ModelAngelo is an automatic atomic model building program for cryo-EM maps.

https://github.com/3dem/model-angelo

with IGBMC HPC

- Upload the protein (and DNA/RNA) sequences as separate fasta files, as well as the map.
- Check the hand of the map and flip it in ChimeraX (*volume flip #1*) if necessary
- Log in to the hpc *ssh <login>@hpc.igbmc.fr*
- Edit the slurm submission script (with *nano* for instance):

```
#!/bin/bash

################################## Slurm options
##################################
### Job name
#SBATCH --job-name=model_angelo
### Limit run time "1-00:00:00"
#SBATCH --time=10:00:00
### Requirements
#SBATCH --partition=gpu
#SBATCH --nodes=1
#SBATCH --ntasks-per-node=1
#SBATCH --mem-per-cpu=40GB
#SBATCH --gres=gpu:1
### Email
#SBATCH --mail-user=xxx@igbmc.fr
#SBATCH --mail-type=ALL
### Output
#SBATCH --output=/shared/mendel/projects/xxx/cryoem/model_angelo/model-
angelo-%j.out
################################################################################
####

echo '########################################'
echo 'Date:' $(date --iso-8601=seconds)
echo 'User:' $USER
echo 'Host:' $HOSTNAME
echo 'Job Name:' $SLURM_JOB_NAME
echo 'Job Id:' $SLURM_JOB_ID
echo 'Directory:' $(pwd)
echo '########################################'
# modules loading
module load model-angelo/1.0.1
# acces database
export TORCH_HOME=/shared/genomes/model_angelo_weights/
#Job command
model_angelo build -v map.mrc -pf prot.fasta -df dna.fasta -rf rna.fasta -o
output
echo 'Done.'
```

```
echo '########################################'
echo 'Job finished' $(date --iso-8601=seconds)
```

Run the script:

```
sbatch model_angelo_script.sh
```

Setup

```
ssh jmwadmin@pollux

bash

conda activate model_angelo
```

Building a map with FASTA sequence

This is the recommended use case, when you have access to a medium-high resolution cryo-EM map (resolutions exceeding 4 Å) as well as a FASTA file with all of your protein sequences.

Let's say the map's name is *`map.mrc`* and the sequence file is *`sequence.fasta`*. To build your model in a directory named *`output`*, you run:

```
model_angelo build -v map.mrc -pf sequence.fasta -o output
```

If you would like to build nucleotides as well, you need to provide the RNA and DNA portions of your sequences in different files like so

```
model_angelo build -v map.mrc -pf prot.fasta -df dna.fasta -rf rna.fasta -o
output
```

If you only have RNA or DNA, you can drop the other input.

If the output of the program halts before the completion of *`GNN model refinement, round 3 / 3`*, there was a bug that you can see in *`output/model_angelo.log`*. Otherwise, you can find your model in *`output/output.cif`*. The name of the mmCIF file is based on the output folder name, so if you specify, for example, *`-o testing/test/model_building`*, the model will be in *`testing/test/model_building/model_building.cif`*.

Building a map with no FASTA sequence

If you have a sample where you do not know all of the protein sequences that occur in the map, you can run *`model_angelo build_no_seq`* instead. This version of the program uses a network that was not trained with input sequences, nor does it do post-processing on the built map.

Instead, in addition to a built model, it provides you with HMM profile files that you can use to search a database such as UniRef with HHblits.

You run this command:

```
model_angelo build_no_seq -v map.mrc -o output
```

The model will be in *output/output.cif* as before. Now there are also HMM profiles for each chain in HHsearch's format here: *output/hmm_profiles*. To do a sequence search for chain A (for example), you should first install HHblits and download one of the databases. Then, you can run

```
hhblits -i output/hmm_profiles/A.hhm -d PATH_TO_DB -o A.hhr -oa3m A.a3m -M first
```

You will have your result as a multiple sequence alignment here: *A.a3m*.