Prof. Dr. Knut Reinert,
Julianus Pfeuffer,
Xiao Liang

Proteomics

WS 2014/15

Exercises 5

1. **Feature finding**

   Download velos datasets from `http://svn.code.sf.net/p/open-ms/code/Tutorials/UM_2014/Example_Data/OpenMS/small/` and perform FeatureFinder-Centroided in Knime/Openms. Inspect the featureXML result in TOPPView.

2. **Averagine model**

   The elemental composition used for averagine was:

   $$C\ 4.9384\ H\ 7.7583\ N\ 1.3577\ O\ 1.4773\ S\ 0.0417.$$

   Assume that a peptide of 1666.881 Da is made up of average amino acids. How many times is this peptide heavier than Averagine. What is possibly the number of sulphur?

3. **Averagine model**

   Given the artificial peptides: PEPVIDEYDANVVK, MACCAMMACCACPK

   (a) Approximation of the molecular formula:
   - For each peptide, determine the molecular formula and monoisotopic weight using the table `http://www.webqc.org/aminoacids.php`.(Mind the condensation reaction!)
   - Determine its molecular formula based on the averagine model.
   - What is the error for each element in terms of abundancy?
     Format: C,H,N,O,S rounded to the nearest full percentage (e.g. 12,2,4,7,22)

   (b) Isotopic error:
   Different molecular formulas give rise to different isotopic patterns. Which peptide exhibits a larger deviation in its isotopic pattern from the averagine model?

4. **Linear map alignment**

   Assume you are given a map $M_1 = \{f1 = (180, 270), f2 = (90, 200), f3 = (1000, 300), f4 = (1100, 250)$ and $M_2 = \{t1 = (80, 200), t2 = (301, 270), t3 = (1529, 300), t4 = (1678, 250)$. Calculate some possible values of coefficients a and b.

   - Given pair (f_1, t_2) and (f_2, t_1), a=?, b =?.

- Given pair (f_2, t_1) and (f_3, t_3), a=? b=?.
- Given pair (f_3, t_3) and (f_1, t_2), $a =?, b =?$.
- Given pair (f_4, t_4) and (f_3, t_3), $a =?, b =?$.

5. **MaxQuant pipeline**

   (a) Given two raw/uncentroided peaks P_1 and P_2 extracted at RT i and $i + 1$,

   ```
   P1:
   Intensities   4.00     8.00      10.00     9.00      5.00
   m/z                  999.98 999.99 1000.00 1000.01 1000.02


   P2:

   Intensities   3.00     6.00      11.00     8.00      1.00
   m/z                  999.98 999.99 1000.00 1000.01 1000.02
   ```

   calculate their centroids and tell if they would be joined by MaxQuant. Why (not)? What would be the centroid of the (two?) 3D feature(s)?

   (b) Why is the average mass difference between two ^{12}C and ^{13}C NOT exactly 1 Da?

   (c) Given the intensities at the centroid of three 3D features (ranging from RT i to i+10) at 1000 Th (F_1), 1000.5002 Th (F_2) and 1001.003 Th (F_3).
   $F_1 \leftarrow c(0, 3, 7, 10, 8, 6, 3, 0, 0, 0, 0)$, $F_2 \leftarrow c(0, 0, 0, 2, 2, 4, 6, 8, 5, 1, 0)$, $F_3 \leftarrow c(0, 5, 8, 9, 10, 7, 6, 4, 2, 0, 0)$
   How would the input graph for the deisotoping procedure look like? Check the cosine correlation along RT first.

   (d) Calculating exact m/z and RT similarities for every pair of features in the map is very expensive. How can one limit the amount of comparisons? That means, give two alternative (easy to calculate) thresholding procedures to prefilter the features to consider for a given feature with a given RT span and centroid m/z.

   (e) An MS1 map is given in fig 1.

   Construct the isotope (sub)graph in MaxQuant-like manner under the following assumptions:

   - assuming the correlation along RT axis are all above 0.6;
   - only charges +1, +2 and +3 are considered;
   - due to high deviation in the centroid masses, deviations from the expected isotope difference up to 0.02 Th are allowed.

   Give all sets of features that will be associated (still connected) after the iterative deisotoping and tell by which charge they were grouped.
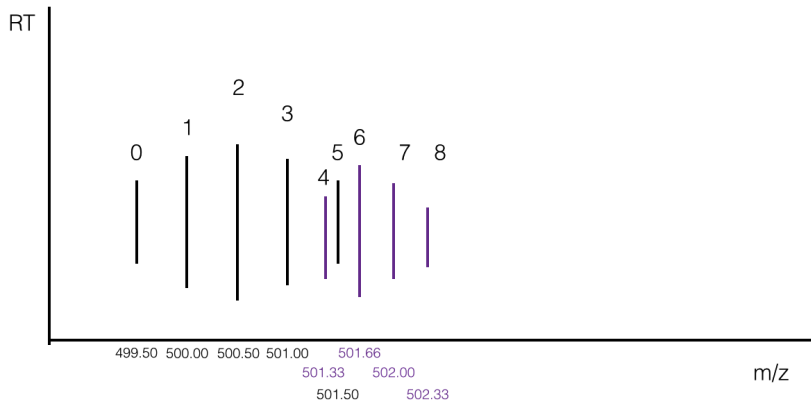
Figure 1: An excerpt of MS1 map

(f) Given that you observe two deisotoped features in a SILAC experiment with fully labeled K and R that are 10 Da apart. Which elements do you have to "add" to the light and heavy isotope patterns respectively, to make their intensities comparable (i.e. so that you compare the same composition)? For which of the isotope patterns (light or heavy) does this result in just a translation by 10 Da?

(g) Now, you have found four corresponding 2D peak pairs in this SILAC pair:

Format (light,heavy): (9, 4), (5.8, 3), (4, 2), (13, 7)

What is the linear least squares solution for the ratio light/heavy?

Why do we need to normalize the ratios afterwards? Is it reasonable to also compare the intensities of different peptides? Why (not)?