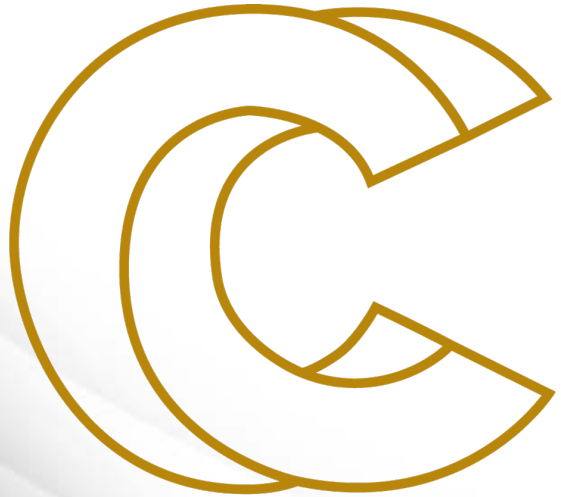

EURO^2

High throughput Virtual Screening with AutoDOCK vina program

Asst.Prof Sefer BADAY

İTÜ Faculty of Computer and Informatics Engineering

ncc@ulakbim.gov.tr

Lesson 3: Distributing molecules to multiple job files

- HPCs often have different queues
- For example on TRUBA “hamsi“ queue allows max 3 days running-time
- Or UHEM altay “defq“ queue allows max 10 days running-time
- You cannot run a docking of millions of molecules in one single run on a typical job submission file
- Thus you need to divide molecules to multiple job submission files

Lesson 3: Distributing molecules to multiple job files

```
number_of_jobs=10
task_per_job=10
ligand_per_task=10
cpu_per_task=4
```

- **Arranging number of jobs**
- You are given prepare_jobs.py file
- You can set the number of jobs, task per job, ligand per task and cpu per task
- In AutoDock Vina program you can choose the number of cpu cores for each run.
- However using many cpu cores for each vina task typically is not efficient.
- For example you can use 128 cores for one Vina task on a node having 128 cores but you will probably get very similar speed using 10 cores (approximation)
- Thus to utilize allocated computing resources it is better in one job file run multiple vina tasks in parallel

Lesson 3: Distributing molecules to multiple job files

- **Example job submission file**

```
#!/bin/bash
#SBATCH -J dock1
#SBATCH -p defq
#SBATCH -A khirri
#SBATCH -n 128
#SBATCH --mail-type=ALL
#SBATCH --mail-user=badays@itu.edu.tr
#SBATCH --output=slurm-%j.out
#SBATCH --error=slurm-%j.err


# python library in UHEM, if you use TRUBA load appropriate module
# python >= 3.6 is needed
module load ANACONDA/Anaconda3-2023.03-python-3.9
source /ari/progs/ANACONDA/Anaconda3-2023.03-python-3.9/etc/profile.d/conda.sh

python task_files/task1.py &
python task_files/task2.py &
python task_files/task3.py &
python task_files/task4.py &
python task_files/task5.py &
python task_files/task6.py &
```

Lesson 3: Distributing molecules to multiple job files

- **Arranging number of jobs**
- You can arrange this numbers depending on the number of available computing nodes, total cpu cores at each node, total days you want to finish computation.
- Before this you can run a small test on couple of ligands to see how it takes to run docking of one molecule
- For example with this setting in 5 day docking of 3.5 million molecules was completed

```
number_of_jobs=11
task_per_job=31
ligand_per_task=10000
cpu_per_task=4
```

Lesson 3: Distributing molecules to multiple job files

- **Preparing files**
- You need to change the following paths

```
main_dir="/Users/seferbaday/Desktop/large_dock/dock"
ligand_dir="/Users/seferbaday/Desktop/large_dock/ligand_1k_v2"
vina_path="/Users/seferbaday/work/autodock_vina_1_1_2_mac_catalina_64bit/bin"
vina_conf="/Users/seferbaday/Desktop/large_dock/dock/prep_materials/vina_conf.txt"
```

- After this if you ran on a terminal
- python prepare_jobs.py
- It will create all job submission files for you

Lesson 4: Distributing molecules to multiple job files

- **Preparing template job submission file**
- Before executing the python prepare_jobs.py make sure the following:
- Put receptor pdbqt under prep_materials folder
- Change template job submission file (run-temp.job) details
- Like account name, queue name, number of total cores available

run-temp.job — prep_materials

```
#!/bin/bash
#SBATCH -J jobname
#SBATCH -p core40q
#SBATCH -A ahtik
#SBATCH -n 40
#SBATCH --mail-type=ALL
#SBATCH --mail-user=badays@itu.edu.tr
#SBATCH --output=slurm-%j.out
#SBATCH --error=slurm-%j.err


# python library in UHEM, if you use TRUBA load appropriate module
# python >= 3.6 is needed
module load Python/python-3.7.4-openmpi-3.0.0-gcc-7.1.0
```

Lesson 3: Distributing molecules to multiple job files

- **Preparing vina configuration file**
- You need to change details for grid center and dimensions

vina_conf.txt — prep_materials

```
receptor=/Users/seferbaday/Desktop/large_dock/dock/prep_materials/mid51_dock.receptor.pdbqt
center_x = -4.83
center_y = -31.47
center_z = -8.97
size_x=30
size_y=30
size_z=30
exhaustiveness=8

```

Lesson 3: Distributing molecules to multiple job files

- **Submitting job files**
- Execution of prepare_jobs.py will create the following folder
- Then you need to run
- ./submit_jobs.sh
- This will submit all jobs
- Outputs of all the jobs will be written to outputs folder
- Each part refers to each job submission

```
[sbaday@altay:~/fimh_hts/dock ]$ ls
outputs         part1   part11  part3  part5  part7  part9          prepare_jobs.py
outputs_backup  part10  part2   part4  part6  part8  prep_materials  submit_jobs.sh
```

References

- Materials can be downloaded from:
- https://drive.google.com/drive/folders/1HgzhXgRK_u0uEfP9SMEoNC-Ncl4M98vo?usp=sharing

Copyright

None

Thanks.

This project has received funding from the European High-Performance Computing Joint Undertaking (JU) under grant agreement No 101101903. The JU receives support from the Digital Europe Programme and Germany, Bulgaria, Austria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, Greece, Hungary, Ireland, Italy, Lithuania, Latvia, Poland, Portugal, Romania, Slovenia, Spain, Sweden, France, Netherlands, Belgium, Luxembourg, Slovakia, Norway, Türkiye, Republic of North Macedonia, Iceland, Montenegro, Serbia