

# Retention Time Prediction for Post-translationally Modified Peptides



Luminita Moruz<sup>1</sup>, An Staes<sup>2</sup>, Joe Foster<sup>3</sup>, Maria Hatzou<sup>1</sup>, Evy Timmerman<sup>2</sup>, Lennart Martens<sup>2</sup>, Lukas Käll<sup>1</sup>

<sup>1</sup>Department of Biochemistry and Biophysics, Stockholm University, Stockholm, Sweden. E-mail: luminita.moruz@scilifelab.se

<sup>2</sup>Department of Medical Protein Research, VIB, Ghent, Belgium. <sup>3</sup>EMBL Outstation, European Bioinformatics Institute, Wellcome Trust Genome Campus Hinxton, Cambridge, UK.

## Overview

We extended our previously reported retention time predictor, Elude [1], to encompass post-translationally modified peptides. The new version of Elude was evaluated on datasets including six different modifications. The results showed that Elude yields equally good performances for modified and unmodified peptides.

## Introduction

Retention time (RT) predictors, that estimate the time that a peptide spends on the chromatographic column using the peptide sequence, are valuable tools for mass spectrometry (MS)-based proteomics. In shotgun proteomics (Fig. 1) they are employed to filter out incorrect identifications [2], and in targeted proteomics to design efficient assays [3]. Nevertheless, these applications are not available for post-translationally modified peptides (PTM) due to the lack of RT predictors for such peptides. Here we propose an extension of our algorithm, Elude, that predicts RT for peptides carrying arbitrary PTMs.

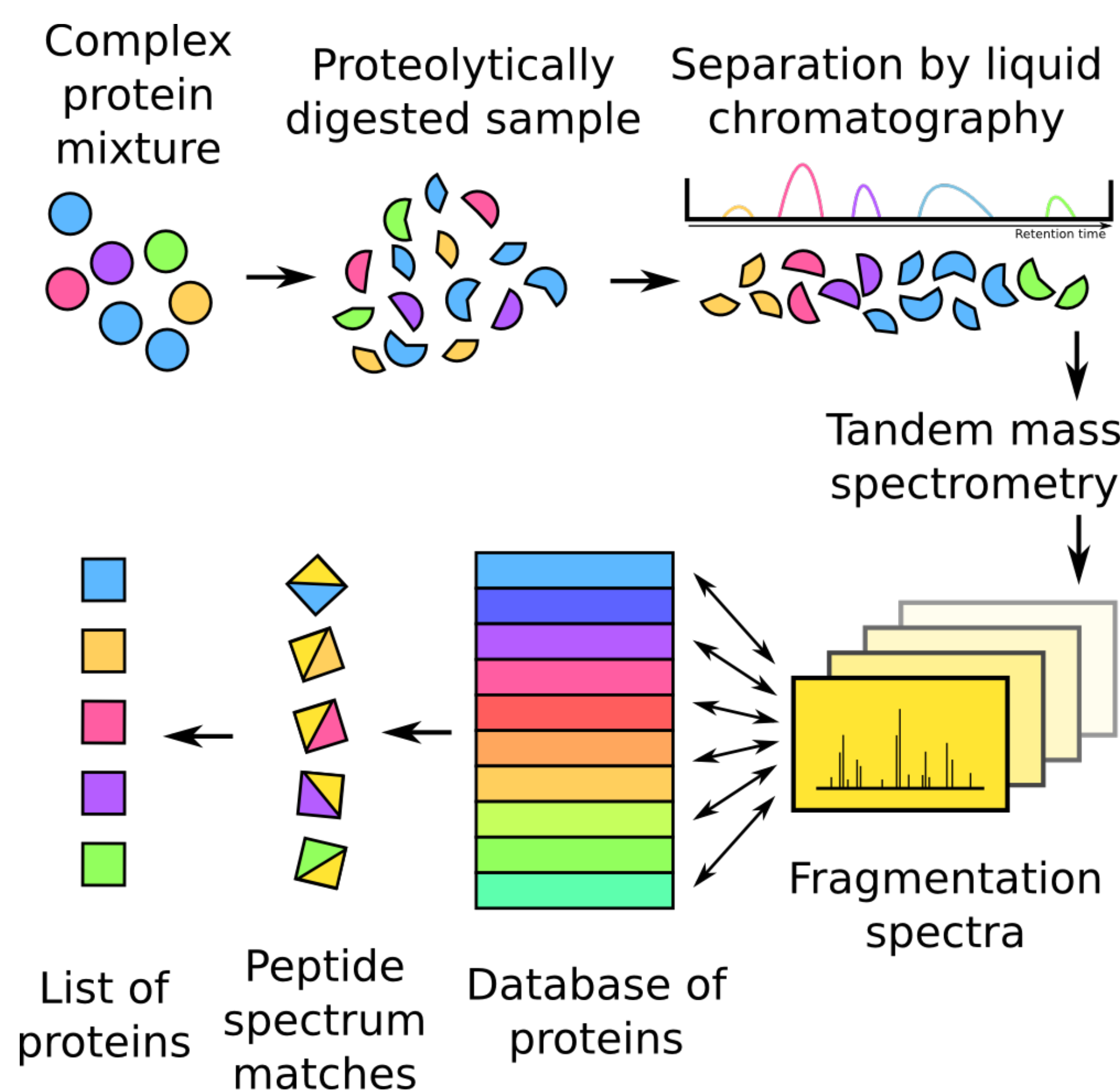
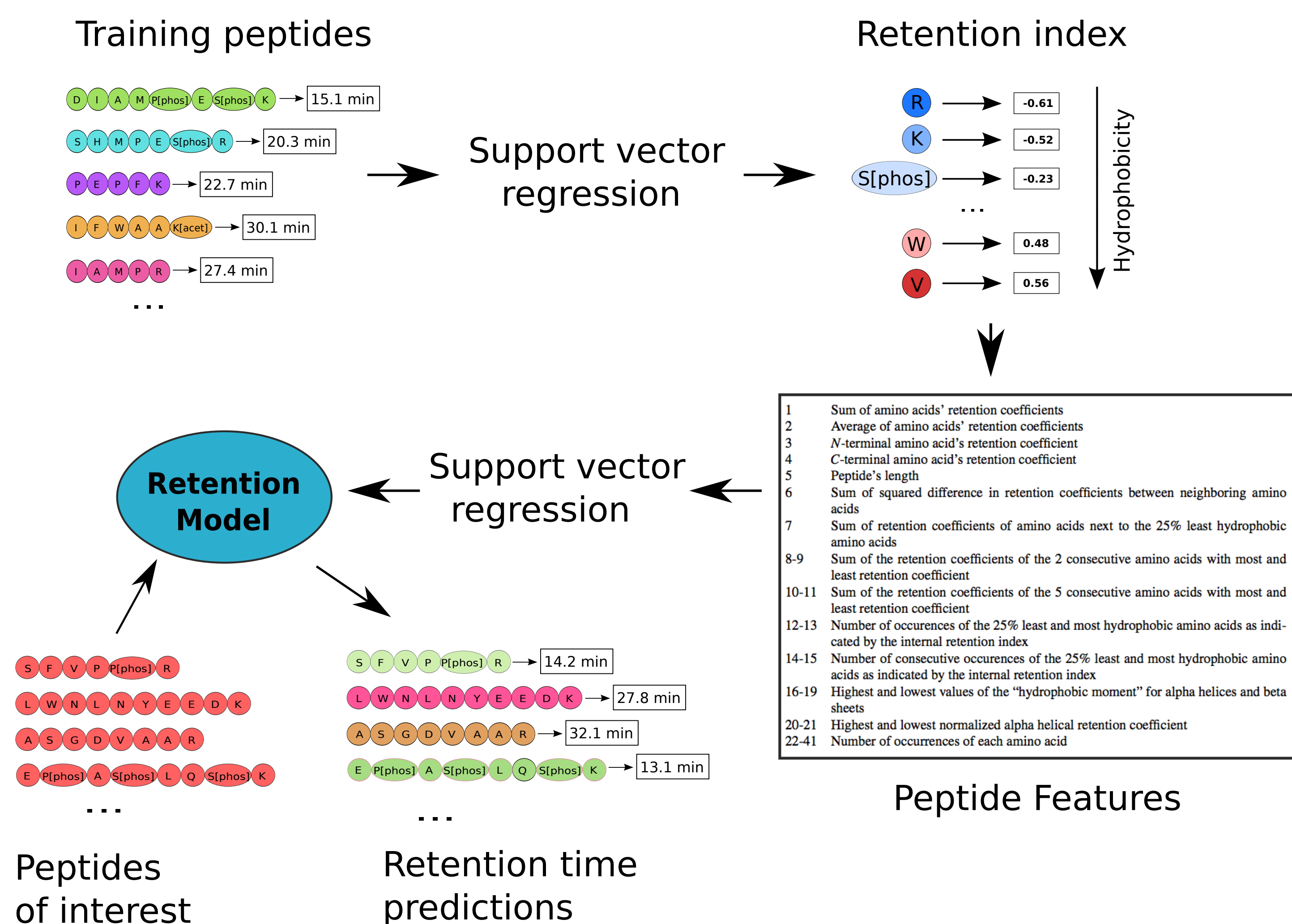


Figure 1. The shotgun proteomics workflow

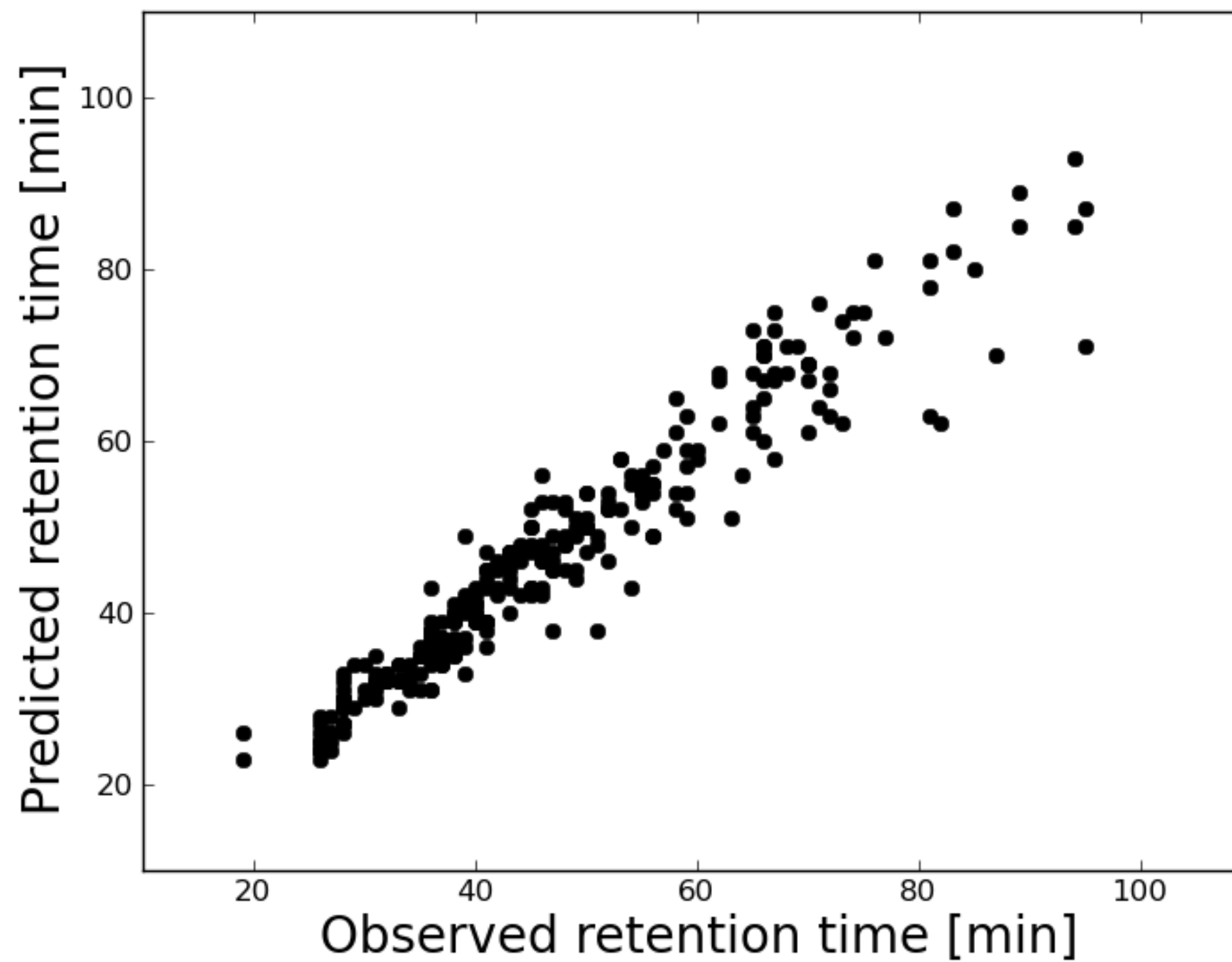
## Elude



**Figure 2. Elude's workflow.** The training procedure in Elude consists of 3 steps: i) a set of retention time coefficients forming a retention time index is derived ii) the features describing each training peptide are computed iii) the final retention time model is built. The resulted model can be subsequently employed to predict RT for other peptides of interest.

## Results

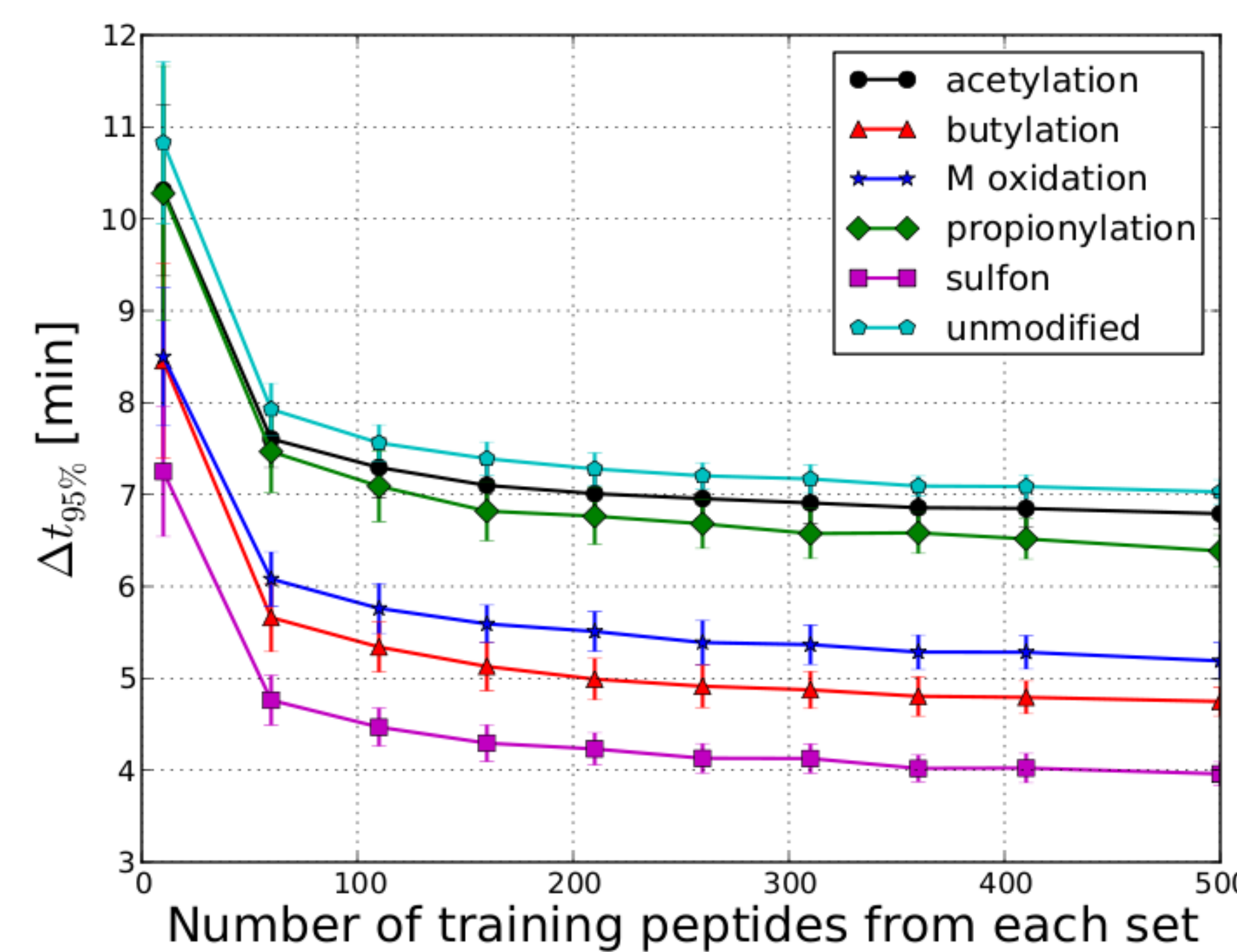
### • Prediction accuracy for phosphorylated peptides



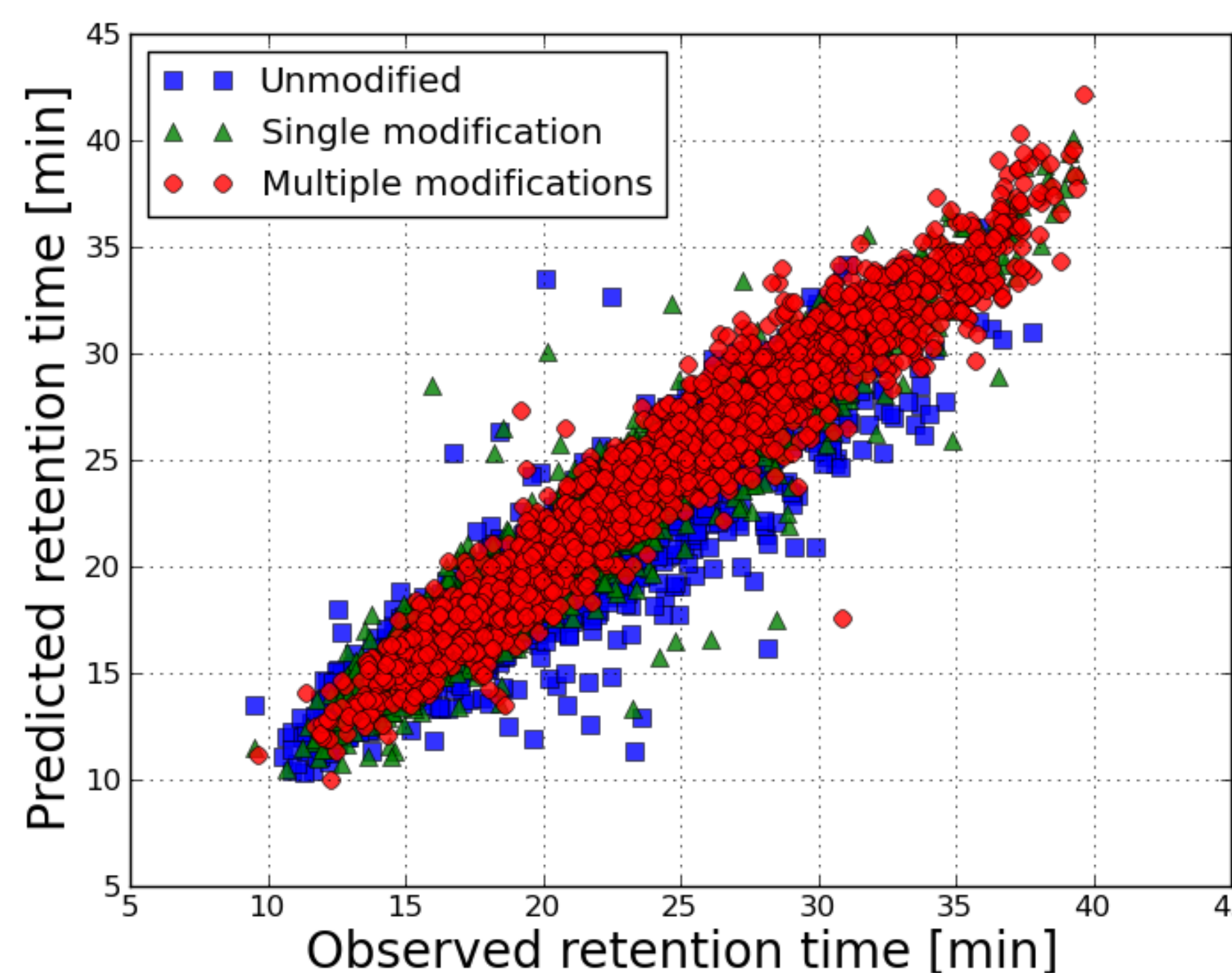
**Figure 3.** Elude was used to predict RT for 72 sets of phosphorylated peptides. The correlation between predicted and observed RTs ranged between 0.89-0.97 (avg 0.95), while the time window around the predicted time that included the observed RT for 95% of the peptides ( $\Delta t_{95\%}$ ) ranged between 12-27% of the total running time (avg 18%). As an example, the observed versus the predicted RTs for the largest dataset (833 peptides) is illustrated.

### • Prediction accuracy for singly modified peptides

**Figure 4.** We further investigated the performance of Elude on a dataset consisting of peptides modified by one of the following modifications: sulfon, acetylation, butylation, propionylation and methionine oxidation. In each case 2/3 of the data was used for training and 1/3 for testing. Here the performance of the algorithm is displayed as a function of the number of training peptides including each modification. Note that one retention model is trained for all the modifications.



### • Prediction accuracy for peptides with multiple modifications



**Figure 5.** The accuracy of Elude was also assessed for datasets consisting of a mixture of unmodified, singly modified and peptides carrying multiple modifications. The performance of the predictor was evaluated on a disjoint set of such 5,633 peptides. Elude's performances were similar to the values illustrated in Fig 4.

## Conclusions

- Elude yields comparable performances for modified and unmodified peptides.
- Elude provides accurate predictions even when little training data is employed.
- Elude can handle peptides with more than one modification.

## Download

Elude can be downloaded at [www.per-colator.com](http://www.per-colator.com) or applied via a web interface at [www.elude.sbc.su.se](http://www.elude.sbc.su.se). For more information please contact Luminita Moruz ([luminita.moruz@scilifelab.se](mailto:luminita.moruz@scilifelab.se)).

## References

- [1] L. Moruz *et al.*, Training, selection, and robust calibration of retention time models for targeted proteomics. *J Proteome Res* 9(10):5209-16, 2010.
- [2] A.A. Klammer *et al.*, Improving tandem mass spectrum identification using peptide retention time prediction across diverse chromatography conditions. *Anal Chem* 79(16):6111-8, 2007.
- [3] A. Bertsch *et al.*, Optimal de novo design of MRM experiments for rapid assay development in targeted proteomics. *J Proteome Res* 9(5):2696-704, 2010.

## Acknowledgements

This project was supported by grants from the Swedish Research Council and the European Molecular Biology Organization (EMBO).