nicolas.housset@vib-ugent.be

# Focus on relatively hydrophilic peptides for targeted proteomics

***Nicolas Housset[1,2], Sven Degroeve[1,2], Lennart Martens[1,2]***

[1]Department of Medical Protein Research, VIB, B-9000 Ghent, Belgium
[2]Department of Biochemistry, Ghent University, B-9000 Ghent, Belgium

## Introduction

Liquid chromatography (LC) coupled to mass spectrometry (MS) is one of the main techniques used in proteomics. LC releases peptides gradually to the mass spectrometer: we can record the moment a peptide is detected after release as the retention time. In general, the higher the hydrophobicity of a peptide, the longer its retention time will be. Targeted proteomics needs this retention time information for selected reaction monitoring (SRM) scheduling: the mass spectrometer will be tuned to specifically look for one peptide in a window around its expected retention time.

Precision of retention time prediction is thus critical: a smaller confidence interval means a smaller window and additional possibilities for the mass spectrometer to target peptides. We make the hypothesis that we can still identify most of an organism's proteome by focusing on their relatively hydrophilic peptides, thus allowing for shorter targeted proteomics experiments, and that the precision of retention time prediction is better in the first part of the gradient, thus increasing the density of transitions schedulable by unit of time. Those two combined effects would lead to an unchanged number of protein identifications in a shorter amount of time.

## Methods

For retention time prediction, we use ELUDE[1] as trained on data stored in our MS-LIMS[2] in-house data repository. The experiments selected to train ELUDE all come from the Thermo-Finigan LTQ-Orbitrap Velos. To test our hypothesis, we will apply the retention time prediction on the study 6 of the CPTAC[3] experiment: 3 different laboratories running in triplicate samples of yeast and increasing concentrations of UPS48, a human protein.

To assess the performance of the retention time, we use the 95% centile of the absolute value of the error in the prediction.

At the LC-run level, we consider a protein to be identified if two unique peptides are identified with a 1% FDR.

## Results

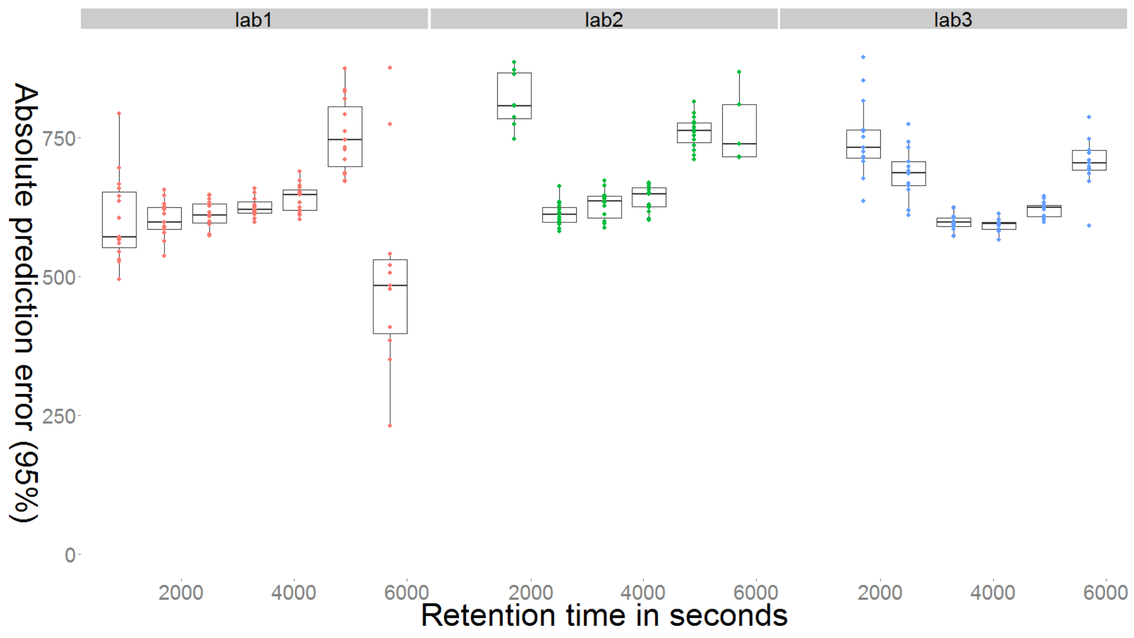


**Fig 1. Error in retention time prediction is lower in the middle of the gradient, but this difference is of limited impact for targeted proteomics**

For each lab, we separate the gradient in 800 seconds interval, and we compute the 95% centile of the error at the sample level. Each boxplot represents those centile by lab and by gradient interval.
We see that the error in the prediction is lower and more stable in the middle of the gradient, but it is partly due to the higher understanding from ELUDE in this part of the gradient of the relationship between peptide properties and retention time.

Moreover, the prediction error does not change that much along the gradient to justify a thinner time width on transition scheduling.

Fig.2 shows for each sample and technical replicate from lab2 the evolution of the number of proteins identified as we move further in the gradient. The retention time shown on the x-axis starts at 900 seconds and stops at 6100 seconds: no further protein is identified between 6100 and 8100 seconds, but it is due to the fact that no MS2 spectra are recorded in this interval.

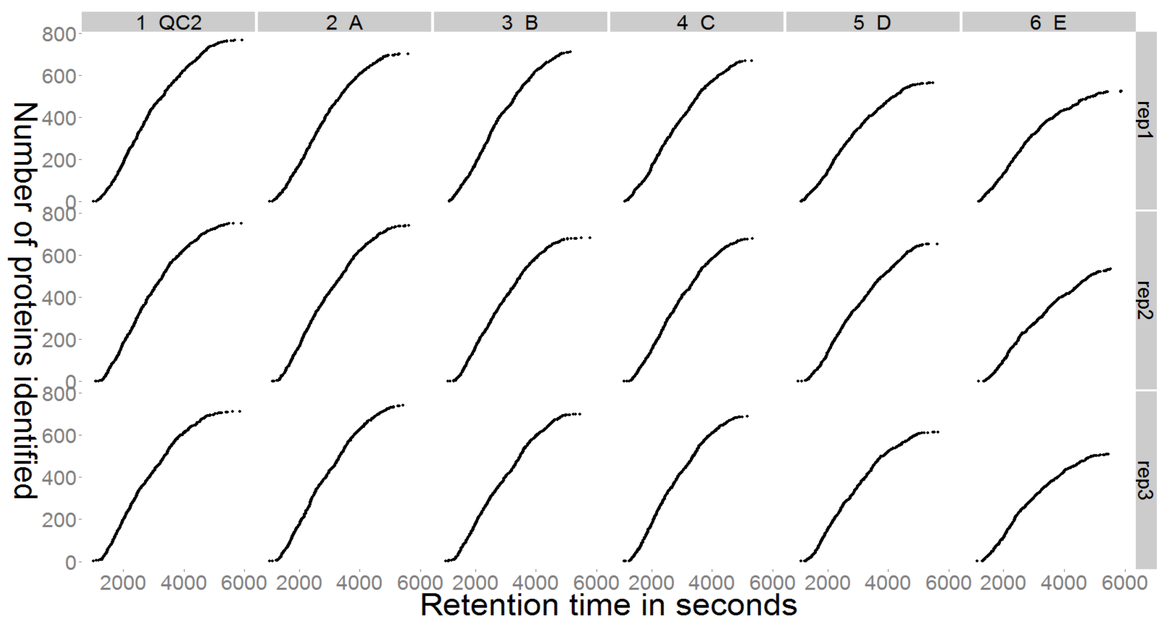


**Fig. 2. Limiting the chromatography gradient reduces the number of proteins identified**

From an experimental point of view, reducing the gradient will affect negatively the proteome coverage. In fact, we could say that the gradient is already reduced because of the absence of MS2 spectra after 100 minutes.

## Future work and outlook

Retention time prediction is an issue that has been existing for 40 years. More than precision, what we really lack in order to make targeted proteomics really efficient and used by companies is an algorithm which is both able to explain what are the peptide/protocol properties with the biggest influence on the retention time AND to adapt to other experimental setups.
Our plan is to develop such an algorithm, and we expect to find differences in the relevant properties at different peptide lengths.

**References**

[1] Moruz L., *et al.* Chromatographic retention time prediction for posttranslationally modified peptides, Proteomics 2012, 12 1151-1159.
[2] Helsens K, *et al*. Ms_lims, a simple yet powerful open source LIMS for mass spectrometry-driven proteomics. Proteomics. 2010 Mar;10(6):1261-4.
[3] Paulovich A. G., *et al.* Interlaboratory Study Characterizing a Yeast Performance Standard for Benchmarking LC-MS Platform Performance. Molecular & Cellular Proteomics 2010, 9.2: 242-254