

```
>cat dbsearch.py
```

```
1  #!/bin/python3
2  # file containing the code submitted last week, updated to correctly
   calculate masses
3  from scripts import *
4  import pandas as pd
5
6  def match_spectrum(wtlist, peaks, ppm=20):
7      # matches a weight list as generated by fragmass with a df of
       peaks identified by call_peaks
8
9      # precalculate tolerances
10     low_thresh = 1 - ppm/1000000
11     up_thresh = 1 + ppm/1000000
12
13     matching = 0
14     # init two sorted iterators across the two lists
15     # uses a sorted-join-algorithm
16     iterwts = iter(wtlist)
17     curwt = next(iterwts)
18     wtlist = sorted(wtlist)
19     try:
20         for _, row in peaks.sort_values(by='m/z').iterrows():
21             #print(row['m/z'], curwt)
22             if curwt*low_thresh < row['m/z'] < curwt*up_thresh:
23                 matching += 1
24                 curwt = next(iterwts)
25             elif curwt*up_thresh < row['m/z']:
26                 curwt = next(iterwts)
27     except StopIteration:
28         pass
29
30     return matching
31
32
33 def get_tryptic_peptide(aaseq: str, peaks: pd.DataFrame):
34     # returns the tryptic peptide of the protein specified by aaseq
       best fitting the spectrum, supplied as a pandas DF
35     tryptics = trypticdigest(aaseq, minwt=0)
36
37     maxp = 0
38     maxtr = None
39     for tr in tryptics:
40         p = match_spectrum(fragmass(tr), peaks)
41         if p > maxp:
42             maxp = p
43             maxtr = tr
44
45     return maxp, maxtr
```

```

46
47 if __name__ == '__main__':
48     import sys
49     df = pd.read_csv(sys.argv[2])
50     seq = 'LHVPLEAGVVLLFK'
51     print("matching", seq, "to", sys.argv[2])
52     print(match_spectrum(fragmass(seq), df, ppm=20))
53
54     print("identifying best-matching peptide from", sys.argv[1], "i",
55           sys.argv[2])
56     with open(sys.argv[1], 'r') as f:
57         f.readline() # skip fasta header
58         seq = ''.join([l.strip() for l in f])
59     print(get_tryptic_peptide(seq, df))

```

```

$ ./dbsearch.py ./Q13740.fasta ./exspec.csv

```

```

1 matching LHVPLEAGVVLLFK to ./exspec.csv
2 10
3 identifying best-matching peptide from ./Q13740.fasta i ./exspec.csv
4 (11, 'VLHPLEGAVVIIFK')

```
