Installing and running CALLC is very easy. The tool provided was tested on windows and linux platforms, but for linux it is assumed you can install the required packages yourself (see the .yml in the install folder). If you have any further questions, feel free to send your questions to: robbin.bouwmeester@ugent.be

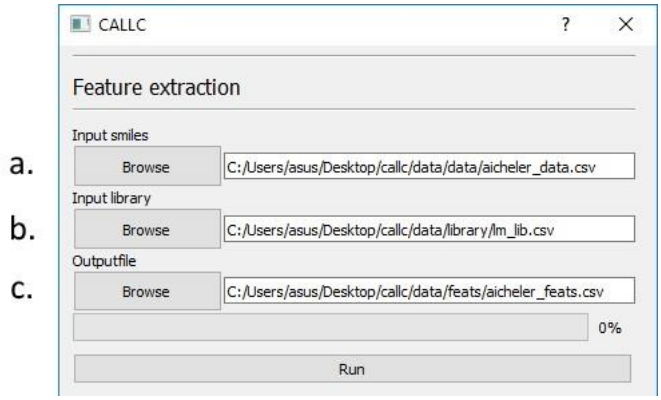
Installation

To install CALLC double click on:

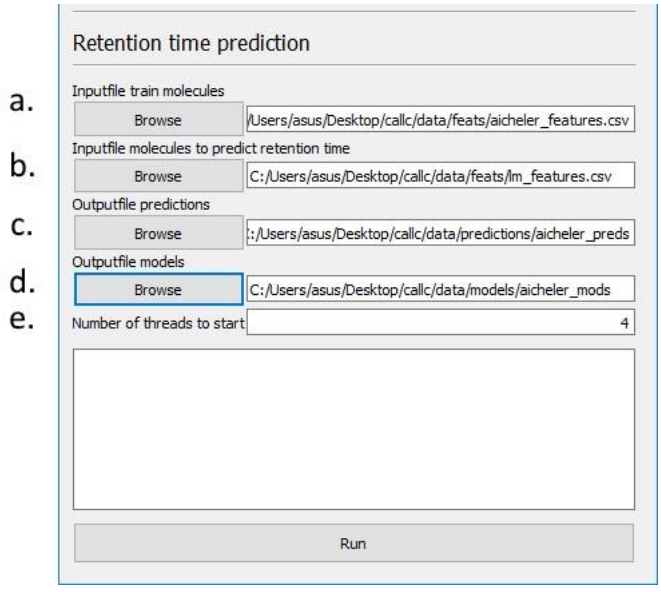
```
install/install.bat
```

Making predictions

1. Run CALCC.vbs in the root (main directory)
2. Extract features for your train set and prediction set you want to make predictions on
    a. Specify the location of the dataset. Train set has three columns: identifiers, SMILES and the experimentally measured retention time (see input and output files examples section for details). The set to make predictions on (e.g. LIPIDMAPS smiles in the folder `rt/datasets/`) has two columns: identifiers and SMILES.
    b. Specify a library with the SMILES structure in the first column and following columns contain the features that are extracted. Defining a library can significantly speed up the process of feature extraction when the process was already performed on a part of the dataset.
    c. Define an output file.



3. Run predictions
    a. Specify the input train set file with features extracted from step one
    b. Specify the dataset to make predictions on from step one
    c. Specify the output file for predictions
    d. Specify the output files for models. If you want to use these models in future for calibration specify it in the folder `rt/mods_l1/`
    e. Specify the number of threads to use for fitting the model parameters



Input and output files examples

Feature extraction examples

Train file:

C:\Users\asus\Desktop\callc\data\data\aicheler_data.csv - Notepad++

File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?

aicheler_data.csv

```
IDENTIFIER,preds,time
LMGP01050041,CCCCCCCCCCCCCCCCCC(=O)OC[C@@H](O)COP(=O)([O-])OCC[N+](C)(C)C,150
LMGP01050045,CCCCCCCCCCCCCCCCCCC(=O)OC[C@@H](O)COP(=O)([O-])OCC[N+](C)(C)C,174
LMGP06050006,CCCCC/C=C\C/C=C\C/C=C\C/C=C\CCCC(=O)OC[C@@H](O)COP(=O)(O)O[C@H]1C(O)C(O)C(O)[C@@H](O)C1O,72
LMGP01010573,CCCCCCCCCCCCCCCCC(=O)O[C@H](COC(=O)CCCCCCCCCCCCCCC)COP(=O)([O-])OCC[N+](C)(C)C,498
LMGP01010976,CCCCCCCCCCCCCCCCCC(=O)OC[C@H](COP(=O)([O-])OCC[N+](C)(C)C)OC(=O)CCCCCCCCCCCCCCCCC,582
LMGP02011213,CCCCCCCCCCCCCCCC(=O)OC[C@H](COP(=O)(O)OCCN)OC(=O)CCCCCCCCCCCCCCC,510
LMSP03010002,CCCCCCCCCCCC/C=C/[C@@H](O)[C@H](COP(=O)([O-])OCC[N+](C)(C)C)NC(=O)CCCCCCCCCC,282
LMGP01050018,CCCCCCCCCCCCCCCC(=O)OC[C@@H](O)COP(=O)([O-])OCC[N+](C)(C)C,102
LMGP01050024,CCCCCCCCCCCCCCCC(=O)OC[C@@H](O)COP(=O)([O-])OCC[N+](C)(C)C,120
LMGP01050026,CCCCCCCCCCCCCCCCC(=O)OC[C@@H](O)COP(=O)([O-])OCC[N+](C)(C)C,132
LMGP01050032,CCCCCCCC/C=C\CCCCCCCC(=O)OC[C@@H](O)COP(=O)([O-])OCC[N+](C)(C)C,108
LMGP01050035,CCCCC/C=C\C/C=C\CCCCCCCC(=O)OC[C@@H](O)COP(=O)([O-])OCC[N+](C)(C)C,90
LMGP01050133,CCCCC/C=C\C/C=C\C/C=C\CCCCCCC(=O)OC[C@@H](O)COP(=O)([O-])OCC[N+](C)(C)C,96
LMGP01050048,CCCCC/C=C\C/C=C\C/C=C\C/C=C\CCCC(=O)OC[C@@H](O)COP(=O)([O-])OCC[N+](C)(C)C,90
LMGP02050002,CCCCCCCCCCCCCCCC(=O)OC[C@H](O)COP(=O)(O)OCCN,108
LMGP02050010,CCCCCC/C=C\CCCCCCCC(=O)OC[C@@H](O)COP(=O)(O)OCCN,90
LMGP02050001,CCCCCCCCCCCCCCCCCC(=O)OC[C@@H](O)COP(=O)(O)OCCN,132
LMGP02050004,CCCCCCCC/C=C\CCCCCCCC(=O)OC[C@@H](O)COP(=O)(O)OCCN,108
LMGP02050011,CCCCC/C=C\C/C=C\CCCCCCCC(=O)OC[C@@H](O)COP(=O)(O)OCCN,90
```

Prediction file:

C:\Users\asus\Desktop\callc\data\data\lm_struct.csv - Notepad++

File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?

aicheler_data.csv | lm_struct.csv

```
LMFA00000001,C=CCCCCC(C)CCCCCC#CCCC(OC)C(=O)OC(=O)C(CCC#CCCCCCC(C)CCCCC=C)OC
LMFA00000002,CCCCCCC[C@H](O)CC(=O)N[C@@H](CO)C(=O)O
LMFA00000003,CCCCCCCCCCCCCCCC(=O)OC(CCCCCCCCCCCCCC)CC(=O)N[C@@H](CCCN)C(=O)O
LMFA00000004,CC/C=C\C/C=C\C/C=C\CCCCCCCC(=O)N[C@@H](CCC(N)=O)C(=O)O
LMFA00000005,CC(C)CCCCCCCCCCCCC(=O)OC(CCCCCCCCCCC(C)C)CC(=O)N[C@H](CO)C(=O)O
LMFA00000006,CC(C)CCCCCCCCCC(O)CC(=O)N[C@@H](CCCCN)C(=O)OCCOC(=O)C(O)CCCCCCCCCC(C)C
LMFA00000007,CCCCCCCCC(=O)NO
LMFA00000008,CCCCCCCC[C@H](O)[C@H](CCCCCCCC(=O)O)OP(=O)(O)O
LMFA00000009,NCCCCCC(=O)NCCCCCC(=O)O
LMFA00000014,CCCCC/C=C\C/C=C\C/C=C\C/C=C\CCCC(=O)OCCN
LMFA00000015,CCCCCCCCCCCCCCCCC/C=C\C(=O)NCCCC[C@H](NC(=O)[C@]1(C)COC(c2ccccc2O)=N1)C(=O)O[C@H](C)CC(=O)N[C@H]1CCCCNC1=O
LMFA00000016,CCCCCCCCC(CCCCCCCCCCCCCOS(=O)(=O)O)OS(=O)(=O)O
LMFA00000017,CCCCCCCCC(OS(=O)(=O)O)C(Cl)CCCCCCCCCCCCOS(=O)(=O)O
LMFA00000018,CCCCCCCC(Cl)C(CCC(Cl)CCCCCCCCCOS(=O)(=O)O)OS(=O)(=O)O
LMFA00000019,CCCCCCC(Cl)C(Cl)C(OS(=O)(=O)O)C(Cl)CC(Cl)CCCCCCCCC(Cl)(Cl)COS(=O)(=O)O
LMFA00000020,CCCCCCCC(Cl)C(OS(=O)(=O)O)C(Cl)CC(Cl)CCCCCCCCC(Cl)(Cl)COS(=O)(=O)O
LMFA00000021,CCCCCCC(Cl)CC(OS(=O)(=O)O)C(Cl)CC(Cl)CCCCCCCCC(Cl)(Cl)COS(=O)(=O)O
LMFA00000022,CCCCCCCCC(CCCCCCCCCCC(Cl)(Cl)COS(=O)(=O)O)OS(=O)(=O)O
LMFA00000023,CCCCCCCCCC(OS(=O)(=O)O)C(Cl)CCCCCCCCCCCCCOS(=O)(=O)O
LMFA00000024,CCCCCCCC(Cl)C(Cl)C(OS(=O)(=O)O)C(Cl)CC(Cl)CCCCCCCCC(Cl)COS(=O)(=O)O
```

Library file:

C:\Users\asus\Desktop\callc\data\library\lm_lib.csv - Notepad++

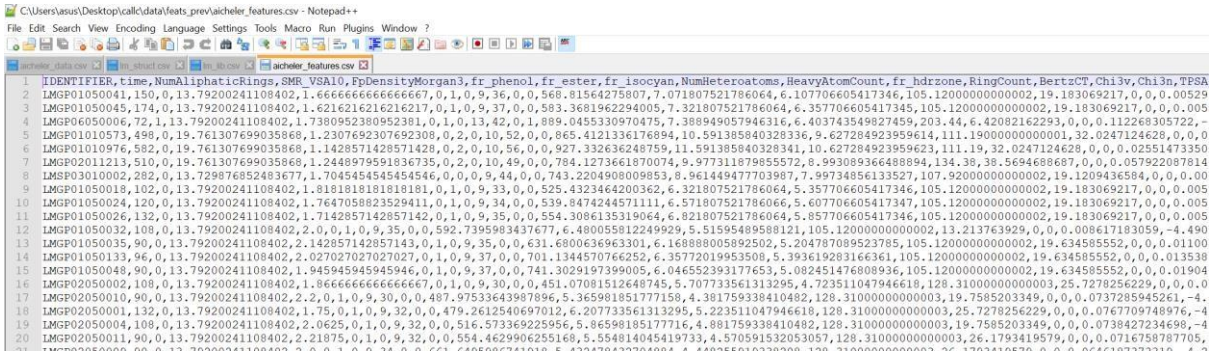
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?

aicheler_data.csv | lm_struct.csv | lm_lib.csv

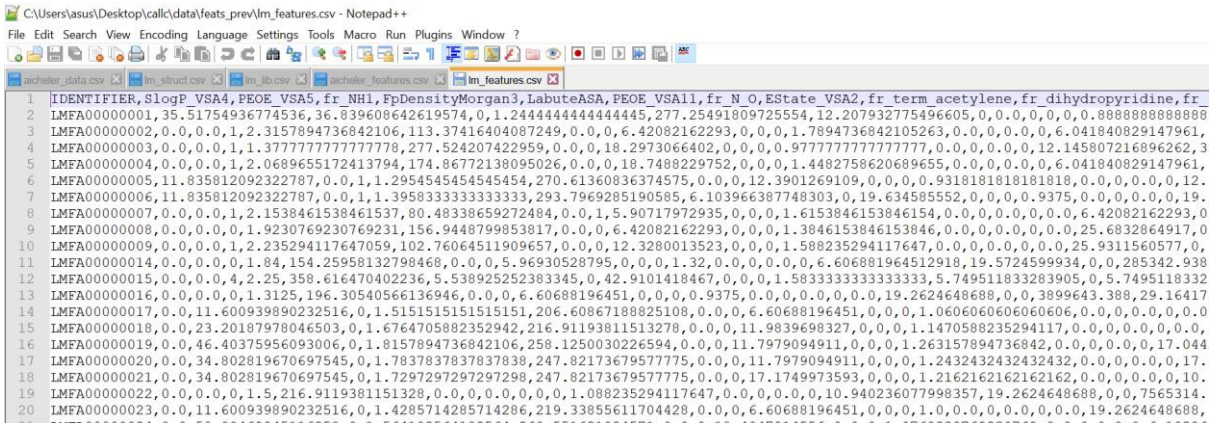
```
IDENTIFIER,fr_furan,fr_imidazole,VSA_EState10,PEOE_VSA6,NOCount,fr_Al_COO,fr_lactone,PEOE_VSA1,SlogP_VSA8,fr_priamide,Chi1v,EState_VSA6,fr_aniline,MaxPartialCharge,
C=CCCCC(C)CCCCC#CCCC(OC)C(=O)OC(=O)C(CCC#CCCCCC(C)CCCC=C)OC,0,0,0,0,90.20773084,5,0,0,14.21058886,0,0,17.83314496,0,0,0.342470887,14.21058886,0,0,38.36613843,25.6
CCCCCCC[C@H](O)CC(=O)N[C@@H](CO)C(=O)O,0,0,0,0,39.02784531,6,1,0,20.63637079,0,0,6.819037056,0,0,0.328121133,10.21305479,2,5.316788604,45.94875638,6.420821623,0,0,0,1
CCCCCCCCCCCCCCC(=O)OC(CCCCCCCCCCCCCC)CC(=O)N[C@@H](CCCN)C(=O)O,0,0,0,0,161.5263717,7,1,0,20.89384643,0,0,19.35746941,0,0,0.32569525,10.47053043,0,11.05045608,41.6428
CC/C=C\C/C=C\C/C=C\CCCCCCCC(=O)N[C@@H](CCC(N)=O)C(=O)O,0,0,0,0,62.64232271,6,1,0,16.15698348,0,1,10.72395623,0,0,0.325710068,10.84019487,0,11.05045608,28.93203297,6.4
CC(C)CCCCCCCCCCCC(=O)OC(CCCCCCCCCCC(C)C)CC(=O)N[C@H](CO)C(=O)O,0,0,0,0,156.1113813,7,1,0,20.26670635,0,0,18.47713298,0,0,0.328121138,9.843390349,1,5.316788604,46.0
CC(C)CCCCCCCCCC(O)CC(=O)N[C@@H](CCCN)C(=O)OCCOC(=O)C(O)CCCCCCCCCC(C)C,0,0,0,0,143.269738,9,0,0,30.73723678,0,0,19.75108668,0,0,0.334575345,15.20739338,2,11.0504560
CCCCCCCCC(=O)NO,0,0,0,0,45.44866694,3,0,0,0,0,0,5.238391115,5.480096598,0,0,2.242828618,5.207253025,0,5.480096598,11.11443275,6.420821623,0,0,0,1,0,9.343315175,1,740.5
CCCCCCCC[C@H](O)[C@H](CCCCCCCC(=O)O)OP(=O)(O)O,0,0,0,0,71.13195343,7,1,0,19.999878,0,0,11.42756729,0,0,0.469440408,19.41709776,1,0,38.17711606,25.68328649,0,0,0,0,0,1
NCCCCCC(=O)NCCCCCC(=O)O,0,0,0,0,12.84164325,5,1,0,16.15698348,0,0,6.35076355,0,0,0.302839566,10.84019487,0,11.05045608,30.07252522,25.93115606,0,0,0,1,0,12.83323327,0
CCCCC/C=C\C/C=C\C/C=C\CCCC(=O)OCCN,0,0,0,0,68.3735413,3,0,0,10.47053043,0,0,9.765151544,0,0,0.305336161,10.47053043,0,5.733667477,19.12094366,19.57245999,0,0,0,
CCCCCCCCCCCCCCCC/C=C\C(=O)NCCCC[C@H](NC(=O)[C@]1(C)COC(c2ccccc2O)=N1)C(=O)O[C@H](C)CC(=O)N[C@H]1CCCCNC1=O,0,0,0,0,115.0239942,13,0,0,35.84740772,0,0,23.05978985,38.1
CCCCCCCCC(CCCCCCCCCCCCCOS(=O)(=O)O)OS(=O)(=O)O,0,0,0,-8.699087761,109.6568832,8,0,0,0,0,0,15.14342587,0,0,0.39711767,13.28858518,0,0,38.65194207,19.26246487,0,0,0,0,0
CCCCCCCCC(OS(=O)(=O)O)C(Cl)CCCCCCCCCCCOS(=O)(=O)O,0,0,0,-2.411051597,103.2360615,8,0,0,0,0,0,15.63141288,0,0,0.397169389,24.88952507,0,0,44.02902994,19.26246487,0,0,
CCCCCCCC(Cl)C(CCC(Cl)CCCCCCCCCCOS(=O)(=O)O)OS(=O)(=O)O,0,0,0,3.916739978,83.97359667,8,0,0,0,0,0,16.10256313,0,0,0.397169395,36.49046496,0,0,49.40611778,25.68328649,0,
CCCCCCC(Cl)C(Cl)C(OS(=O)(=O)O)C(Cl)CC(Cl)CCCCCCCCC(Cl)(Cl)COS(=O)(=O)O,0,0,0,28.27183941,94.33383321,8,0,0,0,0,0,17.91953734,0,0,0.397222282,78.34147465,0,0,64.493647
CCCCCCCC(Cl)C(OS(=O)(=O)O)C(Cl)CC(Cl)CCCCCCCCC(Cl)(Cl)COS(=O)(=O)O,0,0,0,21.9399148,100.7546548,8,0,0,0,0,0,17.43155033,0,0,0.397221372,66.74053476,0,0,59.11655979,25
CCCCCCC(Cl)CC(OS(=O)(=O)O)C(Cl)CC(Cl)CCCCCCCCC(Cl)(Cl)COS(=O)(=O)O,0,0,0,21.93748214,94.33383321,8,0,0,0,0,0,17.41471358,0,0,0.397170268,71.29328463,0,0,59.11655979,3
CCCCCCCCC(CCCCCCCCCCC(Cl)(Cl)COS(=O)(=O)O)OS(=O)(=O)O,0,0,0,2.983519025,126.4379413,8,0,0,0,0,0,15.98442607,0,0,0.39711767,36.49046496,0,0,42.98529618,19.26246487,0,
CCCCCCCCCC(OS(=O)(=O)O)C(Cl)CCCCCCCCCCCCOS(=O)(=O)O,0,0,0,-2.374109272,116.0777048,8,0,0,0,0,0,16.63141288,0,0,0.397169389,24.88952507,0,0,44.02902994,19.26246487,0,
```

Prediction examples

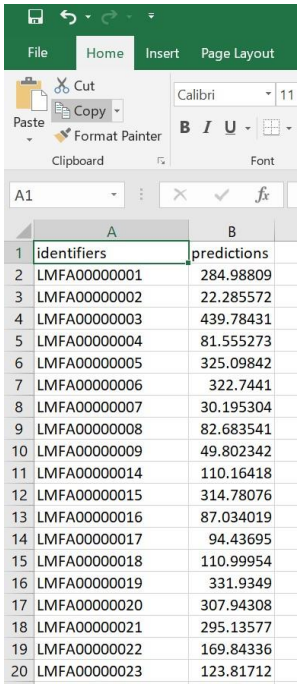
Input file train:



Input file predictions:



Output file predictions:



	A	B
1	identifiers	predictions
2	LMFA00000001	284.98809
3	LMFA00000002	22.285572
4	LMFA00000003	439.78431
5	LMFA00000004	81.555273
6	LMFA00000005	325.09842
7	LMFA00000006	322.7441
8	LMFA00000007	30.195304
9	LMFA00000008	82.683541
10	LMFA00000009	49.802342
11	LMFA00000014	110.16418
12	LMFA00000015	314.78076
13	LMFA00000016	87.034019
14	LMFA00000017	94.43695
15	LMFA00000018	110.99954
16	LMFA00000019	331.9349
17	LMFA00000020	307.94308
18	LMFA00000021	295.13577
19	LMFA00000022	169.84336
20	LMFA00000023	123.81712