



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

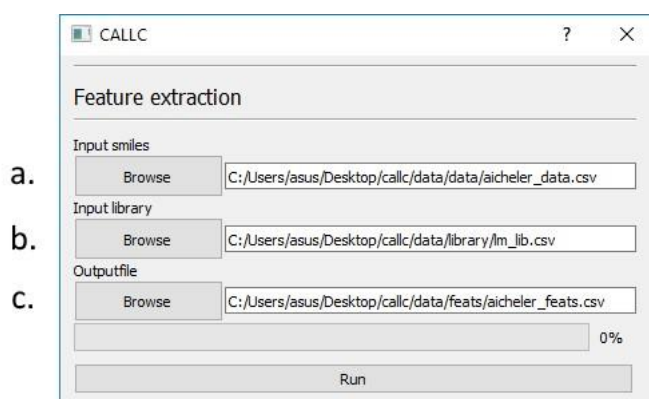
Installation

To install CALLC double click on:

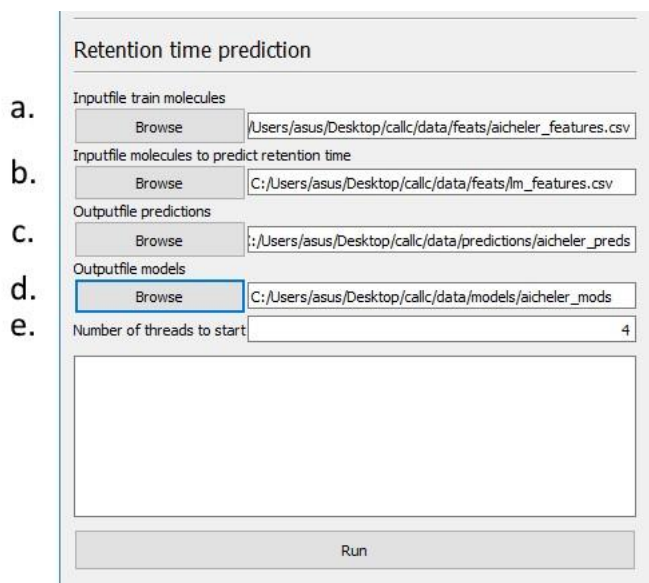
`install/install.bat`

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



Feature extraction examples

Train file:

```
C:\Users\asus\Desktop\call\data\data/aicheir_data.csv - Notepad++  
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?  
 aicheir_data.csv  
1 IDENTIFIER,preds,time  
2 LMGP01050041,CCCCCCCCCCCCCCCCCCCC(=O)OC[C@H](O)COP(=O)([O-])OCC[N+](C)(C)C,150  
3 LMGP01050045,CCCCCCCCCCCCCCCCCCCC(=O)OC[C@H](O)COP(=O)([O-])OCC[N+](C)(C)C,174  
4 LMGP06050006,CCCC/C=C\C/C=C\C\C/C=C\C\CCCC(=O)OC[C@H](O)COP(=O)O[C@H]1C(O)C(O)[C@H](O)C1O,72  
5 LMGP01010573,CCCCCCCCCCCCCCCCCCCC(=O)O[C@H](COC(=O)CCCCCCCCCCCCCCCOP(=O)([O-])OCC[N+](C)(C)C,498  
6 LMGP01010976,CCCCCCCCCCCCCCCCCCCC(=O)OC[C@H](COP(=O)([O-])OCC[N+](C)(C)C)OC(=O)CCCCCCCCCCCCCCCCCCCC,582  
7 LMGP02011213,CCCCCCCCCCCCCCCCCCCC(=O)OC[C@H](COP(=O)(O)OCCN)OC(=O)CCCCCCCCCCCCCCCCCCCC,510  
8 LMSPO3010002,CCCCCCCCCCCCC/C=C/[C@H](O)[C@H](COP(=O)([O-])OCC[N+](C)(C)C)NC(=O)CCCCCCCCCCCC,282  
9 LMGP01050018,CCCCCCCCCCCCCCCCCCCC(=O)OC[C@H](O)COP(=O)([O-])OCC[N+](C)(C)C,102  
10 LMGP01050024,CCCCCCCCCCCCCCCCCCCC(=O)OC[C@H](O)COP(=O)([O-])OCC[N+](C)(C)C,120  
11 LMGP01050026,CCCCCCCCCCCCCCCCCCCC(=O)OC[C@H](O)COP(=O)([O-])OCC[N+](C)(C)C,132  
12 LMGP01050032,CCCCCCCC/C=C\CCCCCCCC(=O)OC[C@H](O)COP(=O)([O-])OCC[N+](C)(C)C,108  
13 LMGP01050035,CCCC/C=C\C/C=C\CCCCCCCC(=O)OC[C@H](O)COP(=O)([O-])OCC[N+](C)(C)C,90  
14 LMGP01050133,CCCCC/C=C\C/C=C\C/C=C\CCCCCCCC(=O)OC[C@H](O)COP(=O)([O-])OCC[N+](C)(C)C,96  
15 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  
16 LMGP02050002,CCCCCCCCCCCCCCCCCCCC(=O)OC[C@H](O)COP(=O)(O)OCCN,108  
17 LMGP02050010,CCCCC/C=C\C\CCCCCCCC(=O)OC[C@H](O)COP(=O)(O)OCCN,90  
18 LMGP02050001,CCCCCCCCCCCCCCCCCCCC(=O)OC[C@H](O)COP(=O)(O)OCCN,132  
19 LMGP02050004,CCCCCCCC/C=C\CCCCCCCC(=O)OC[C@H](O)COP(=O)(O)OCCN,108  
20 LMGP02050011,CCCCC/C=C\C/C=C\C\CCCCCCCC(=O)OC[C@H](O)COP(=O)(O)OCCN,90
```

Prediction file:

```
C:\Users\asus\Desktop\call\data\data\lm_struct.csv - Notepad++  
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?  
  
aichecker_data.csv | lm_struct.csv  
1 LMFA00000001, C=CCCCC (C) CCCCC#CCCC (OC) C (=O) OC (=O) C (CCC#CCCCCCC (C) CCCCC=C) OC  
2 LMFA00000002, CCCCCC[C@H] (O) CC (=O) N[C@H] (CO) C (=O) O  
3 LMFA00000003, CCCCCCCCCCCCCC (=O) OC (CCCCCCCCCCCCC) CC (=O) N[C@H] (CCCN) C (=O) O  
4 LMFA00000004, CC/C=C\C/C=C\C/C=C\CCCCC (=O) N[C@H] (CCC(N)=O) C (=O) O  
5 LMFA00000005, CC (C) CCCCCCCCCCCCC (=O) OC (CCCCCCCCCCCCC (C) C) CC (=O) N[C@H] (CO) C (=O) O  
6 LMFA00000006, CC (C) CCCCCCCCCC (O) CC (=O) N[C@H] (CCCN) C (=O) OCCO (=O) C (O) CCCCCCCCCC (C) C  
7 LMFA00000007, CCCCCCCCC (=O) NO  
8 LMFA00000008, CCCCCC[C@H] (O) [C@H] (CCCCCCC (=O) O) OP (=O) (O) O  
9 LMFA00000009, NCCCCC (=O) NCCCCC (=O) O  
10 LMFA00000014, CCCC/C=C\C/C=C\C/C=C\C/C=C\CCCC (=O) OCCN  
11 LMFA00000015, CCCCCCCCCCCCCC/C=C\C (=O) NCCC[C@H] (NC (=O) [C@H] 1 (C) COC (c2cccc20)=N1) C (=O) O[C@H] (C) CC (=O) N[C@H] 1CCCCN1=O  
12 LMFA00000016, CCCCCCCC (CCCCCCCCCCCCCOS (=O) (=O) O) OS (=O) (=O) O  
13 LMFA00000017, CCCCCCCC (OS (=O) (=O) O) C (Cl) CCCCCCCCCCOS (=O) (=O) O  
14 LMFA00000018, CCCCCC(Cl) C (CC(Cl) CCCCCCCCCCOS (=O) (=O) O) OS (=O) (=O) O  
15 LMFA00000019, CCCCCC(Cl) C (Cl) C (OS (=O) (=O) O) C (Cl) CC (Cl) CCCCCC(Cl) (Cl) COS (=O) (=O) O  
16 LMFA00000020, CCCCCC(Cl) C (OS (=O) (=O) O) C (Cl) CC (Cl) CCCCCC(Cl) (Cl) COS (=O) (=O) O  
17 LMFA00000021, CCCCCC(Cl) CC (OS (=O) (=O) O) C (Cl) CC (Cl) CCCCCC(Cl) (Cl) COS (=O) (=O) O  
18 LMFA00000022, CCCCCCCC (CCCCCCCCCCCC (Cl) (Cl) COS (=O) (=O) O) OS (=O) (=O) O  
19 LMFA00000023, CCCCCCCCCC (OS (=O) (=O) O) C (Cl) CCCCCCCCCCCCCOS (=O) (=O) O  
20 LMFA00000024, CCCCCC(Cl) C (Cl) C (OS (=O) (=O) O) C (Cl) CC (Cl) CCCCCC(Cl) COS (=O) (=O) O
```

Library file:

[illegible]

Prediction examples

Input file train:

```
C:\Users\asus\Desktop\callc\data\feats_prev\aiCheler_features.csv - Notepad++
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
aiCheler_data.csv lm_struct.csv lm_ib.csv aiCheler_features.csv
1 IDENTIFIER,time,NumAliphaticRings,SMR_VSA10,FpDensityMorgan3,fr_phenol,fr_ester,fr_isocyan,NumHeteroatoms,HeavyAtomCount,fr_hdrzone,RingCount,BertzCT,Chi3v,Chi3n,TPSA
2 LMGP01050041,150,0,13.79200241108402,1.666666666666667,0,1,0,9,36,0,0,568.81564275807,7.071807521786064,6.107706605417346,105.120000000000002,19.183069217,0,0,0.00529
3 LMGP01050045,174,0,13.79200241108402,1.6216216216216217,0,1,0,9,37,0,0,583.3681962294005,7.321807521786064,6.357706605417345,105.120000000000002,19.183069217,0,0,0.005
4 LMGP06050006,72,1,13.79200241108402,1.7380952380952381,0,1,0,13,42,0,1,889.0455330970475,7.388949057946316,6.403743549827459,203.44,6.42082162293,0,0,0.112268305722,-
5 LMGP01010573,498,0,19.761307699035868,1.2307692307692308,0,2,0,10,52,0,0,865.4121336176894,10.591385840328336,9.627284923959614,111.190000000000001,32.0247124628,0,0,0
6 LMGP01010976,582,0,19.761307699035868,1.1428571428571428,0,2,0,10,56,0,0,927.332636248759,11.591385840328341,10.627284923959623,111.19,32.0247124628,0,0,0.02551473350
7 LMGP02011213,510,0,19.761307699035868,1.2448979591836735,0,2,0,10,49,0,0,784.1273661870074,9.977311879655572,8.993089366488894,134.38,38.5694688687,0,0,0.057922087814
8 LMGP03010002,282,0,13.729876852483677,1.7045454545454546,0,0,0,9,44,0,0,743.2204908009853,8.961449477703987,7.99734856133527,107.920000000000002,19.1209436584,0,0,0.000
9 LMGP01050018,102,0,13.79200241108402,1.8181818181818181,0,1,0,9,33,0,0,525.4323464200362,6.321807521786064,5.357706605417346,105.120000000000002,19.183069217,0,0,0.005
10 LMGP01050024,120,0,13.79200241108402,1.7647058823529411,0,1,0,9,34,0,0,539.8474244571111,6.571807521786066,5.607706605417347,105.120000000000002,19.183069217,0,0,0.005
11 LMGP01050026,132,0,13.79200241108402,1.7142857142857142,0,1,0,9,35,0,0,554.3086135319064,6.821807521786064,5.857706605417346,105.120000000000002,19.183069217,0,0,0.005
12 LMGP01050032,108,0,13.79200241108402,2.0,0,1,0,9,35,0,0,592.7395983437677,6.480055812249929,5.51595489588121,105.120000000000002,13.213763929,0,0,0.008617183059,-4.490
13 LMGP01050035,90,0,13.79200241108402,2.142857142857143,0,1,0,9,35,0,0,631.6800636963301,6.16888005892502,5.204787089523785,105.120000000000002,19.634585552,0,0,0.01100
14 LMGP01050133,96,0,13.79200241108402,2.027027027027027,0,1,0,9,37,0,0,701.1344570766252,6.35772019953508,5.393619283166361,105.120000000000002,19.634585552,0,0,0.013538
15 LMGP01050048,90,0,13.79200241108402,1.945945945945946,0,1,0,9,37,0,0,741.3029197399005,6.046552393177653,5.082451476808936,105.120000000000002,19.634585552,0,0,0.01904
16 LMGP02050002,108,0,13.79200241108402,1.866666666666667,0,1,0,9,30,0,0,451.07081512648745,5.707733561313295,4.723511047946618,128.310000000000003,25.727826229,0,0,0.0
17 LMGP02050010,90,0,13.79200241108402,2.2,0,1,0,9,30,0,0,487.97533643987896,5.36598185177158,4.381759338410482,128.310000000000003,19.7585203349,0,0,0.0737285945261,-4.
18 LMGP02050001,132,0,13.79200241108402,1.75,0,1,0,9,32,0,0,479.2612540697012,6.207733561313295,5.223511047946618,128.310000000000003,25.727826229,0,0,0.0767709748976,-4
19 LMGP02050004,108,0,13.79200241108402,2.0625,0,1,0,9,32,0,0,516.573369225956,5.86598185177716,4.881759338410482,128.310000000000003,19.7585203349,0,0,0.0738427234698,-4
20 LMGP02050011,90,0,13.79200241108402,2.21875,0,1,0,9,32,0,0,554.4629906255168,5.554814045419733,4.570591532053057,128.310000000000003,26.1793419579,0,0,0.0716758778705,
LMGP06050006,72,1,13.79200241108402,1.7380952380952381,0,1,0,13,42,0,1,889.0455330970475,7.388949057946316,6.403743549827459,203.44,6.42082162293,0,0,0.112268305722,-
```

Input file predictions:

```
C:\Users\asus\Desktop\callc\data\feats_prev\lm_features.csv - Notepad++
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
aiCheler_data.csv lm_struct.csv lm_ib.csv aiCheler_features.csv lm_features.csv
1 IDENTIFIER,SlogP_VSA4,PEOE_VSA5,fr_NH1,FpDensityMorgan3,LabuteASA,PEOE_VSA11,fr_N_O,ESTate_VSA2,fr_term_acetylene,fr_dihydropyridine,fr
2 LMFA00000001,35.51754936774536,36.839608642619574,0,1.2444444444444445,277.25491809725554,12.207932775496605,0,0,0,0,0,0.888888888888888
3 LMFA00000002,0,0,0,0,1,2.3157894736842106,113.37416404087249,0,0,0,6.42082162293,0,0,0,1.7894736842105263,0,0,0,0,0,0,6.041840829147961,
4 LMFA00000003,0,0,0,0,1,1.3777777777777778,277.524207422959,0,0,0,18.2973066402,0,0,0,0.9777777777777777,0,0,0,0,0,0,12.145807216896262,3
5 LMFA00000004,0,0,0,0,1,2.0689655172413794,174.86772138095026,0,0,0,18.7488229752,0,0,0,1.4482758620689655,0,0,0,0,0,0,6.041840829147961,
6 LMFA00000005,11.835812092322787,0,0,1,1.2954545454545454,270.61360836374575,0,0,0,12.3901269109,0,0,0,0.9318181818181818,0,0,0,0,0,0,12.
7 LMFA00000006,11.835812092322787,0,0,1,1.3958333333333333,293.7969285190585,6.103966387748303,0,19.634585552,0,0,0,0.9375,0,0,0,0,0,0,19.
8 LMFA00000007,0,0,0,0,1,2.1538461538461537,80.48338659272484,0,0,1,5.90717972935,0,0,1.6153846153846154,0,0,0,0,0,0,6.42082162293,0
9 LMFA00000008,0,0,0,0,1.9230769230769231,156.9448799853817,0,0,0,6.42082162293,0,0,1.3846153846153846,0,0,0,0,0,0,25.6832864917,0
10 LMFA00000009,0,0,0,0,1,2.235294117647059,102.76064511909657,0,0,0,12.3280013523,0,0,1.588235294117647,0,0,0,0,0,0,25.9311560577,0
11 LMFA00000014,0,0,0,0,1.84,154.25958132798468,0,0,0,5.96930528795,0,0,1.32,0,0,0,0,0,0,6.606881964512918,19.5724599934,0,0,285342.938
12 LMFA00000015,0,0,0,0,4,2.25,358.616470402236,5.538925252383345,0,42.9101418467,0,0,1.5833333333333333,5.749511833283905,0,5.7495118332
13 LMFA00000016,0,0,0,0,1.3125,196.30540566136946,0,0,0,6.60688196451,0,0,0.9375,0,0,0,0,0,0,19.2624648688,0,3899643.388,29.16417
14 LMFA00000017,0,0,0,0,11.600939890232516,0,1.5151515151515151,206.60867188825108,0,0,0,6.60688196451,0,0,1.0606060606060606,0,0,0,0,0,0
15 LMFA00000018,0,0,0,0,23.20187978046503,0,1.6764705882352942,216.91193811513278,0,0,0,11.9839698327,0,0,1.1470588235294117,0,0,0,0,0,0
16 LMFA00000019,0,0,0,0,46.40375956093006,0,1.8157894736842106,258.1250030226594,0,0,0,11.7979094911,0,0,1.263157894736842,0,0,0,0,0,0,17.044
17 LMFA00000020,0,0,0,0,34.802819670697545,0,1.7837837837837838,247.82173679577775,0,0,0,11.7979094911,0,0,1.2432432432432432,0,0,0,0,0,0,17.
18 LMFA00000021,0,0,0,0,34.802819670697545,0,1.7297297297297298,247.82173679577775,0,0,0,17.1749973593,0,0,1.2162162162162162,0,0,0,0,0,0,10.
19 LMFA00000022,0,0,0,0,1.5,216.9119381151328,0,0,0,0,0,0,1.088235294117647,0,0,0,0,0,0,10.940236077998357,19.2624648688,0,7565314.
20 LMFA00000023,0,0,0,0,11.600939890232516,0,1.4285714285714286,219.33855611704428,0,0,0,6.60688196451,0,0,1.0,0,0,0,0,0,19.2624648688,
```

Output file predictions:

File		Home	Insert	Page Layout
Cut		Calibri 11		
Copy		B <i>I</i> <u>U</u>		
Format Painter		Clipboard		
A1		fx		
	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		