

# ComMap Manual

## Brief Introduction

ComMap is a software for protein complexes structural mapping for cross-linking mass spectrometry (CXMS). The cross-linking peptides is identified from LC-MS/MS data by other identification software, such as pLink2, SpotLink et. al. ComMap can be used for both targeted protein complex analysis and large-scale protein-structure mapping analysis.



Icon explanation:  
Complex structure of  
Acetohydroxyacid Synthase

# The configuration

## ## Hardware configuration

**CPU:** Intel or AMD processor with 64-bit support; 2.3 GHz or faster processor with at least 8 cores is recommended

**RAM:** 16GB or higher is recommended

**ROM:** 5G or higher is recommended

**OS:** Windows

**Other:** Internet access for online mode

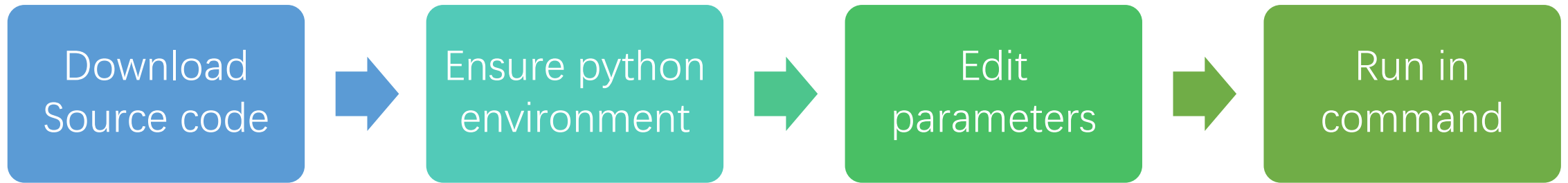
## ## Runtime configuration

**OS:** Windows

**Runtime:** python 3.8

**External runtime library:** numpy, pandas

## How to run ComMap?



## Download Source code

Download ComMap source code on GitHub.  
Unzip the code in suitable folder. (C drive is not recommended.)

### About



CXMS generated cross-linking pairs  
mapping to structures

Readme

0 stars

1 watching

0 forks

### Releases 1

**ComMap v1.0** Latest  
4 minutes ago



## ComMap v1.0

Latest

DICP1810 released this 6 minutes ago



ComMap



e47ef58

The ComMap software.

### ▼ Assets 2



[Source code \(zip\)](#)



[Source code \(tar.gz\)](#)



Ensure python  
environment

If you are new to python:

Please download and install Anaconda3-2021.05-Windows-x86\_64.exe from [Anaconda](#).

Else:

Please ensure python version 3.8 with installed numpy and pandas.



```
Python 3.8.8 (default, Apr 13 2021, 15:08:03) [MSC v.1916 64 bit (AMD64)]
Type 'copyright', 'credits' or 'license' for more information
IPython 7.22.0 -- An enhanced Interactive Python. Type '?' for help.
```

```
In [1]: import numpy
```

```
In [2]: numpy.__version__
Out[2]: '1.20.1'
```

```
In [3]: import pandas
```

```
In [4]: pandas.__version__
Out[4]: '1.2.4'
```

## Edit parameters

Please copy Configuration.commap file from source code to your project folder.  
Then open and edit it with any text editor.  
(such as Windows Notepad or [Visual Studio Code](#))

```
1 #####
2 ### Basic ###
3 #####
4
5 # fasta file for identifying cross-links (abssolute path)
6 A3_PATH_FASTA =
7
8 # calculation type (1 for on-line mode, 0 for off-line mode)
9 C60_CALCULATION_TYPE=1
10
11 # (ONLY FOR OFF-LINE MODE) protein structure map (absolute path)
12 C60_TYPE0_PROTEIN_STRUCTURE_MAP=
13
14 # PDB structure files (default location: "Data" folder in the source code) (absolute path)
15 C60_PDB_STRUCTURE_PATH=
16
17 # result output folder (abssolute path)
18 E1_PATH_EXPORT =
19
20 # input cross-linking result type (0 for generic TAB delimited format; 1 fpr SpotLink format; 2 for pLink2 cross-link format; 3 for pLink2 loop-link format)
21 C61_INPUT_LINK_RESULT_TYPE=
22
23 # input cross-linking result file (abssolute path)
24 C61_INPUT_LINK_RESULT_FILE=
25
26 #####
27 ### Advanced ###
28 #####
29
30
31 # (ONLY FOR ON-LINE MODE) max protein structure query times
32 C62_MAX_DOWNLOAD_TRY_TIME=3
33
34 # the arm length of chosen cross-linker (in angstrom)
35 C65_LINKER_ARM_LENGTH=11.4
36
37 # the max acceptable arm length of structure mapping for filter usage
38 C65_LINKER_MAX_ARM_LENGTH=50
39
40 # parallel computation number
41 P1_NUMBER_THREAD=10
```

Run in  
command

Run ComMap with command:

```
python COMMAP_LOCATION COMMAP_FILE_LOCATION
```

COMMAP\_LOCATION

The location of ComMap.py file.

If you are unfamiliar with command operation, please filled with the absolute path of this file.

COMMAP\_FILE\_LOCATION

The location of .commap file modified in the previous step.

If you are unfamiliar with command operation, please filled with the absolute path of this file.

```
[ComMap] Copyright © 2021. All rights reserved.
2022-04-25 14:43:40.833374

[ComMap] ComMap whole workflow.
2022-04-25 14:43:40.839376
[ComMap] Start to read fasta file and cross-link result file.
2022-04-25 14:43:40.840374
[ComMap] Start to download Uniprot information.
2022-04-25 14:43:40.897374
[ComMap] End downloading Uniprot information.
2022-04-25 14:43:40.898374
[ComMap] Start to download Protein Data Bank information.
2022-04-25 14:43:40.898374
[ComMap] End downloading Protein Data Bank information.
2022-04-25 14:43:40.917373
[ComMap] Start sequence alignment.
2022-04-25 14:43:40.917373
[ComMap] End sequence alignment.
2022-04-25 14:43:49.399178
[ComMap] Start structure distance calculation.
2022-04-25 14:43:49.399178
[ComMap] End structure distance calculation.
2022-04-25 14:43:54.607179
[ComMap] Organize and output.
2022-04-25 14:43:54.607179
[ComMap] Start quality control.
2022-04-25 14:43:54.624180

[ComMap] Finished!
2022-04-25 14:43:54.631180
```



# Q & A

- Which format of protein structure files does ComMap support?

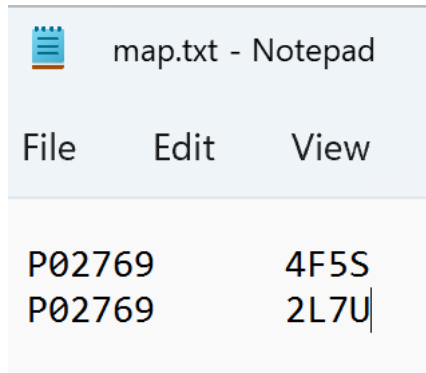
mmCIF is a flexible and extensible tag-value format for representing macromolecular structural data. The set of mmCIF tags, which determine the classes of information present in a given mmCIF format file, are defined in an mmCIF dictionary. The structure of the dictionary is in turn defined by a ddl. The standard mmCIF dictionary - now stable since being ratified by the IUCr in 1997 - was pioneered by Paula M. Fitzgerald, Helen Berman, Phil Bourne, Brian McMahon, Keith Watenpaugh, and John Westbrook. However, work is still ongoing fine-tuning dictionary extensions. More information on mmCIF can be found at the [mmCIF Resource pages](#) at the RCSB.

- What is the difference between on-line mode and off-line mode?

The online mode will obtain protein structure information from the [Protein Data Bank](#) database, and the offline mode will obtain protein structure information from the protein structure folder provided by the user.

# Q & A

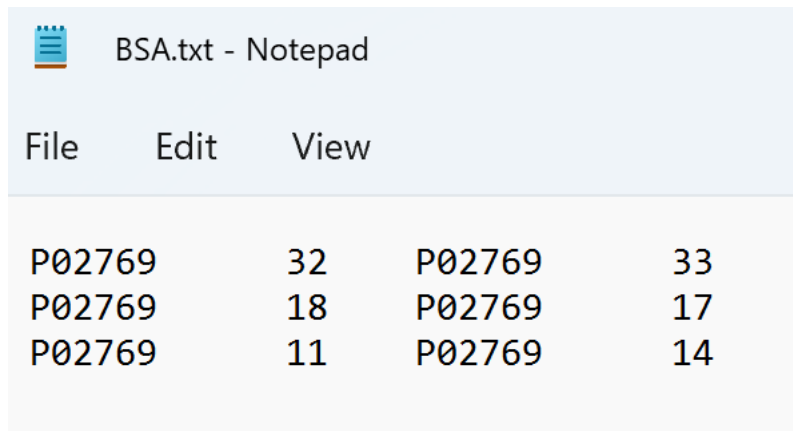
- What is the protein structure map file looks like in the off-line mode?



map.txt - Notepad	
File	Edit View
P02769	4F5S
P02769	2L7U

The first column in the file stores the name of the protein, and the second column stores the name of the protein structure file. (Tab delimited)  
Please make sure protein names match with the name in fasta file and the name of the protein structure file exists in the structure folder.

- What is the generic TAB delimited format looks like?



BSA.txt - Notepad			
File	Edit	View	
P02769	32	P02769	33
P02769	18	P02769	17
P02769	11	P02769	14

The first and third columns of the file store the names of interacting proteins, and the second and fourth columns represent the sites of interacting proteins. (Tab delimited)

Please make sure protein names match with the name in fasta file and the cross-linking sites match with sequence in the input fasta file.

## Q & A

- What is the input SpotLink file looks like?

Please select the SpotLink-generated file with the name result\_filtered.csv as input.

- What is the input pLink2 file looks like?

Please select the pLink2-generated file with the name filtered\_cross-linked\_spectra.csv or filtered\_loop-linked\_spectra.csv as input.

- What is the meaning of each parameter in the .commap file ?

The meaning of the parameters is explained in detail in the comments of the .commap file in the source code

- What do the files output by ComMap contain?

ComMap generated three main files as outputs, including a comprehensive file saved all PPI and structural distance information, and a categorized file of PPIs based on its structure, and a score file contained the ComMap score for PPI-structure. The categorized file offered pymol scripts for each protein structure for visualization.

# Q & A

- ValueError: numpy.ndarray size changed, may indicate binary incompatibility. Expected 88 from C header, got 80 from PyObject

Please try pip install --upgrade numpy or pip uninstall numpy pip install numpy

**Thanks!**