

Application of SILAC data in antibiotic discovery

Student:

Sergey Chernov

Scientific Advisor:

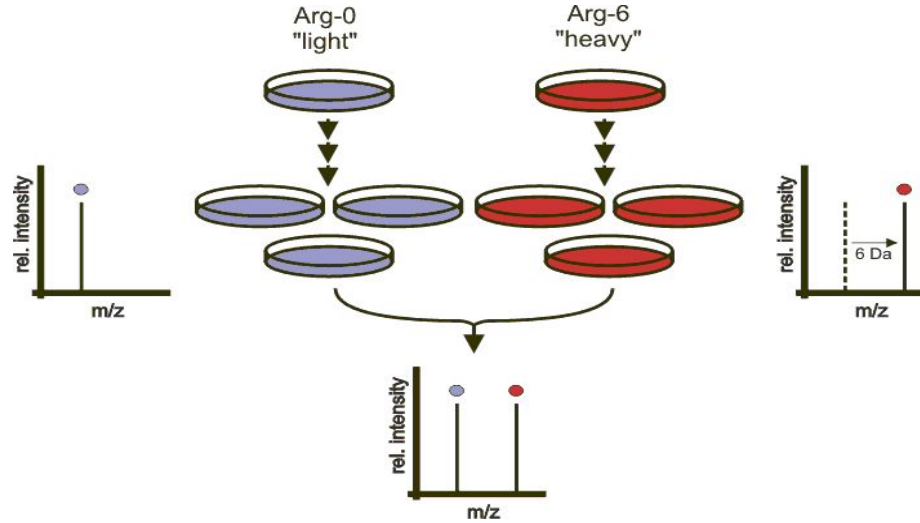
Alexey Gurevich

Goals

- Think how to use SILAC data for improving antibiotic discovery
- Implement scoring procedure for pair of spectra and a given compound
- Implement the full pipeline
- Test on real SILAC data and tune the parameters
- Support RiPPquest output

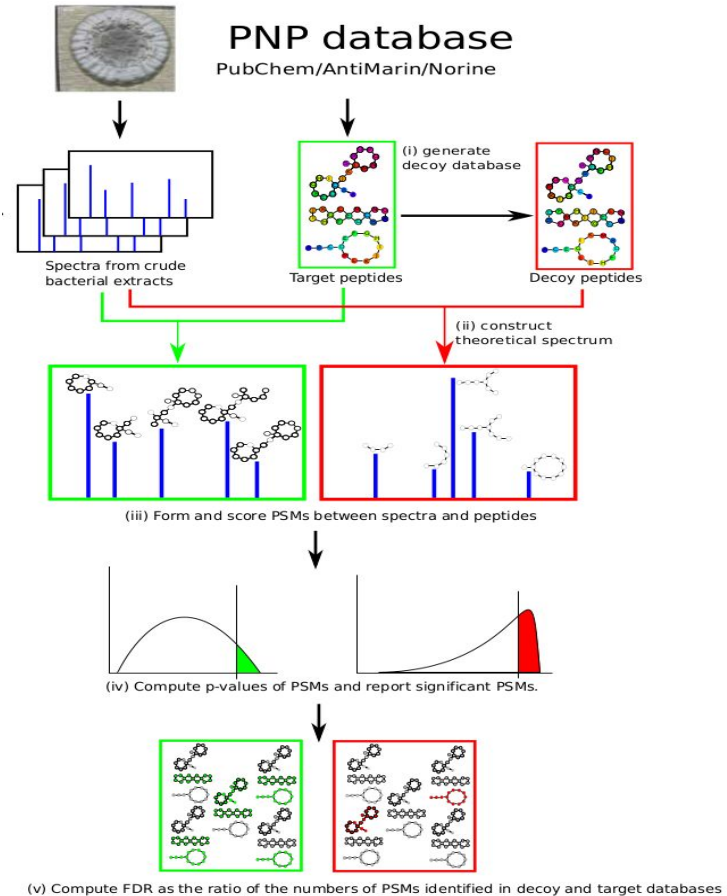
SILAC

Stable isotope labeling with amino acids

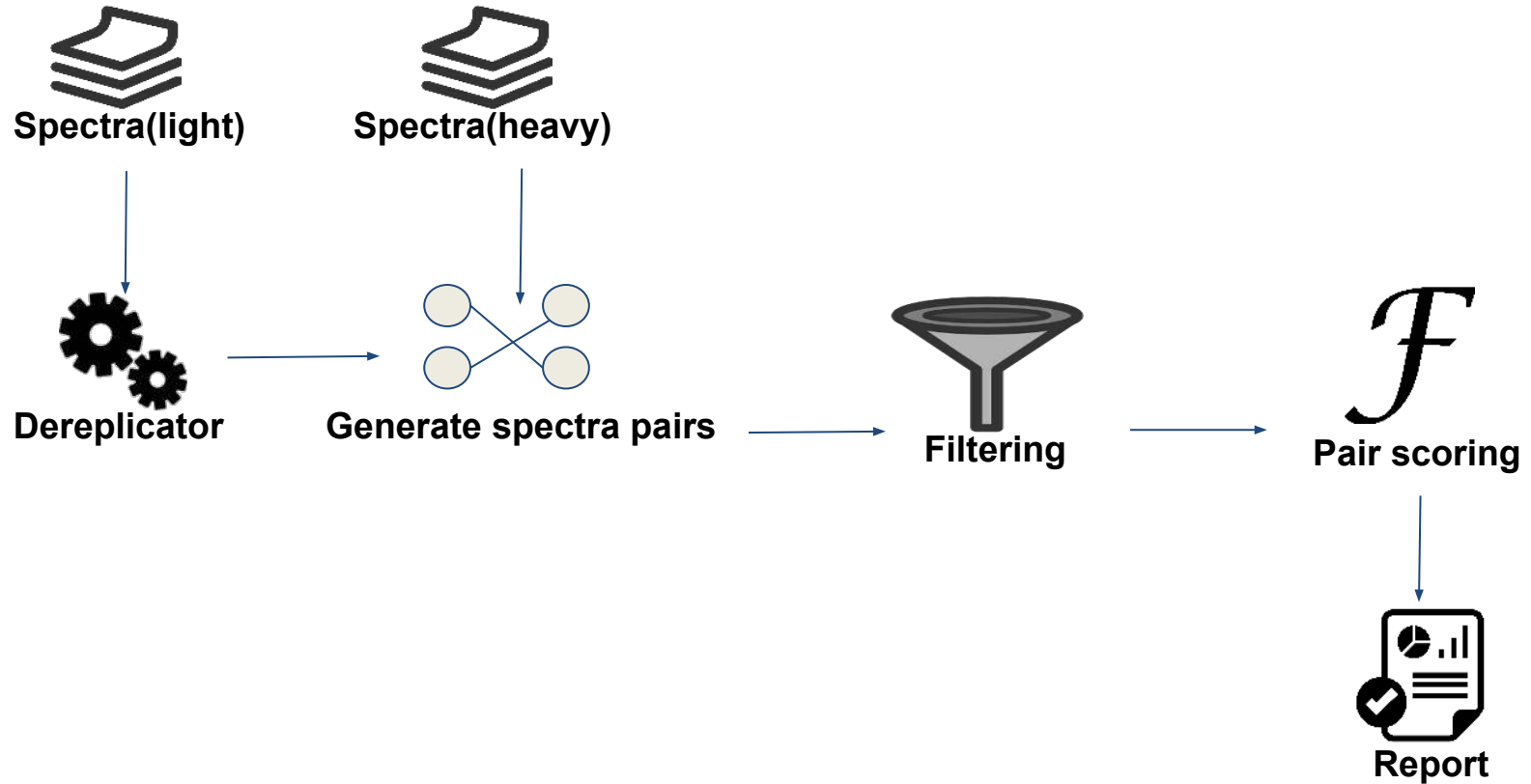


- Cells are grown in almost identical culture media
- One contains a 'light'(^{12}C), when the other a 'heavy'(^{13}C) form of a particular amino acid

Dereplicator pipeline



Pipeline

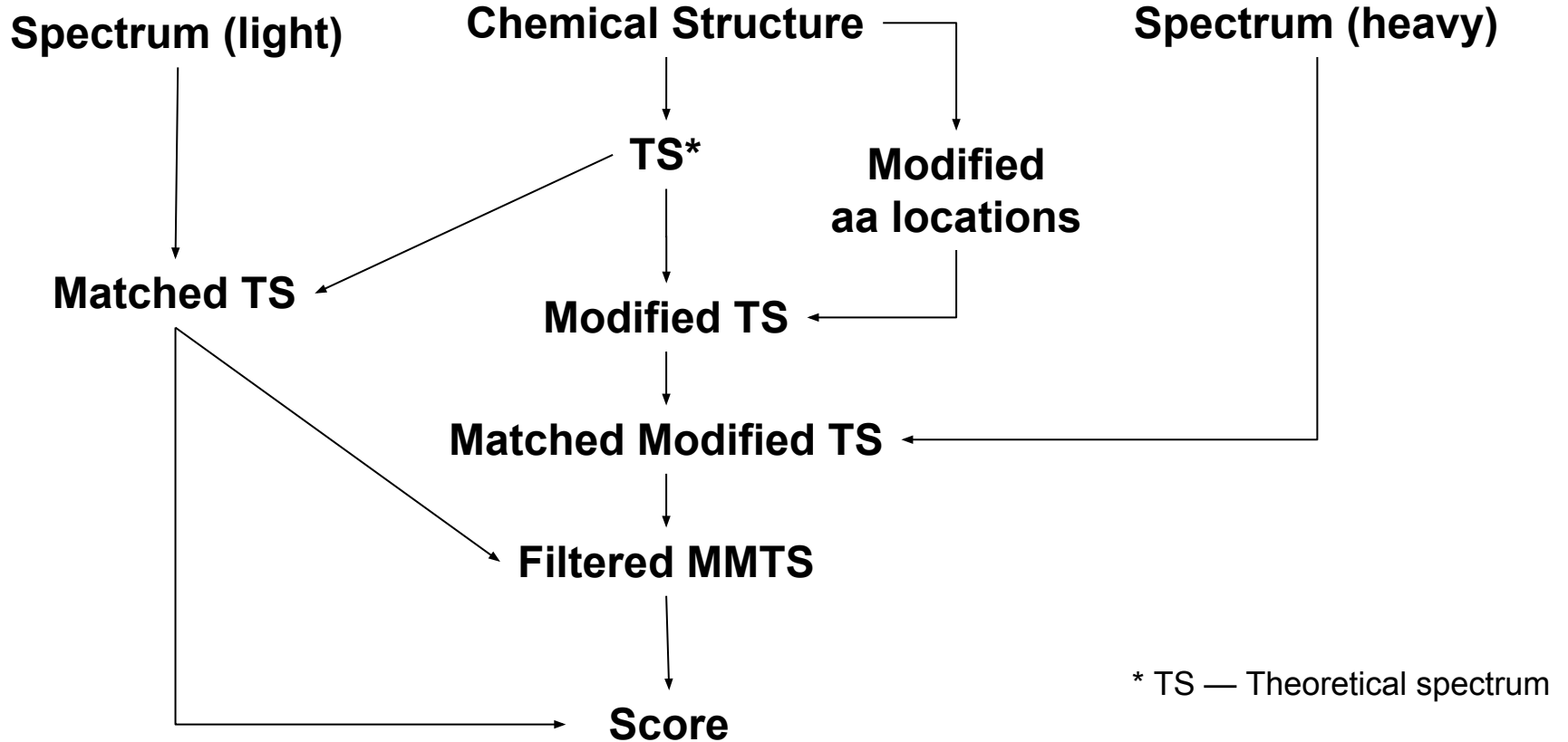


Spectra pairs and filtering

- Get all spectra from:
 - Files with ^{12}C (reported by Dereplicator)
 - Files with ^{13}C (all)
- Find such pairs of spectra for which the mass difference is a multiplier of 6 Da (\pm threshold)
- Drop pairs that do not have enough shared & *shifted** peaks

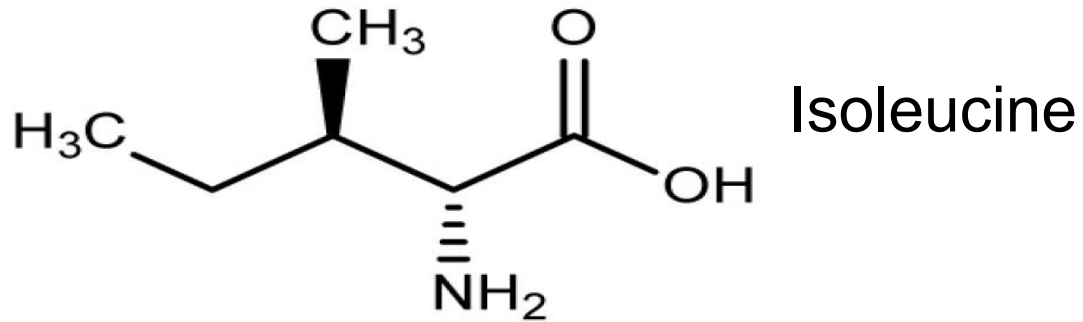
* *shifted* means peaks different by $k * 6$ Da

Scoring pipeline



Modified aa finding

- Standard amino acids may be represented as binary trees!

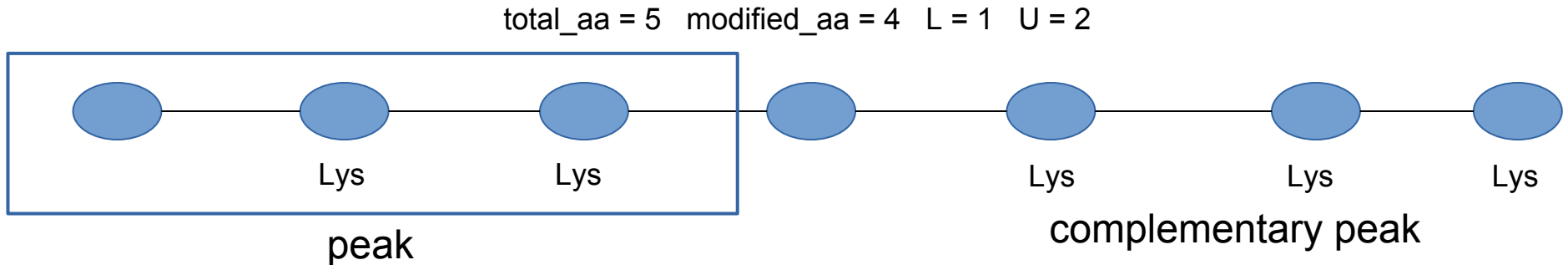


For the aa and each candidate in PNP

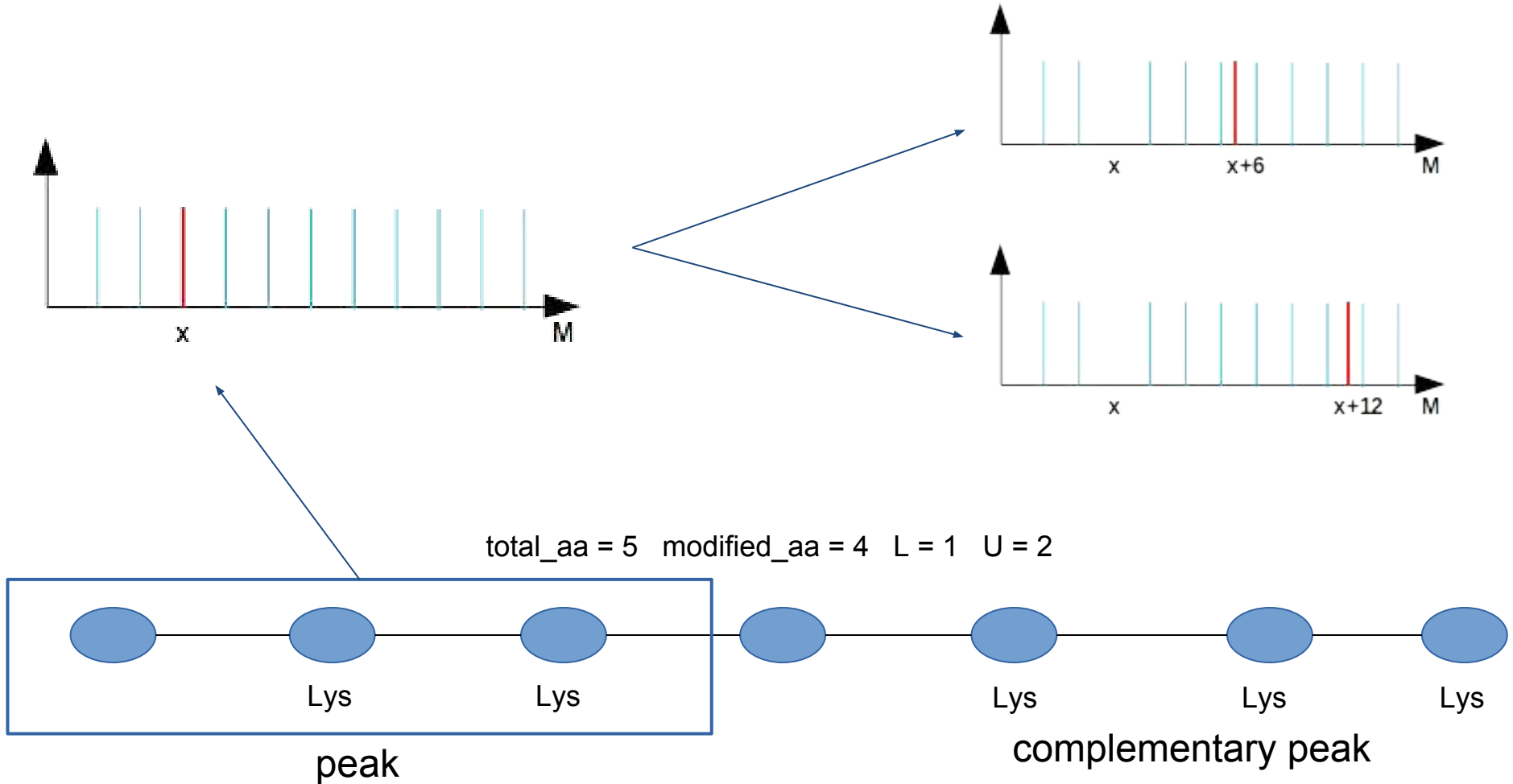
- 1) Remove Hs and reroot at N
- 2) Check isomorphism in $O(V)$

TS modification

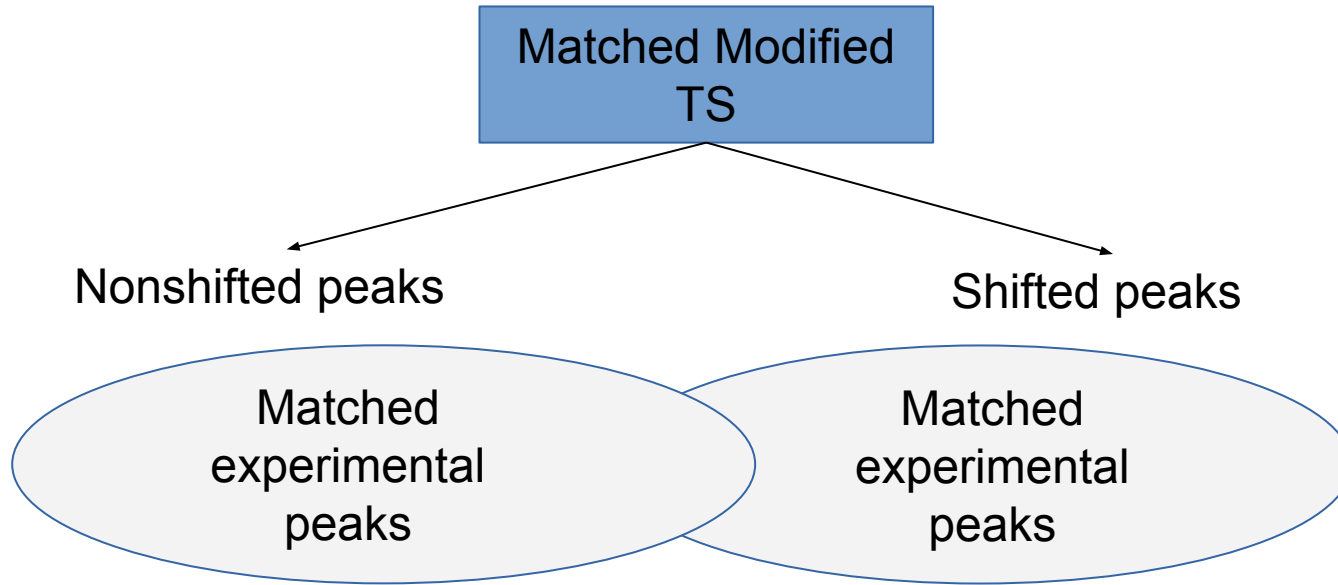
- Input: TS, *Lys* locations, *modified_aa*, *total_aa*
- For each **peak** in TS:
 - How many *Lys* included in **peak** (*aa_in_peak*)
 - $L = \max[0, \text{modified_aa} - \text{aa_in_comp_peak}]$
 - $U = \min[\text{aa_in_peak}, \text{modified_aa}]$
 - Add new *shifted* peak for each *i* in $[L, U]$



TS modification

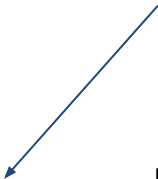
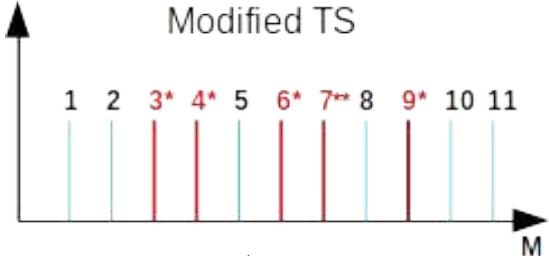
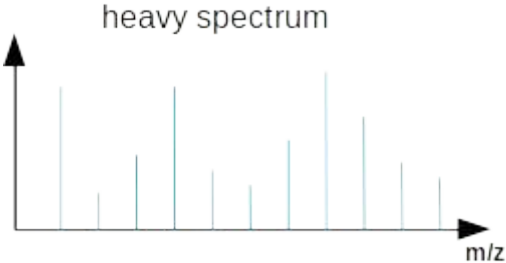


Matched Modified TS filtration



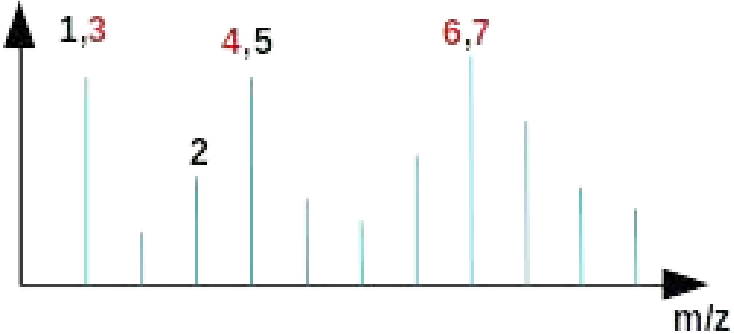
Remove *shifted* peaks matched with experimental peaks explained by *nonshifted* peaks

Matched Modified TS filtration

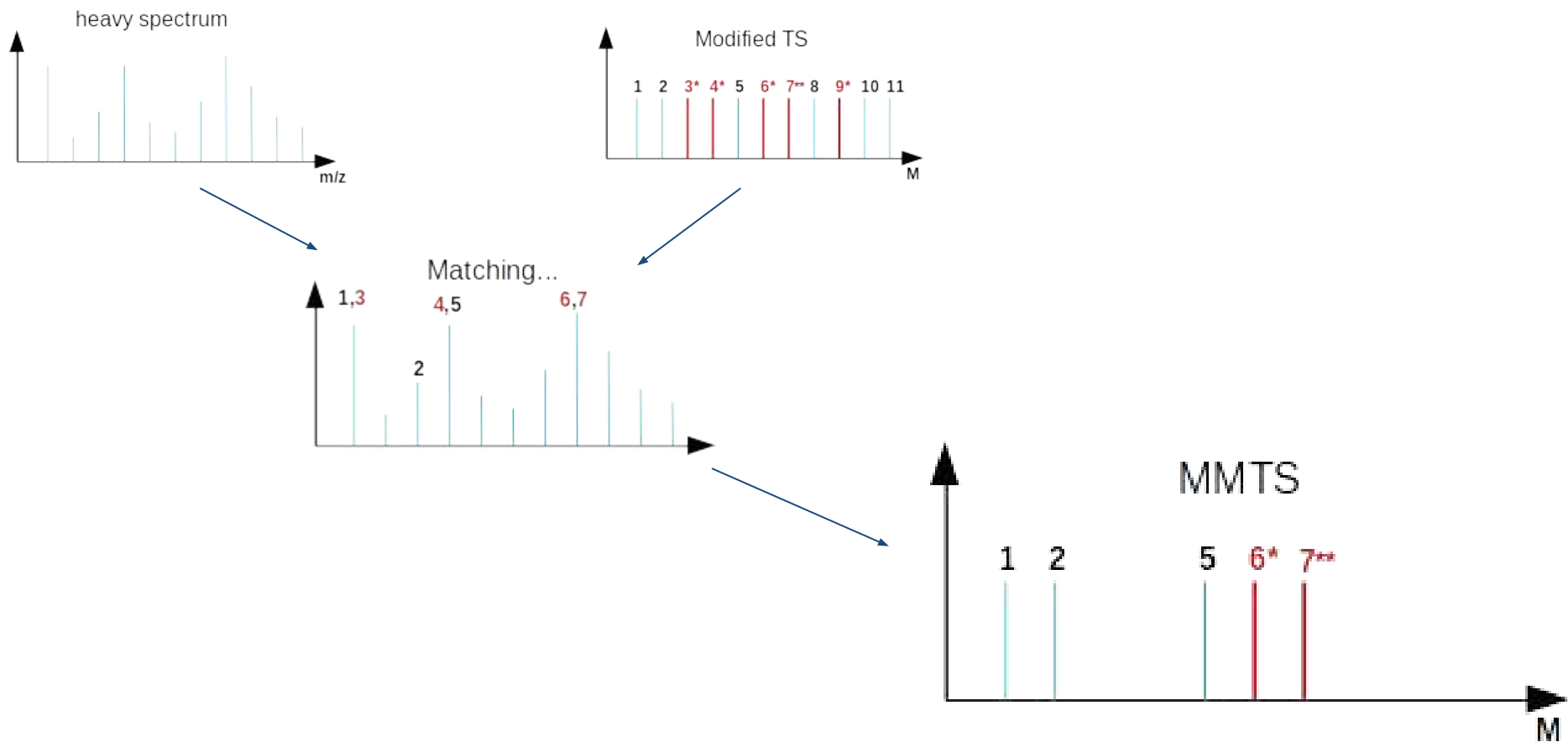


MMTS

Peaks **3** and **4** will be removed!

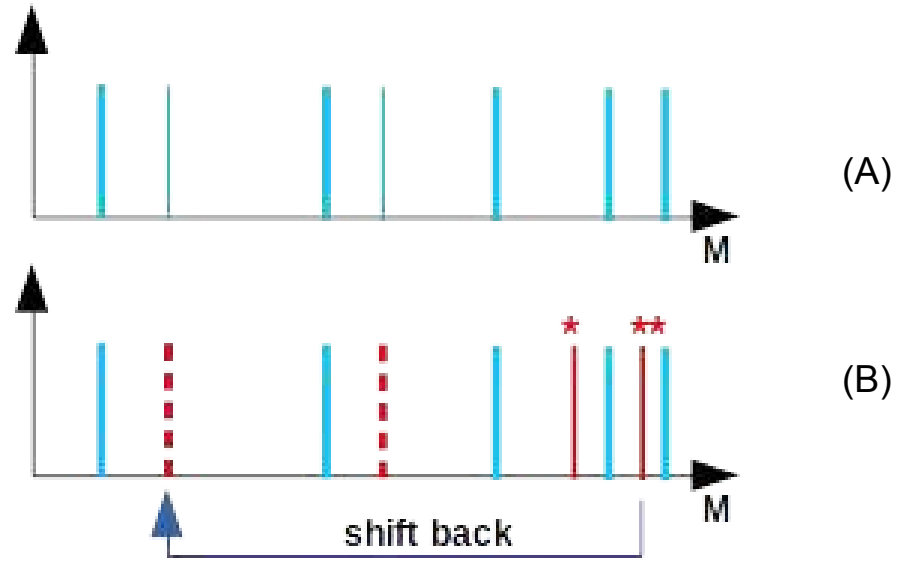


Matched Modified TS filtration



Scoring

- Input: Matched TS (A), Matched Modified TS (B)
- Score = (N, S)
 - N is nonshifted score
 - S is shifted score



- N – # nonshifted peaks from B having equivalent peaks in A
- S – # shifted peaks from B having *shifted-back equivalent* peaks in A

Results

- Dataset 1 (Isoleucine-labeled):
 - 77498 heavy spectra and 70451 light spectra
 - 43618 spectra pairs
 - 2477 pairs after filtering
 - 2358 scored pairs
 - 1648 with shifted score ≥ 10
 - 679 with shifted score in $[5, 10)$

Results

- Dataset 2 (Lysine-labeled):
 - 14067 heavy spectra and 70451 light spectra
 - 12222 spectra pairs
 - 177 pairs after filtering
 - 108 scored pairs
 - 18 with shifted score ≥ 10
 - 58 with shifted score in $[5, 10)$

Future plans

- Implement FDR & p-value computation -- ongoing
- Pipeline performance improvement
- Support all standard amino acids
- Support RiPPquest output

- Test on Dataset 3 from UCSD (20 strains vs 20 modified aa)

Thank you.