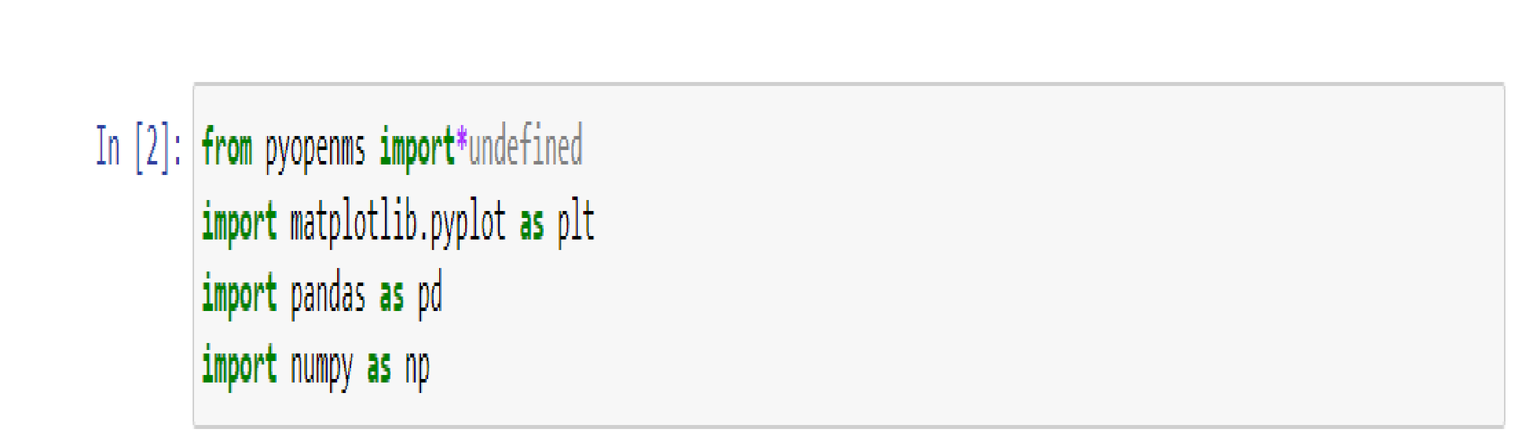
Leukemia proteomics

-Zeinab Alaa
-Sammar Alaa
-Mai Mohamed
-Yasmeen Hossam
-Ahmed Ali
-Mohamed Allam

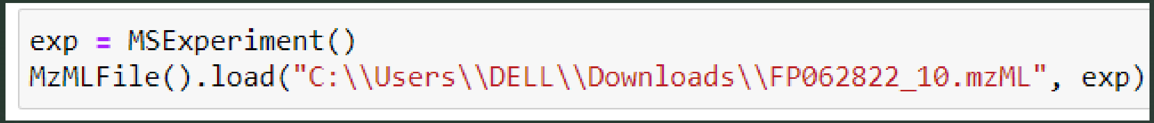
Import pyopenms



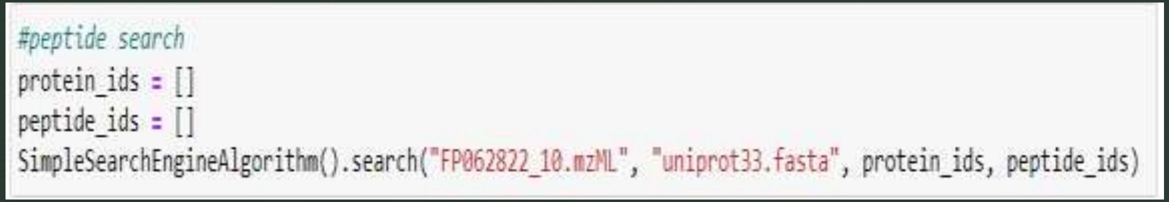
```
In [2]: from pyopenms import*undefined
        import matplotlib.pyplot as plt
        import pandas as pd
        import numpy as np
```

Load raw file.

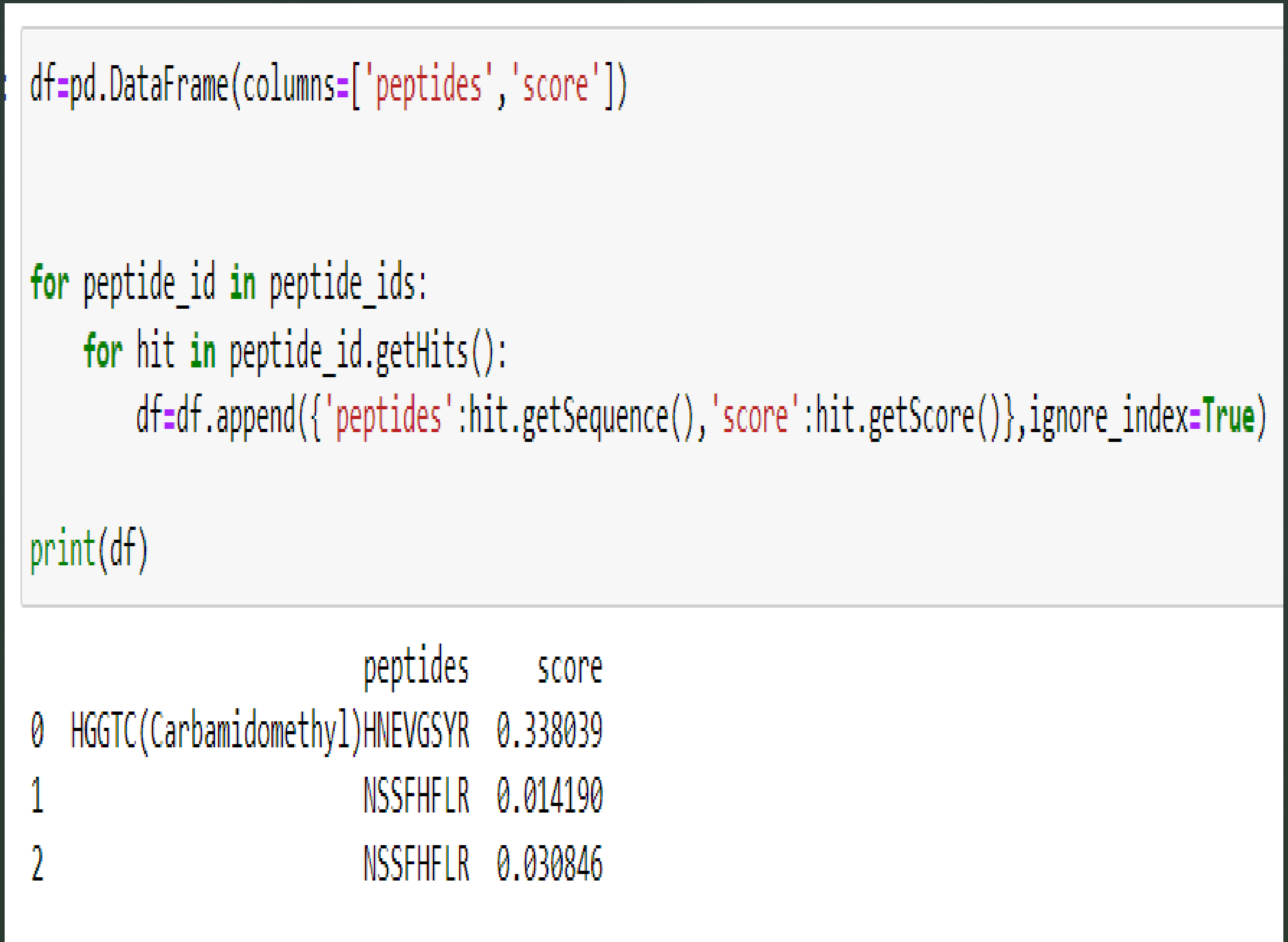
Search for peptide in the raw file .



```
exp = MSExperiment()
MzMLFile().load("C:\\Users\\DELL\\Downloads\\FP062822_10.mzML", exp)
```

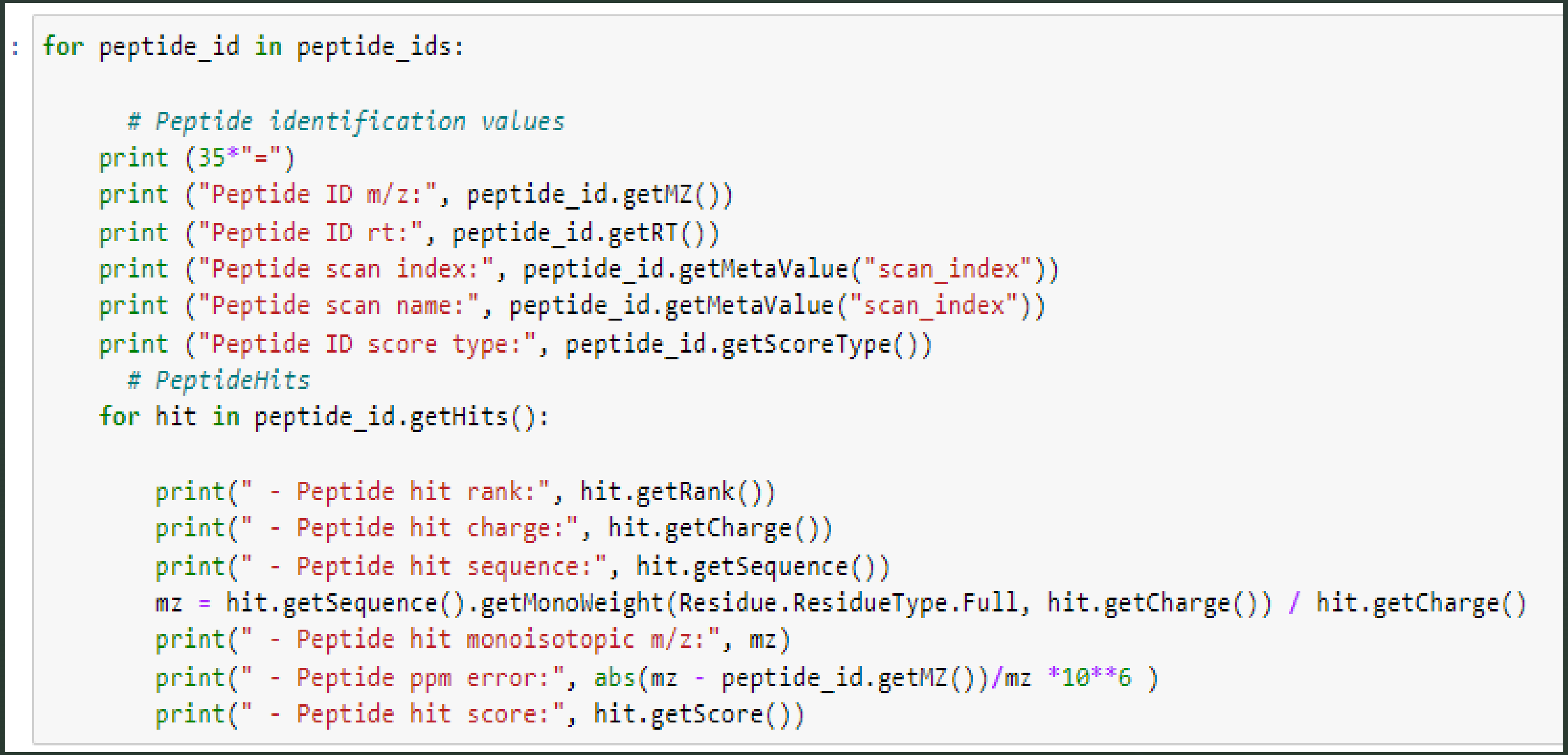


```
#peptide search
protein_ids = []
peptide_ids = []
SimpleSearchEngineAlgorithm().search("FP062822_10.mzML", "uniprot33.fasta", protein_ids, peptide_ids)
```

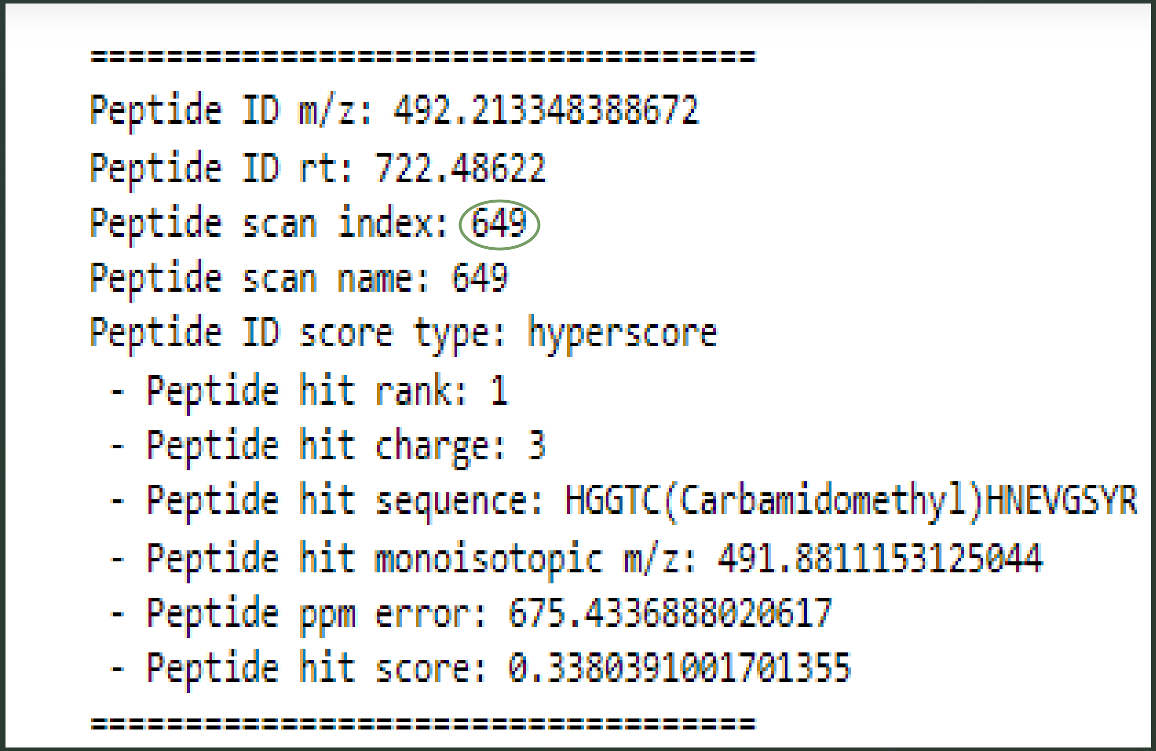


```python
df=pd.DataFrame(columns=['peptides','score'])



for peptide_id in peptide_ids:
    for hit in peptide_id.getHits():
        df=df.append({'peptides':hit.getSequence(),'score':hit.getScore()},ignore_index=True)


print(df)
```

```
                         peptides     score
0  HGGTC(Carbamidomethyl)HNEVGSYR  0.338039
1                        NSSFHFLR  0.014190
2                        NSSFHFLR  0.030846
```

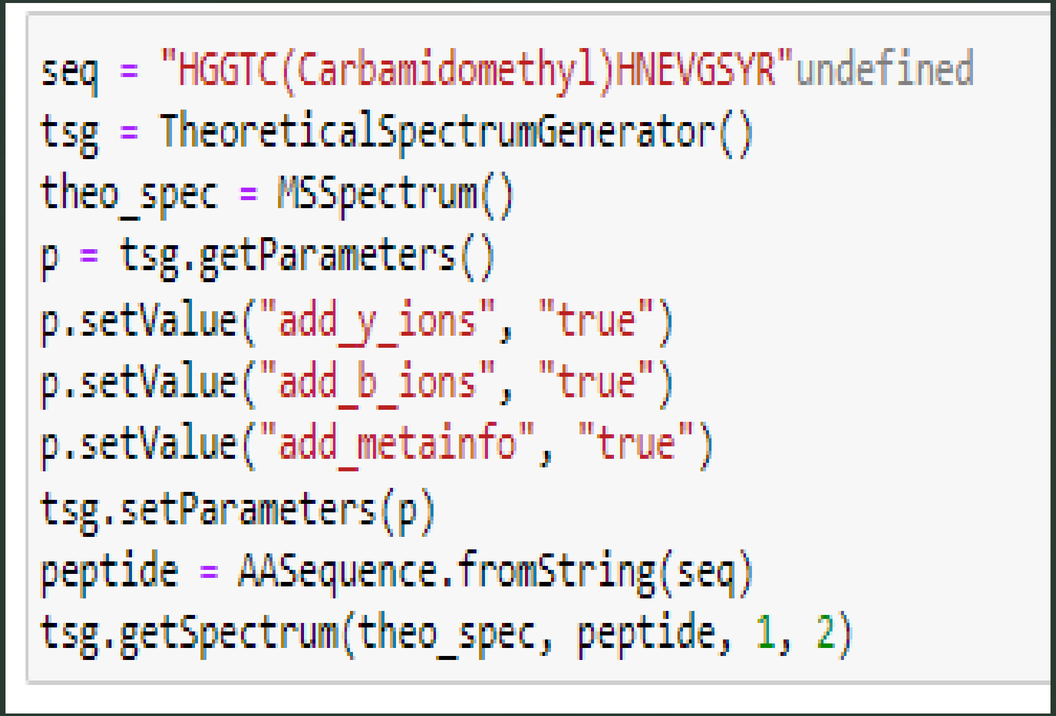


```
for peptide_id in peptide_ids:

      # Peptide identification values
    print (35*"=")
    print ("Peptide ID m/z:", peptide_id.getMZ())
    print ("Peptide ID rt:", peptide_id.getRT())
    print ("Peptide scan index:", peptide_id.getMetaValue("scan_index"))
    print ("Peptide scan name:", peptide_id.getMetaValue("scan_index"))
    print ("Peptide ID score type:", peptide_id.getScoreType())
      # PeptideHits
    for hit in peptide_id.getHits():

        print(" - Peptide hit rank:", hit.getRank())
        print(" - Peptide hit charge:", hit.getCharge())
        print(" - Peptide hit sequence:", hit.getSequence())
        mz = hit.getSequence().getMonoWeight(Residue.ResidueType.Full, hit.getCharge()) / hit.getCharge()
        print(" - Peptide hit monoisotopic m/z:", mz)
        print(" - Peptide ppm error:", abs(mz - peptide_id.getMZ())/mz *10**6 )
        print(" - Peptide hit score:", hit.getScore())
```



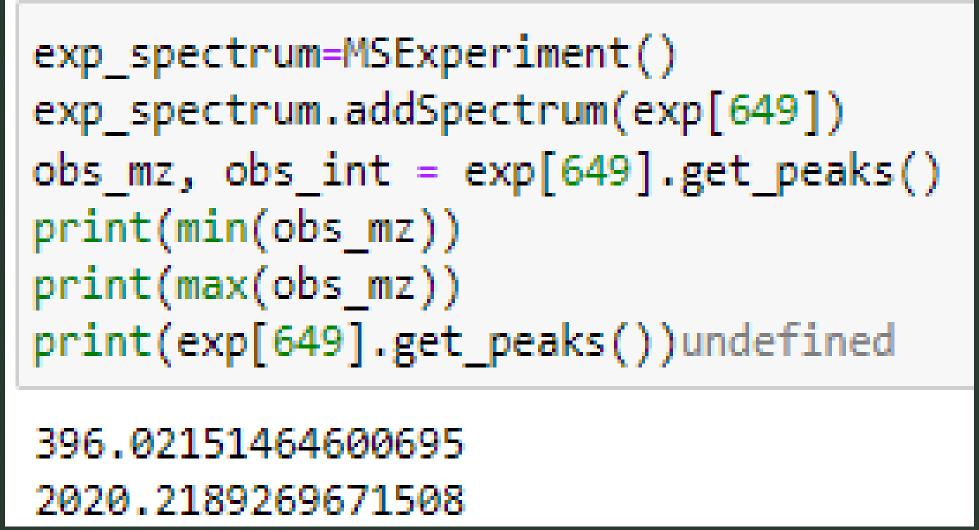
```
===================================
Peptide ID m/z: 492.213348388672
Peptide ID rt: 722.48622
Peptide scan index: 649
Peptide scan name: 649
Peptide ID score type: hyperscore
 - Peptide hit rank: 1
 - Peptide hit charge: 3
 - Peptide hit sequence: HGGTC(Carbamidomethyl)HNEVGSYR
 - Peptide hit monoisotopic m/z: 491.8811153125044
 - Peptide ppm error: 675.4336888020617
 - Peptide hit score: 0.3380391001701355
===================================
```

Generate theoretical spectrum of the first peptide



```
seq = "HGGTC(Carbamidomethyl)HNEVGSYR"undefined
tsg = TheoreticalSpectrumGenerator()
theo_spec = MSSpectrum()
p = tsg.getParameters()
p.setValue("add_y_ions", "true")
p.setValue("add_b_ions", "true")
p.setValue("add_metainfo", "true")
tsg.setParameters(p)
peptide = AASequence.fromString(seq)
tsg.getSpectrum(theo_spec, peptide, 1, 2)
```

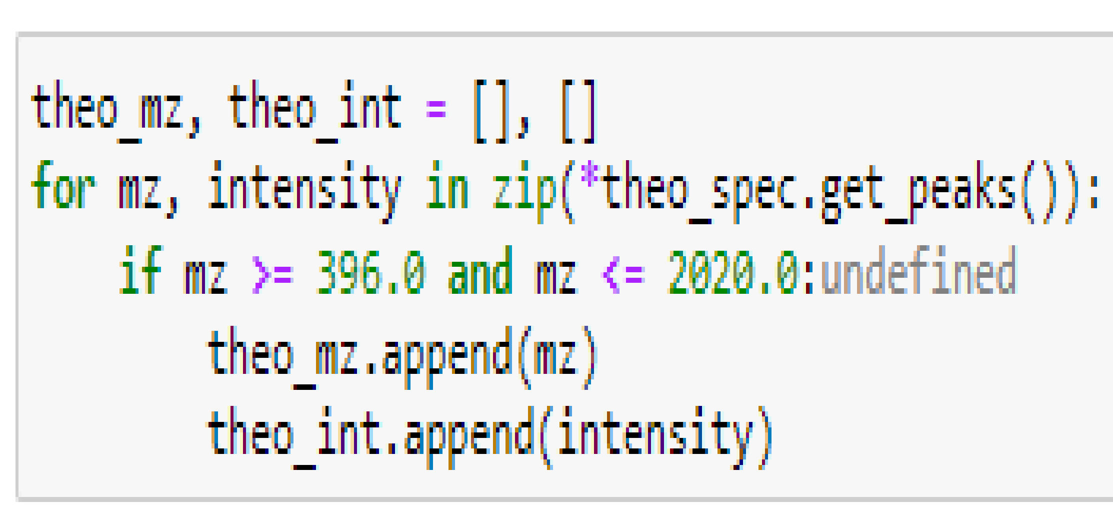
Experimental spectrum



```
exp_spectrum=MSExperiment()
exp_spectrum.addSpectrum(exp[649])
obs_mz, obs_int = exp[649].get_peaks()
print(min(obs_mz))
print(max(obs_mz))
print(exp[649].get_peaks())undefined
```

```
396.02151464600695
2020.2189269671508
```

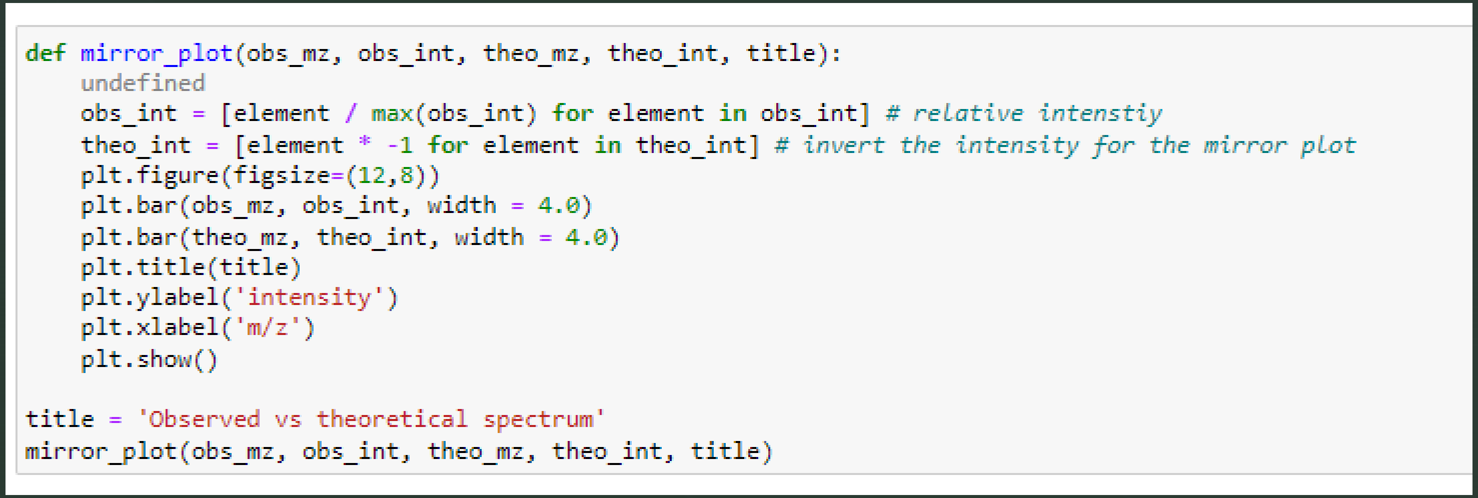
The mass to charge of our observed spectrum ranges 396-2020



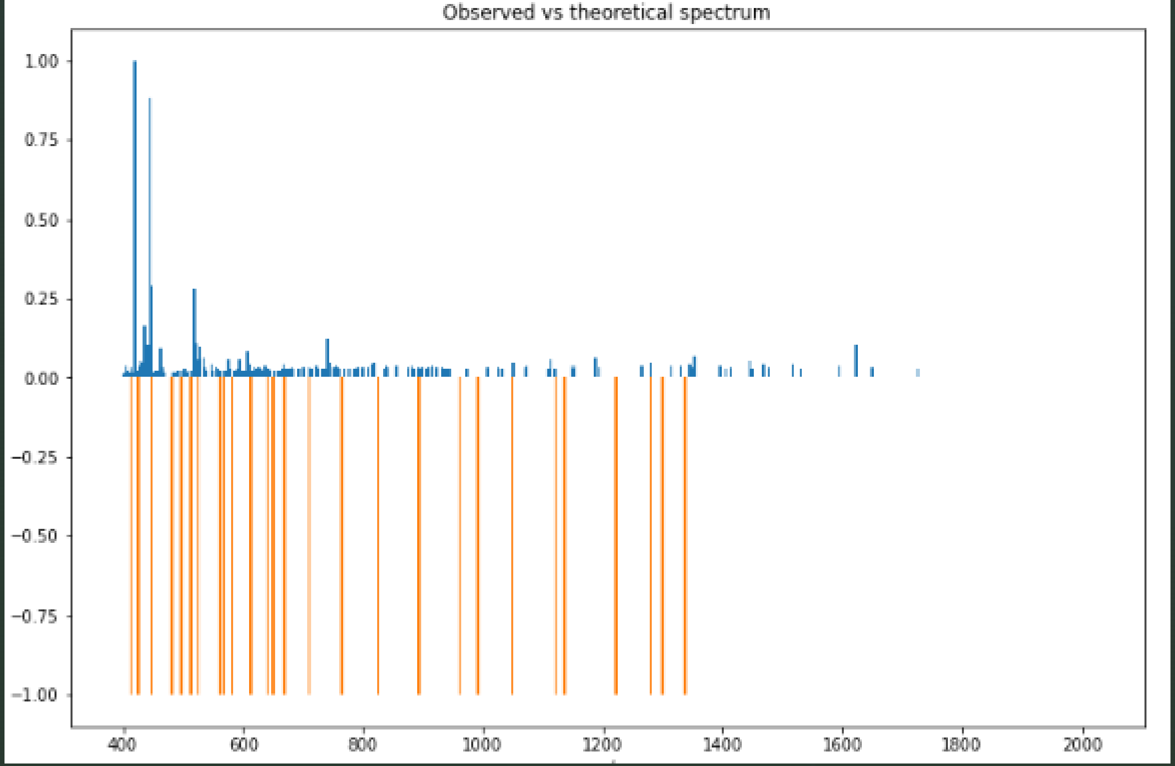
```
theo_mz, theo_int = [], []
for mz, intensity in zip(*theo_spec.get_peaks()):
    if mz >= 396.0 and mz <= 2020.0:undefined
        theo_mz.append(mz)
        theo_int.append(intensity)
```

We filter the peaks of the theoretical spectrum to fit the range(to reduce image complexity)

Mirror Plot



```
def mirror_plot(obs_mz, obs_int, theo_mz, theo_int, title):
    undefined
    obs_int = [element / max(obs_int) for element in obs_int] # relative intenstiy
    theo_int = [element * -1 for element in theo_int] # invert the intensity for the mirror plot
    plt.figure(figsize=(12,8))
    plt.bar(obs_mz, obs_int, width = 4.0)
    plt.bar(theo_mz, theo_int, width = 4.0)
    plt.title(title)
    plt.ylabel('intensity')
    plt.xlabel('m/z')
    plt.show()

title = 'Observed vs theoretical spectrum'
mirror_plot(obs_mz, obs_int, theo_mz, theo_int, title)
```


Observed vs theoretical spectrum
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
400
600
800
1000
1200
1400
1600
1800
2000

Thank You