# 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. -H2O

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 y-ion series of a peptide between 500 and 1000 Da?
    a. Transition settings → Filter → Ion charge: 2; product ions: From: ion 1; To: last ion
    b. Transition settings → Library → deactivate "If a library spectrum is available, pick the most intense ions"
    c. Transition settings → Instrument → Min m/z: 500; Max m/z: 1000

# 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? (Assuming that no additional time is needed to switch between transitions.) Are those two cycle times suitable and what is important to consider in this context?
   a. 5 ms = 313 x 5 ms = 1.57 s
   b. 20 ms = 313 x 20 ms = 6.26 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 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?
   a. Peptides without spectrum:
      i. HLIDDALK
      ii. LLGSVSSGLLR
      iii. GILTVSVAVSEGKPTEK

   b. Spectrum triggered, but not identified in search → relax search criteria, use fasta file just containing the peptides for the search.
   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.

# Tutorial 4 – Parameters

A) Collision Energy Optimization (CEO)

1. How many steps were used ? What is the step count ?
   a. 11 steps. Hence the step count is 5 (2*5+1=11).

2. What was the step size used ?
   a. Step size is 2.

3. What is the increment used for Q3 ?
   a. The increment for Q3 (which is meant to enable the concurrent measurement of these transitions at various CEs) is 0.01.

4. For which peptide has CEO generated the most significant improvement in signal intensity ?
   a. The peptide AQAATAGIDDLRPALIR had the most significant improvement in signal intensity (mostly attributed to an improved signals for the +2 fragment ions).

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.
   a. You see 7 steps instead of 11 steps. Since the increment in Q3 masses is 0.01, changing the mass tolerance to 0.01 would mean that two consecutive steps are no longer separated.

6. Although we have optimized the CE per transition, we will not use it in the next tutorials. Why ?
   a. We plan to use the dot product score. Optimized CEs were not used in generation of the library , and therefore are avoided in the next tutorials.

7. 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 ?
   a. The difference between the default and optimized CE can go up to 10eV , for a few transitions. By increasing the step size and/or step count one could explore a larger range of CEs. However, for most transition this "window" of CEs is sufficient.
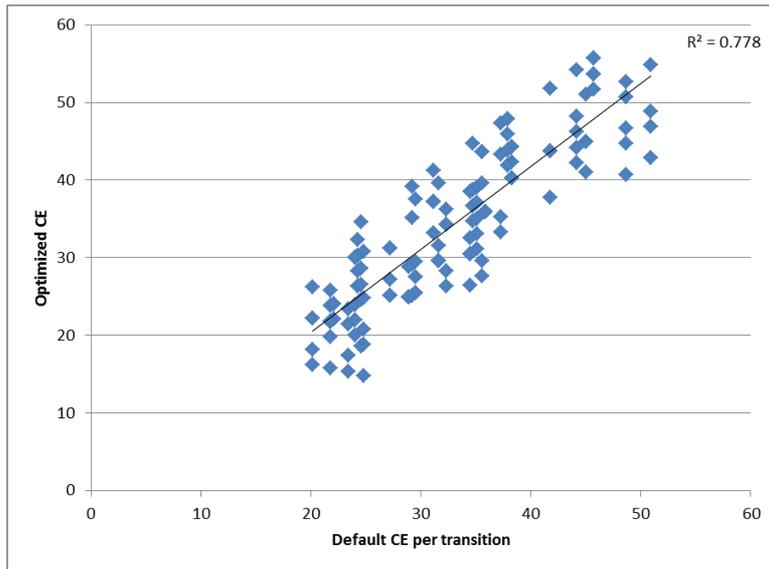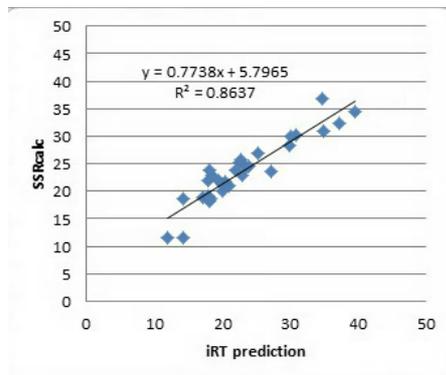
B) Dwell time / Cycle time

1. 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. (for simplicity, use only the formula given).
   a. For the dwell times 10,20,40,60,100ms, the cycle time would be roughly #transitions*dwelltime, and the estimated number of points would be 30/cycle time (where 30 stands for an estimated width of 30 seconds per peak).

| dwell time | cycle time | number of points |
|---|---|---|
| 10 | 1.43 | 20.97902 |
| 20 | 2.86 | 10.48951 |
| 40 | 5.72 | 5.244755 |
| 60 | 8.58 | 3.496503 |
| 100 | 14.3 | 2.097902 |

2. In this tutorial we changed the dwell time, and thus increased the cycle time. How could be change only the cycle time or only the dwell time ?
   a. By changing the number of transitions per method we could change only one parameter while keeping the other parameter unchanged.

# Tutorial 5 – Scheduling

1. Compare the iRT values that you generated manually in Excel with the iRT values generated automatically within Skyline (you can copy-paste the Skyline values out of the "Retention Time Calculator" table).

2. iRT standard peptide GTFIIDPGGVIR seems to be an outlier in the retention time linear regression viewer of file *arielb_Q130619_017.wiff* (in Skyline file SRMcourse_20140211_iRT.sky). Can you explain what is the issue here?

3. 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..)

4. 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 ≥30% of gradient).

5. 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. R2 = 0.86 (see graph)



6. 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

7. 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.

8. What is the difference between SSRCalc 3.0 (100A) and SSRCalc 3.0 (300A)?
    a. Pore size of C18 beads in angstrom.

# Tutorial 6 – ManualAnalysis

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 peps 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 be around 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!