Example pep_lc

```
from lc_pep import LCPep
from feat_extractor import FeatExtractor

# Native imports
import pickle
import sys
import os
import random
import itertools

# Pandas
import pandas as pd

# Matplotlib
from matplotlib import pyplot as plt

# Numpy
import numpy as np


# Read the input data to make predictions for
df = pd.read_csv("datasets/seqs_exp.csv",sep=",")

# Generate some identifiers, any kind of identifiers will do
df.index = ["Pep_"+str(dfi) for dfi in df.index]

# Make a feature extraction object; you can skip this if you do not want to use the default settings
# for pep_lc. Here we want to use a model that does not use RDKit features so we skip the chemical
# descriptor making procedure.
f_extractor = FeatExtractor(chem_descr_feat=False,
                            verbose=False)

# Make the pep_lc object that will handle making predictions and calibration
pepper = LCPep(config_file=config_file,
               path_model=os.path.join(os.getcwd(),"mods/lcpep_synt.pickle"),
               f_extractor=f_extractor,
               verbose=False)

# Since we use the same data transform the retention times so we can calibrate
df["tr"] = df["tr"]**0.85

# Calibrate the original model based on the new retention times
pepper.calibrate_preds(seq_df=df)

# Make predictions; calibrated and uncalibrated
print("Predictions (calibrated): ",pepper.make_preds(seq_df=df))
print("Predictions (uncalibrated): ",pepper.make_preds(seq_df=df,calibrate=False))

# compare calibrated and uncalibrated predictions
plt.scatter(df["tr"],pepper.make_preds(seq_df=df),label="Calibrated",s=1)
plt.scatter(df["tr"],pepper.make_preds(seq_df=df,calibrate=False),label="Uncalibrated",s=1)
plt.legend()
plt.show()
```

Example input seqs_exp.csv

```
seq,modifications,tr
AAGPSLSHTSGGTQSK,,12.1645
AAINQKLIETGER,6|Acetyl,34.095
AANDAGYFNDEMAPIEVKTK,12|Oxidation|18|Acetyl,37.3765
AANDAGYFNDEMAPIEVKTK,18|Acetyl,41.943999999999996
AANMLQQSGSKNTGAK,4|Oxidation|11|Acetyl,16.14
AANMLQQSGSKNTGAK,4|Oxidation,9.498750000000001
AANMLQQSGSKNTGAK,11|Acetyl,21.488000000000003
AAQASDLEKIHLDEK,9|Acetyl,32.957
AELNKAVDTK,5|Acetyl,26.41
```

Example input mod_to_struct.csv RDKit features

```
,BalabanJ,Chi0n,LabuteASA,MinPartialCharge,MolLogP
Acetyl,2.1874960973678963,1.9082482904638631,19.265715242269923,-0.2912343143503756,0.11599999999999999
Biotin,1.7030406543934982,8.939543180393162,98.35218009520547,-0.4812303897702795,0.7968
Butyryl,2.4017152226171965,3.399812122026584,31.99559947106314,-0.30340309725779463,0.9854
Crotonyl,2.9244181981729174,3.1402990980327403,31.305996438878882,-0.2985764972892379,0.7614000000000001
Deamidated,2.1874960973678963,1.4328121551534467,17.695006880069332,-0.483466609975982,-0.29919999999999997
Dimethyl,1.0,2.0,15.104193142226158,-0.06826238342059164,1.0262
Formyl,2.0,1.1153550716504106,12.900773127873316,-0.3070970261864102,-0.1849
GG,3.490729660390813,5.072731091493441,56.63534195611968,-0.48029079053921064,-1.9699999999999998
Malonyl,3.046489080910934,3.0481672268038573,34.58641320890631,-0.48091550528760457,-0.34000000000000014
Methyl,0.0,0.0,8.739251027829551,-0.07755789119307553,0.6361
Nitro,2.1874960973678963,1.2637101764276841,16.817313149337416,-0.19360570618031736,-0.4136000000000001
Oxidation,0.0,0.5,6.84923118831809,-0.4115095219374226,-0.8247
Phospho,3.4720664956001186,2.119172062391505,28.30662084449607,-0.7900715423358614,-2.1926
Propionyl,2.2968352619191386,2.6927053408400363,25.630657356666532,-0.30341423261619294,0.5952999999999999
Succinyl,2.9670528832697762,3.755274007990405,40.95135532330292,-0.48122076640195605,0.05009999999999998
Trimethyl,1.6329931618554523,2.7071067811865475,21.469135256622767,-0.06564544121023959,1.4163
```

Example input unimod_to_formula.csv

```
name,formula_pos,formula_neg
Acetyl,"H(2) C(2) O",""
Amidated,"H N","O(-1)"
Biotin,"H(14) C(10) N(2) O(2) S",""
Carbamidomethyl,"H(3) C(2) N O",""
Carbamyl,"H C N O",""
Carboxymethyl,"H(2) C(2) O(2)",""
Deamidated,"O","H(-1) N(-1)"
ICAT-G,"H(38) C(22) N(4) O(6) S",""
ICAT-G:2H(8),"H(30) 2H(8) C(22) N(4) O(6) S",""
```