

02-725 HW4

Anushka Sinha

1 DATABASE SEARCH

1.1 PROBLEM STATEMENT

Since we are trying to implement an algorithm that finds the best match between the provided spectrum (S) and the database, the input is a proteomics database file and the spectrum file, the output is to find the peptide from the proteomics database that best matches the spectrum provided, and the goal is to find the peptide from the proteomics database whose spectrum explains the maximum number of b-ions and y-ions in the spectrum file.

1.2 ALGORITHM DESIGN

Algorithm design for the goal stated above is as follows:

Step 1: Get the trypsin digested fragments for each peptide from the database.

Step 2: Generate b-ions and y-ions for each fragment and combine them and sort them to get a spectrum for each fragment.

Step 3: Find the best match by looping through each spectrum and use the tolerance to find the peptide that best explains the b-ions and the y-ions of the peptide of interest.

1.3 TEST IMPLEMENTATION

The best match found using the above algorithm design is

IKRKAMAKEEFVRTKPHVNIGTIGHVDHGKTTLTAAISKVLNEKLTSEAVKSFDQIDNAPEEKERGITINSAHI
EYETEKRYAHVDCPGHADYVKNMVTGAAQMDGAILVCAATDGPMPQTREHVLLARQVNVPRLVVFLNKC
DMVDDEEMLELVEMELREILEQYGYEEDTPIVRGSALGALNGVEKVVKSVELMDTVDEWIQEPEREIDKPFLL
MPIEDVFSITGRGT VATGRIETGRCKVGDEVQLLGLGEDKKSVITGVEMFRKILAEGEAGDNVGLLLRGIDKAEV
KRGMVVVHPGAI T PHDHF KASIYVLKKEEGGRHTPF GNKYRPQFYLR TMDCTGEIKLPEGVEMVMPGDNVEI
EVELIYKVALNEGLRFAIREGGRTVGSGQIT TILDDIK

2. MODIFICATION DISCOVERY

2.1 PROBLEM STATEMENT

The input is the proteomics database and the spectrum (S), the output is the modified version of the peptide(s) from the database, and the goal is to find a modification of peptide that matches the mass of S and best explains the b-ions and y-ions of the spectrum (S).

2.2 ALGORITHM DESIGN

Algorithm design for the goal stated above is as follows:

Step 1: For each peptide in the database, obtain its trypsin digested fragments.

Step 2: Find all possible masses for those fragments by considering every combination of T-18 and S-18 modifications.

Step 3: Find the modifications of the peptides that match the mass of the peptide of interest with a tolerance of 0.02.

Step 4: For each modified peptide that does match the mass of the peptide of interest, find their spectrum by calculating their b-ions and y-ions and combining them and sorting them.

Step 5: After obtaining the spectrums for the modified peptides, loop through each spectrum to find the ones that best explain the b-ions and y-ions of the spectrum (S).

2.3 TEST IMPLEMENTATION

The peptides and their modifications that matches the mass of the peptide of interest are CPARRRCS-18AT-18T-18CPRT-18PWWR and CRPWPDRSDPPISGPWPLR. Neither of these peptides matched the spectrum (S).