COMPUTIONAL BIOLOGY

Under Supervisor:
Prof: Ibrahim El-Awdi
Eng: Ghada Samir

TEAM MEMBER:

- **Rana Mohamed Abdelkhalek**

- **Reem Ashraf Zaki**

- **Shahd Medhat Abdelhakim**

- **Salma Osama Ahmed**

INTRODUCTION :

Advances in instrumentation and data acquisition modes, such as the recently developed real-time database search (RTS), can reduce interference .However, no diagnostic tool is available currently to rapidly and accurately assess ion interference. To address this need, we developed a multiplexed TMT-based standard, termed the triple knockout (TKO). This standard is comprised of three yeast proteomes in triplicate, each from a strain deficient in a highly abundant protein (Met6, Pfk2, or Ura2). Specifically, we expand the TKO standard to incorporate the TMTpro18-plex reagents (TKO18). We also construct a variant thereof which has been digested only with LysC (TKO18L). We compare proteome coverage and interference levels of TKO18 and TKO18L data that were acquired under different data acquisition modes and analyzed using TVT2.5. Our data illustrate that RTS reduces interference while improving proteome coverage and suggest that digesting with LysC alone only modestly reduces interference, albeit at the expense of proteome depth.

FASTA SEQUENCE:

>sp|P39723|SPC72_YEAST Spindle pole component SPC72 OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) OX=559292 GN=SPC72 PE=1 SV=2
MVRRWIPSGRHLRNNDNTGDDDDSEFTNSMDSGMSIPSLRDSMTTRSSHNDPIKPALMND
SNKVKNLEKELTNAKIKIQVLYEYIRRIPNKDGNAPSLGNDTDFRNSIIEGLNLEINKLK
QDLKAKEVEYQDTLQFVQENLENSESIVNTINHLLSFILTHFNEQDENAHLLDKEERETL
EETLELSSDYVLEKMDTLSKFIIQFLQDFLHSKSRAESKQDKEEFLSLAQSSPAGSQLES
RDSPSSKEENTDGGYQNDEIHDSNNHIDTENVMANSTSLPISAVESRFEKTLDTQLEIVI
ENLHKEYDQFINSIRLKFEKSQKLEKIIASKLNEQSHLLDSLELEENSSSVIEKQDHLIS
QLKEKIESQSVLINNLEKLKEDIIKMKQNEKVLTKELETQTKINKLKENNWDSYINDLEK
QINDLQIDKSEEFHVIQNQLDKLDLENYQLKNQLNTLDNQKLILSQYESNFIKFNQNLLL
HLDSIFNILQKILQESSIAQFDRKMKSIKSVPNALKNLNLIQPKLESLYTFIETALESII
NSYISSLISMETPEQPHQQGNELTATPNKELTLRIEELQRRWISERERRKLDANASEARI
KALEQENESLRSKLFNLSINNP

```python
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]
    theo_int = [element * -1 for element in theo_int]
    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("ea07029.mzML","yeast.fasta", protein_ids, peptide_ids)


print(len(peptide_ids),len(protein_ids))
```

```
2 1
```

```python
for peptide_id in peptide_ids:
    print (35*"=")
    print ("Peptide Index m/z:", peptide_id.getMetaValue("scan_index"))
    for hit in peptide_id.getHits():
        index=str(peptide_id.getMetaValue("scan_index"))
        seq=str( hit.getSequence())
        print(" - Peptide hit sequence:", hit.getSequence())
        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)

        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("ea07029.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)

        spa.getSpectrumAlignment(alignment, theo_spec, spectra)
```

```python
if(len(alignment)!=0):
    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())
    if(max(obs_int)!=0):
        title ="The Seq :"+ seq+" & The index:"+index
        mirror_plot(obs_mz, obs_int, theo_mz, theo_int, title)
```

FIRST FILE MZ

Protein

Peptide

EXP 1:

NLNLIQPK

The Seq :NLNLIQPK & The index:6610

PROTEIN

Peptide

Exp 2

ILQESSIAQFDR

SECOND FILE MZ

Protein

Exp 1

Peptide

ELETQTK

The Seq :ELETQTK & The index:4882

PROTEIN

Peptide

Exp 2

TLDTQLEIVIENLHK

THIRD FILE MZ

Protein

Exp 1

Peptide

LILSQYESNFIK

The Seq :LILSQYESNFIK & The index:1758

The Seq :IESQSVLINNLEK & The index:5113

PROTEIN

Peptide

Exp 2

IESQSVLINNLEK

PROTEIN

Peptide

Exp 3

TLDTQLEIVIENLHK

PROTEIN

Peptide

Exp 4

ILQESSIAQFDR

The Seq :ILQESSIAQFDR & The index:7230

The Seq :TLDTQLEIVIENLHK & The index:16656

PROTEIN

Peptide

Exp 5

TLDTQRLEIVIENLHK
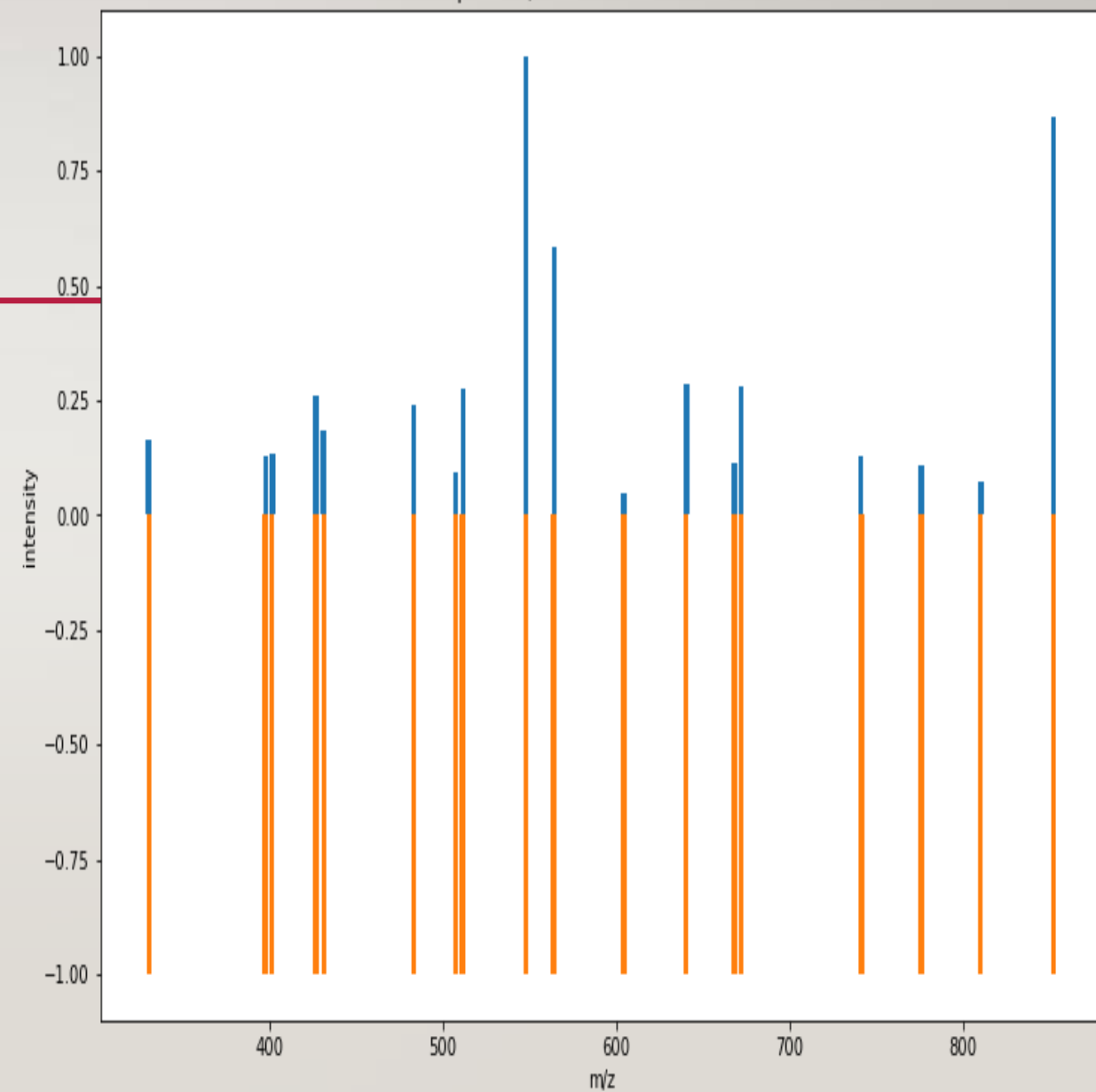
FOURTH FILE MZ

Protein

Peptide

Exp 1

TLDTQRLEIVIENLHK

FIFTH FILE MZ

Protein

Peptide

Exp 1

NLNLIQPK

The Seq :NLNLIQPK & The index:3110

PROTEIN

Peptide

Exp 2

IESQSVLINNLEK
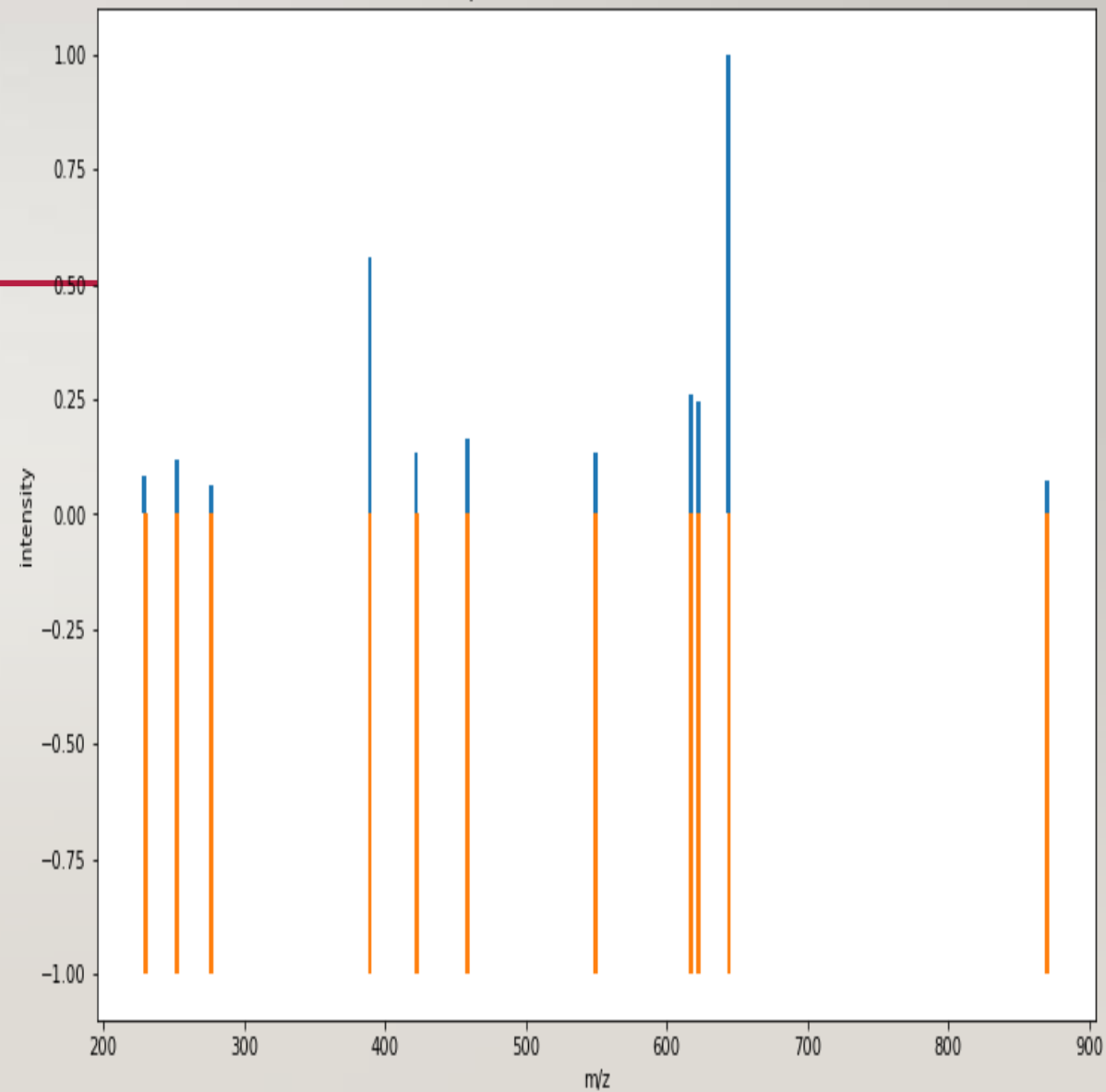
PROTEIN

Peptide

Exp 3

IQVLYEYIR
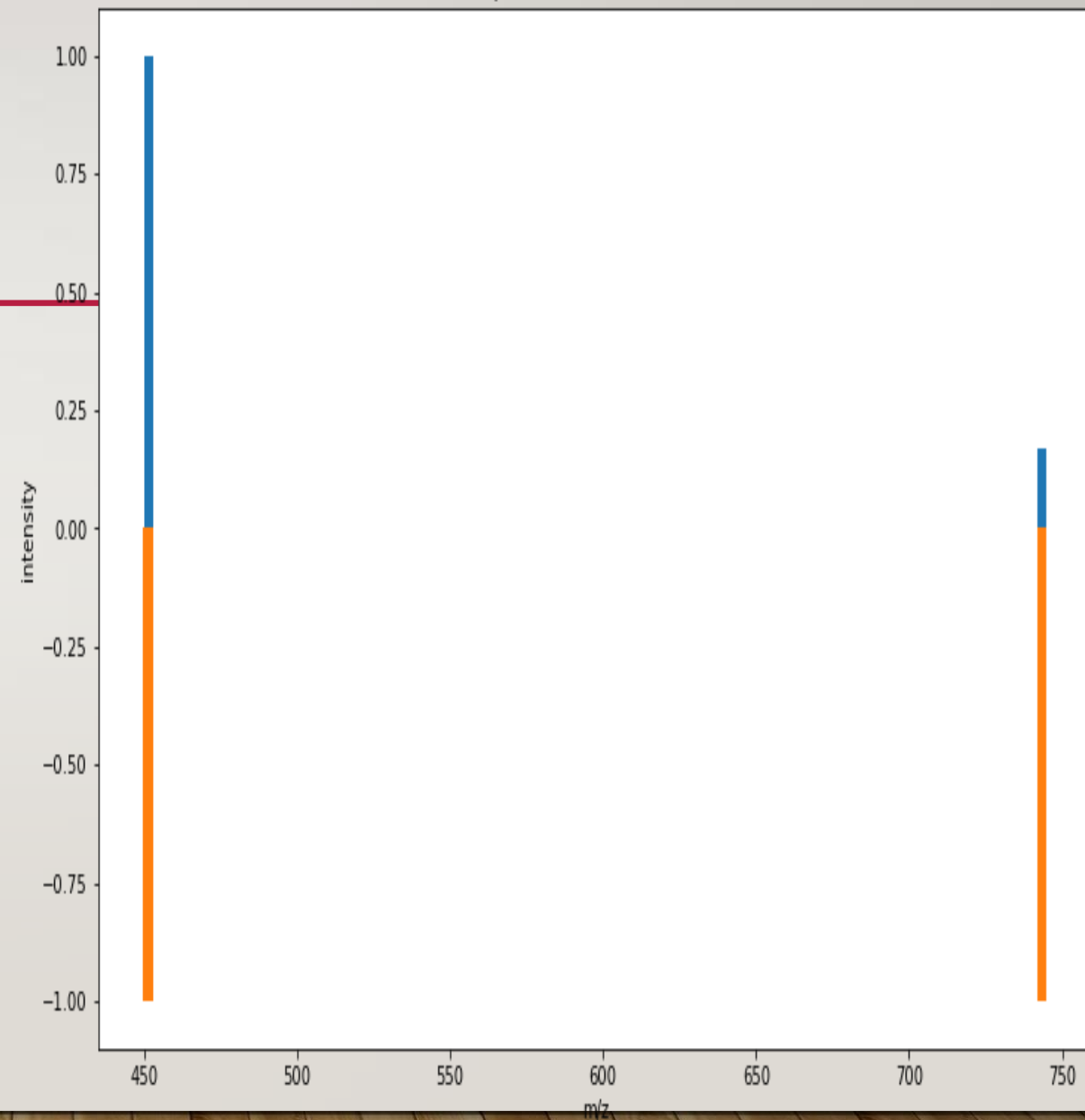
SIXTH FILE MZ

Protein   ## Peptide

Exp 1   LDLENYQLK

The Seq :LDLENYQLK & The index:1584
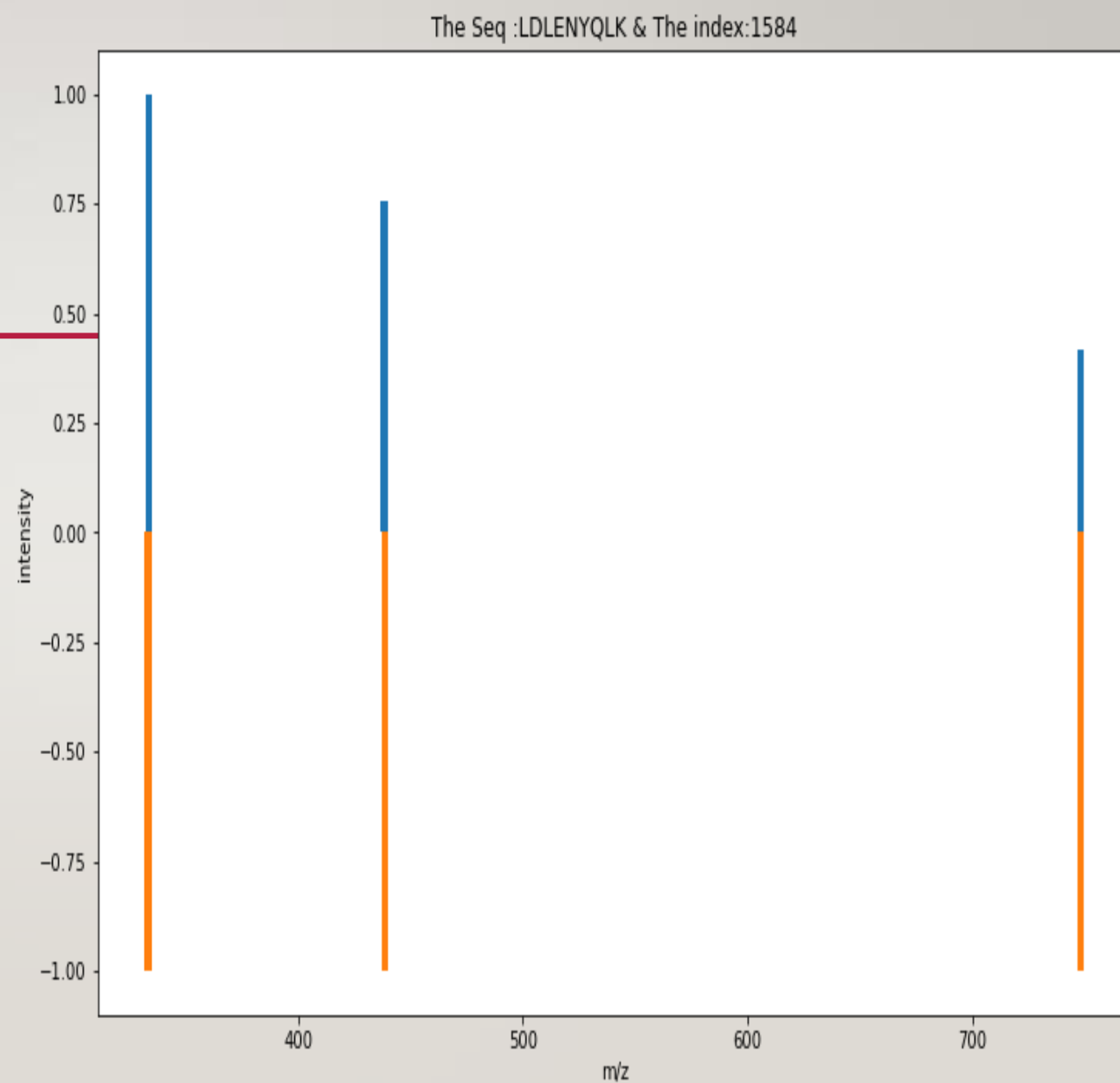
PROTEIN

Peptide

Exp 2
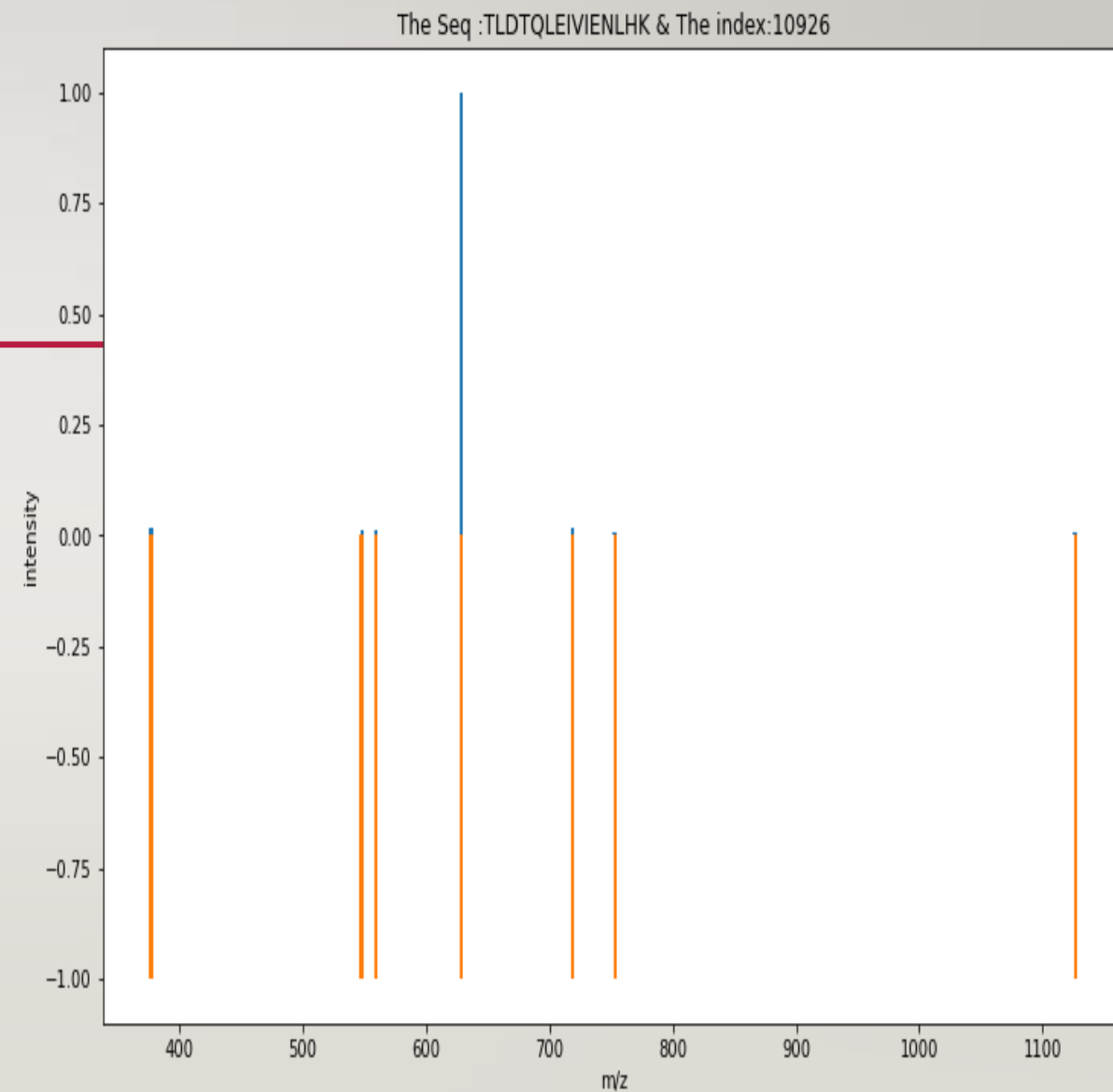
TLDTQLEIVIENLHK

PROTEIN

Peptide

Exp 3

TLDTQLEIVIENLHK

The Seq :TLDTQLEIVIENLHK & The index:17640
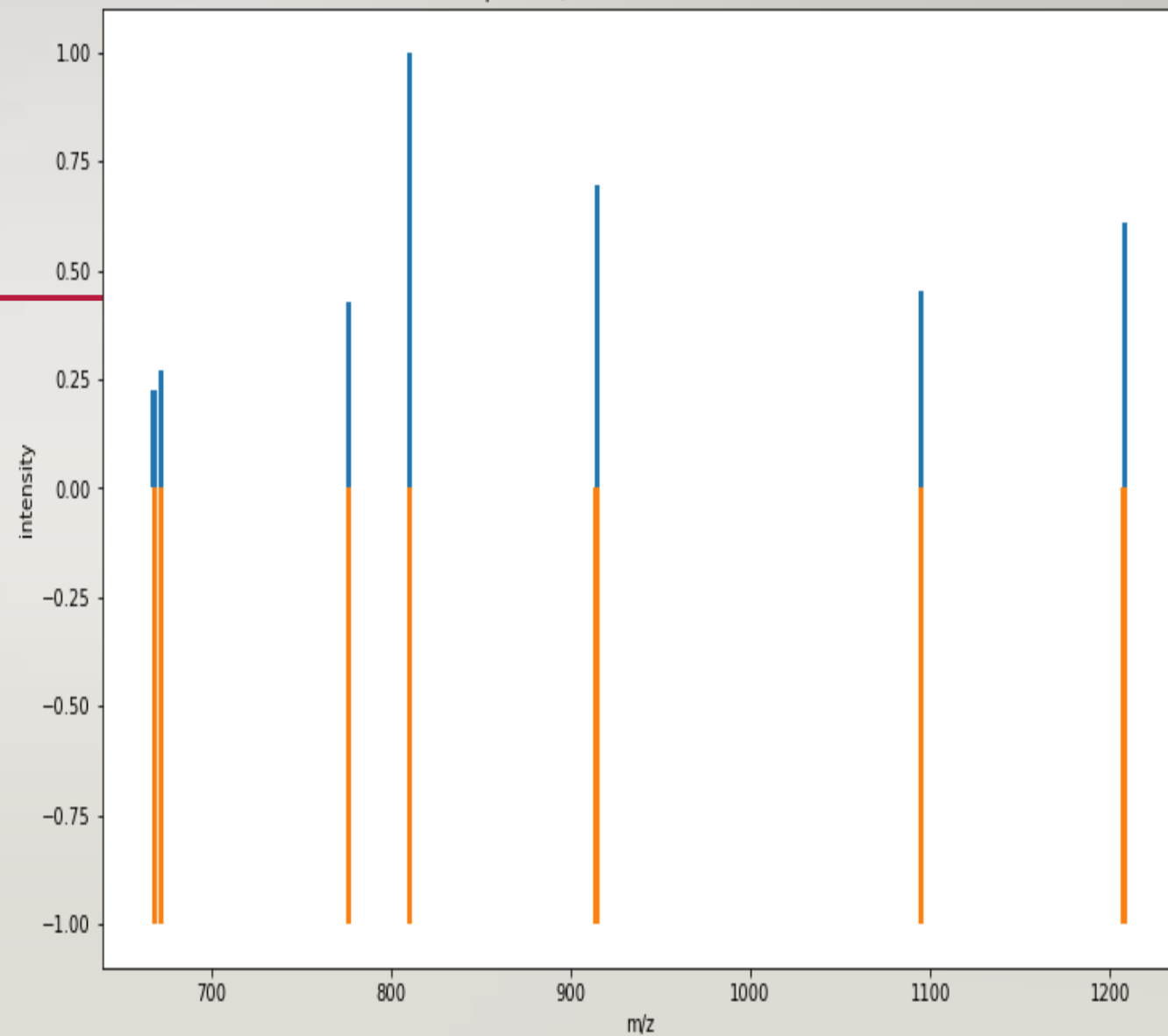
SEVENTH FILE MZ

Protein

Peptide

Exp 1

EEFLSAQSSPAGSQLESR
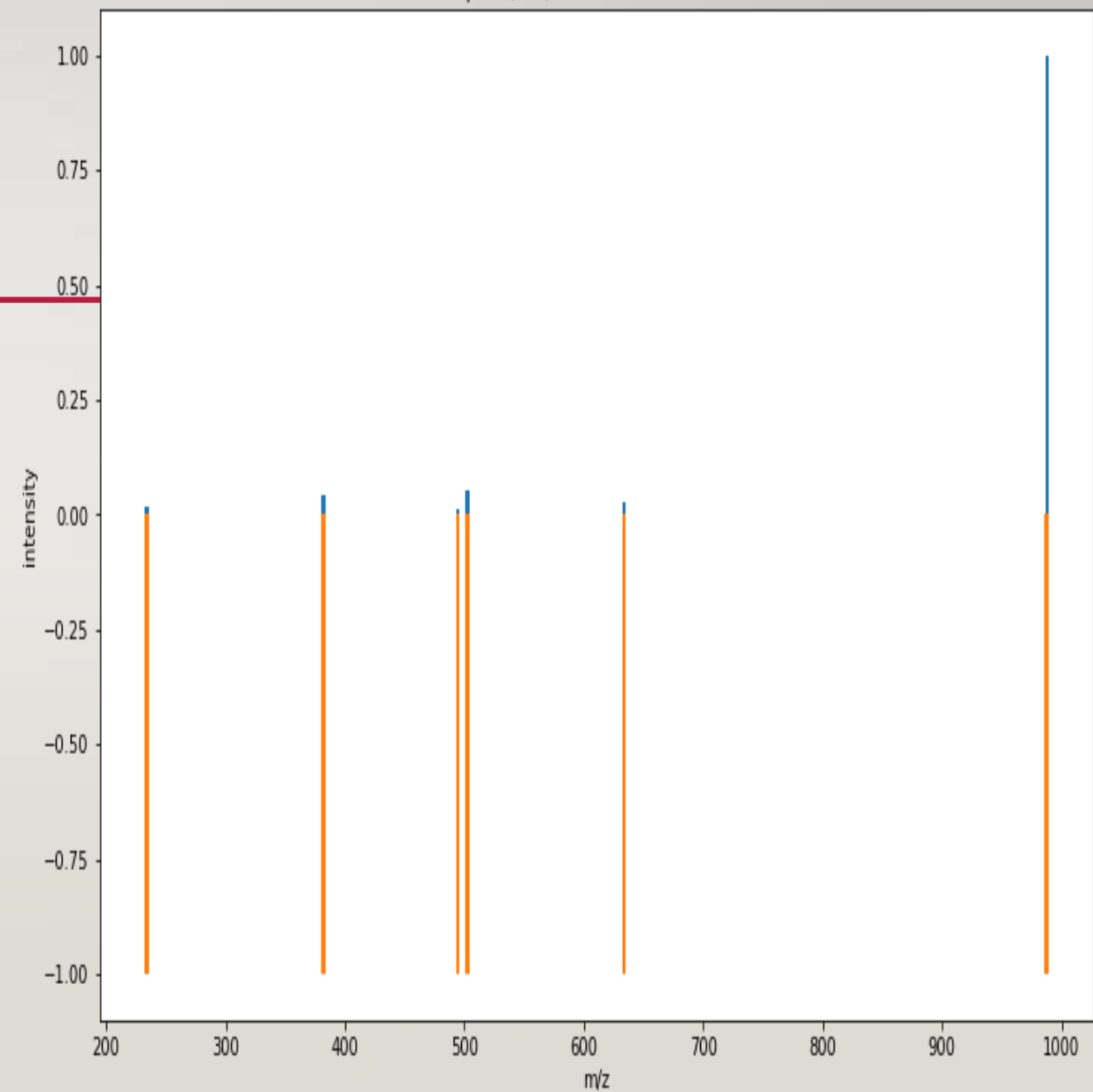

The Seq :EEFLSLAQSSPAGSQLESR & The index:8899
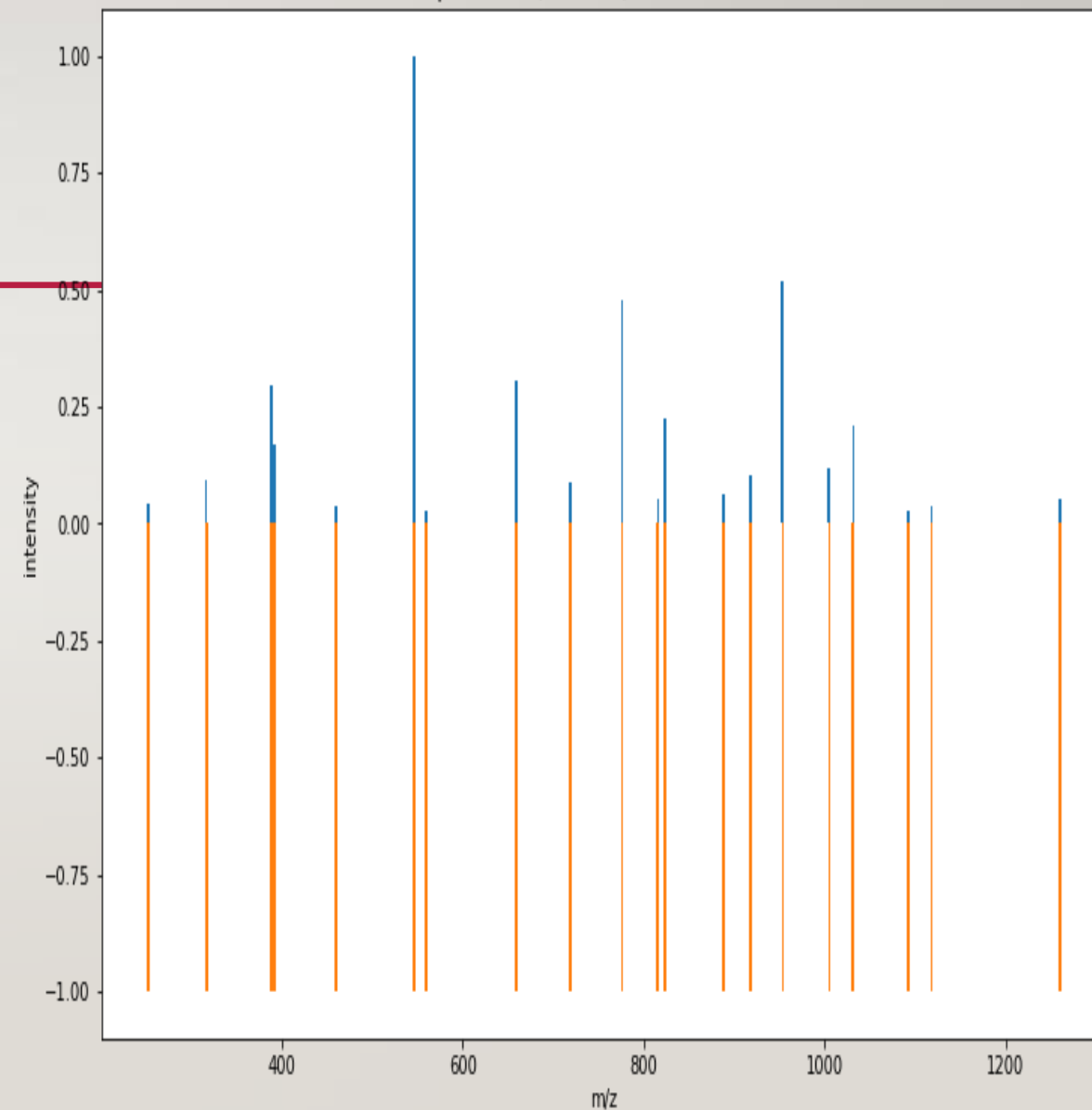
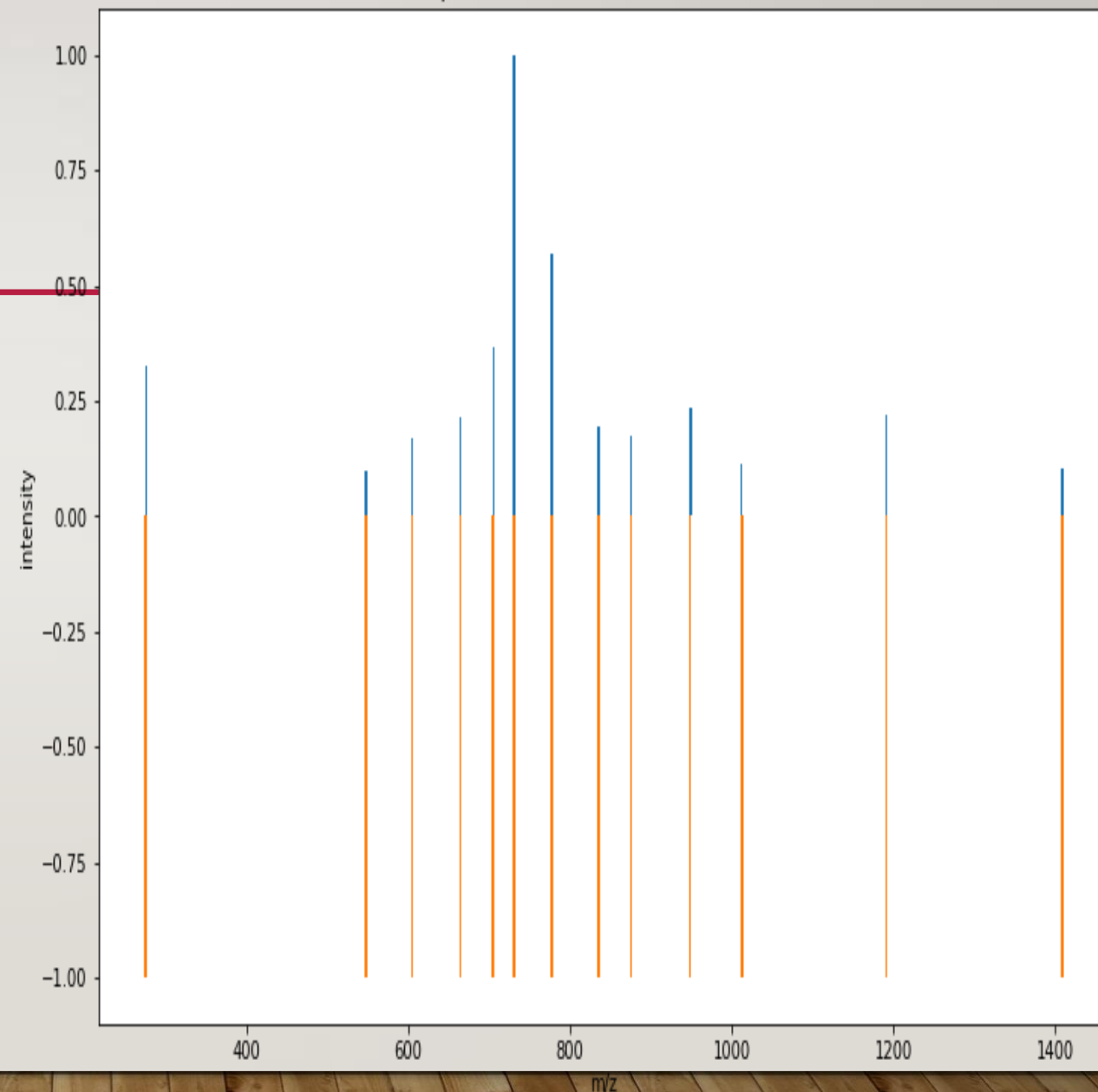PROTEIN

Peptide

Exp 2

FIIQFLQDFLHSK
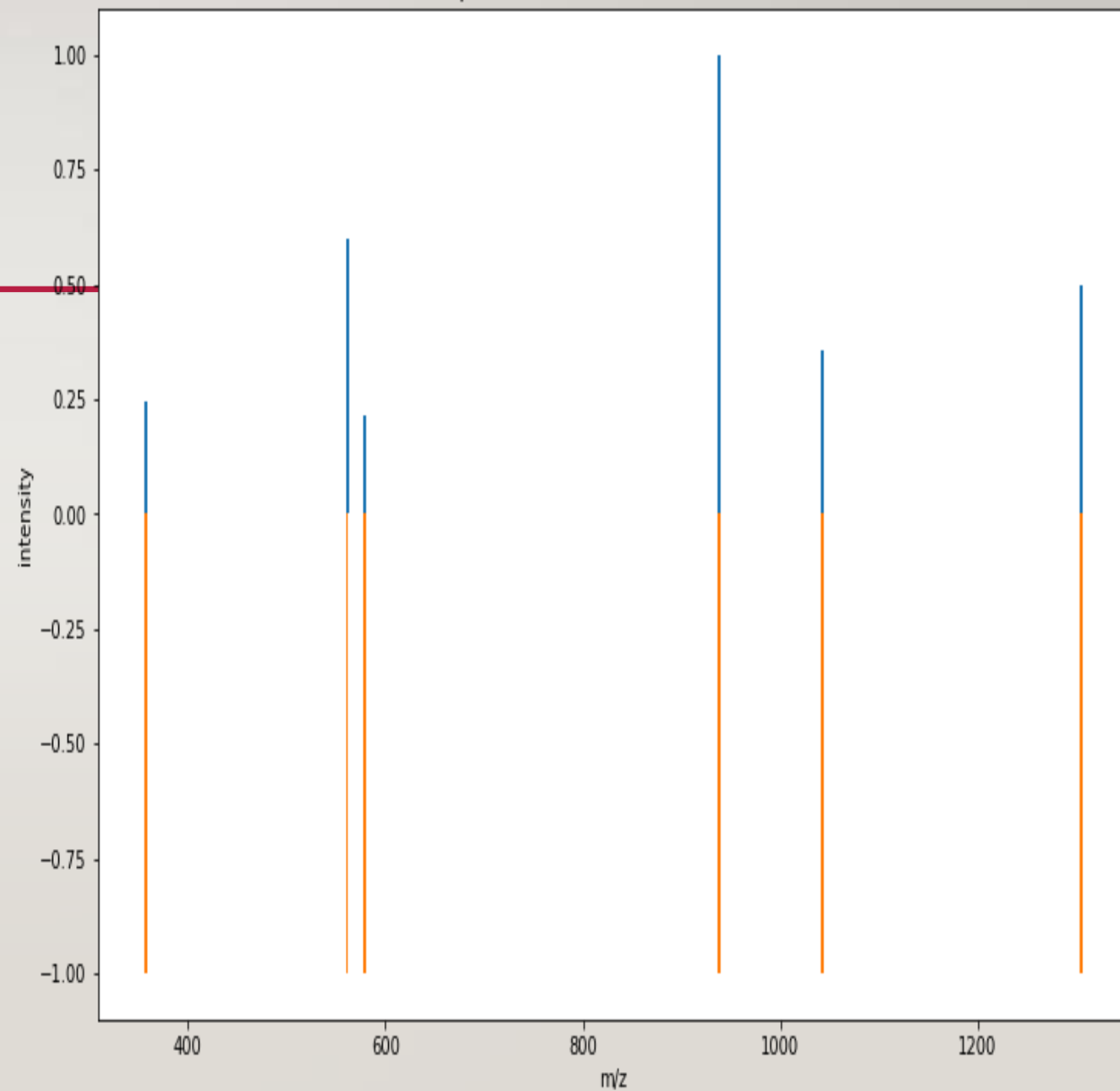
EIGHTH FILE MZ

Protein

Exp 1

Peptide

FNQNLLLHLDSIFNIL

The Seq :FNQNLLLHLDSIFNILQK & The index:22543

LINKS OF RAW AND FASTA FILE:

- RAW FILE LINK:

- https://www.ebi.ac.uk/pride/archive/projects/PXD029458

- FASTA FILE LINK:

- https://www.uniprot.org/uniprot/P39723