TEAM 20

TEAM MEMBERS

- احمد محمد محمد فرغلي
- ايهاب راشد مسعود رضوان
- عبد الرحمن ماهر
- ايه حسني سعيد
- يوسف خليفة

LINKS

- RAW DATA(10 Files):https://www.ebi.ac.uk/pride/archive/projects/PXD029586

- Fasta file:
- >sp|Q99527|GPER1_HUMAN G-protein coupled estrogen receptor 1 OS=Homo sapiens OX=9606 GN=GPER1 PE=1 SV=1
- MDVTSQARGVGLEMYPGTAQPAAPNTTSPELNLSHPLLGTALANGTGELSEHQQYVIGLF
- LSCLYTIFLFPIGFVGNILILVVNISFREKMTIPDLYFINLAVADLILVADSLIEVFNLH
- ERYYDIAVLCTFMSLFLQVNMYSSVFFLTWMSFDRYIALARAMRCSLFRTKHHARLSCGL
- IWMASVSATLVPFTAVHLQHTDEACFCFADVREVQWLEVTLGFIVPFAIIGLCYSLIVRV
- LVRAHRHRGLRPRRQKALRMILAVVLVFFVCWLPENVFISVHLLQRTQPGAAPCKQSFRH
- AHPLTGHIVNLAAFSNSCLNPLIYSFLGETFRDKLRLYIEQKTNLPALNRFCHAALKAVI
- PDSTEQSDVRFSSAV

CODE

In [48]:

```
from pyopenms import *
import matplotlib.pyplot as plt
def mirror_plot(obs_mz, obs_int, theo_mz, theo_int, title):
  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=3.0)
  plt.bar(theo_mz, theo_int, width=3.0)
  plt.title(title)
  plt.ylabel('intensity')
  plt.xlabel('m/z')
  plt.show()
protein_ids = []
peptide_ids = []
SimpleSearchEngineAlgorithm().search("cc3.mzML", "cancer.fasta", protein_ids, peptide_ids)
for peptide_id in peptide_ids:
    print ("_____")
    print ("Peptide Index m/z:", peptide_id.getMetaValue("scan_index"))
    for hit in peptide_id.getHits():
        print(" - Peptide hit sequence:", hit.getSequence())
        string= hit.getSequence().toString()
        tsg = TheoreticalSpectrumGenerator()
        theo_spec = MSSpectrum()
        p = Param()
        p.setValue("add_y_ions", "true")
        p.setValue("add_b_ions", "true")
        p.setValue("add_metainfo", "true")
        tsg.setParameters(p)
        peptide = AASequence.fromString(hit.getSequence().toString())
        tsg.getSpectrum(theo_spec, peptide, 1, 2)
        # Iterate over annotated ions and their masses
        print("Spectrum 1 of", peptide, "has", theo_spec.size(), "peaks.")
        for ion, peak in zip(theo_spec.getStringDataArrays()[0], theo_spec):
            print(ion.decode(), "is generated at m/z", peak.getMZ())
        exp = MSExperiment()
        MzMLFile().load("cc3.mzML", exp)
        spectra = exp.getSpectrum(peptide_id.getMetaValue("scan_index"))
        alignment = []
        spa = SpectrumAlignment()
        p = spa.getParameters()
        p.setValue("tolerance", 0.5)
        p.setValue("is_relative_tolerance", "false")
        spa.setParameters(p)
        # align both spectra
        spa.getSpectrumAlignment(alignment, theo_spec, spectra)

        # Print matching ions and mz from theoretical spectrum
        print("Number of matched peaks: " + str(len(alignment)))
        print("ion\ttheo. m/z\tobserved m/z")

        for theo_idx, obs_idx in alignment:
            ion_name = theo_spec.getStringDataArrays()[0][theo_idx].decode()
            ion_charge = theo_spec.getIntegerDataArrays()[0][theo_idx]
            print(ion_name + "\t" + str(ion_charge) + "\t"
                  + str(theo_spec[theo_idx].getMZ())
                  + "\t" + str(spectra[obs_idx].getMZ()))

        theo_mz, theo_int, obs_mz, obs_int = [], [], [], []
        for theo_idx, obs_idx in alignment:
            theo_mz.append(theo_spec[theo_idx].getMZ())
            theo_int.append(theo_spec[theo_idx].getIntensity())
            obs_mz.append(spectra[obs_idx].getMZ())
            obs_int.append(spectra[obs_idx].getIntensity())
        title = f'{string},{peptide_id.getMetaValue("scan_index")}'
        flag=False
        for i in obs_int:
            if(i>0):
                flag=True
                break
        if(flag==True):
           mirror_plot(obs_mz, obs_int, theo_mz, theo_int, title)
        else:
           flag=False
```

MIRROR PLOT FUNCTION

```
from pyopenms import *
import matplotlib.pyplot as plt
def mirror_plot(obs_mz, obs_int, theo_mz, theo_int, title):
  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=3.0)
  plt.bar(theo_mz, theo_int, width=3.0)
  plt.title(title)
  plt.ylabel('intensity')
  plt.xlabel('m/z')
  plt.show()
```

SIMPLE SEARCH ENGINE

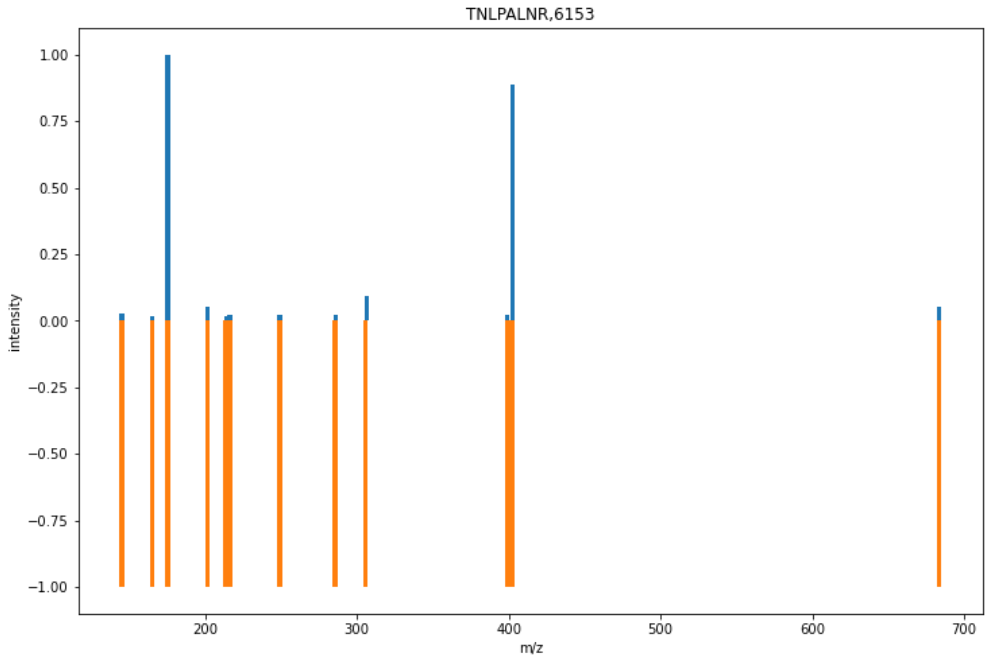
```
peptide_ids = []
SimpleSearchEngineAlgorithm().search("cc3.mzML", "cancer.fasta", protein_ids, peptide_ids)
for peptide_id in peptide_ids:
    print ("_____")
    print ("Peptide Index m/z:", peptide_id.getMetaValue("scan_index"))
    for hit in peptide_id.getHits():
        print(" - Peptide hit sequence:", hit.getSequence())
        string= hit.getSequence().toString()
        tsg = TheoreticalSpectrumGenerator()
        theo_spec = MSSpectrum()
        p = Param()
        p.setValue("add_y_ions", "true")
        p.setValue("add_b_ions", "true")
        p.setValue("add_metainfo", "true")
```

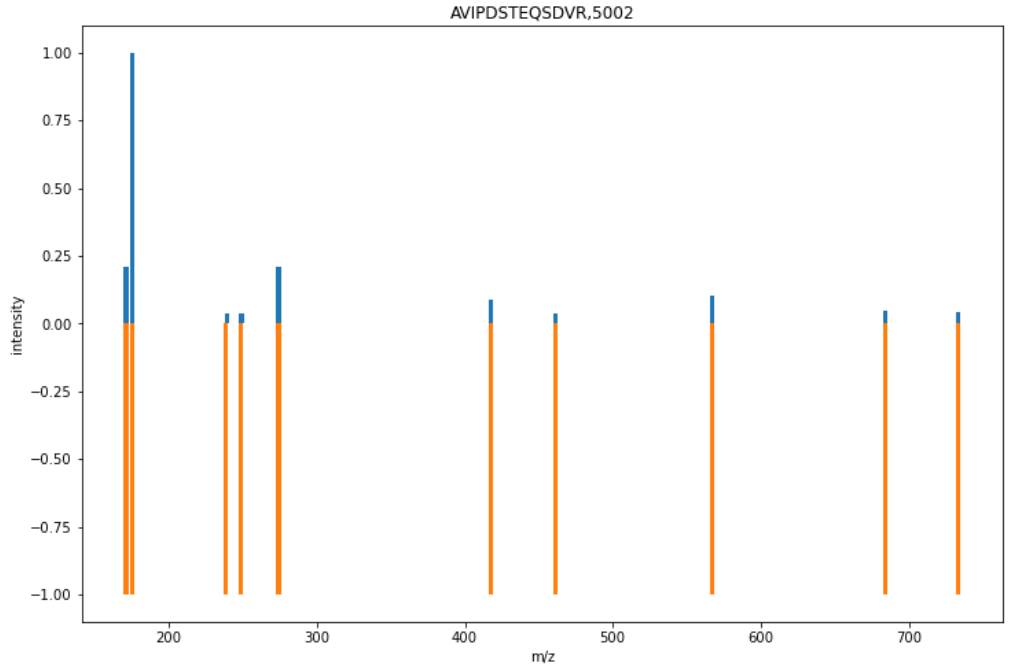
```
_____
Peptide Index m/z: 11517
 - Peptide hit sequence: YYDIAVLC(Carbamidomethyl)TFMSLFLQVNM(Oxidation)YSSVFFLTWM(Oxidation)SFDR
Spectrum 1 of YYDIAVLC(Carbamidomethyl)TFMSLFLQVNM(Oxidation)YSSVFFLTWM(Oxidation)SFDR has 126 peaks.
y1++ is generated at m/z 88.06311469007102
y2++ is generated at m/z 145.576586769821
b2++ is generated at m/z 164.07060575387098
y1+ is generated at m/z 175.11895291337l
y3++ is generated at m/z 219.110793913371
b3++ is generated at m/z 221.58407783362097
y4++ is generated at m/z 262.62680849312096
b4++ is generated at m/z 278.12611000907094
y2+ is generated at m/z 290.1458970728710l
b5++ is generated at m/z 313.64466708882094
b2+ is generated at m/z 327.133935040971
y5++ is generated at m/z 336.14450850167094
b6++ is generated at m/z 363.17887423237096
b7++ is generated at m/z 419.72090640782096
y6++ is generated at m/z 429.18416516117094
y3+ is generated at m/z 437.214311359971
b3+ is generated at m/z 442.16087920047096
y7++ is generated at m/z 479.7080047728209
b8++ is generated at m/z 499.736230900421
y4+ is generated at m/z 524.246340519471
y8++ is generated at m/z 536.2500369482709
b9++ is generated at m/z 550.260070512071
b4+ is generated at m/z 555.244943551371
y9++ is generated at m/z 609.7842440918209
b10++ is generated at m/z 623.794277655621
b5+ is generated at m/z 626.282057710871
y5+ is generated at m/z 671.2817405365709
y10++ is generated at m/z 683.318451235371
b11++ is generated at m/z 689.314520164171
b6+ is generated at m/z 725.350471997971
b12++ is generated at m/z 732.830534743921
```

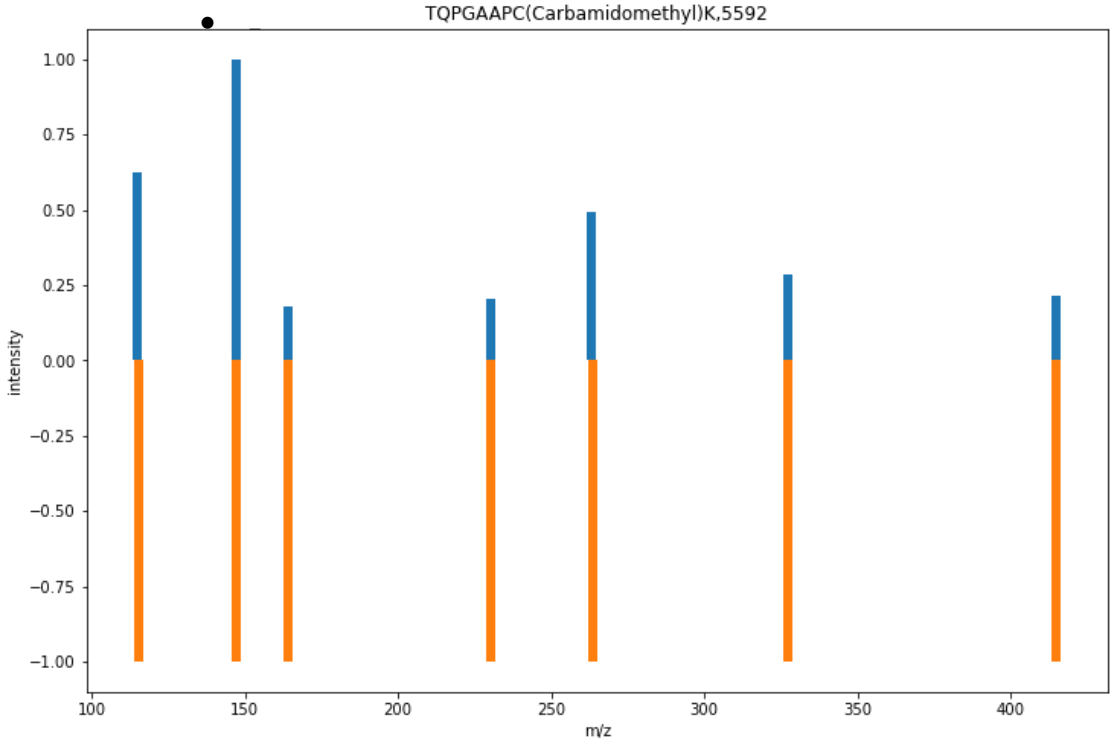
EXPERIMENNTAL AND ALIGN PEAKS

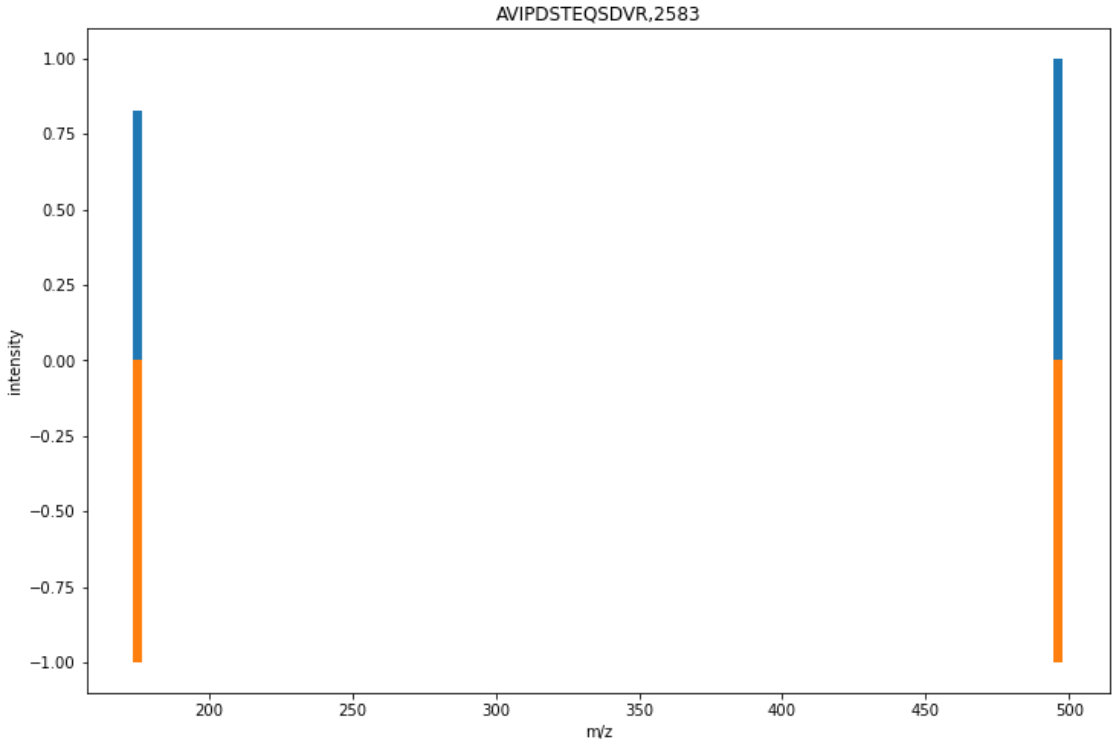
```
# Iterate over annotated ions and their masses
print("Spectrum 1 of", peptide, "has", theo_spec.size(), "peaks.")
for ion, peak in zip(theo_spec.getStringDataArrays()[0], theo_spec):
    print(ion.decode(), "is generated at m/z", peak.getMZ())
exp = MSExperiment()
MzMLFile().load("cc3.mzML", exp)
spectra = exp.getSpectrum(peptide_id.getMetaValue("scan_index"))
alignment = []
spa = SpectrumAlignment()
p = spa.getParameters()
p.setValue("tolerance", 0.5)
p.setValue("is_relative_tolerance", "false")
spa.setParameters(p)
# align both spectra
spa.getSpectrumAlignment(alignment, theo_spec, spectra)

# Print matching ions and mz from theoretical spectrum
print("Number of matched peaks: " + str(len(alignment)))
print("ion\ttheo. m/z\tobserved m/z")

for theo_idx, obs_idx in alignment:
    ion_name = theo_spec.getStringDataArrays()[0][theo_idx].decode()
    ion_charge = theo_spec.getIntegerDataArrays()[0][theo_idx]
    print(ion_name + "\t" + str(ion_charge) + "\t"
          + str(theo_spec[theo_idx].getMZ())
          + "\t" + str(spectra[obs_idx].getMZ()))

theo_mz, theo_int, obs_mz, obs_int = [], [], [], []
for theo_idx, obs_idx in alignment:
    theo_mz.append(theo_spec[theo_idx].getMZ())
    theo_int.append(theo_spec[theo_idx].getIntensity())
    obs_mz.append(spectra[obs_idx].getMZ())
    obs_int.append(spectra[obs_idx].getIntensity())
title = f'{string},{peptide_id.getMetaValue("scan_index")}'
flag=False
for i in obs_int:
    if(i>0):
        flag=True
        break
if(flag==True):
   mirror_plot(obs_mz, obs_int, theo_mz, theo_int, title)
else:
   flag=False
```

RAW FILE 1 (AVIPDSTEQSDVR),(TNLPALNR)

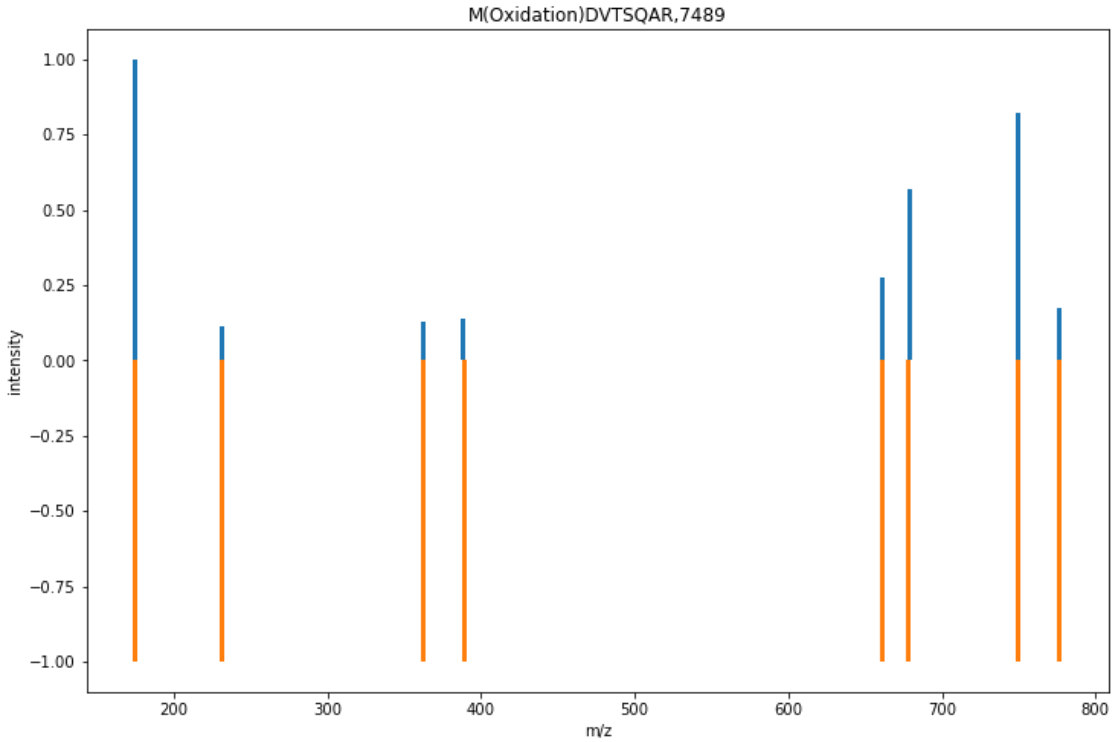


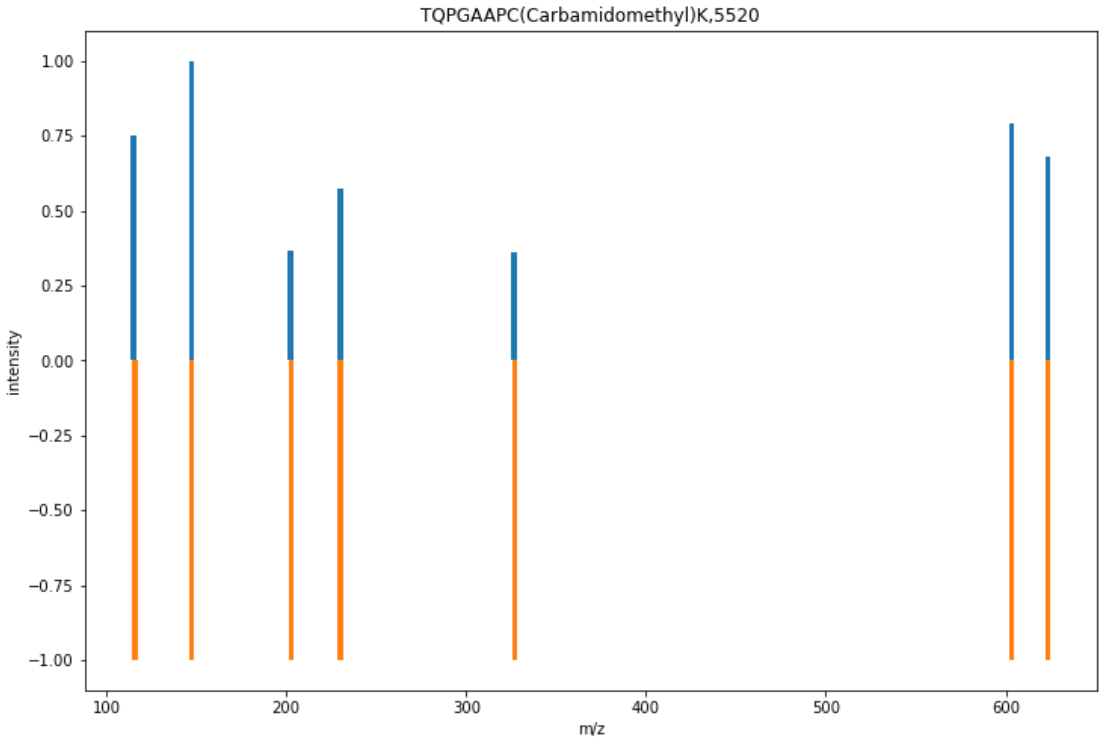






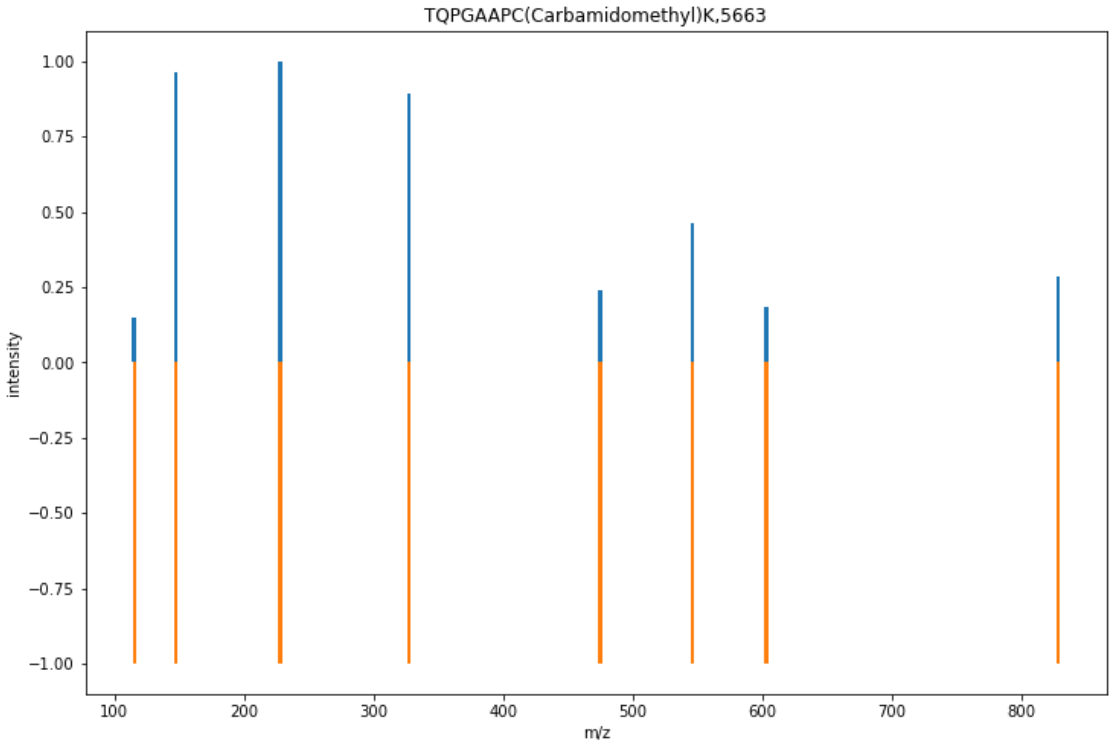
RAW FILE 1 (TQPGAAPCK),(AVIPDSTEQSDVR)

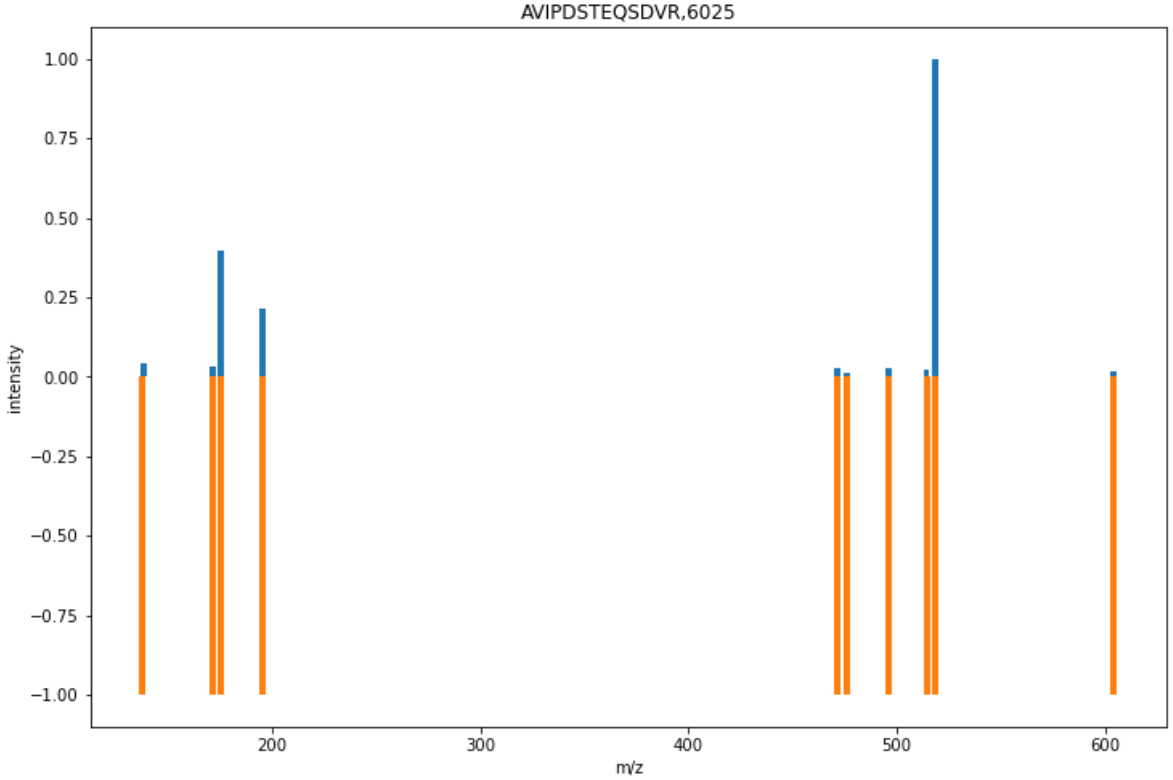




RAW FILE 2

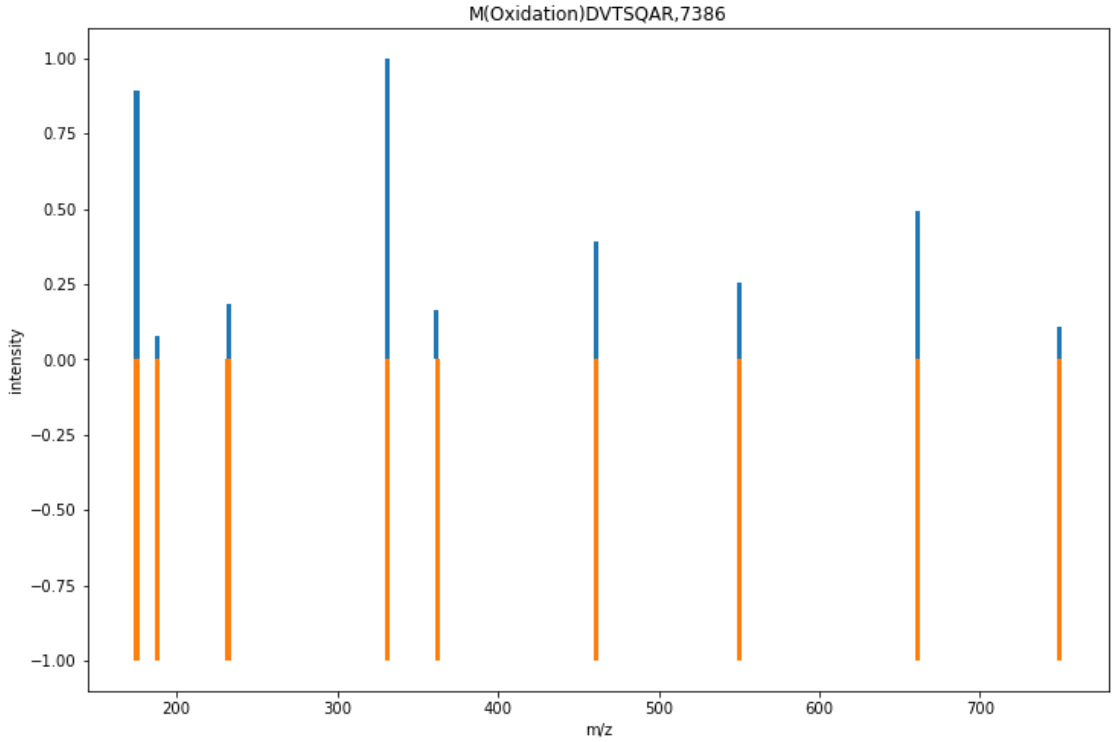
(MDVTSQAR),(TQPGAAPC(CARBAMIDOMETHYL)K)

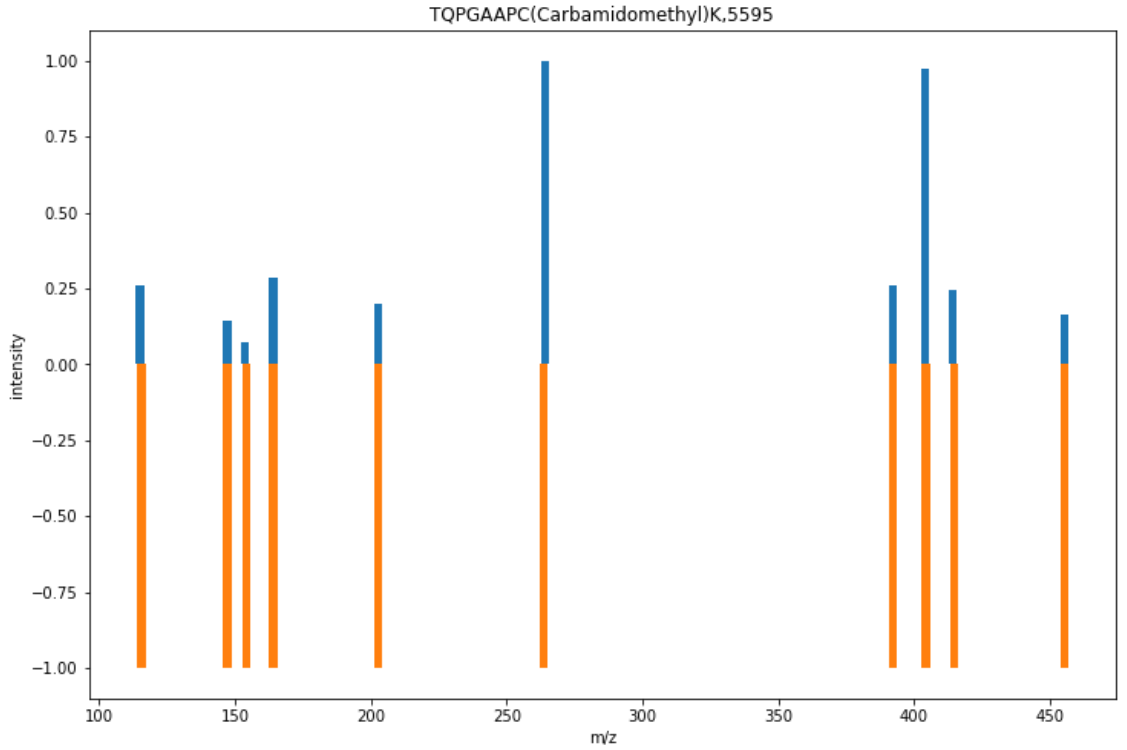




RAW FILE3

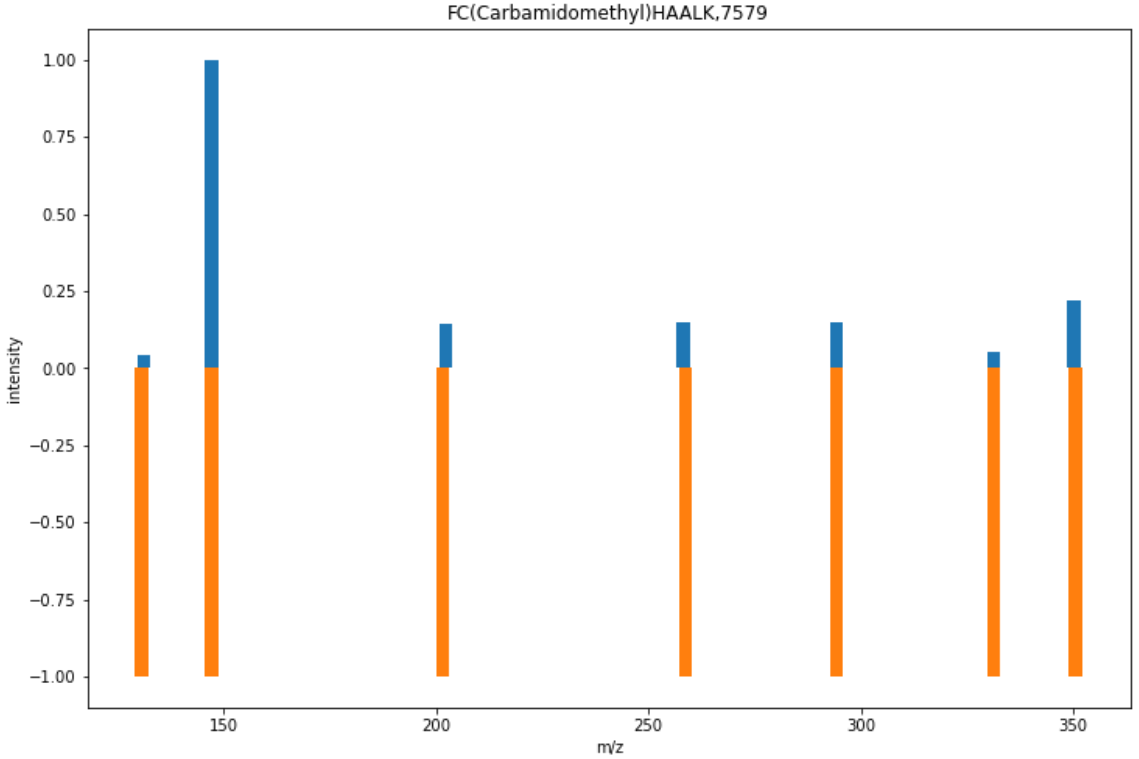
(AVIPDSTEQSDVR),(TQPGAAPC(CARBAMIDOMETHYL)K)

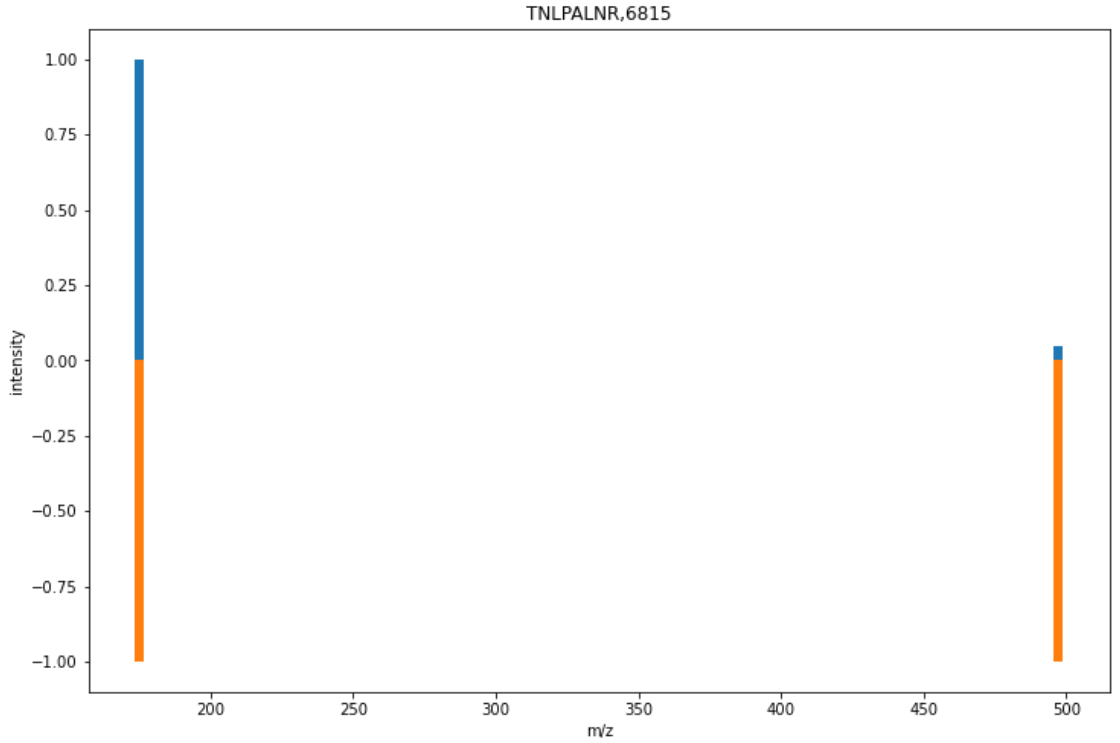




RAW FILE 4

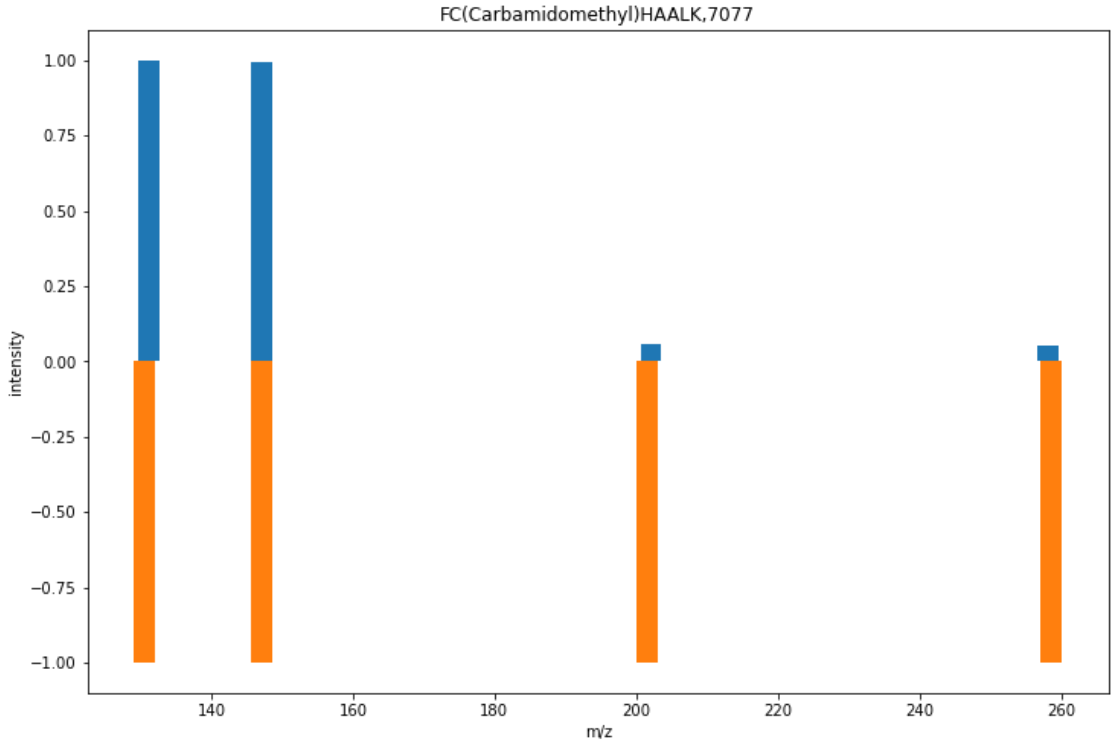
(M(OXIDATION)DVTSQAR),(TQPGAAPC(CARBAMIDOMETHYL)K)

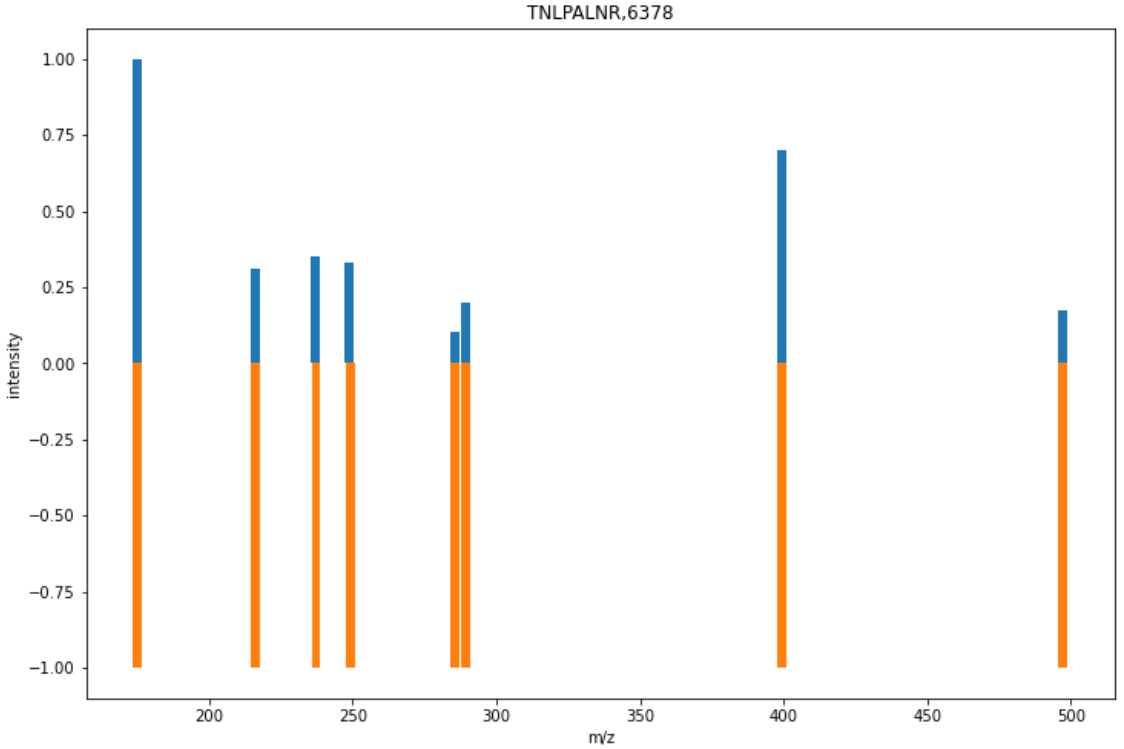




RAW FILE 4

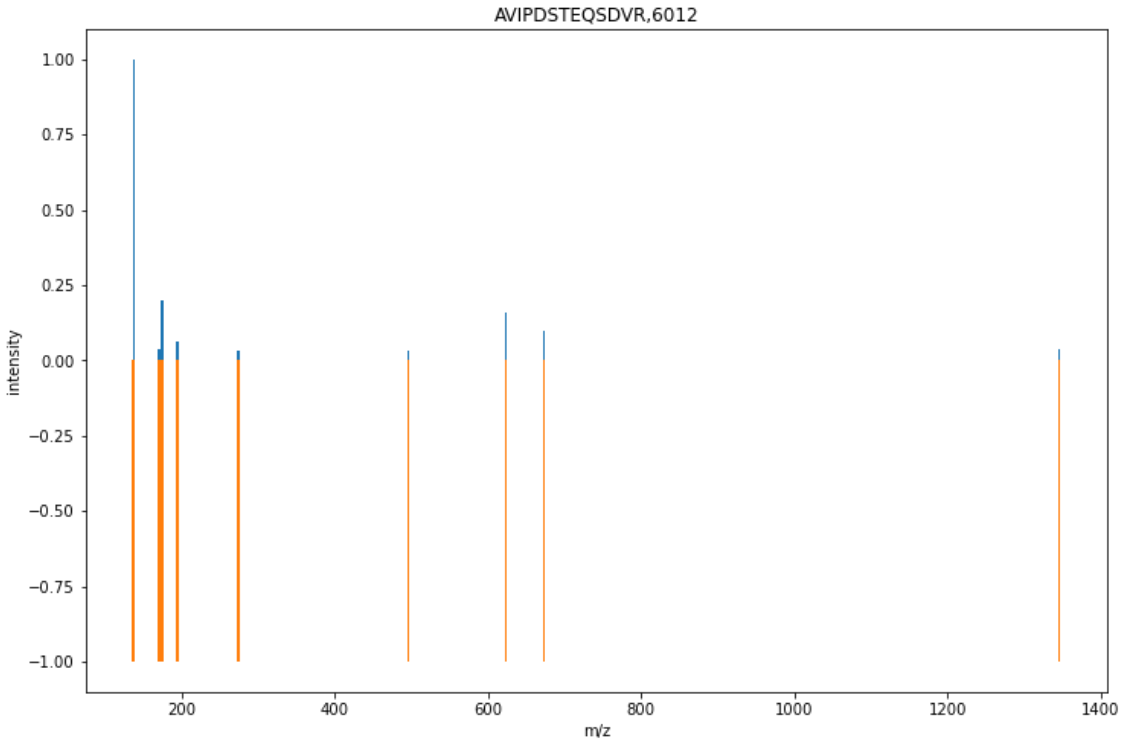
(FC(CARBAMIDOMETHYL)HAALK),(TNLPALNR)

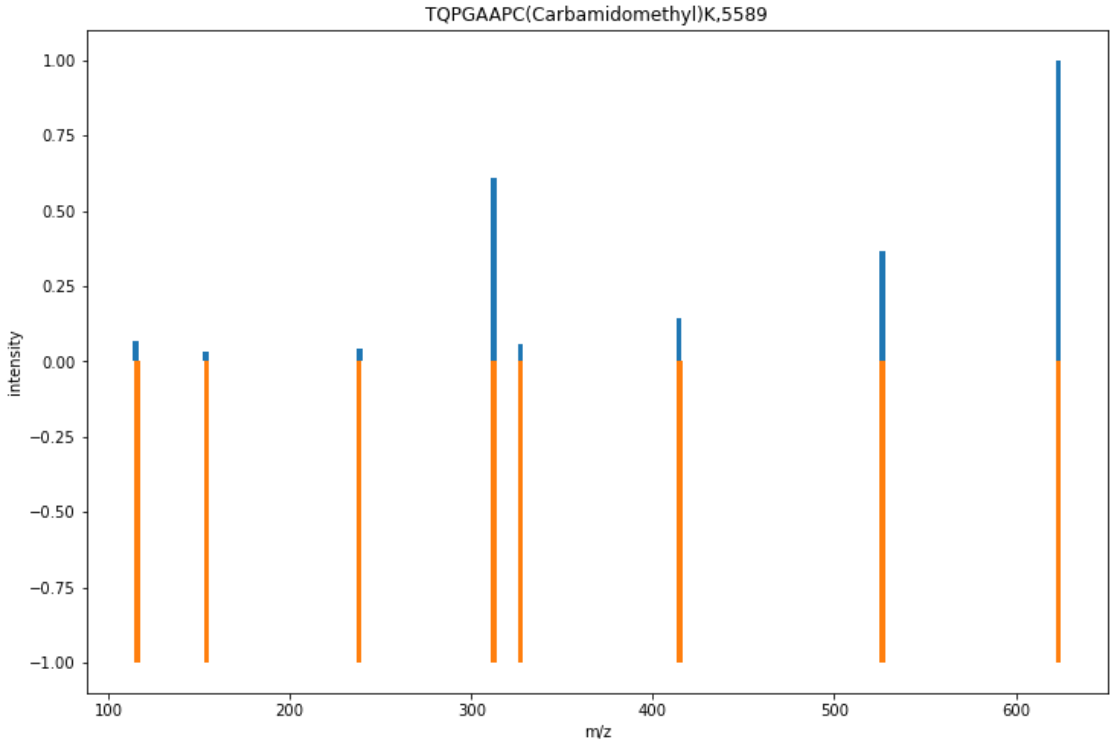




RAW FILE 5

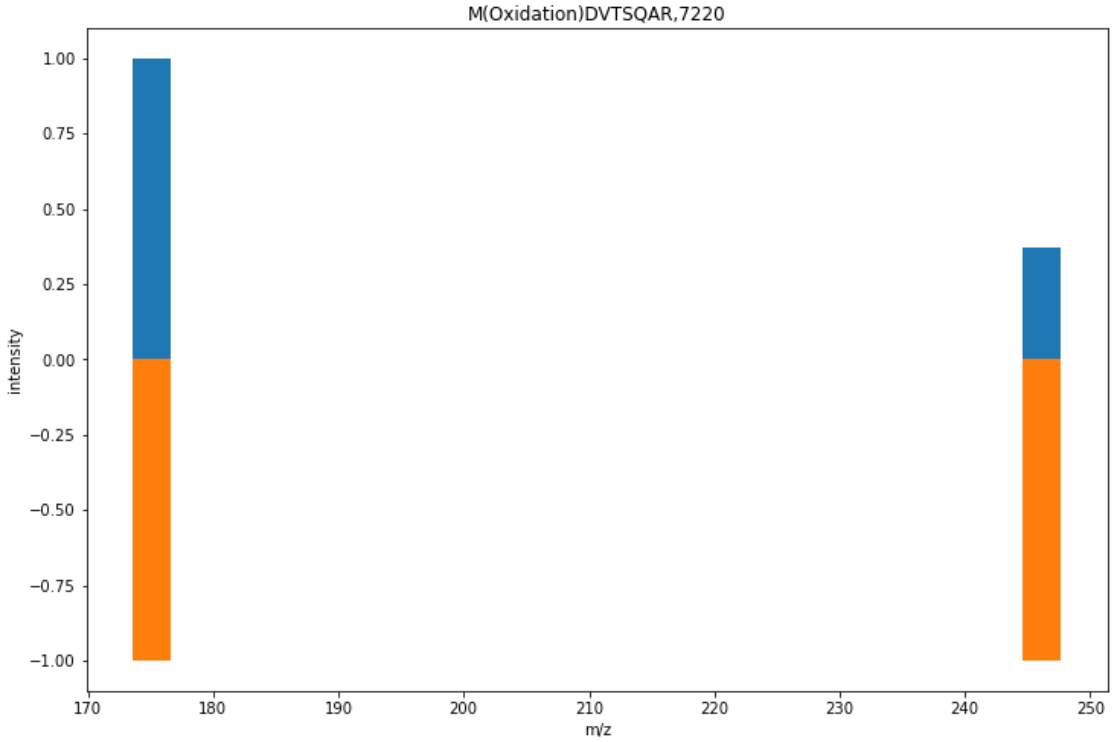
(FC(CARBAMIDOMETHYL)HAALK),(TNLPALNR)

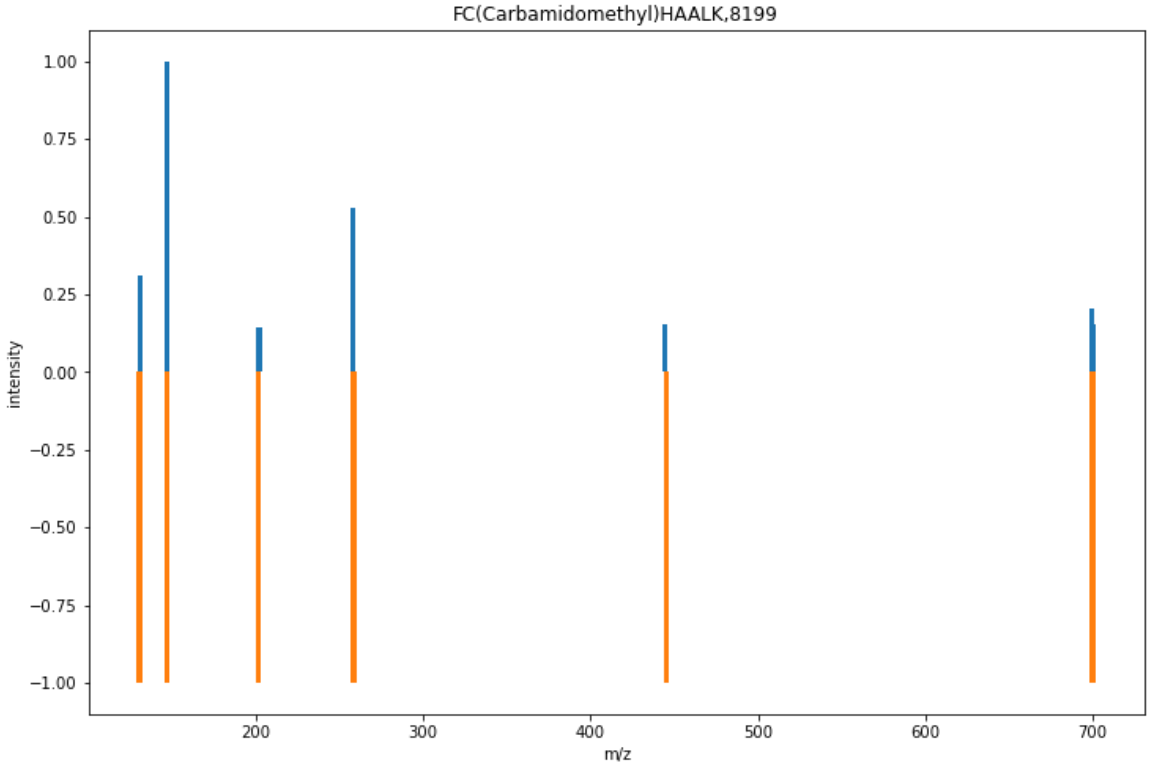




RAW FILE 5

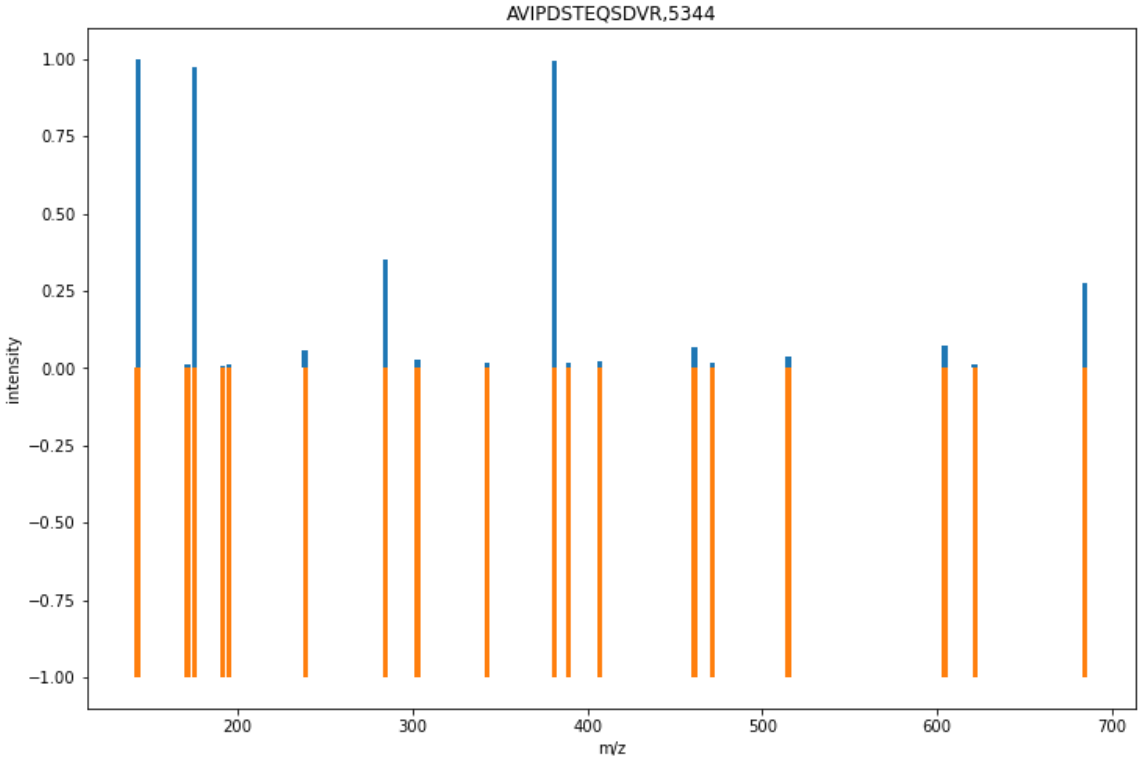
(AVIPDSTEQSDVR),(TQPGAAPC(CARBAMIDOMETHYL)K)

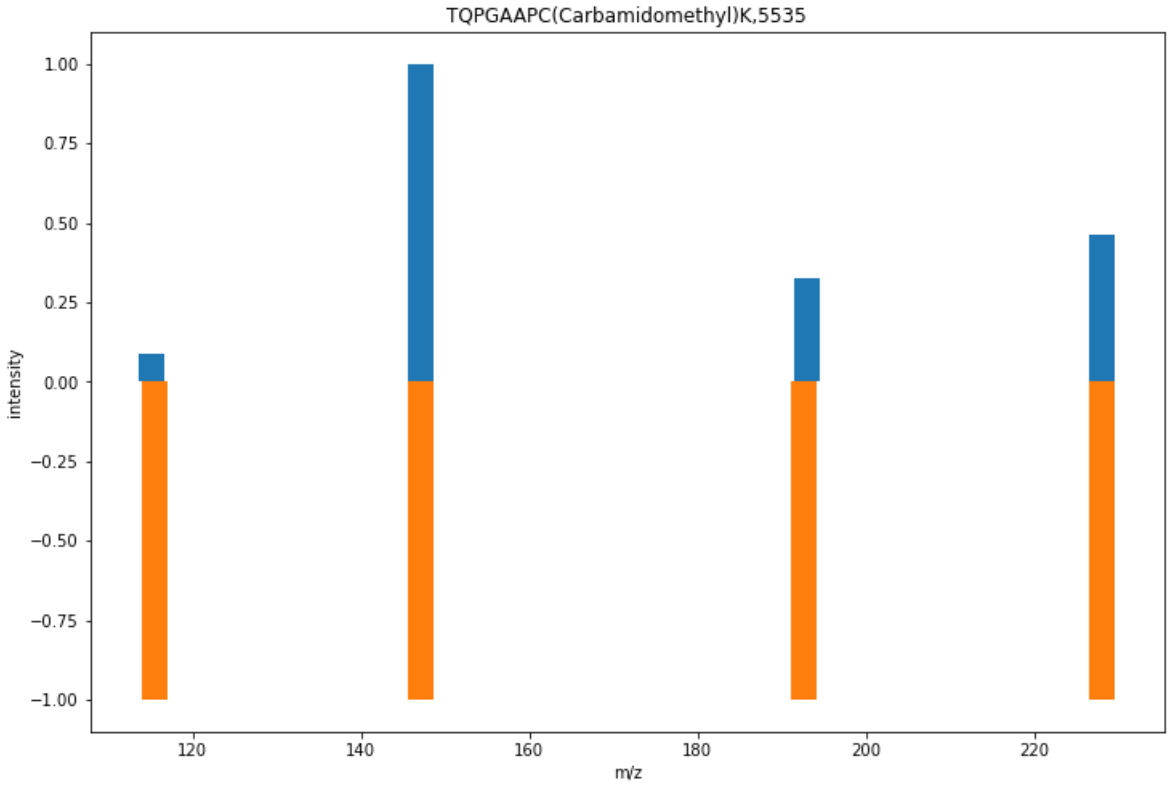




RAW FILE 5

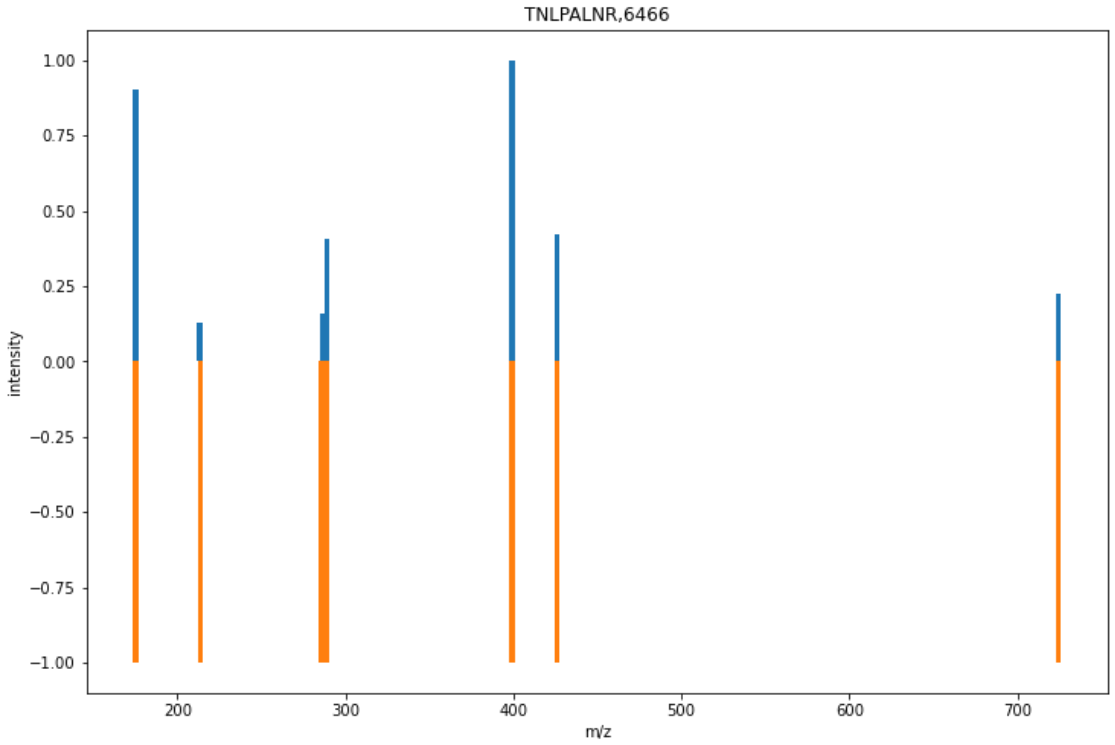
(M(OXIDATION)DVTSQAR),(FC(CARBAMIDOMETHYL)HAALK)

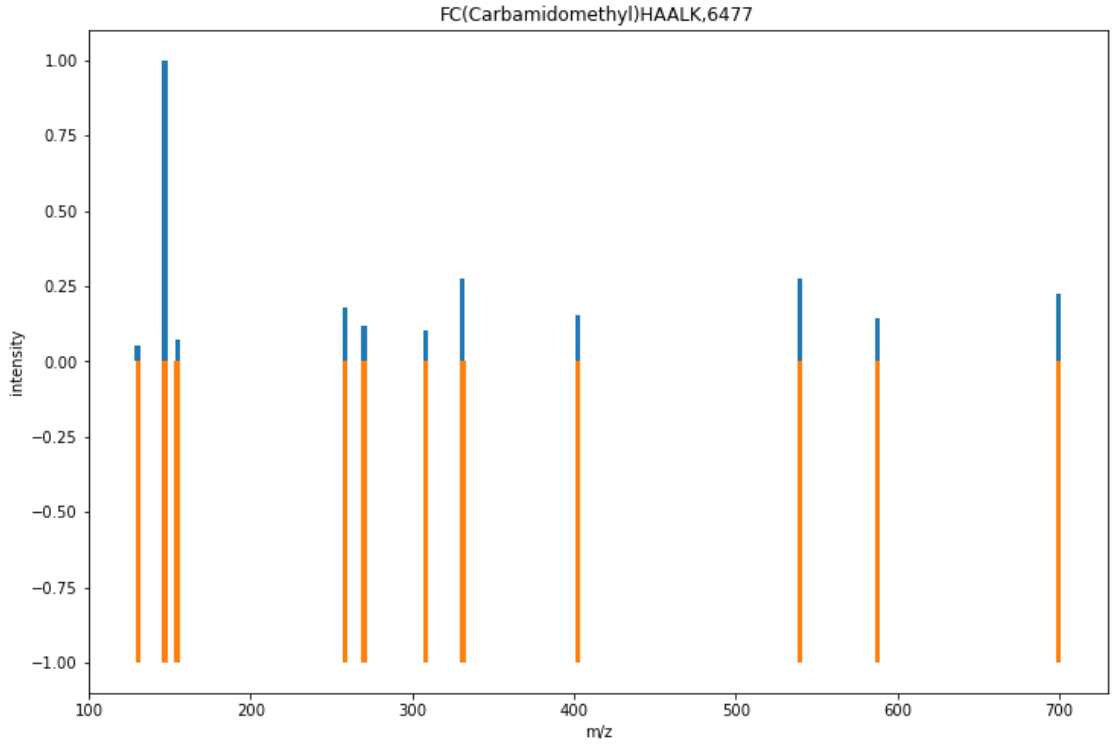




RAW FILE 6

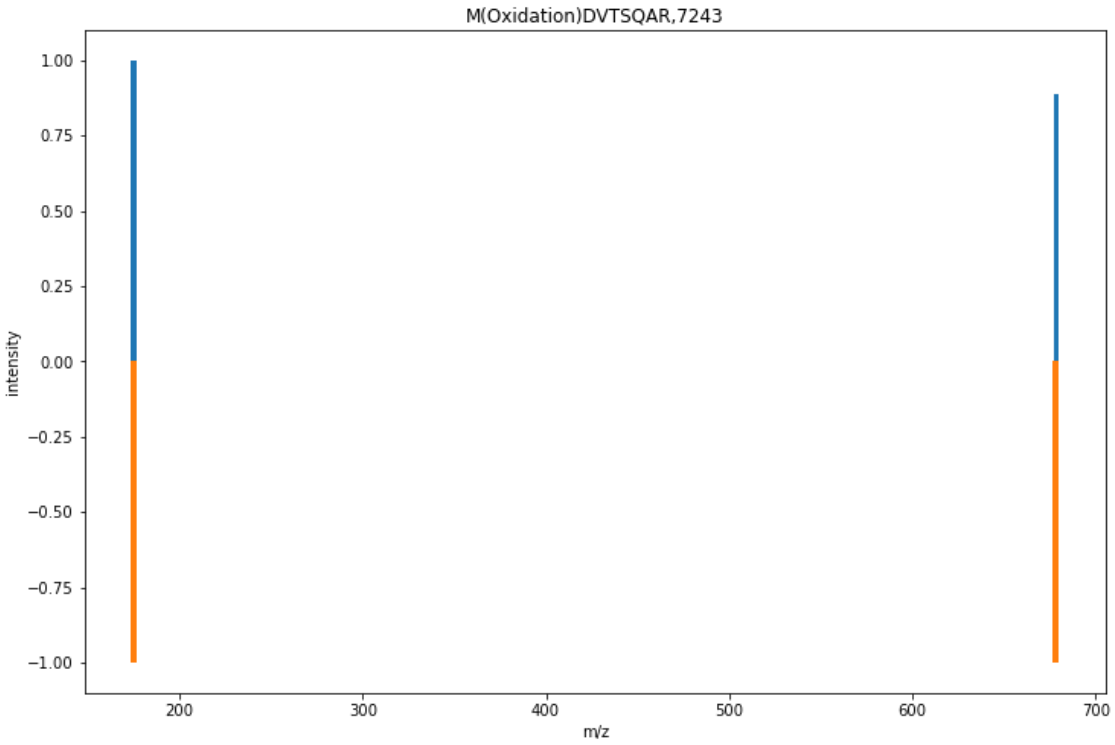
(AVIPDSTEQSDVR),(TQPGAAPC(CARBAMIDOMETHYL)K)

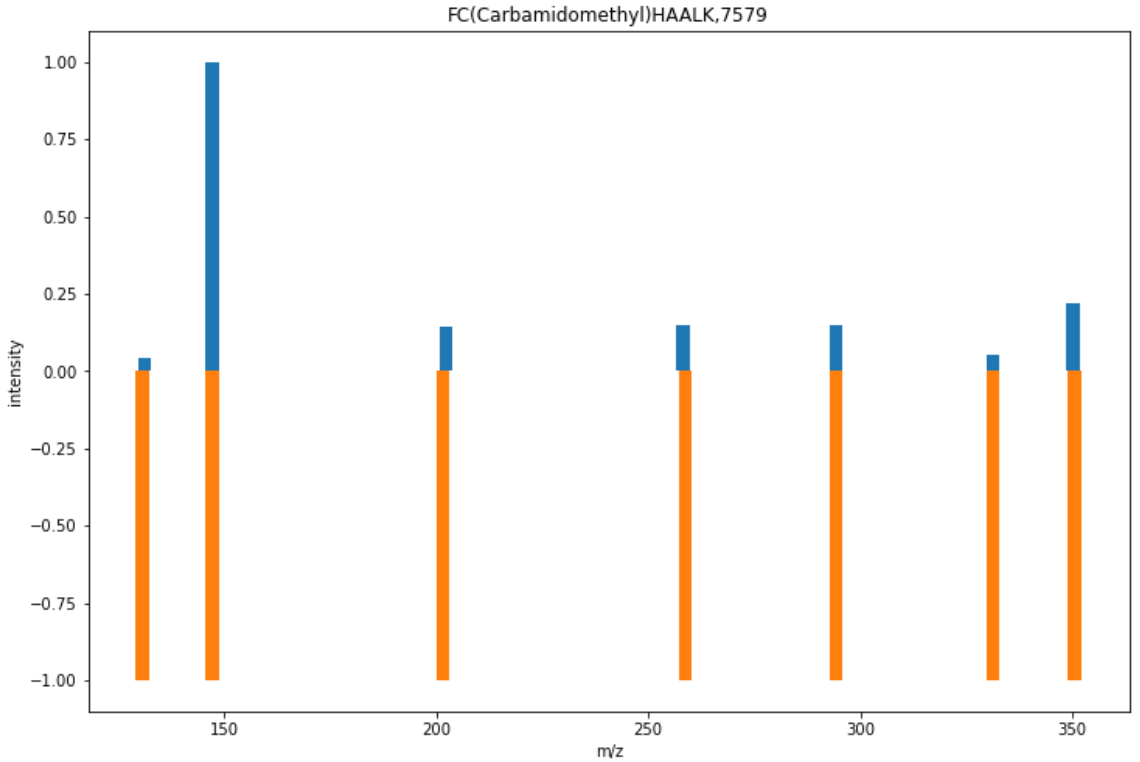




RAW FILE 6

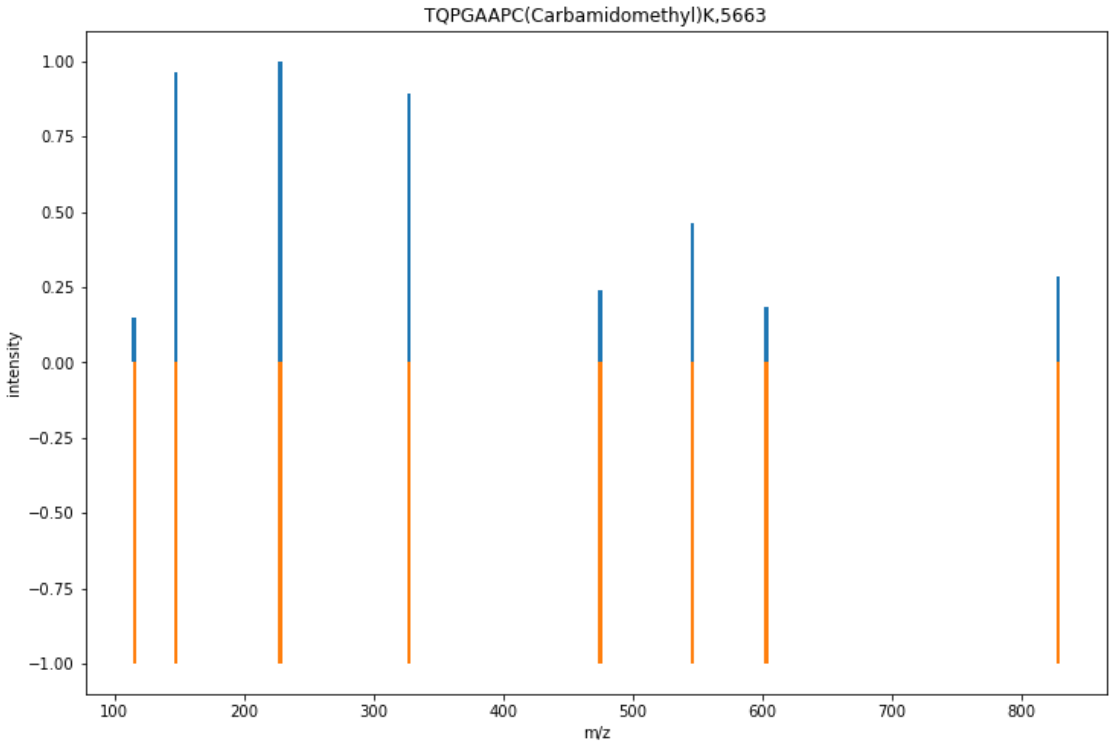
(TNLPALNR),(FC(CARBAMIDOMETHYL)HAALK)

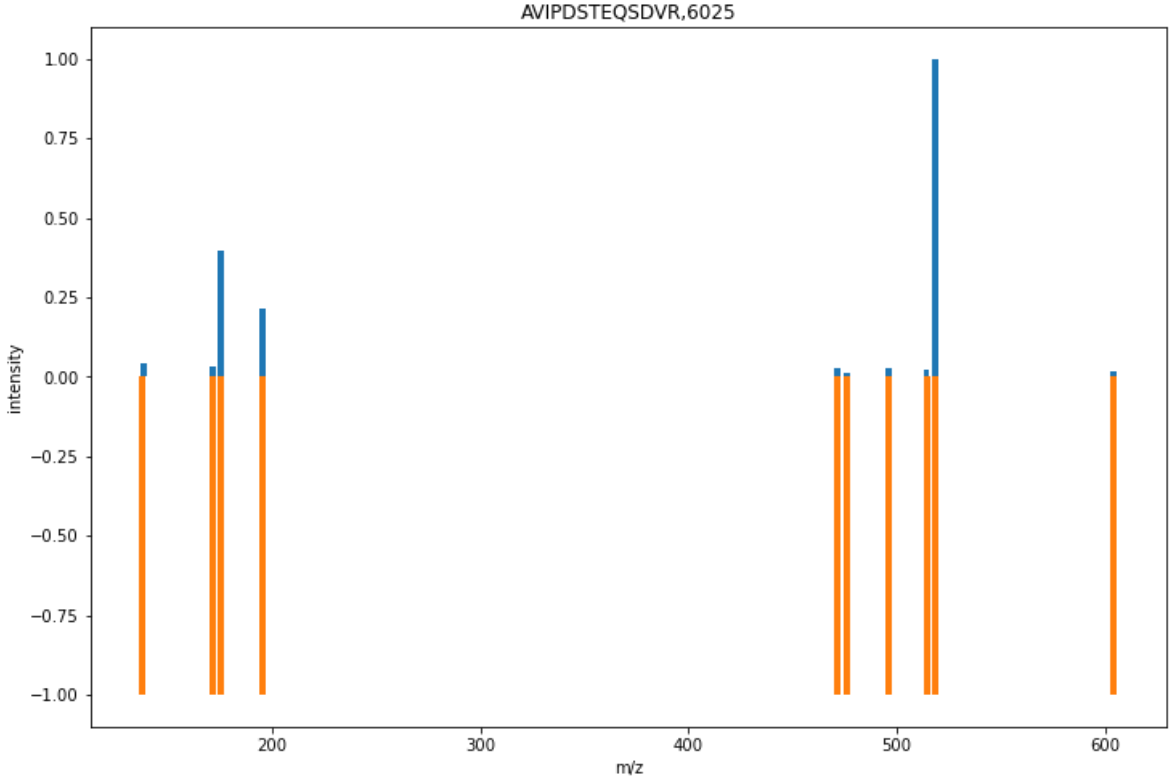




RAW FILE 6

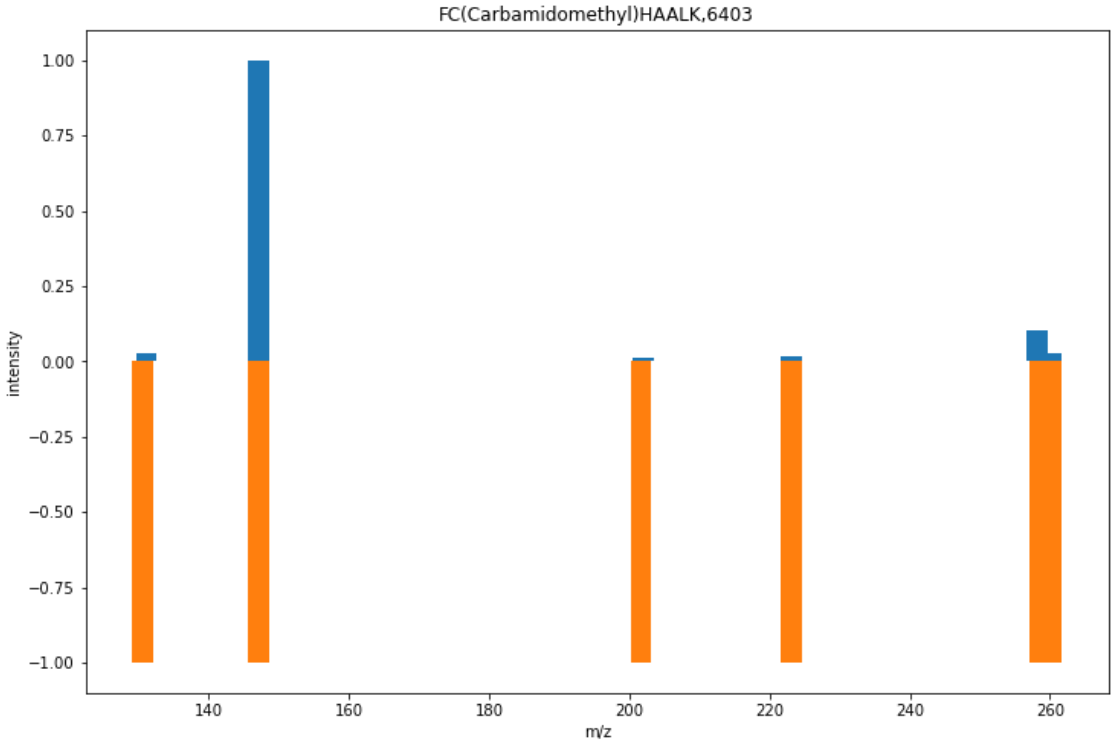
(M(OXIDATION)DVTSQAR),(FC(CARBAMIDOMETHYL)HAALK)

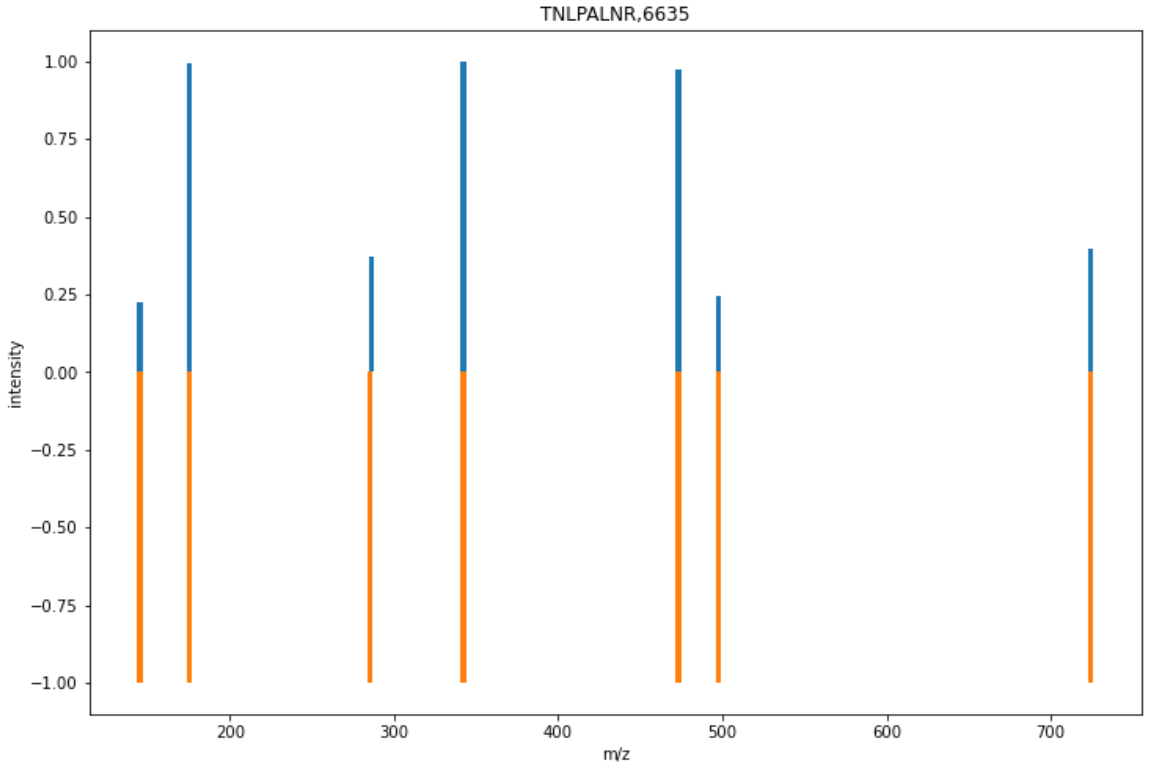




RAW FILE 7

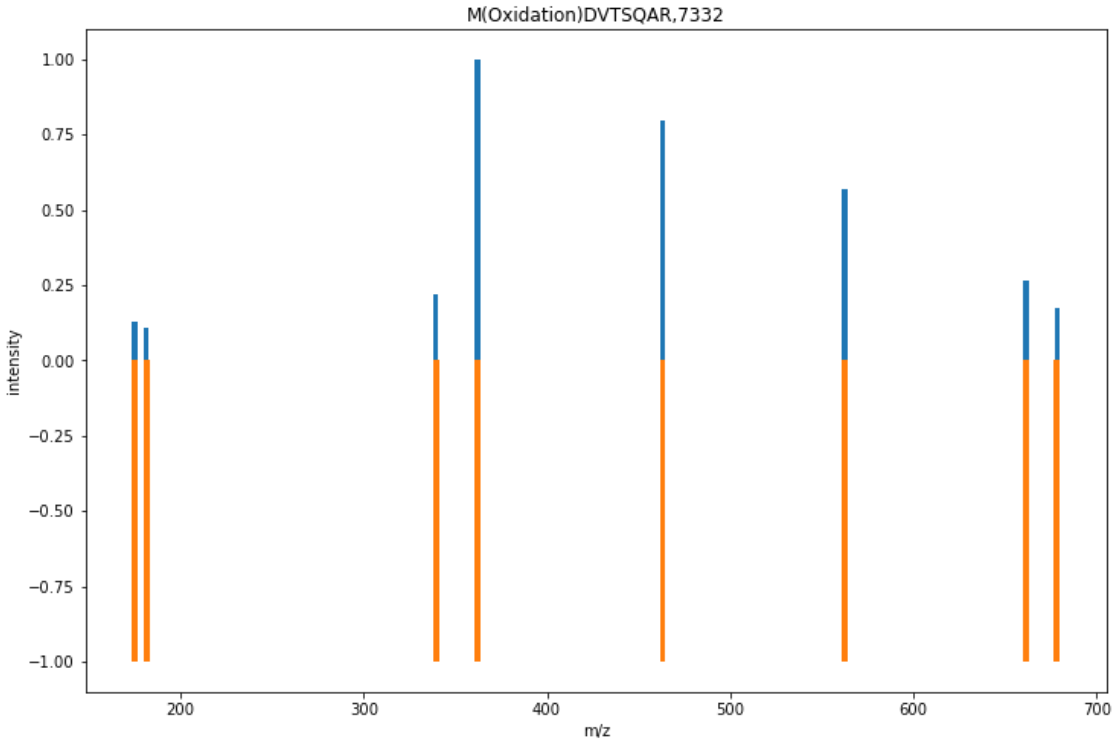
(TQPGAAPC(CARBAMIDOMETHYL)K),(AVIPDSTEQSDVR)

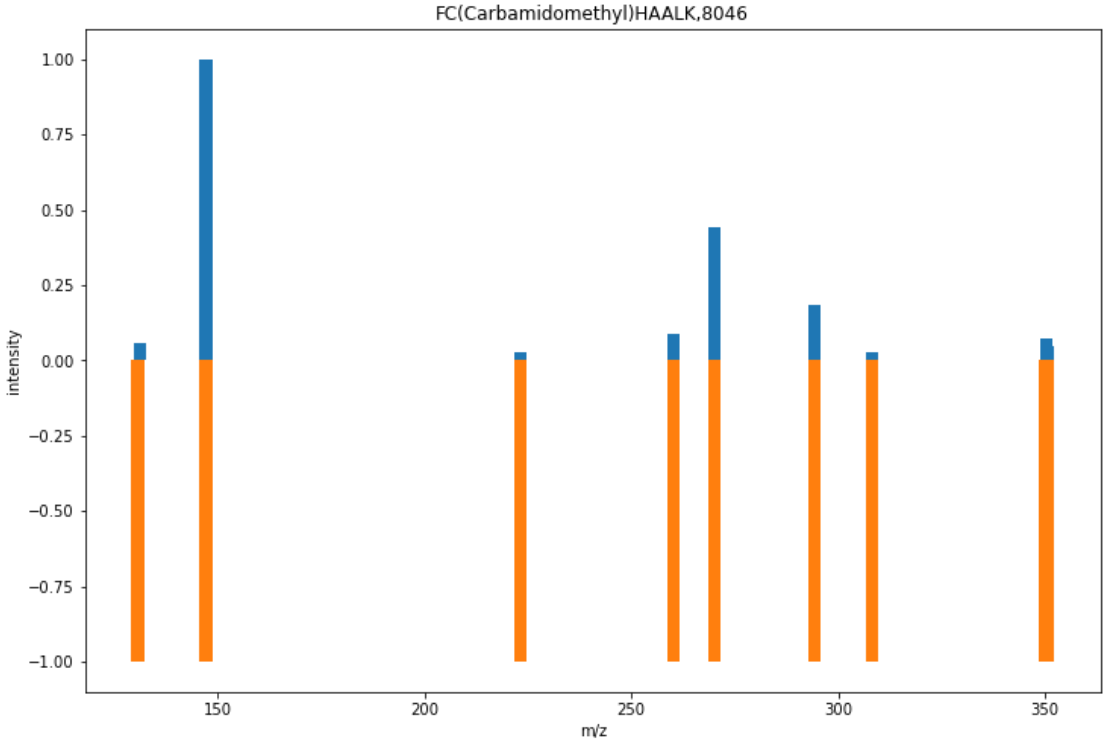




RAW FILE 7

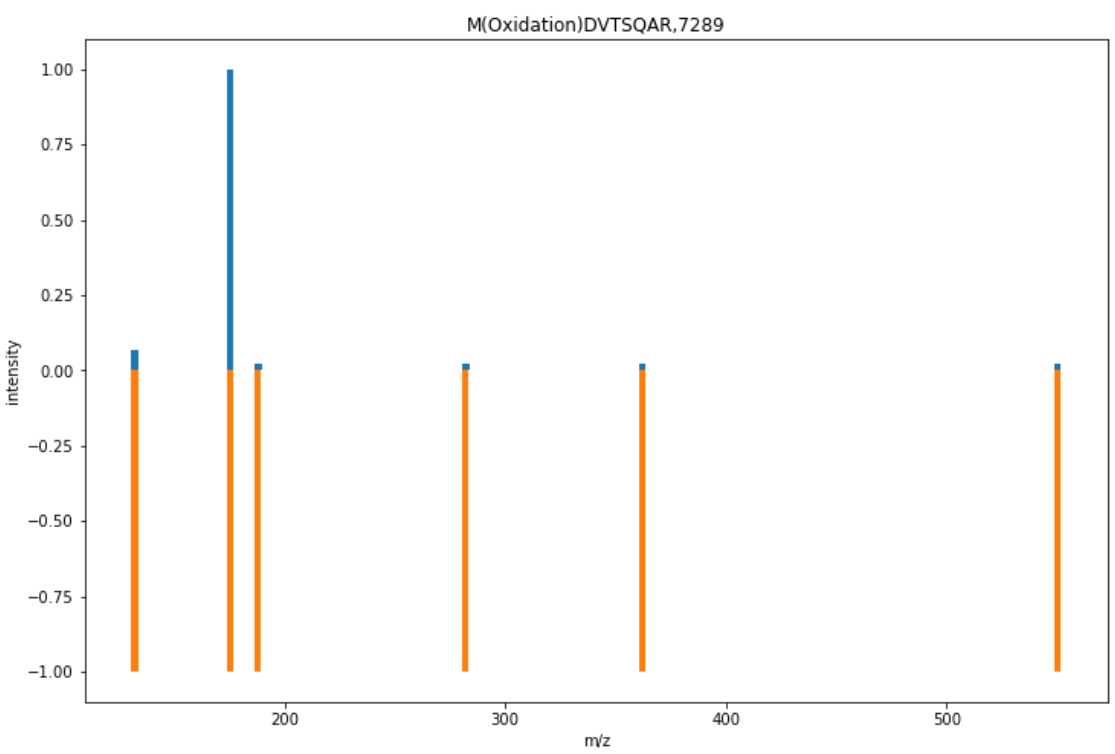
(FC(CARBAMIDOMETHYL)HAALK),(TNLPALNR)

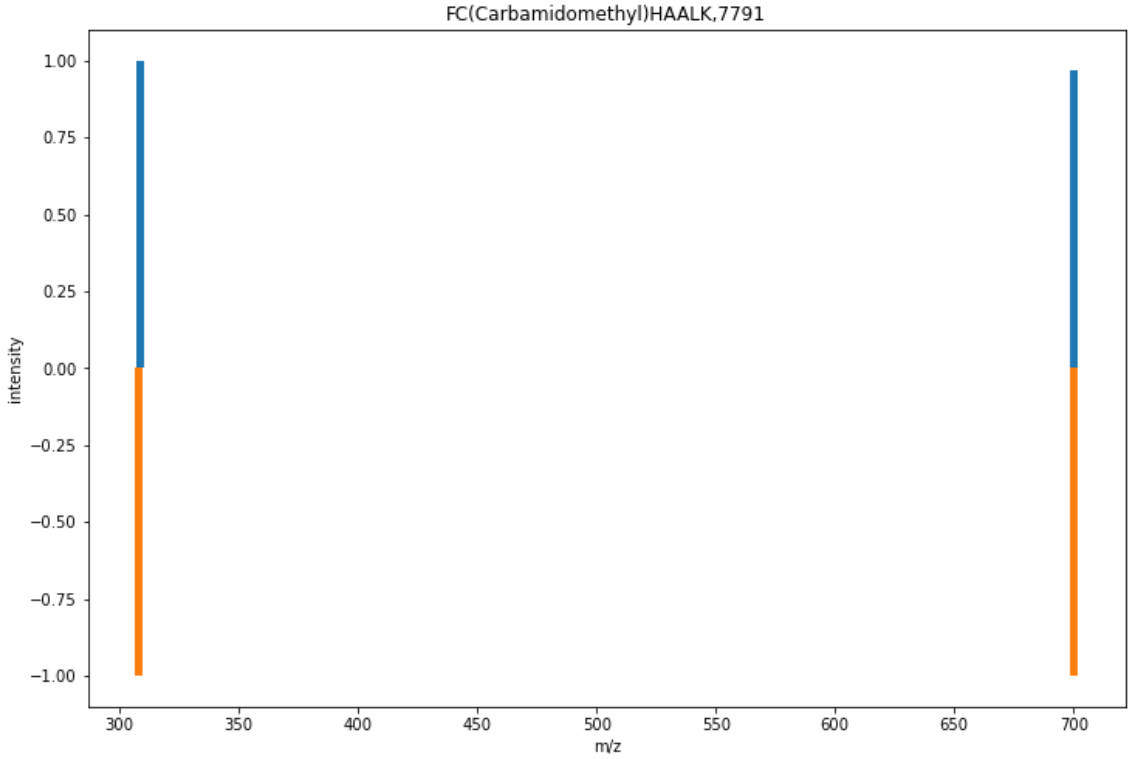




RAW FILE 7

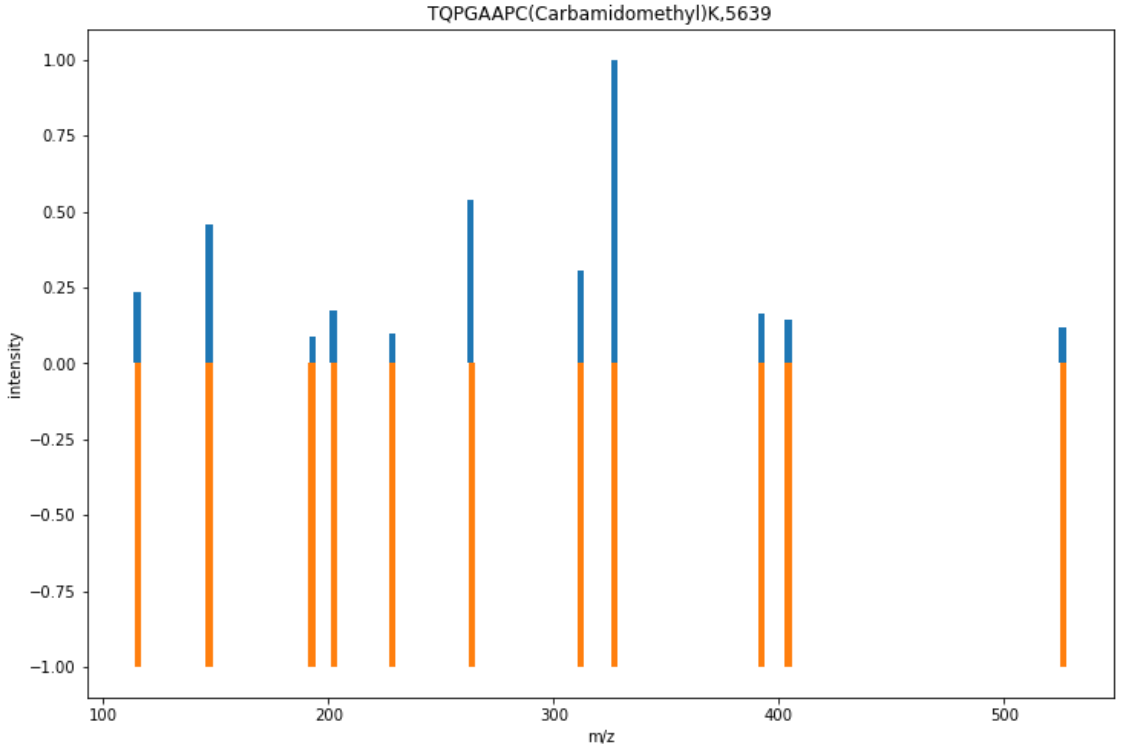
(M(OXIDATION)DVTSQAR),(FC(CARBAMIDOMETHYL)HAALK)

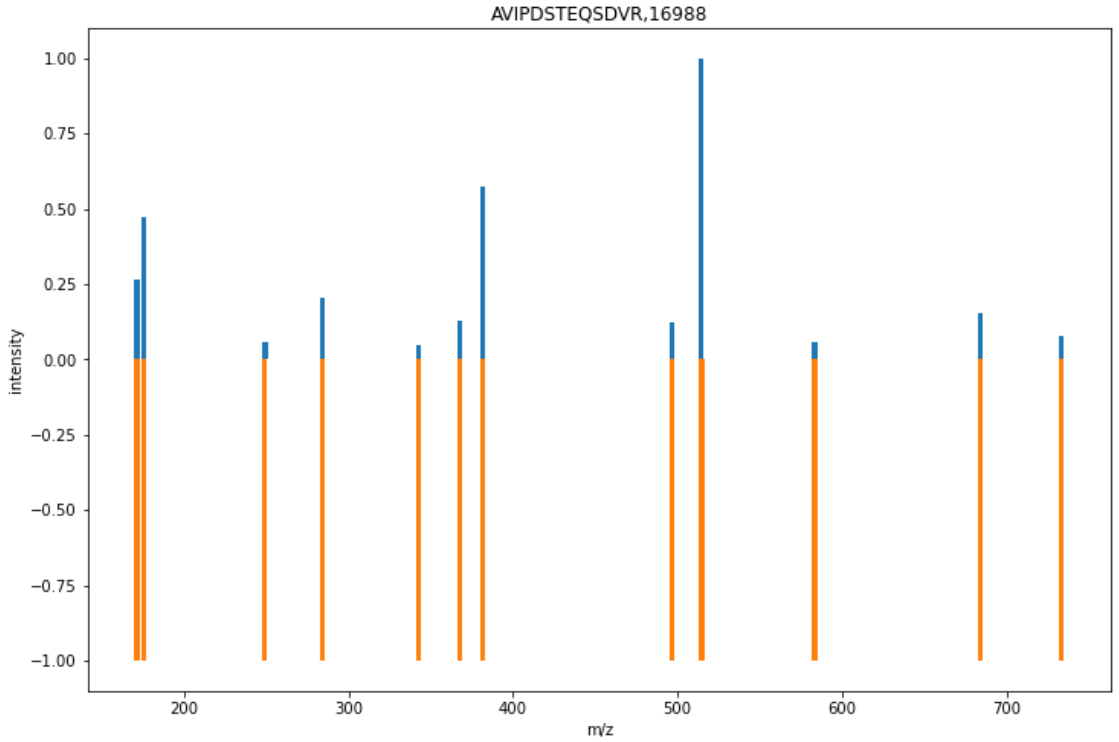




RAW FILE 8

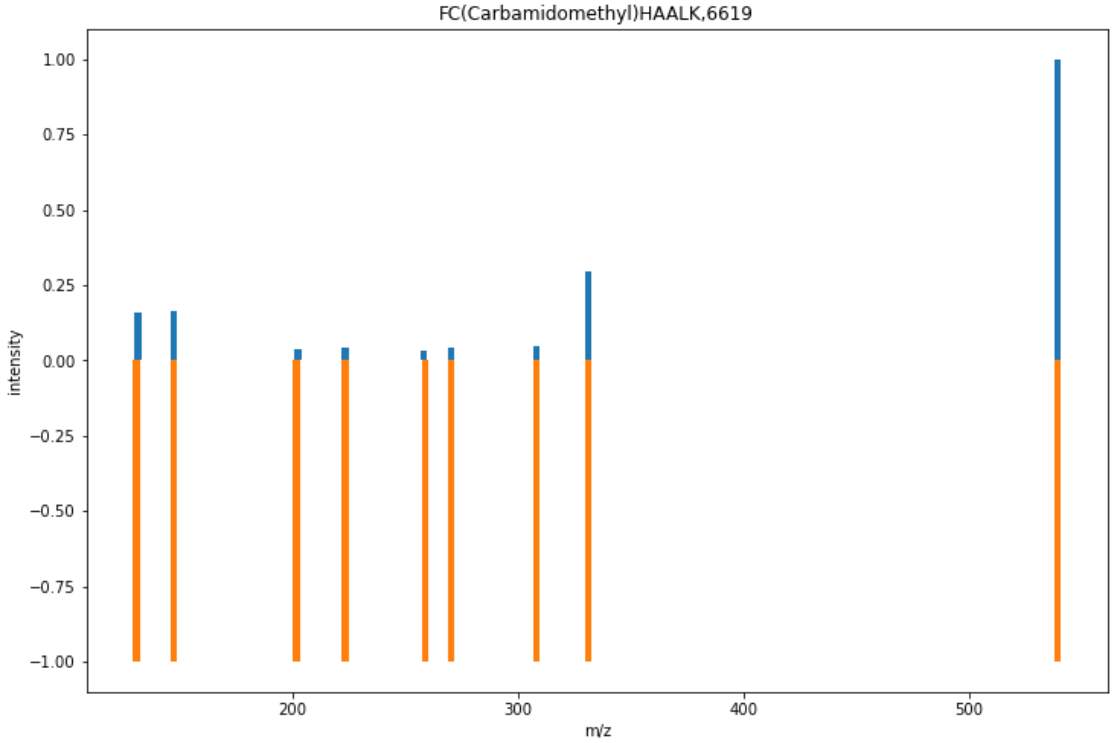
(M(OXIDATION)DVTSQAR),(FC(CARBAMIDOMETHYL)HAALK)

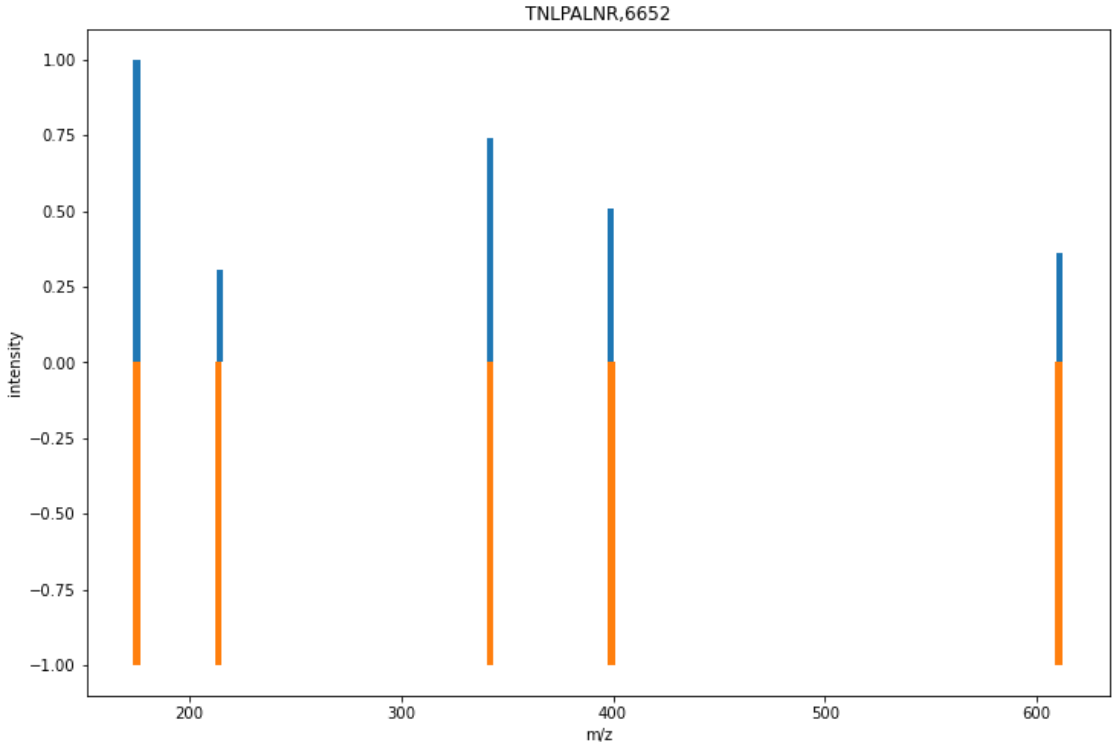




RAW FILE 8

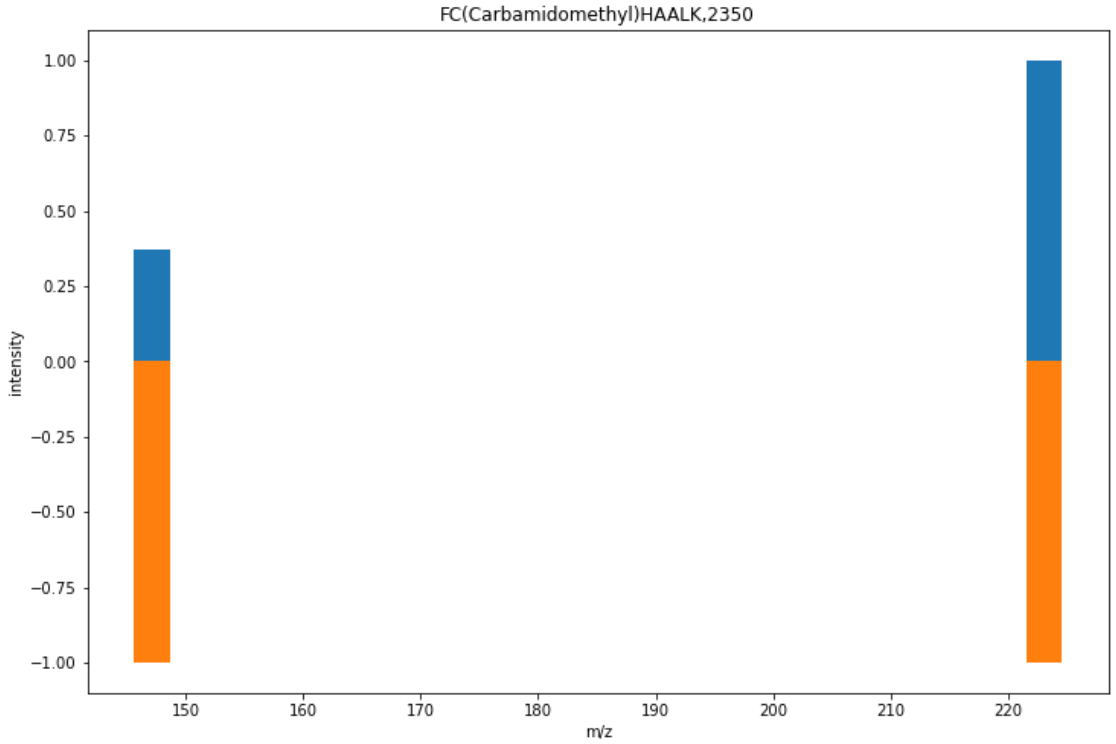
(TQPGAAPC(CARBAMIDOMETHYL)K),(AVIPDSTEQSDVR)

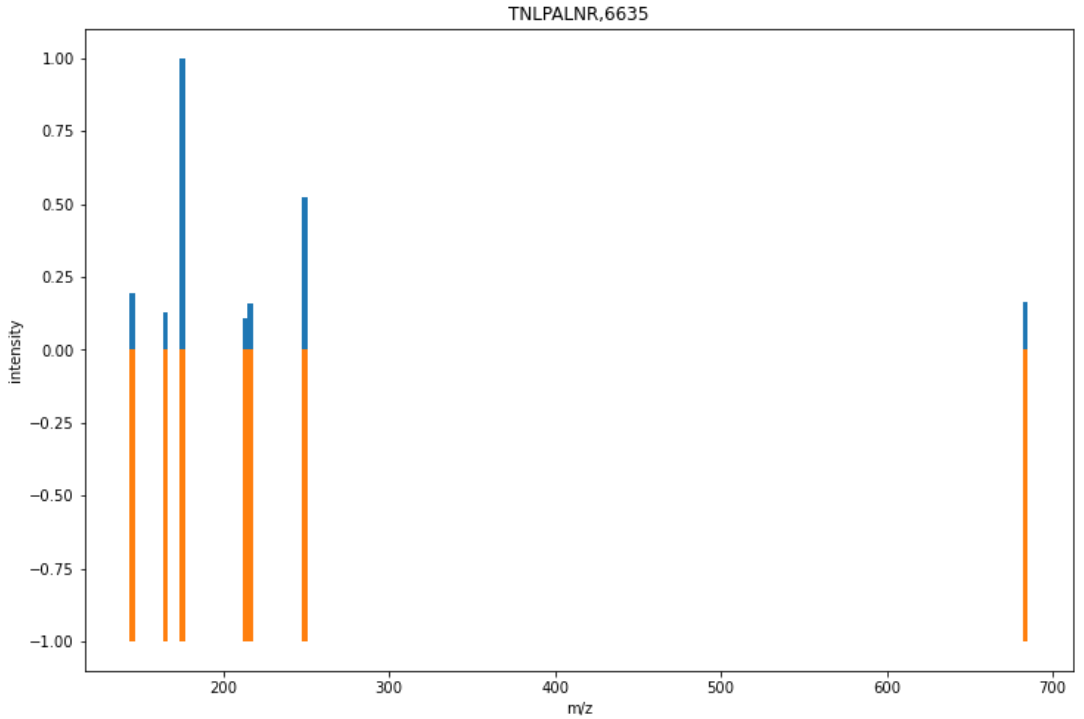




RAW FILE 8

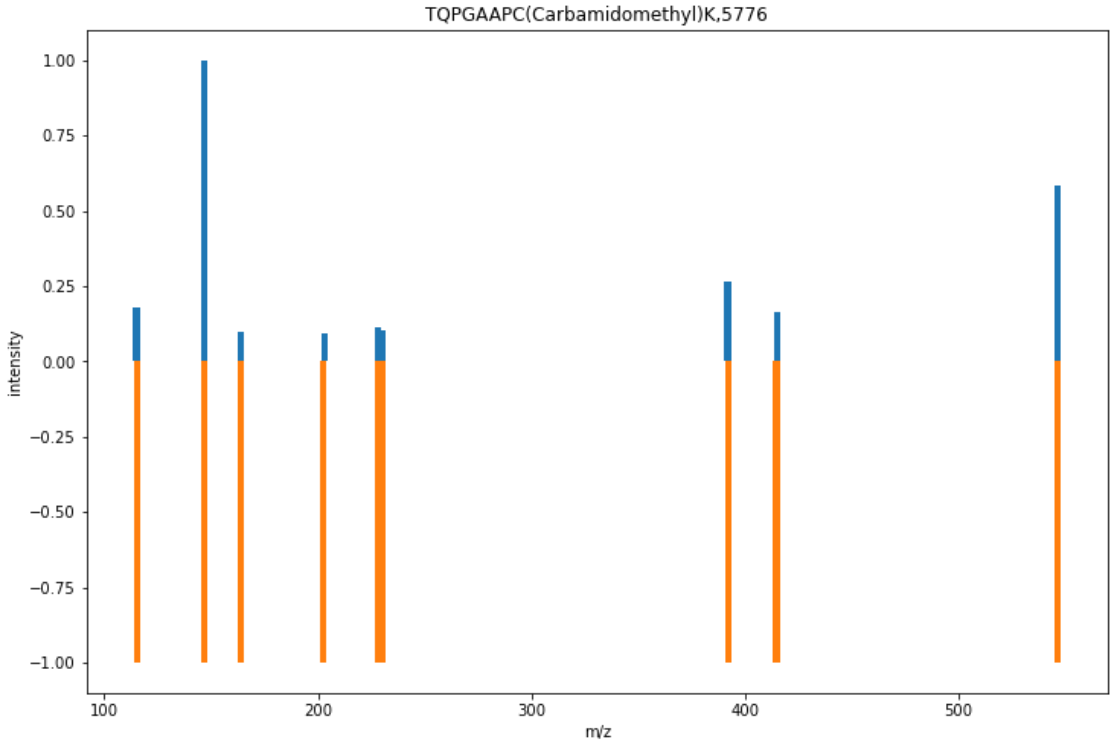
(FC(CARBAMIDOMETHYL)HAALK),(TNLPALNR)

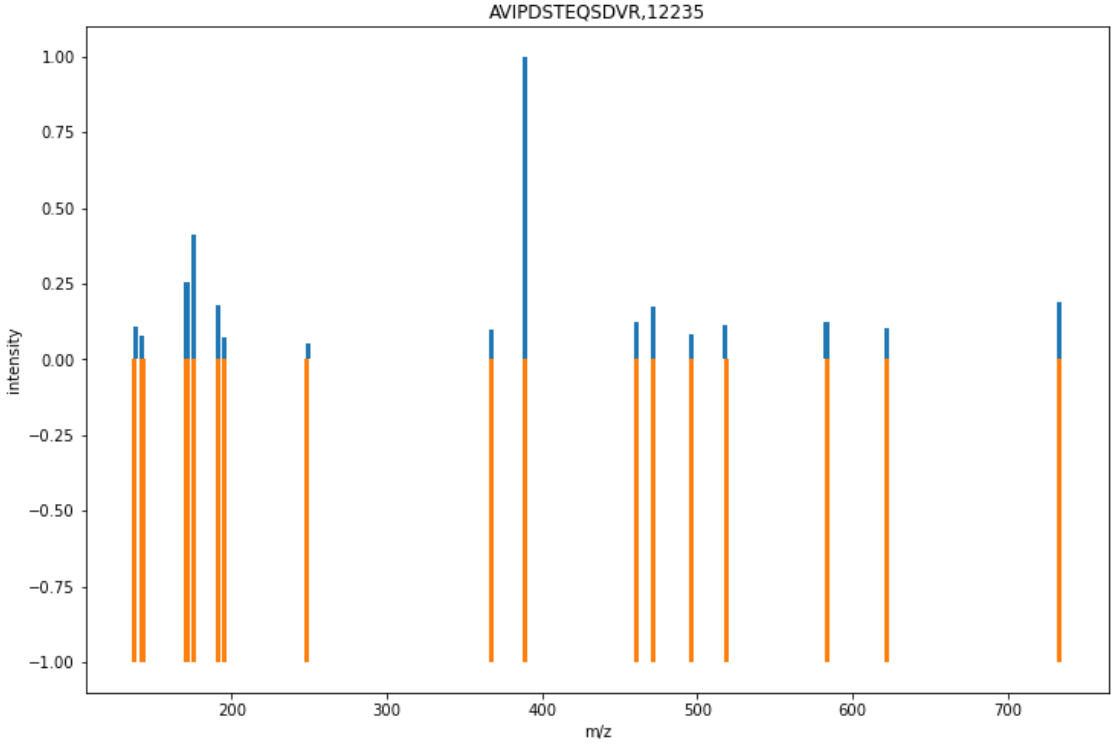




RAW FILE 9

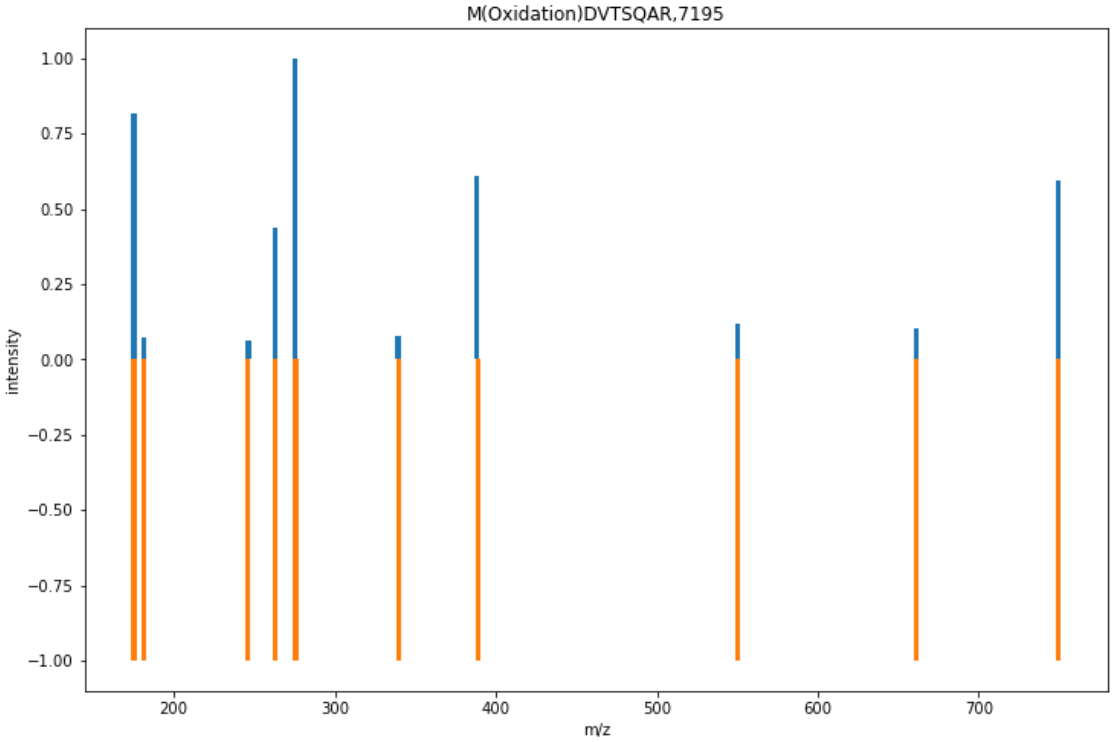
(FC(CARBAMIDOMETHYL)HAALK),(TNLPALNR)

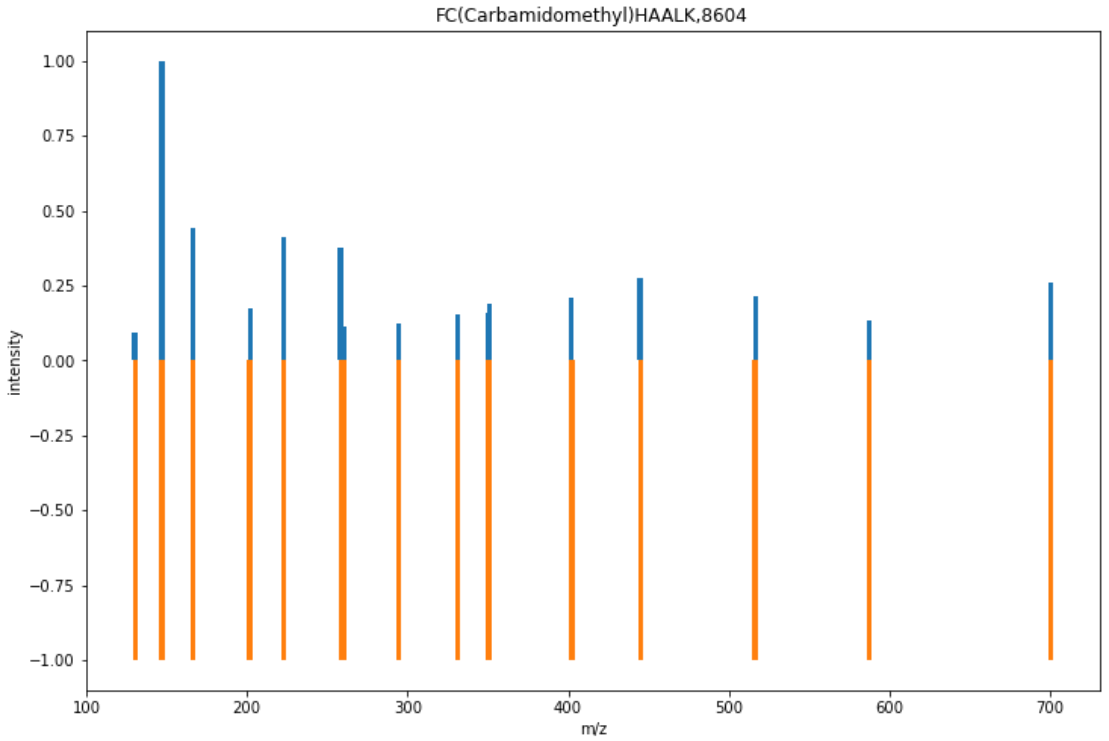




RAW FILE 9

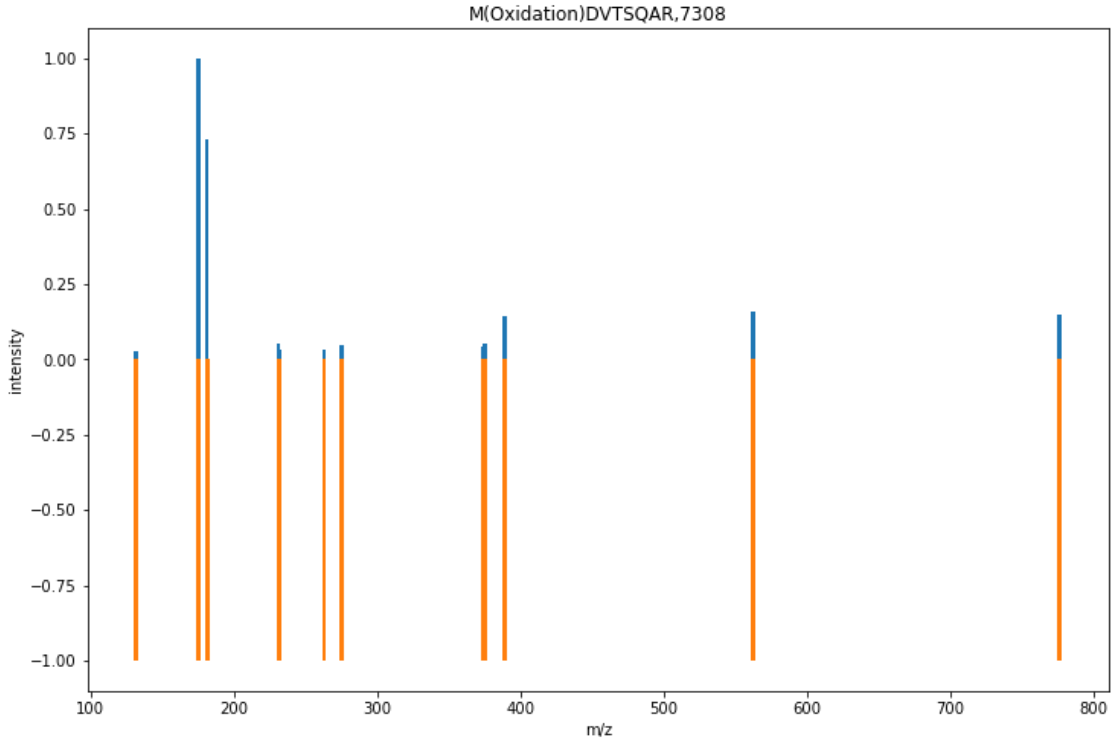
(TQPGAAPC(CARBAMIDOMETHYL)K),(AVIPDSTEQSDVR)

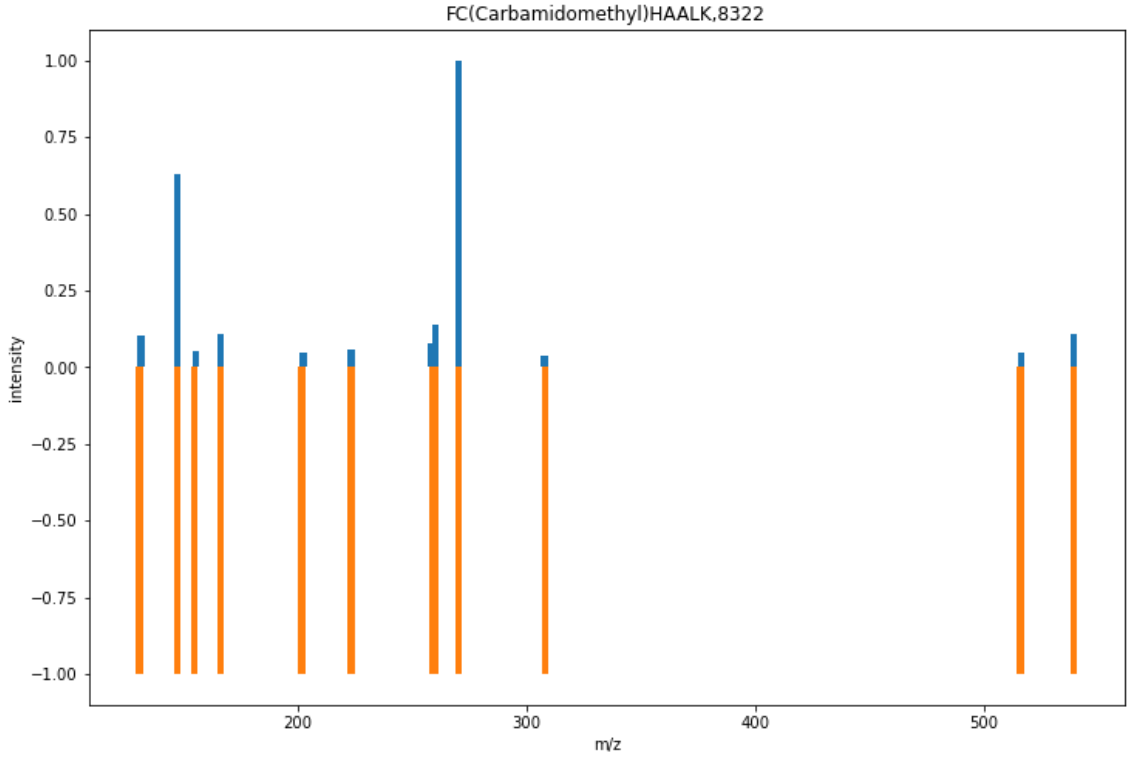




RAW FILE 9

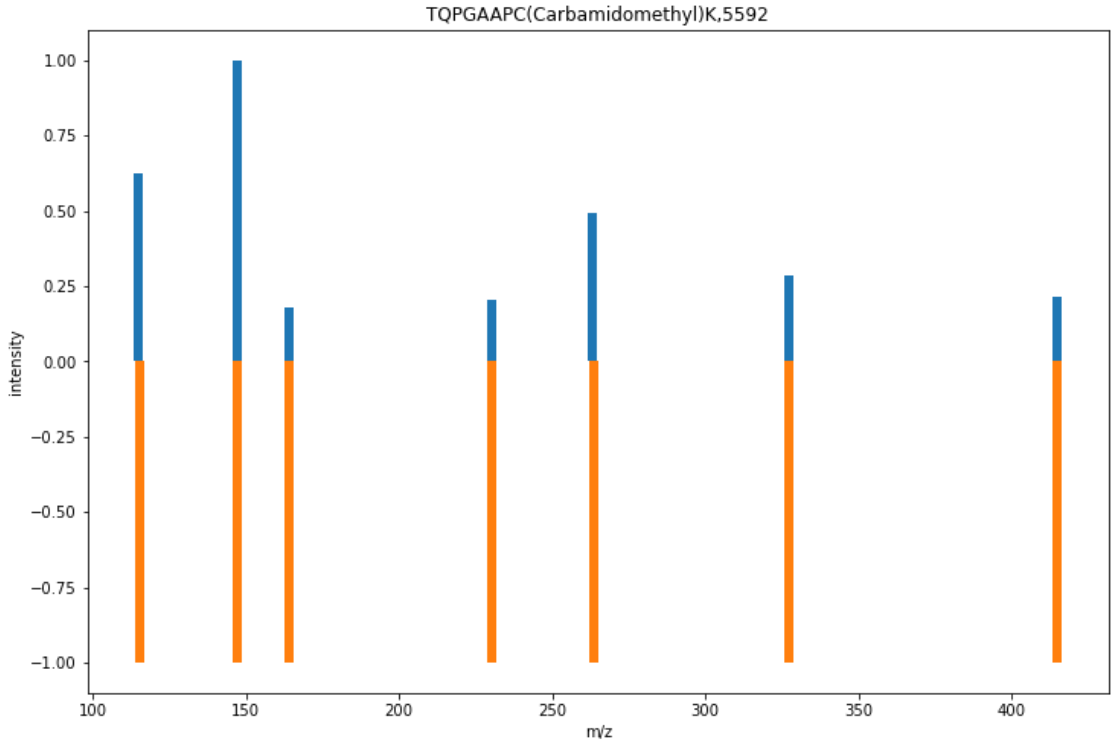
(M(OXIDATION)DVTSQAR),(FC(CARBAMIDOMETHYL)HAALK)

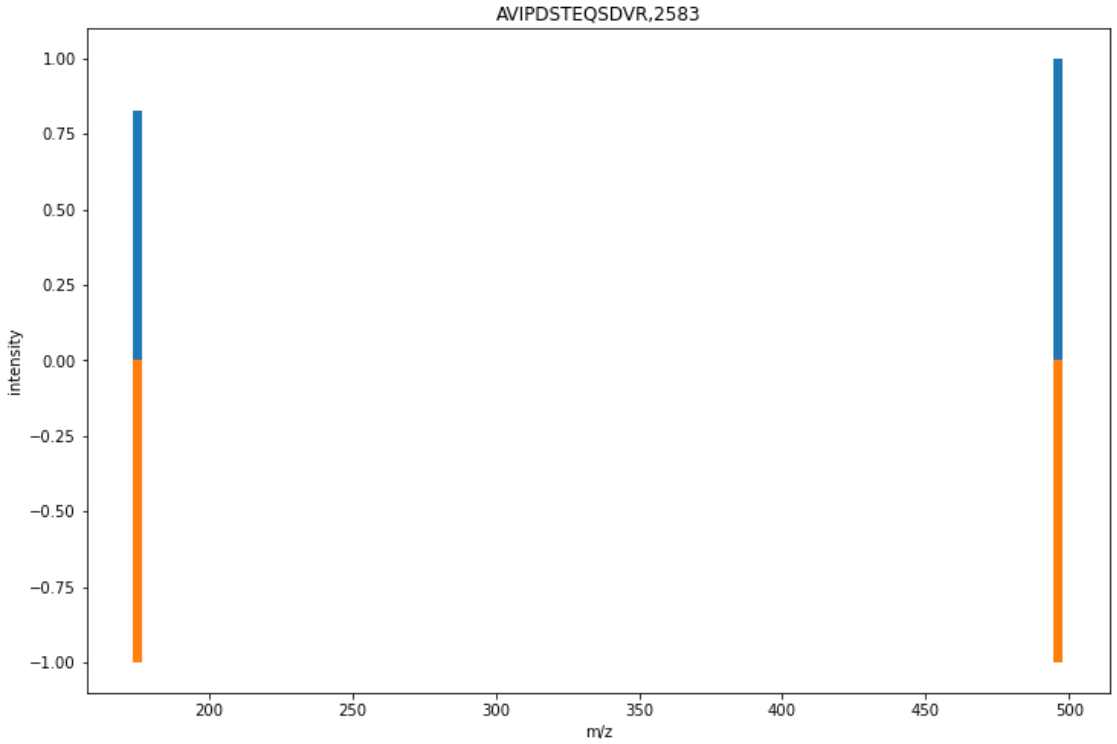




RAW FILE 10

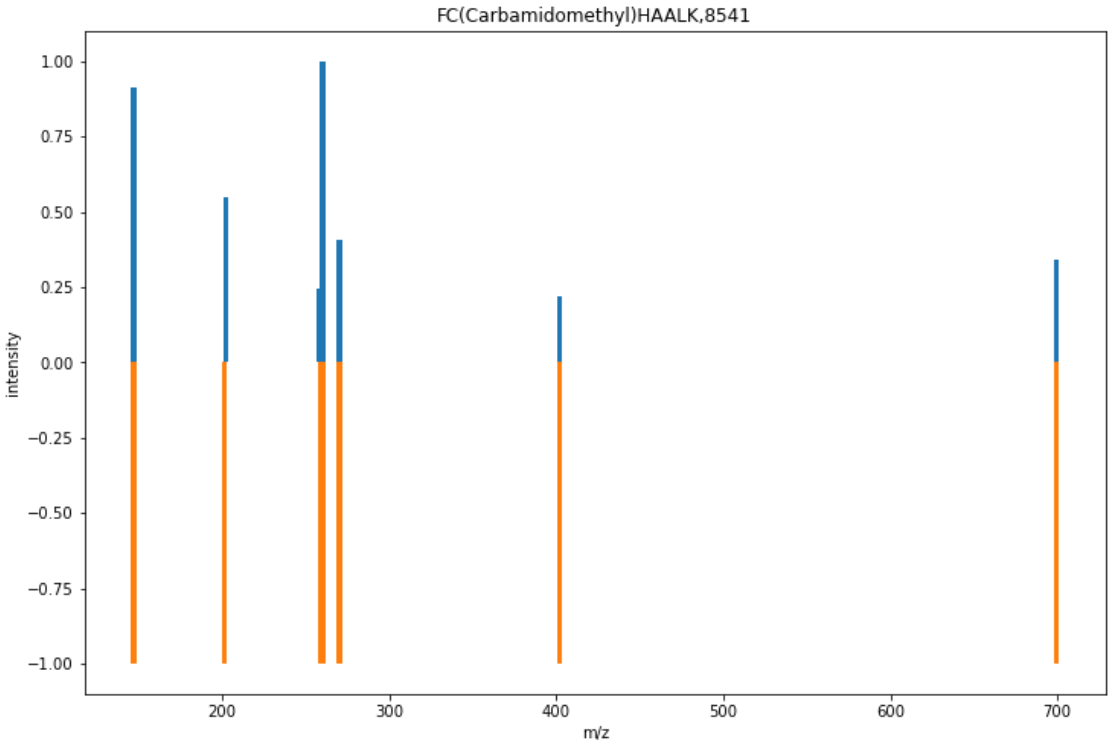
(M(OXIDATION)DVTSQAR),(FC(CARBAMIDOMETHYL)HAALK)

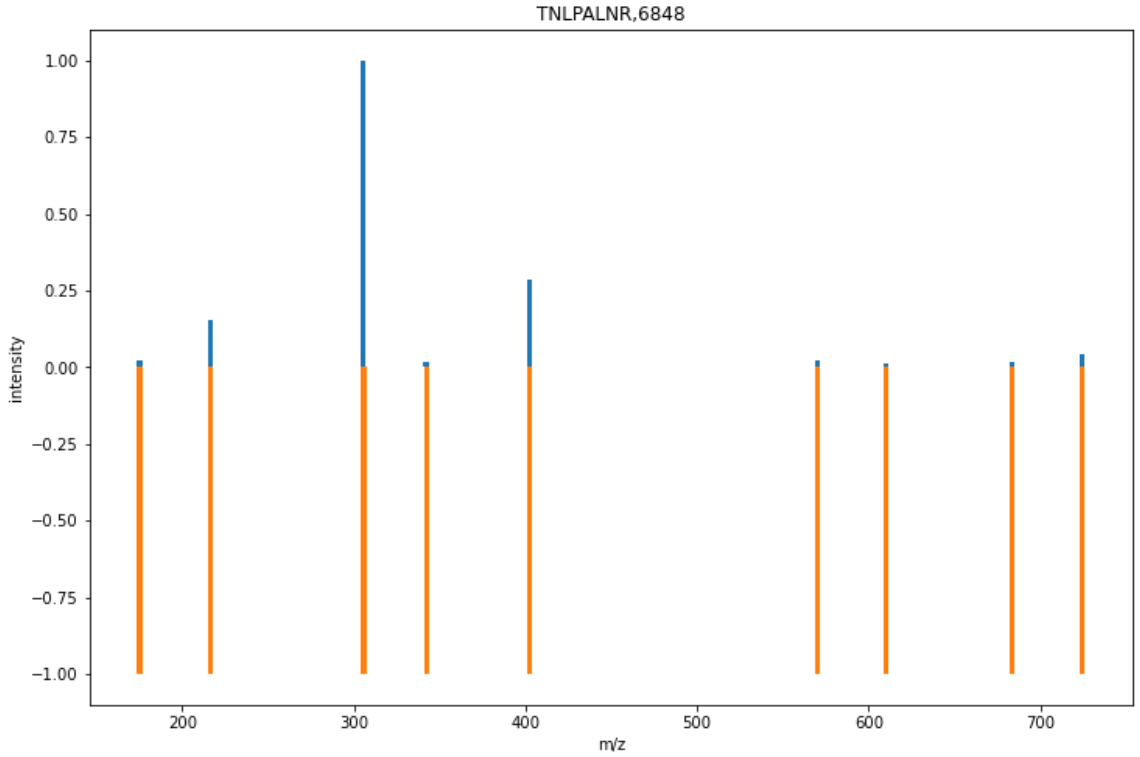




RAW FILE 10

(TQPGAAPC(CARBAMIDOMETHYL)K),(AVIPDSTEQSDVR)





RAW FILE 10

(FC(CARBAMIDOMETHYL)HAALK),(TNLPALNR)