**Instructions for Database\_construction**

1. Download and open PyCharm

2. Open Database\_construction.py and configure the required runtime environment

3. Create three folders for the results of the metabolic prediction (your path/ BioTransformer(superbio); CTS; sygma)

4. Unified structural formula

4.1 Modify the path for reading files

# csv\_data1 = pd.read\_csv('your path/BioTransformer(superbio)/C1/bio\_1.csv')

# csv\_data2 = pd.read\_csv('your path/CTS/C1/cts\_1.csv')

4.2 Modify the path where the results are saved

# csv\_data1.to\_csv("your path/C1/BIO\_1.csv", index=False, encoding="utf-8")

# csv\_data2.to\_csv("your path/C1/CTS\_1.csv", index=False, encoding="utf-8")

5. Combine the predicted results of each prototype into a list

5.1 Modify the path for reading files

# csv\_CTS = pd.read\_csv('your path/C1/CTS\_1.csv')

# csv\_BIO = pd.read\_csv('your path/C1/BIO\_1.csv')

# csv\_sygma = pd.read\_csv('your path/C1/sygma1.csv')

5.2 Modify the path where the results are saved

# df.to\_csv("your path/C1/C1.csv")

6. All the predicted results are integrated into a total metabolite list

6.1 Modify the path for reading files

# csv1 = pd.read\_csv('your path/C1/C1.csv')

# csv2 = pd.read\_csv('your path/C2/C2.csv')

6.2 Modify the path where the results are saved

# df.to\_csv("your path/C1\_C2.csv")

7. Construct a database that can be used for MS-FINDER

7.1 Modify the path for reading files

# csv = pd.read\_csv('your path/C1\_C2.csv')

7.2 Modify the path where the results are saved

# df.to\_csv("your path/C1\_C2\_database1.csv")

8. Remove duplicates from the database

8.1 Modify the path for reading files

# csv = pd.read\_csv('your path/C1\_C2\_database1.csv')

8.2 Modify the path where the results are saved

# df.to\_csv("your path/C1\_C2\_database2.csv")

9. Determine Molecular formula

9.1 Modify the path for reading files

# csv1 = pd.read\_csv('your path/C1\_C2\_database1.csv')

# csv = pd.read\_csv('your path/C1\_C2\_database2.csv')

9.2 Modify the path where the results are saved

# df.to\_csv("your path/C1\_C2\_database3.csv")

10. Construct a unified structural formula for the data obtained from literature

10.1 Modify the path for reading files

# csv\_data1 = pd.read\_csv('your path/literature\_database1.csv')

10.2 Modify the path where the results are saved

# csv\_data1.to\_csv("your path/literature\_database2.csv", index=False, encoding="utf-8")

11. Add molecular weight and other information to the literature data

11.1 Modify the path for reading files

# csv = pd.read\_csv('your path/literature\_database2.csv')

11.2 Modify the path where the results are saved

# df.to\_csv("your path/literature\_database3.csv")

12. Literature data are manually combined with metabolite prediction data, and then duplicates are removed

12.1 Modify the path for reading files

# csv = pd.read\_csv('your path/C1\_C2\_literature\_database1.csv')

12.2 Modify the path where the results are saved

# df.to\_csv("your path/C1\_C2\_literature\_database2.csv")

13. Get the m/z information in the database for running Targeted\_extraction\_main.py

13.1 Modify the path for reading files

# csv1 = pd.read\_csv('your path/C1\_C2\_literature\_database2.csv')

13.2 Modify the path where the results are saved

# df1.to\_csv("your path/mz.csv")