AlphaFold 2.1.1 (2021/11/6)

Webpage

https://github.com/deepmind/alphafold

Version

2.1.1 (downloaded on Nov 6, 2021; databases were prepared on Nov. 5, 2021)

References

- Notes about initial installation on RCCS (Jul 20, 2021): https://ccportal.ims.ac.jp/en/node/2946
- Notes about update on Aug 19, 2021: https://ccportal.ims.ac.jp/en/node/2972
- Citation (jump to official page): https://github.com/deepmind/alphafold#citing-this-work
    - (paper about multimer was added)

Installation notes

- alphafold itself was installed in the same procedure as the previous verisions.
    - (patch to alphafold/common/residue_constants.py is no longer necessary!)

update python env

Python environment built on initial installation (Jul, 2021) was also used.
However, "pandas" was not available in the original environment. We thus installed it via conda.

The environment was a mixture of conda and pip, but we can install pandas without severe conflict or vast updates, fortunately.
Already-existing alphafold versions won't be affected by this update.

Updated databases

Four databases listed below were updated/installed. For the other ones, we employed ones from Aug 19, 2021 installation.

- alphafold model parameters (multimer models were added)
- PDB (mmCIF format)
- PDB seqres
    - these two pdb data were downloaded from Japanese mirror site on Nov 5, 2021.
    - (according to the official page, these two must be of the same date.)
    - (the latest record seems to be Oct 20, 2021.)
- UniProt (newly installed for multimer)

For Uniclust30, 2021/6 version is employed as in Aug 19, 2021 version.

scatter load of uniprot database (for Lustre filesystem)

```
[root@ccfep4 uniprot]$ lfs migrate -c 20 uniprot.fasta
```

update of wrapper script

Wrapper script was also updated, since there were some changes on optional parameters.
run_alphafold21x_rccs.sh

sample script (multimer)

In case of multimer, multi-FASTA files must be specified for -f option amd model_preset(-m) parameter must be "multimer".

```
#!/bin/sh
#PBS -l select=1:ncpus=12:mpiprocs=1:ompthreads=12:jobtype=core
#PBS -l walltime=72:00:00

# NOTE: sample for alphafold 2.1.x (downloaded Nov 2021)
#       this may be imcompatible with previous versions!
```

```
# at least 8 cpu cores will be requested internally.
# in this sample, we employ 12 cores to get enough amount of memory.
# not sure how much is necessary/required, though.

# note about available memory:
# Available memory amount is proportional to ncpus value.
# If you need more memory, please increase ncpus in the header.

if [ ! -z "${PBS_O_WORKDIR}" ]; then
  cd "${PBS_O_WORKDIR}"
fi

AF2ROOT=/local/apl/lx/alphafold2
RUNAF2=${AF2ROOT}/run_alphafold21x_rccs.sh

# pass "-a $AF2DIR" to $RUNAF2 if you want to change alphafold version
#AF2DIR=/local/apl/lx/alphafold2/alphafold-20211106

# load miniconda environment (where necessary binaries reside)
. ${AF2ROOT}/conda_init.sh

# Required:
# -o [output directory]
# -f [sequence file (FASTA)]

# Optional arguments:
# -a [alphafold code root]
#    (default: /local/apl/lx/alphafold2/alphafold-20211106)
# -d [database root]
#    (default: /local/apl/lx/alphafold2/databases-20211105)
# -t [max template date] (default: 2021-11-05)
#    in the original example 2020-05-14 was used (CASP14 target case?)
# -m [model_preset] (default: monomer)
#    model control. available values: monomer, monomer_casp14, monomer_ptm, multimer
# -p [db_preset] (default: full_dbs)
#    database control. available values: full_dbs, reduced_dbs
# -Q
#    change model names to _ptm version; additional scores will be available
#    in output file. plDDT score will be in B-factor column of output pdbs.
#    (same as "-p monomer_ptm")
${RUNAF2} \
  -o ./multimer_test/ \
  -m multimer \
  -f complex.fasta
```