The instructions provided in this document will be give step-by-step instructions to run the MF-LOGP algorithm that was published in “Dimensionally Reduced Machine Learning Model for Predicting Single Component Octanol-Water Partition Coefficients”.

The instructions are designed for Windows operating systems. It is encouraged that users read through all steps, regardless of coding experience, to avoid confusion in later steps.

Some steps will have bolded text at the end of the statement. These texts are either directional instructions for the computer or code that can be directly copied into the console window.

For questions about running this program, please contact the principal investigator:

Dr. Andrew R. Teixeira

[arteixeira@wpi.edu](mailto:arteixeira@wpi.edu)

**Download Anaconda**

1. Download and install the newest version of Anaconda ([here](https://www.anaconda.com/))
2. Once installed, choose the environment that you would like to operate in. The code was developed using Spyder and thus the remaining instructions will be written for Spyder.

Graphical user interface, application

Description automatically generated

**Downloading MFLOGP Files from GitHub**

1. Go to the Teixeira Research Lab GitHub and enter the MF-LOGP\_Development- branch ([here](https://github.com/TeixeiraResearchLab/MF-LOGP_Development-))
2. Download a copy of the repository in a zip file format: **Code > Download ZIP**

Graphical user interface, text, application, email

Description automatically generated

1. The file will download as MF-LOGP\_Development—main. Unzip and extract the file to a local directory.



1. Open the file and ensure you have the following files:
   1. MFLOGP.zip
   2. MFLOGP\_Run\_Code.py
   3. scale\_X.sav
   4. scale\_y.sav

NOTE: Additional files will be downloaded but the four listed above are the core files needed to predict partition coefficients

1. Due to size constrains the published MF-LOGP algorithm had to be compressed to be published on GitHub. It is important that the MFLOGP.zip be unzipped and contents extracted to be with the other files in MF-LOGP\_Development -- main folder

**Installing Relevant Python Packages**

1. Launch the Spyder environment.
2. Within the console (bottom right), type the following commands and press enter
   1. **pip install sklearn**
   2. **pip install chemparse**
   3. **pip install joblib**

A screenshot of a computer

Description automatically generated with medium confidence

1. Once packages are installed, return to the directory noted in **Step 5** and open the script with the Spyder application

Text

Description automatically generated

1. Once opened, scroll down to the **User Inputs** section:

Text

Description automatically generated

* 1. First, choose if you would like to analyze single compound or multiple compounds via an excel spreadsheet. The 1 and 0 are logical statements **(1 = Yes, 0 = No)**
  2. If looking at a single compound (**single\_compound = 1)**, simply change the molecular formula to the compound of interest.

NOTE: This algorithm can only accept compounds with (C, H, N, O, S, P, F, Cl, Br, I) without further training.

* 1. Table

     Description automatically generatedIf looking at multiple compounds (**compound\_list = 1**), add the file directory for the excel document containing the list of compound formulas under **file\_dir**
     1. If running multiple compounds, ensure that the compound list is in Excel, the sheet name matches the Excel document, and the list has a header ‘**Formula’**
  2. Save and close the excel document

1. Save the workspace (**ctrl+S**)
2. Run the code, the answer will populate in the console window

End of Instructions