This folder contains the required information to analyze the ClusPRO results in the same manner as described in (Liu et al., 2023).

When you run the code in this file using UCSF Chimera and your preferred IDE (Spyder through Anaconda in this case), you should end up with a file that contains the number of times each LAP amino acid appears within 5Å of internalin. The number in these files is the number of models where that amino acid was interacting. The processed files do have some unnecessary lines, such as the names of files, but these do not impact the frequencies of the amino acids found. This was used as part of the investigation as to which amino acids might be the most influential when LAP and internalin interact.

Contents:

1. Read\_me.docx
   1. This document with instructions on using the contents of this folder
2. code
   1. close\_residues.py
      1. Code file to be run through UCSF Chimera. This code identifies LAP amino acids with 5Å of the internalin protein.
   2. Value\_counts\_example.py
      1. Python file to be run in your preferred IDE (in this case Spyder using Anaconda).
      2. Analyzes the results from close\_residues.py and identifies frequencies of interaction
3. example\_data
   1. This folder contains 30 .pdb files that came from ClusPRO. These are the files that get analyzed using close\_residues.py
4. example\_output
   1. close.csv
      1. example of the results that come from running close\_residues.py
   2. close\_processed.csv
      1. example of what is exported from running Value\_counts\_example.py

To use this file:

1. Install required software
   1. Chimera
      1. UCSF Chimera, version 1.15 was used in this project. This version is no longer supported by UCSF Chimera, which may lead to the analysis breaking if run on newer versions. This is especially true of analyzing files using Chimera X.
      2. <https://www.cgl.ucsf.edu/chimera/>
   2. Python
      1. Install Python. This was written in Python 3.8.12
      2. <https://www.python.org/>
   3. Anaconda
      1. Install Anaconda. This was written in Anaconda 22.9.0.
      2. <https://anaconda.org/>
   4. Spyder
      1. Install Spyder through the Anaconda platform. This was written using Spyder 5.2.1.
   5. Libraries
      1. It is recommended that you create a virtual environment in Anaconda before conducting any analysis. In the new environment, install pandas and Matplotlib.
      2. Pandas version 1.3.5
      3. Install pandas

>>> conda install pandas

* + 1. Install matplotlib 3.5.1

>>> conda install matplotlib

1. Find the amino acids from LAP within 5Å of internalin for each .pdb file in the example data
   1. Ensure that the code, example\_data, and example output are in locations that you can find on your computer.
   2. Open Anaconda
   3. Open Spyder
   4. Open close\_residues.py
   5. Change the line that says “outf = open(PATH\TO\close.csv','w')” to the location where you want your data to be exported
   6. Change the line that says “os.chdir(PATH\TO\example\_data')” to the location where your example\_data is stored.
   7. Open UCSF Chimera
   8. In the command bar type “open” and press enter
   9. Browse to the location where you have saved close\_residues.py
   10. Run the program. This will take a minute or so, depending on how many .pdb files are in the data folder.
   11. This will generate a file in the location where you specified your outputs to go.
2. Analyze the output file
   1. Ensure that the code, example\_data, and example output are in locations that you can find on your computer.
   2. Open Anaconda
   3. Open Spyder
   4. Value\_counts\_example.py
   5. In the section labeled “# In[RUN THE CODE]” change the file paths to the desired locations.
      1. path\_in is where the file generated from close\_residues.py is located
      2. frequency.to\_csv should be set to where you want the output files to export.