User Guide for ChemFluor

We use the patch of OLED as an example.

[Note]

1. Because the OLED molecule is less in the database, the direct prediction will not show good results, so we prepare such a patch.

2. We encourage the users to send us some patches, and we will put these patches on our website (www.chemfluor.top) after checking the correctness (please attach with the reference).

3. Due to the limit funding, the prediction on the website will be very slow. (Most time is spent on the transfer of Fingerprints, about 4s/mol, but if transfer on the personal computer, just need 0.1s/mol)

4. If there is any question, please contact chengwei.ju99@gmail.com .

How to use ChemFluor?

1. Install Python and related libraries.

Install PaDEL[1](http://www.yapcwsoft.com/dd/padeldescriptor/).

2. We recommend storing the relevant data in the form of OLED_20200407.xlsx.

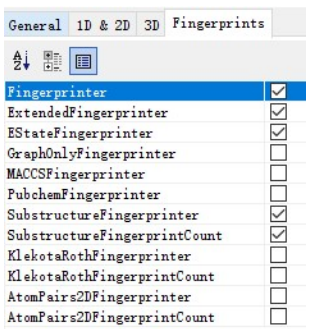
SMILES could be achieved by ChemDraw(CTRL+ALT+C) or some other software.

3. Save the SMILES as xxx.smi.(eg, oled.smi)

4. Use PaDEL to change SMILES to fingerprints, and the output file should be looked at as Fingerprints_OLEDpatch.csv.

General-Descriptor-Fingerprints

Fingerprints-choose selected



5. Insert related solvents descriptor[2] (five columns, Et30, SP, SdP, SA, SB) and change the first column as molecular names. The first column in Database_Train_OLEDpatch.csv is the value of emission wavelength.

6. Put such document [3] (.csv) to ChemFluor\put_your_predict_file_here.

7. Use some Python editor, such as jupyter, vscode, pycharm, etc.

When you run main.py, it will show something like this.

```
please input the job type you want to do.
1: predict EM;
2: predict ABS;
3: predict QY;
0: quit
```

If you don't want to use a patch, you can just predict it without retrain.

How to Retrain ML-model?

Step1-4 is the same as above.

5. Insert related descriptor and change the first column as the value of PLQY/emission wavelength/absorption wavelength, such as Database_Train_OLEDpatch.csv.

6. Put Database_Train_OLEDpatch.csv to ChemFluor\put_your_train_file_here.

And put yours predict file [3] to ChemFluor\put_your_predict_file_here.

7. Use some Python editor, such as jupyter, vscode, pycharm, etc..

When you run main.py, it will show something like this.

```
please input the job type you want to do.
1: predict EM;
2: predict ABS;
3: predict QY;
0: quit
```

Input 1 at the second step, it will cost about 3-5minutes when you predict wavelength and about 1 minute when you predict PLQY.

```
please input the job type you want to do.
1: predict EM;
2: predict ABS;
3: predict QY;
0: quit
 2
New models? 0: No; 1:Yes
 1
training
```

[1] Yap, Chun Wei. "PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints." *Journal of computational chemistry* 32.7 (2011): 1466-1474.

[2] The solvent descriptor could be found in the following papers.

(a) Reichardt, Christian. "Solvatochromic dyes as solvent polarity indicators." *Chemical reviews* 94.8 (1994): 2319-2358.

(b) Catalán, Javier. "Toward a generalized treatment of the solvent effect based on four empirical scales: dipolarity (SdP, a new scale), polarizability (SP), acidity (SA), and basicity (SB) of the medium." *The Journal of Physical Chemistry B* 113.17 (2009): 5951-5960.

[3] The file used for prediction and training should be look like follows.

Predict File: Name1, Et30 SdP,, SP, SA, SB, CDKFingprints(1024bits), ExtFingprints(1024bits), EstateFingprints(79bits), SubFingprintsPrecence(307bits), SubFingprintsCounts(307bits)

Training File: Value, Et30, SP, SdP, SA, SB, CDKFingprints(1024bits), ExtFingprints(1024bits), EstateFingprints(79bits), SubFingprintsPrecence(307bits), SubFingprintsCounts(307bits)