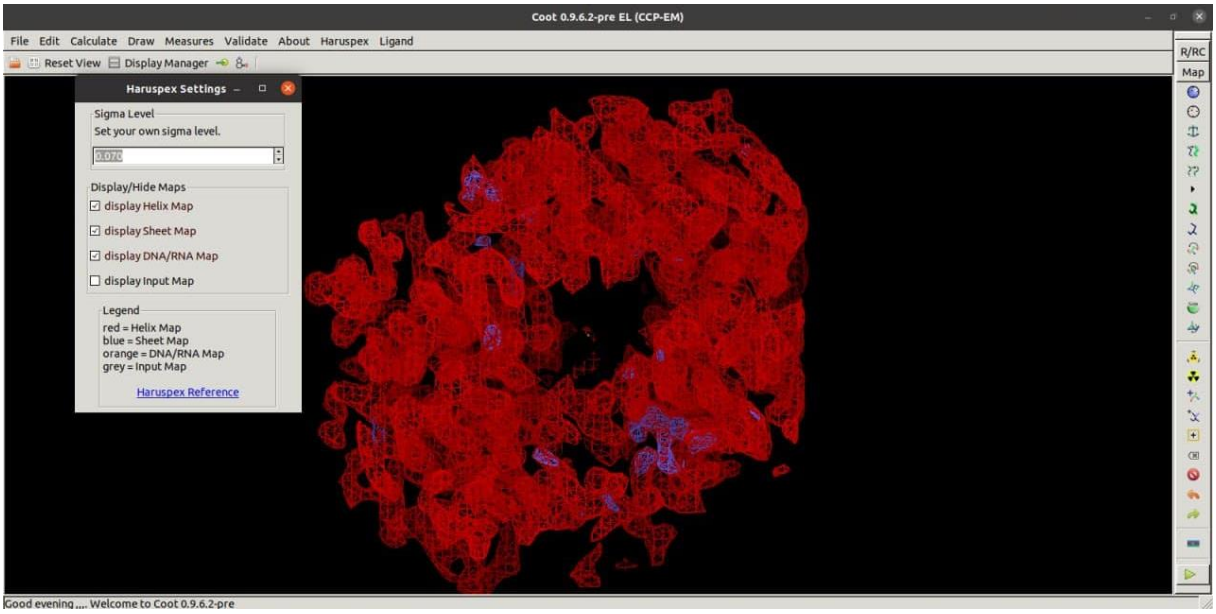
The Coot window displays the anotated density map. Red regions represent regions annotated as α-helices, blue stands for β-sheets, and orange shows map regions annoted with DNA/RNA.



You can modify the visualization by clicking on Haruspex at the top and selecting “Haruspex Settings”. This will pop up a small window.



Here you can toggle the displays of the annotation calsses and set your own sigma level. You can also toggle the display of the input map.



Adjusting the sigma level modifies the display of density by adjusting a cutoff value. This can help to find certain annotations which cover only small regions. This does not affect the underlying data.

Haruspex references:

Mostosi, Philipp, et al. "Haruspex: a neural network for the automatic identification of oligonucleotides and protein secondary structure in cryo-electron microscopy maps." Angewandte Chemie 132.35 (2020)

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