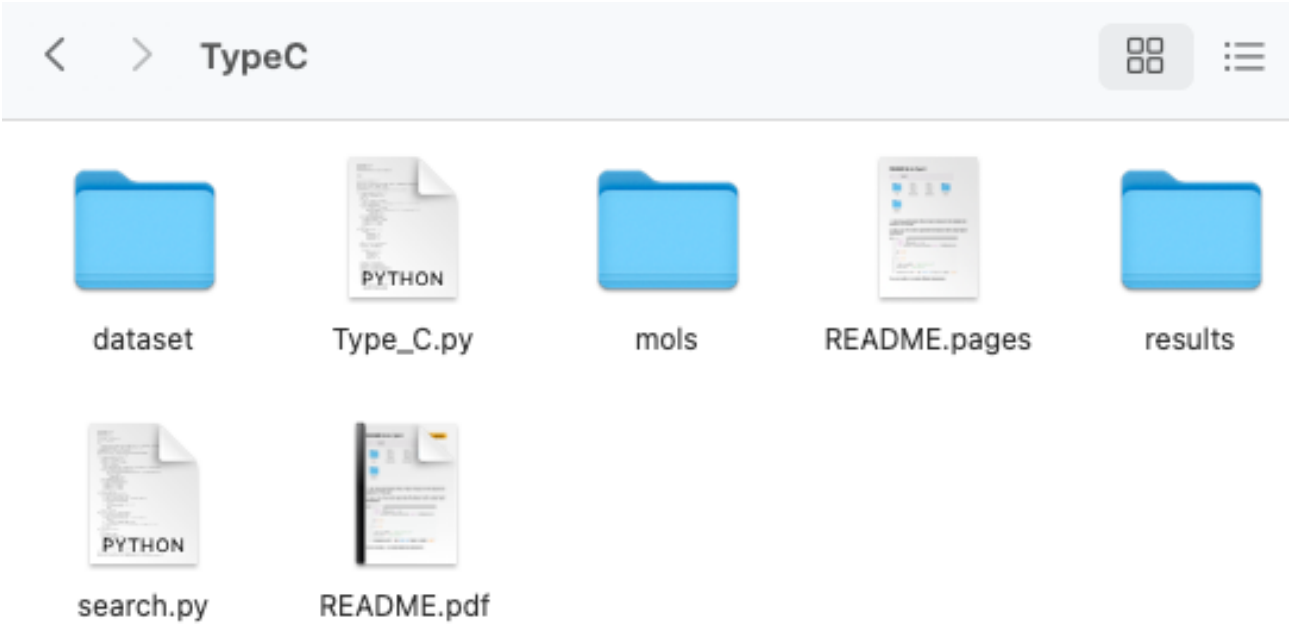
README file for Type C



1. mols: the conformation files of each molecule in the dataset are stored in XYZ format.

2. Type_C.py: the code to generate the feature matrix using Type C descriptors.

```
Type_D.py
3  import os
4  import networkx as nx
5  from sklearn.preprocessing import MinMaxScaler
6
7  a = 2.3
8  b = 4.1
9
10 n = 50
11
12 list_cc_path = "mols/list_cc"
13 mols_dir = "mols/mols"
14
15 molecule_info = pd.read_csv(list_cc_path, sep='
```

You can modify n to obtain different descriptors.

Variables in the code	Variables in the article
a	ϵ_1
b	ϵ_n
n	n

The table above shows the correspondence between the variables in the code and those in the article.

3. search.py: the code to explain the reason why we set $\epsilon_1 = 2.3$ and $\epsilon_n = 4.1$, which is detailed in Section 4 of the SI.

4. results: the results of descriptive statistics measured on the graph.

5. dataset: the feature matrices corresponding to different values of n are used for model selection.