========================================================================

**Step 1.** Pairwise MAFFT alignment

========================================================================

Download the fasta files for 3CC2, 4W2E, 5DM6, 5NGM, 8HKU, 6SKF, 6SPB, 6V39, 7JI1, 7NHK, 7OOD, 7S0S, 7S9U, 7SFR, 8A57, 8FMW, 7K00, and 4YBB. Remove all chains except 23S. Concatenate into file ‘pdb\_sequences\_construct.fa’.

Paste the sequences from table ‘rRNA\_pdb\_23S\_sequence.csv’ (the output of distances matrix generation part) into ‘pdb\_sequences\_structure.fa’

Install MAFFT and add to PATH, e.g.

wget <https://mafft.cbrc.jp/alignment/software/mafft_7.520-1_amd64.deb>

sudo dpkg -i mafft\_7.520-1\_amd64.deb

Input

* pdb\_sequences\_structure.fa
* pdb\_sequences\_construct.fa
* pairwise\_mafft\_alignment.py

Run script

python .\pairwise\_mafft\_alignment.py .\pdb\_sequences\_construct.fa .\pdb\_sequences\_structure.fa .\pdb\_sequences\_structure\_projected.fa

Output

pdb\_sequences\_structure\_projected.fa

========================================================================

**Step 2.** Expand the matrices according to MAFFT

========================================================================

Input

* pdb\_sequences\_structure\_projected.fa
* hs\_raw\_map\_\*\*\*\*.csv

Run

python .\expand\_acc\_to\_mafft.py

Output

\*\*\*\*\_mod\_iter1.csv

========================================================================

**Step 3.** Expand the matrices according to Infernal MSA

========================================================================

Align the pdb\_sequences\_construct.fa with the 23S rRNA Infernal MSA, get ‘pdb\_sequences\_infernal\_msa.afa’.

Input

* pdb\_sequences\_construct.fa
* pdb\_sequences\_infernal\_msa.afa
* \*\*\*\*\_mod\_iter1.csv

Run

python .\expand\_acc\_to\_infernal\_msa.py

Output

* \*\*\*\*\_mod\_iter2.csv

========================================================================

**Step 4.** Truncate the matrices according to GNN MSA

========================================================================

Input

* pdb\_sequences\_infernal\_msa.afa
* \*\*\*\*\_mod\_iter2.csv

Run

python .\truncate\_acc\_to\_infernal\_msa\_7k00.py

Output

\*\*\*\*\_aligned.csv

File ‘7k00\_aligned.csv’ is the input for the GNN model graph generation. The rest of the distance matrices are generated for the analysis purpose only.