

Manual of Unicon3D

2016/07/20

1. **Github access:** <https://github.com/bioinformatics-tools/UniCon3D>

2. **Files and Directories (All under root):**

1. examples/: contains the test files to run Unicon3D
2. scripts/: contains the scripts for preparing the input files of Unicon3D
3. lib/: contains the trained model
4. src/: contains the source codes of Unicon3D project
5. bin/: contains binaries for linux and osx
6. test/: contains test fasta sequence that used in this manual
7. README: the installation and usage manual
8. Manual.pdf: The detailed guide of how to run the Unicon3D

3. **Run Unicon3D:**

Follow the below steps:

Step 1: Go through the README file.

Step 2: Run the following commands to execute the program with example data:

```
> cd examples/  
> ../bin/linux/UniCon3D -i test -f test.fa -s test.ss -c test.cm -m  
../lib/UniCon.iohmm -d 10
```

Note: 10 decoys are generated in the examples folder starting from test_000001.pdb to test_000010.pdb and test_stats.txt contains the folding statistics.

Step 3 (optional): Use a test protein of your choice. Get the fasta file, secondary structure and contact matrix. Then run the UniCon3D program with it to sample decoys.

4. **Recommended tools:**

1) **Secondary structure prediction:**

Unicon3D project originally applied SSpro5 in the package 'SCRATCH-1D release 1.1' to predict the secondary structure, which can be downloaded from <http://download.igb.uci.edu/>, and it is easy to install and use.

Please try following steps after you have 'SCRATCH-1D release 1.1' installed.

- a. > cd UniCon3D/test/
- b. > <SCRATCH-1D 1.1 path>/bin/run_SCRATCH-1D_predictors.sh test.fa
test.out 1

- c. The expected output should be as follows:

```
test.fa test.out.acc test.out.acc20 test.out.ss test.out.ss8
```

Note: Unicon3D project used 8-class secondary structure information in the training models. So please use file 'test.out.ss8' as secondary structure input for Unicon3D

2) Contact map prediction:

Unicon3D project originally applied in-house contact prediction pipeline tools to get contact map prediction which haven't released to community at this moment. Users can use any residue-residue contact prediction tool (ex. DNcon) to get the predicted residue contacts and convert the residue contact predictions to contact map matrix. Or users can also applied contact map prediction tool (ex.NNcon) directly to get the predicted contact map.

- Here we gave an example to show how to use '**NNcon**' to generate contact map prediction:

a. Download 'NNcon1.0' from MULTICOM-Toolbox
(http://sysbio.rnet.missouri.edu/multicom_toolbox/tools.html)

b. Follow the installation manual to install the tool

c. > cd UniCon3D/test/

d. > mkdir ConMap/

e. > perl <nncon1.0 path>/bin/predict_ss_sa_cm.sh test.fa ConMap/

f. The expected output should be as follows:

```
test_fa.align test_fa.cm12a test_fa.cm8a test_fa.pssm
```

You can choose either 'test_fa.cm8a' or 'test_fa.cm12a' as contact map input for Unicon3D

g. **Run Unicon3D:**

```
>../bin/linux/UniCon3D -i test -f test.fa -s test.out.ss8 -c  
ConMap/test_fa.cm8a -m ../lib/UniCon.iohmm -d 10
```

Note: 10 decoys are generated in the examples folder starting from test_000001.pdb to test_000010.pdb and test_stats.txt contains the folding statistics.

- Here we gave an example to show how to use '**DNcon**' to generate contact residue-residue prediction and convert it to contact map prediction:

a. Download 'DNcon1.0' from MULTICOM-Toolbox
(http://sysbio.rnet.missouri.edu/multicom_toolbox/tools.html)

b. Follow the installation manual to install the tool

c. > cd UniCon3D/test/

d. > mkdir ConRes/

e. > perl <dncon1.0 path>/run_dncon.pl <dncon1.0
path>/local_submit_job.cgi <UniCon3D path>/test/FASTA/
<UniCon3D path>/test/ConRes/ 'I' CPU

f. The expected output 'test.DNcon' should be generated

Note: It is better to format residue-residue contact to CASP format with 5 column, please refer to
(<http://predictioncenter.org/casproll/index.cgi?page=format>)

- g.** Convert DNcon results to CASP format:
> perl ../scripts/dncon2casp.pl ConRes/test.DNcon
ConRes/test.DNcon.casp
- h.** Generate contact map prediction:
> python2 ../scripts/rr2cm.py --fasta test.fa --rr ConRes/test.DNcon.casp
> ConRes/test.DNcon.casp.cm
- i.** **Run Unicon3D:**
> ../bin/linux/UniCon3D -i test -f test.fa -s test.out.ss8 -c
ConRes/test.DNcon.casp.cm -m ../lib/UniCon.iohmm -d 10

Note: 10 decoys are generated in the examples folder starting from test_000001.pdb to test_000010.pdb and test_stats.txt contains the folding statistics.

3) (Option) All-atom model generation: PULCHRA software can be used to generate all-atom model from UniCon3D models.

- a.** Download 'PULCHRA' from <http://www.pirx.com/pulchra/index.shtml>
- b.** Follow the installation manual to install the tool
- c.** > cd UniCon3D/test/
- d.** > <pulchra path>/pulchra -g -v test_000010.pdb
- e.** The full-atom model 'test_000010.pdb.rebuilt.pdb' will be generated

5. Built-in Functions

We wrote several simple program to call the tools to generate inputs for users to run Unicon3D smoothly. All the scripts can be copied and modified to users' preference.

> cd <UniCon3D path>/test

1) Get secondary structure prediction:

> python2 ../scripts/Getsecondary.py --fasta <UniCon3D path>/test/test.fa --id test

Note:

- a.** Please change the '**SCRATCH_1D**' installation path to yours in script 'scripts/Getsecondary.py'.
- b.** Please give the full path of fasta file
- c.** The predicted secondary structures are inside folder '**secondary/**'

2) Get contact map prediction from NNcon:

> python2 ../scripts/GetNNcon.py --fasta <UniCon3D path>/test/test.fa

Note:

- a.** Please change the '**NNcon**' installation path to yours in script 'scripts/GetNNcon.py'.
- b.** Please give the full path of fasta file

- c. The predicted contact map files are inside folder
'NNcon_contact/test_fa.cm8a'

3) Get contact map prediction from DNcon:

```
> python2 ../scripts/GetDNcon.py --fasta <UniCon3D path>/test/test.fa --range  
'1' --device CPU
```

Note:

- a. Please change the '**NNcon**' installation path to yours in script
'scripts/GetNNcon.py'.
- b. Please give the full path of fasta file
- c. The predicted contact files are inside folder
'DNcon_contact/output/test.DNcon'
- d. Convert DNcon results to CASP format:

```
> perl ../scripts/dncon2casp.pl DNcon_contact/output/test.DNcon  
DNcon_contact/output/test.DNcon.casp
```
- e. Generate contact map prediction:

```
> python2 ../scripts/rr2cm.py --fasta test.fa --rr  
DNcon_contact/output/test.DNcon.casp >  
DNcon_contact/output/test.DNcon.casp.cm
```

4) Run Unicon3D on NNcon contact map and secondary structure

```
> ../bin/linux/UniCon3D -i test -f test.fa -s secondary/test.ss8 -c  
NNcon_contact/test_fa.cm8a -m ../lib/UniCon.iohmm -d 10
```

Note: 10 decoys are generated in the examples folder starting from test_000001.pdb to test_000010.pdb and test_stats.txt contains the folding statistics.

5) Run Unicon3D on DNcon contact map and secondary structure

```
> ../bin/linux/UniCon3D -i test -f test.fa -s secondary/test.ss8 -c  
DNcon_contact/output/test.DNcon.casp.cm -m ../lib/UniCon.iohmm -d 10
```

Note: 10 decoys are generated in the examples folder starting from test_000001.pdb to test_000010.pdb and test_stats.txt contains the folding statistics.