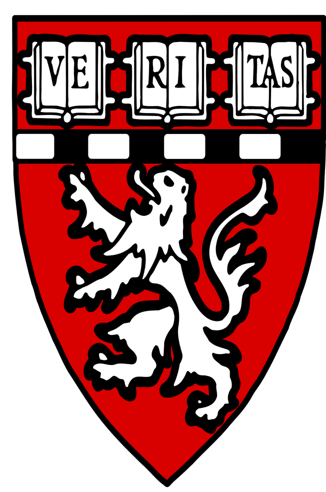


Finding the appropriate LC setup for proteomics experiments depending on the sample complexity using chip-based columns

Jan Muntel[1,2]; Helena Svobodova[3]; Gary A. Valaskovic[3]; Saima Ahmed[1,2]; Kevin Broadbelt[1,2]; Omar Barnaby[1,2]; Hanno Steen[1,2]

1. *Proteomics Center at Boston Children's Hospital, Boston, MA*
2. *Department of Pathology Boston Children's Hospital and Harvard Medical School, Boston, MA*
3. *New Objective, Inc., Woburn, MA*

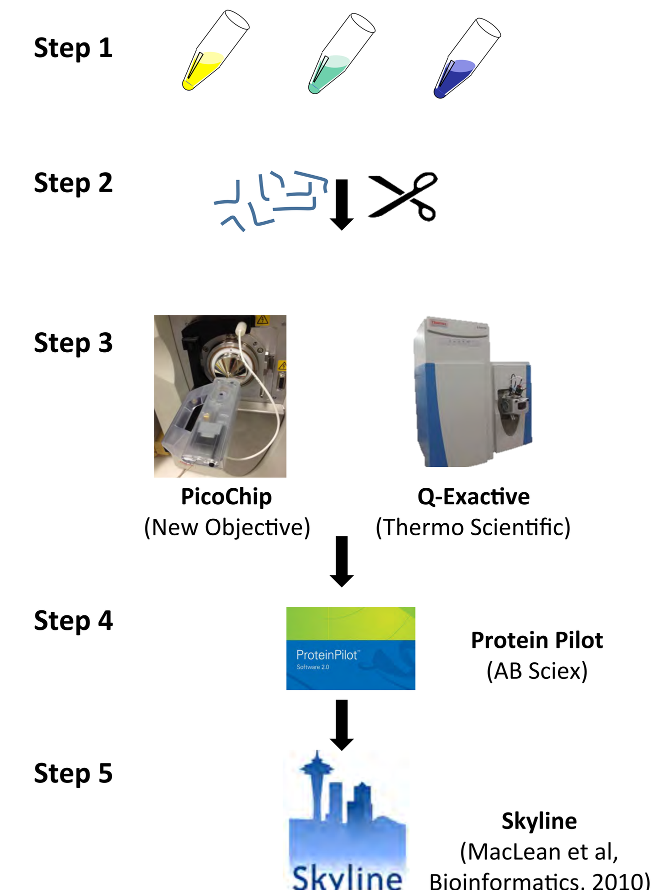
Introduction

LC-MS/MS has become the method of choice for comprehensive proteome analysis. Due to the large variety of sample types, LC-MS/MS has to deal with a wide range of sample complexities, ranging from 10-100 proteins in a purified protein complex to more than 10,000 proteins in tissue lysates.
In order to cope with such variety in sample type and especially complexity, the choice of an appropriate LC setup is crucial to identify and to subsequently quantify as many peptides and proteins as possible. The number of samples in large scale experiments are also increasing making effective use of instrument time increasingly important.
To maximize the information content for a given sample, we introduce a guideline to choose the appropriate LC setup, regarding: 1) gradient length, 2) column length and most important 3) sample complexity.

We tested the following parameters/samples.
1) gradient length: 30 min., 60 min., 120 min., 240 min.
2) column length: 10.5 cm, 15 cm
3) sample complexity:
 a) urine (diluted low complex, ~500 proteins)
 b) bacterial cell lysate (*Mycobacterium tuberculosis,* ~2,000 proteins)
 c) human cell lysate (K562, >10,000 proteins)

Methods

We prepared urine and *Mycobacterium tuberculosis* (Mtb) samples by tryptic digestion and used a digest of a human cell lysate sample (K562, Promega) as a highly complex sample. To compare peak widths in different setups, we added an equimolar 8-peptide mixture (PicoSure™, New Objective, Inc.) in a final concentration of 50 fmol/µl. These samples were analyzed by 30, 60, 120 and 240 min. gradients on chip-based columns (5-28% buffer B, PicoChip® - New Objective, Inc.) packed with BEH, 1.7 µm, C18 resin (Waters) connected by a PicoClear™ union (New Objective, Inc.) to an Eksigent LC system (AB Sciex). Data were acquired on a Q-Exactive mass spectrometer (Thermo Scientific). Data were searched with ProteinPilot (AB Sciex) and a 1% FDR cut-off was applied. MS1 signal intensities of the standard peptides were extracted using Skyline.



Step 1
three samples of different complexity

Step 2
trypsinization and spike-in of equimolar 8-peptide mix as standard

Step 3
LC-MS/MS analysis using 4 different gradients from 5 to 28% buffer B loading: 500 ng of Mtb and human digest; ~50 ng of urine digest

Step 4
database search (ProteinPilot)

Step 5
determination of peak widths of standard peptides (Skyline)

Results of 10.5 cm column

Initially we started to compare the influence of the gradient length on the number of identified proteins/peptides for the 10.5 cm column. Within the 30 min. gradient, we identified in the human sample 1,971 proteins/9,100 peptides, for the Mtb sample 756/5281 and 332/2597 in the urine sample. Doubling the gradient led to the identification of 24% more proteins and 45% more peptides for the human sample and 19%/28% more identifications for the Mtb sample. Any extension of the gradient for the urine sample did not improve the number of identified proteins (Fig. 1a). This can be explained by an increase in peak width from 0.11 to 0.34 min. as determined for the eight spiked-in standard peptides (FWHM, Fig. 1b).

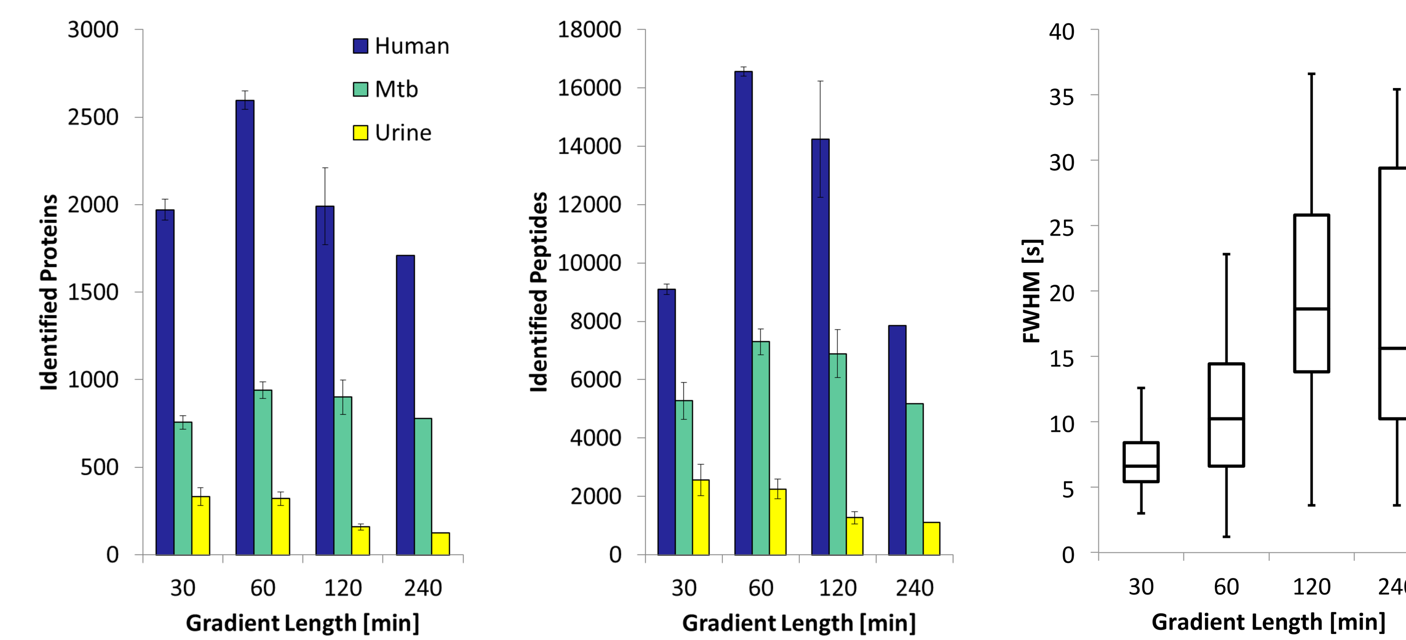


Figure 1a: Overview of number of identified proteins and peptides by increasing the separation time from 30 min., 60 min., 120 min. to 240 min. for 3 samples (human, Mtb, urine) on a 10.5 cm chip-based column.

Figure 1b: Overview of peak width (FWHM) for the different gradient lengths for 8 spiked-in peptides.

Results of 15 cm column

In the next step we analyzed the samples under the same conditions with a 15 cm column. We saw a steady increase in the number of identified proteins/peptides with increasing gradient length (Fig. 2a). Especially for the human samples an increase in gradient length from 60 min. to 2 hrs. increased the number of identified peptides by 26% (14% more proteins). Roughly the same increase in identified peptides was observed by extending the gradient to 240 min. For the Mtb sample the best results were achieved with a 120 min. gradient, whereas for the very low complex urine sample a maximum in identified proteins and peptides was achieved with a 30 min. gradient. Although the peak capacity of the 15 cm column is higher, the peak width still increases with gradient length (Fig. 2b) explaining the low identification result for the urine sample.

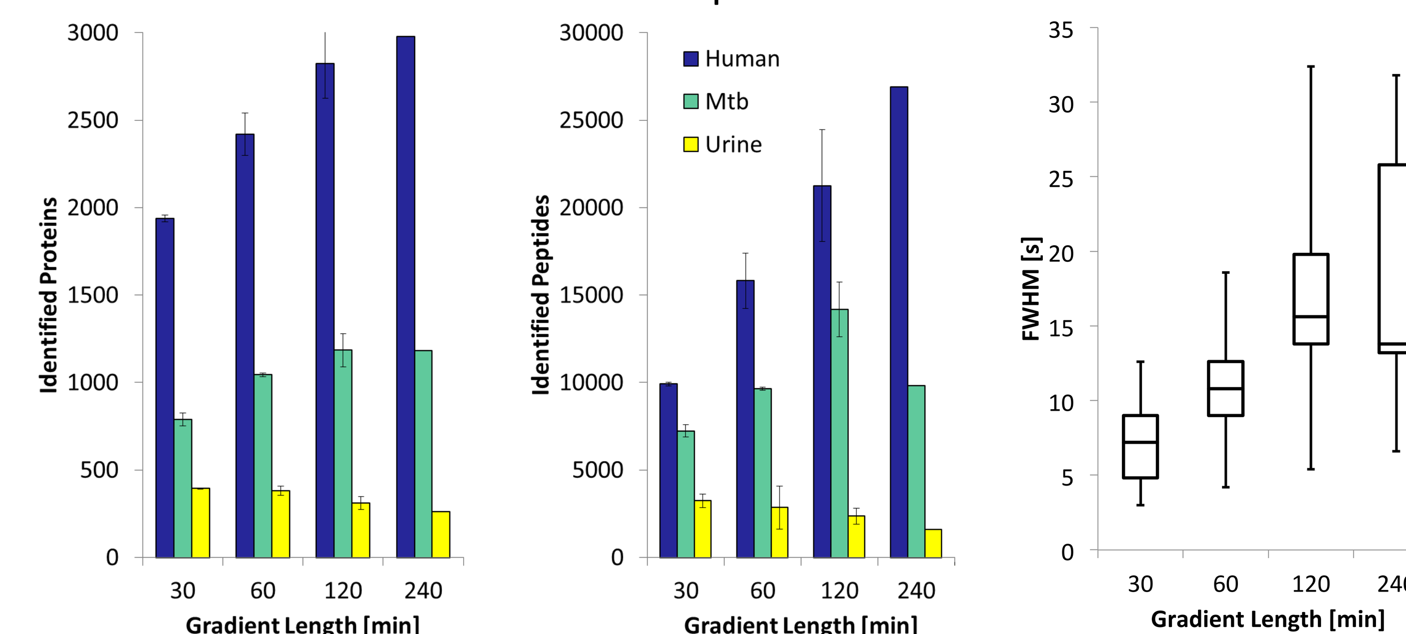


Figure 2a: Overview of number of identified proteins and peptides by increasing the separation time from 30 min., 60 min., 120 min. to 240 min. for 3 samples (human, Mtb, urine) on a 15 cm chip-based column.

Figure 2b: Overview of peak width (FWHM) for the different gradient lengths for 8 spiked-in peptides.

Choosing the appropriate LC setup

Low complex/low abundant sample

In this experiment we used a low abundant urine sample as an example. Independent of the column length, a short, **30 min.** gradient (Fig. 3a) resulted in the best identification result. Additionally a longer column increased the number of identified peptides/protein only very little (10%), so that a **10.5 cm** long column is sufficient for this kind of sample.

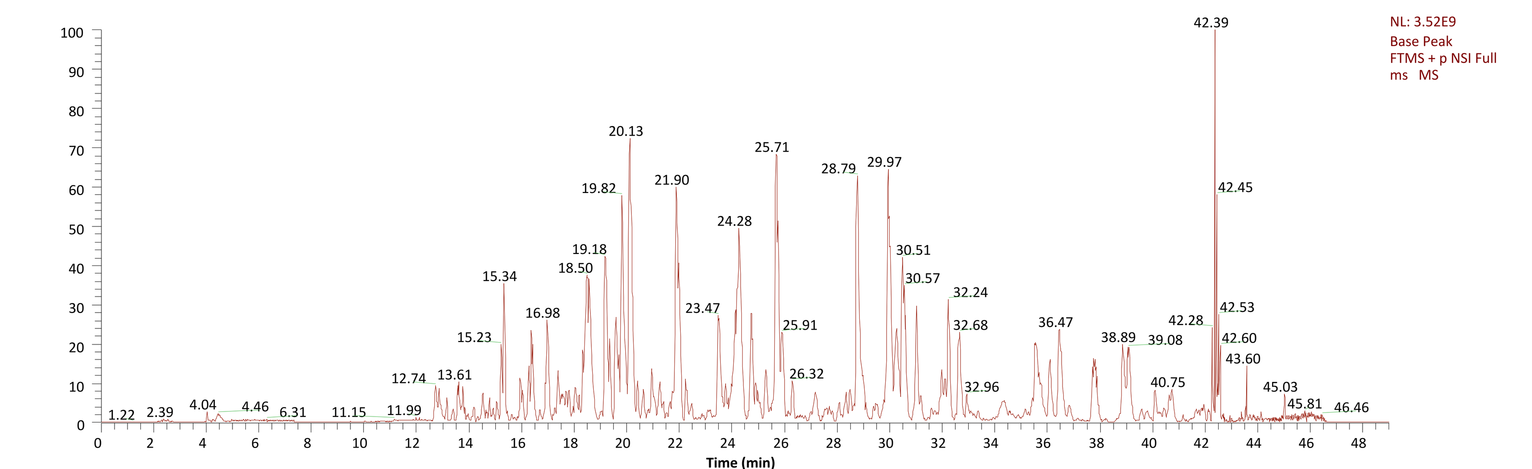


Figure 3a: Base peak chromatogram of a urine protein digest separated on a 10.5 cm column by a 30 min. gradient.

Medium complex sample

A cell lysate of *Mycobacterium tuberculosis* (Mtb) was analyzed as a medium complex sample. For this sample an optimum of identified proteins and peptides was achieved with a **120 min.** gradient on the **15 cm** column (Fig. 3b). Longer separation times didn't lead to more identification due to a decreased chromatographic sensitivity. Further optimization of the gradient will further increase the number of identifications.

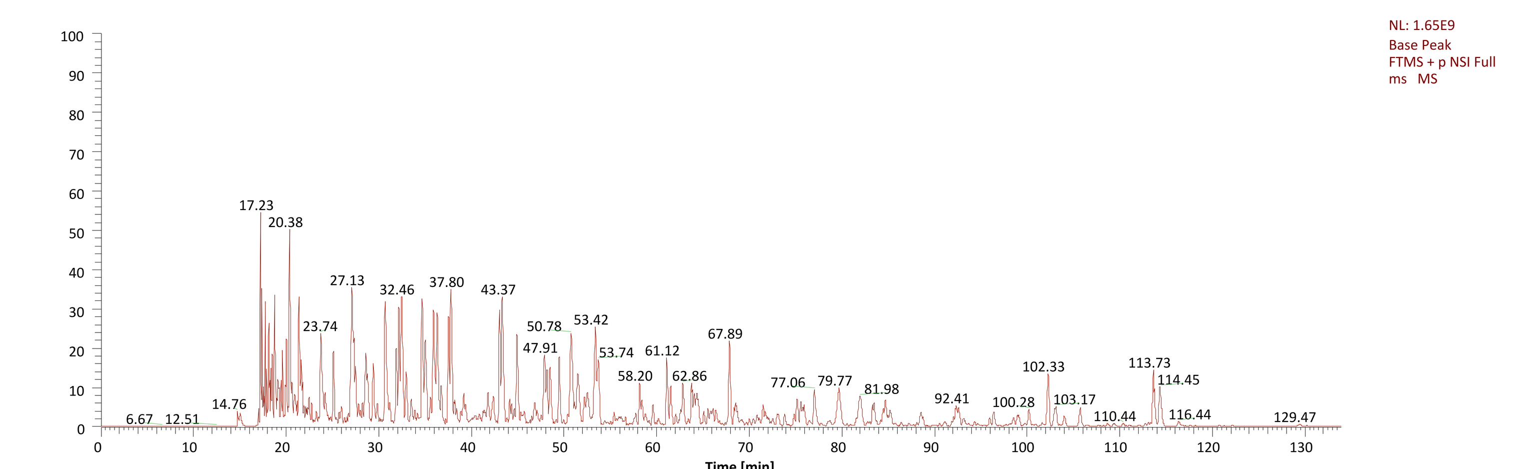


Figure 3b: Base peak chromatogram of a Mtb cell lysate separated on a 15 cm column by a 120 min. gradient.

Highly complex sample

The benefit of applying the longer column (**15 cm**) was highest for the human cell lysate digest. The increased peak capacity significantly improved the identification result. The highest number of peptides and proteins were identified in the 240 min. run. However, this modest increase in identified peptides (21%) came at the expense of doubled gradient length. Thus, if instrument time is limited, we recommend a **120 min.** run (Fig. 3c). Longer columns with an even higher peak capacity will probably circumvent the throughput disadvantage of gradients longer than 120 min. and allow for the identification of even more than 3,000 proteins per run.

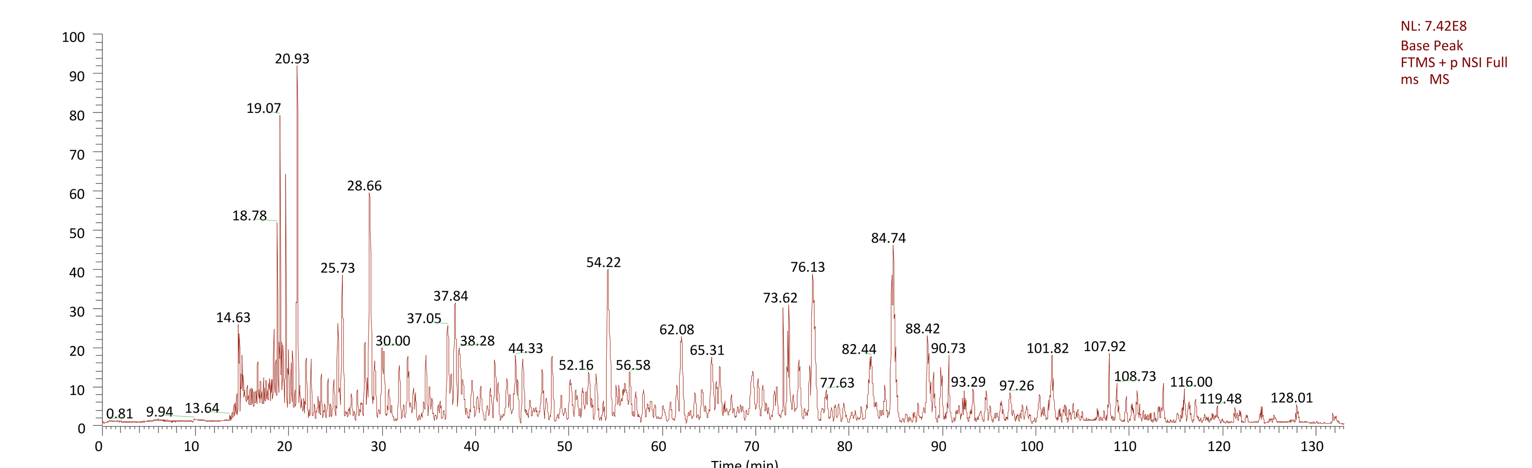


Figure 3c: Base peak chromatogram of a human cell lysate separated on a 15 cm column by a 120 min. gradient.