

TUTORIAL EXERCISES WITH ANSWERS

Tutorial 1 – Settings

1. What is the exact monoisotopic mass difference for peptides carrying a ^{13}C (and NO additional ^{15}N) labelled C-terminal lysine residue?
 - a. 6.020129
2. What is the chemical formula of the loss occurring in pyro-glutamate formation on N-terminal glutamic acid?
 - a. $-\text{H}_3\text{N}$
3. How many proteins are in the fasta file and how many are in the Skyline background proteome? What could be the reason for the discrepancy?
 - a. The fasta file contains 4022 protein coding genes, 40 of which have an identical counterpart, so there remain 3982 different proteins which are shown in Skyline.
4. Chymotrypsin cleaves C-terminal to which amino acids?
 - a. FWYL, but if followed by P
5. What is the slope for triply charged precursor ions defined in the collision energy equation of the ABI 5500 Qtrap Instrument?
 - a. 0.0544
6. Which transition settings do you need to apply to automatically select the complete doubly-charged γ -ion series of a peptide between 500 and 1000 Da?
 - a. Transition settings \rightarrow Filter \rightarrow Ion charge: 2; product ions: From: ion 1; To: last ion
 - b. Transition settings \rightarrow Library \rightarrow deactivate "If a library spectrum is available, pick the most intense ions"
 - c. Transition settings \rightarrow Instrument \rightarrow Min m/z: 500; Max m/z: 1000

Caution! Do not save any changes you made now for the Exercises. Change these values by back to what we discussed above before you continue.

Tutorial 2 – TransitionList

1. If a dwell time of 5 ms or 20 ms is used for each transition, what would be the resulting cycle times for the generated transition list in a non-scheduled SRM run? Are those two cycle times suitable and what is important to consider in this context?
 - a. $5 \text{ ms} = 313 \times 5 \text{ ms} = 1.57 \text{ s}$
 - b. $20 \text{ ms} = 313 \times 20 \text{ ms} = 6.26 \text{ s}$
 - c. Depending on the LC system, SRM peak widths are in the range of 15 to 30 seconds. Lets assume 15 seconds, hence 1.57 seconds cycle time would lead to 9.5 measurement points over the peak which is acceptable to reconstitute the peak. 6.2 seconds cycle time lead to only 2.5 points per peak, which is not enough. For this transition list, a dwell time above 10 ms, and approx. 10 data points per chromatographic peptide peak, i.e. a cycle time $< 1.5 \text{ s}$, the transition list should either be split over two injections, or the transition list should be run by scheduled SRM.
2. Into how many methods do you need to split this transition list to end up with a dwell time of 15 ms and a maximal cycle time of 1.5 s? Generate those transition lists with Skyline using the “Multiple methods” option in the transition list export.
 - a. You need 4 transition lists to have less than 100 transitions per method.
3. Why is for the peptide VPAARPDVAVLDVR only a single transition automatically selected? Is this suitable for an SRM measurement? What can you do?
 - a. One could manually add b-ions or other y-ions (right-click on precursor m/z → Pick Children → click filter sign → manually select some transitions).
4. For which proteins are two directly flanking peptides selected?
 - a. Hover mouse over protein name, or check numbers in brackets behind peptide sequence.
 - b. Rv1996_Rv1996
 - c. Rv2031c_hspX
 - d. Rv2031c_hspX
 - e. Rv2626c_hrp1

Tutorial 3 – Library

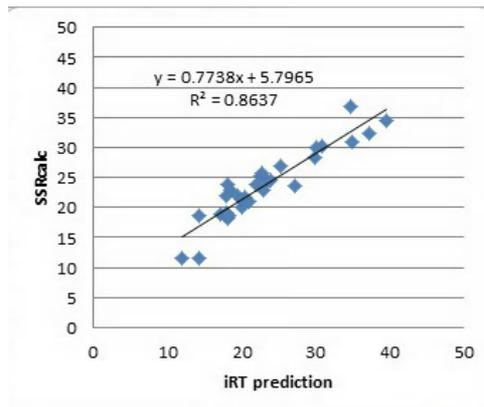
1. For which peptides did the custom-made SRM-triggered MS2 library not lead to an identified SRM spectrum? What could be the reasons for this? What could you do to get MS2-information?
 - a. Peptides without spectrum:
 - i. HLIDDALK
 - ii. LLGSVSSGLLR
 - iii. GILTVSVAVSEGKPTK
 - b. Spectrum triggered, but not identified in search → relax search criteria, use fasta file just containing the peptides for the search.
 - c. No spectrum triggered → try again with higher peptide concentration, longer dwell time, more transitions for triggering, try less complex peptide mix, etc.
2. How can you also include b-ions into the automatic transition selection? How many b-ions show up among the best 5 transitions?
 - a. 86 out of 460 transitions.
3. What are the reasons for the discrepancy in the number of transitions that are picked for the light and heavy peptide versions.
 - a. m/z exceeds only in the heavy version 1250 Da
 - b. m/z is within ± 2.5 Da of the precursor m/z.
4. Manually curate all peptides with less than 5 transitions. How do you choose which other transitions to add?
 - a. Take the ones annotated in the spectrum, if no other peaks are annotated, select the singly charged y-ions above the precursor.
5. In file *SRMcourse_20130715_SRM-MS2*, why do some precursors have >2 transitions?
 - a. Transition settings: always add N-terminal to Pro and C-terminal to Glu/Asp
6. For SRM-triggered MS2, why do we select transitions above the precursor and always add transitions N-terminal to Pro and C-terminal to Glu/Asp.
 - a. Selectivity: singly charged chemical background cannot result in fragments with higher m/z than the precursor
 - b. Sensitivity: if present in a peptide, those transitions tend to be the most intense ones.

Tutorial 4 – Parameters

1. How many steps were used ? What is the step count ?
2. What was the step size used ?
3. What is the increment used for Q3 ?
4. For which peptide has CEO generated the most significant improvement in signal intensity ?
5. To clarify the use of Q3 increments, set the Match tolerance m/z to 0.01 (Settings → Transition settings → Instrument). How many steps do you see now ? Re-adjust the value to 0.055.
6. Since you know the dwell times we used for each file, how long was the cycle time in each file ? Given the cycle time, estimate how many points are measured on average.
7. Although we have optimized the CE per transition, we will not use it in the next tutorials. Why ?
8. Homework: Further explore CEO by plotting the default CE and the optimized CE per transition in Excel. Use the two transition lists that you exported. Plot the default CE against the optimized CE per transition. Approximately how big is the difference ? How could this information affect the choice of step count and step size for future CEO experiments ?
9. Homework: Further explore relationships between quantification and cycle time by loading the Results grid (in View) and browsing through the precursors. Have a look at the background and peak areas for example.

Tutorial 5 – Scheduling

1. What is the difference between a “Retention time calculator” and a “Retention time predictor”?
 - a. Calculator:
 - i. SSRCalc algorithm calibrated on few peptides
 - ii. iRT database for your target peptides calibrated into iRT units on 11 iRT peptides
 - b. Predictor: Instruction on how to apply the calculator to the current dataset (window size, use results..)
2. What is the difference between SSRCalc 3.0 (100A) and SSRCalc 3.0 (300A)?
 - a. Pore size of C18 beads in angstrom.
3. Which retention time window size would you apply if you wanted to run a single scheduled method with maximally 150 concurrent transitions? Could you apply this to SSRCalc as well as iRT prediction?
 - a. 3 minutes (10% of 35 min gradient), possible for well calibrated iRTs, but quite small for SSRCalc (we recommend window of $\geq 30\%$ of gradient).
4. To compare the predicted retention times from SSRcalc and iRT calculator, generate a graph showing the correlation between both predictors. Which correlation coefficient (R^2) is achieved?
 - a. E.g. $R^2 = 0.86$ (see graph)



5. In a scheduled SRM method, not the dwell time but the cycle time is kept constant. This means that the dwell time for a given transition can change over time depending on the number of concurrent transitions. To which minimal dwell time does this approximately lead when using the transition list generated above with a 4 min retention time window size and a fixed cycle time of 2 seconds? At which time of the gradient does this minimal dwell time approximately occur?
 - a. Peak time 20 minutes, ca. 200 concurrent transitions = ca. 10 ms dwell time
6. Why do continuously changing dwell times in scheduled SRM runs not lead to observable differences in signal intensities? (longer dwell times should increase signal intensity, shorter dwell times decrease signal intensity)
 - a. Intensities are given in counts per second (cps), meaning that all intensities are normalised by the dwell time that was used to acquire them.

Tutorial 6 – Manual Analysis

1. What are advantages and disadvantages of label-free and label-based SRM data analysis?
 - a. Pro label-based
 - i. Isotope labelling adds a lot of additional information (more transitions, compare intensity of transitions between H/L) → more confidence in peak picking (identification)
 - ii. Labelled reference peptides can be spiked in at concentrations that are easily detectable → more confidence in peak picking (identification)
 - iii. Labelled ref pepts allow to pick noise for the endogenous peptide if not expressed → no missing values. (also possible with label-free, but runs have to be “aligned”...)
 - iv. With heavy peptides normalization in every run possible, you can account for LC and MS instrument variations (depending on the sample preparation workflow also for variations in there). This should lead to smaller standard deviations/CVs for the isotope-labelled analysis.
 - b. Pro label-free
 - i. Half the number of transitions to be measured → more transitions or longer dwell times.
 - ii. Cheaper and faster
2. How important is the careful placement of the peak boundaries in label-free compared to label-based data analysis?
 - a. Much more in label-free, because in isotope-labelling it is likely that both peaks get affected by a too short or too long peak boundary in a similar way so that the final ratio output is not affected.
3. Why should the ratio between a spike-in and endogenous peptide ideally be around the value 1 for accurate protein quantification?
 - a. Linear dynamic quantification range is best.
 - b. Same peak width, integrations boundaries.
 - c. No bleeding effect of light impurities from a heavy synthetic peptide possible!

Tutorial 7 – mProphet

1. Which are the scores that mProphet takes into account for the discrimination of true and false peak groups for the label-based and the label-free workflow, respectively?
Check parameter files:
 - a. Label
 - i. log10_total_xic
 - ii. intensity_correlation_with_assay
 - iii. xcorr_coelution_score
 - iv. xcorr_shape_score
 - v. weighted_xcorr_shape_score
 - vi. light_heavy_correlation
 - vii. light_heavy_shape_score (**main score**)
 - viii. light_heavy_coelution_score
 - ix. abs_delta_iRT
 - b. Label-free
 - i. log10_total_xic
 - ii. intensity_correlation_with_assay
 - iii. xcorr_coelution_score
 - iv. abs_Tr_deviation
 - v. xcorr_shape_score (**main score**)
2. What is the typical range in delta_iRT of the best-scoring peak group?
 - a. Typical iRT deviation of best peak group approximately 1-2 iRT units.
3. Which d_score cut-off would you choose to get good results for the label-based and label-free analysis, respectively? Which FDR and sensitivity would you get according to the suggested cut-off (mProphet_raw_stat.xls)? Is the FDR estimated by mProphet what you would expect? Why/why not?
 - a. Label: The peak groups above 1.2 – 1.5 look all good. This would correspond to an FDR of 0.02% and a sensitivity of 76% according to mProphet_raw_stat.xls.
 - b. Label-free: The peak groups above 2.5 look all good. This would correspond to an FDR of 0.1% and a sensitivity of 69% according to mProphet_raw_stat.xls.
 - c. The FDR given by mProphet is not very accurate because there are too few data points to train the model well. The bigger the data set, the more accurate this estimation will be.
4. Look at a few examples where mProphet picked the wrong peak and try to explain why this happened.
 - a. E.g. AQAATA.. in file A1-s2 (label) → 3rd is the right one. E.g. delta_iRT is quite big...
5. How would the results change if you used as an input for mProphet a transition list which has been refined before in Skyline (i.e. containing only non-interfered transitions)?
 - a. Identification more sensitive and quantification more accurate, however, how trustworthy are FDR and discriminant score values when an assay has been tweaked and optimized before to make it fit perfectly to the chromatogram you think it is the correct one...