# Spectrum Library Retention Time Prediction Based on Endogenous Peptide Standards

*Barbara Frewen,[1] Bin Fang,[2] Scott Peterman,[1] John Koomen,[2] Jiannong Li,[2] Eric Haura,[2]*
*Bryan Krastins,[1] David Sarracino,[1] Mary Lopez,[1] Amol Prakash[1]*
*[1]Thermo Fisher Scientific, BRIMS Center, Cambridge, MA; [2]Moffitt Cancer Center, Tampa, FL*

Thermo SCIENTIFIC

## Overview

**Purpose:** Accurately estimate peptide retention based on spectrum library data utilizing commonly observed peptides in place of synthetic standards.

**Methods:** We consolidate many months' worth of LC-MS/MS data into a library of MS/MS spectra. Our automated analysis selects endogenous peptides to act as standards which are used to predict retention times of any peptide in the library.

**Results:** Seventeen peptides were identified as appropriate endogenous standards. Relative retention time information stored in the library allowed us to predict the retention times of 1750 peptides more accurately than predictions based on hydrophobicity.

## Introduction

Spectrum libraries are an invaluable starting point for developing targeted assays (e.g. SRM, PRM) because they provide information about fragmentation patterns and retention times. When library data are collected under a variety of LC conditions, the use of synthetic peptide standards can greatly improve the ability to accurately predict retention time in new experiments. Unfortunately, any samples not including those peptide standards cannot be used in the predictions. We present a method for selecting peptides endogenous to a sample to act as standards and demonstrate their use for predicting retention times of other peptides including those with chemical modifications, which indicate portability to both unmodified and post-translationally modified peptides.

## Methods

### Sample Preparation

Activity-based protein profiling (ABPP) was performed on various human lung cancer cells and five pairs of tumor and adjacent control human tissue samples. Thermo Scientific™ Pierce™ ActivX ™ desthiobiotin ATP probes were used to interact with ATP utilizing enzymes and lysine close to the active sites were labeled with desthiobiotin.

### Liquid Chromatography and Mass Spectrometry

Trypsin-digested samples were run on one of three gradients (2 hr on HPLC, 2 hr on UPLC, 4 hr on UPLC). The validation experiment used a 4 hr gradient on UPLC. Spectra were acquired on a Thermo Scientific™ LTQ Orbitrap™ MS using data-dependent acquisition.

### Data Analysis

Peptide identification was done in Thermo Scientific™ Proteome Discoverer™ (PD) software. The spectrum library was built using the Crystal node for PD version 1.4. A custom script was written to analyze the library entries and find appropriate endogenous peptides to use as standards.

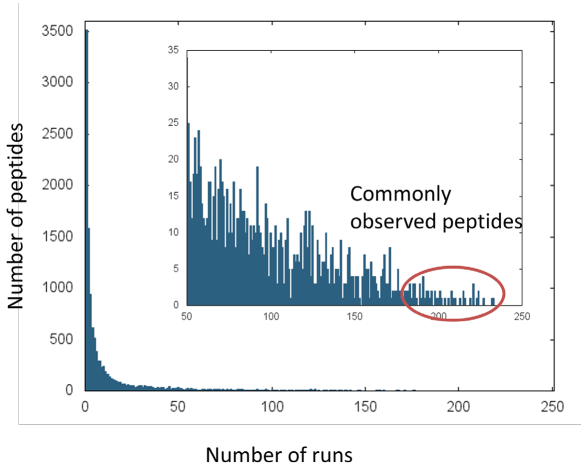
# Results

### Peptide Frequency in the Spectrum Library

Assembly of the Crystal spectrum library collected the retention time information into one resource. The library contained 220,542 spectra from 250 LC-MS runs including 9,109 peptide sequences (12,063 total with modified forms). As these samples did not contain a synthetic peptide standard, we first sought appropriate endogenous peptides.

The best candidates for peptides to act as retention time landmarks are those most commonly seen from run to run. We looked at the frequency of peptides in the 250 runs used to build the library. No peptides were observed in every run, the most commonly seen peptide having 233 appearances. (Figure 1) We selected the 50 most commonly seen peptides which were seen in no fewer than 185 runs.

**FIGURE 1. Frequency of peptide observation. The library collected spectra from 250 DDA runs. Peptides were observed with varying frequency, between 1 and 233 runs. We focused on the 50 most frequently seen peptides (circled in inset).**



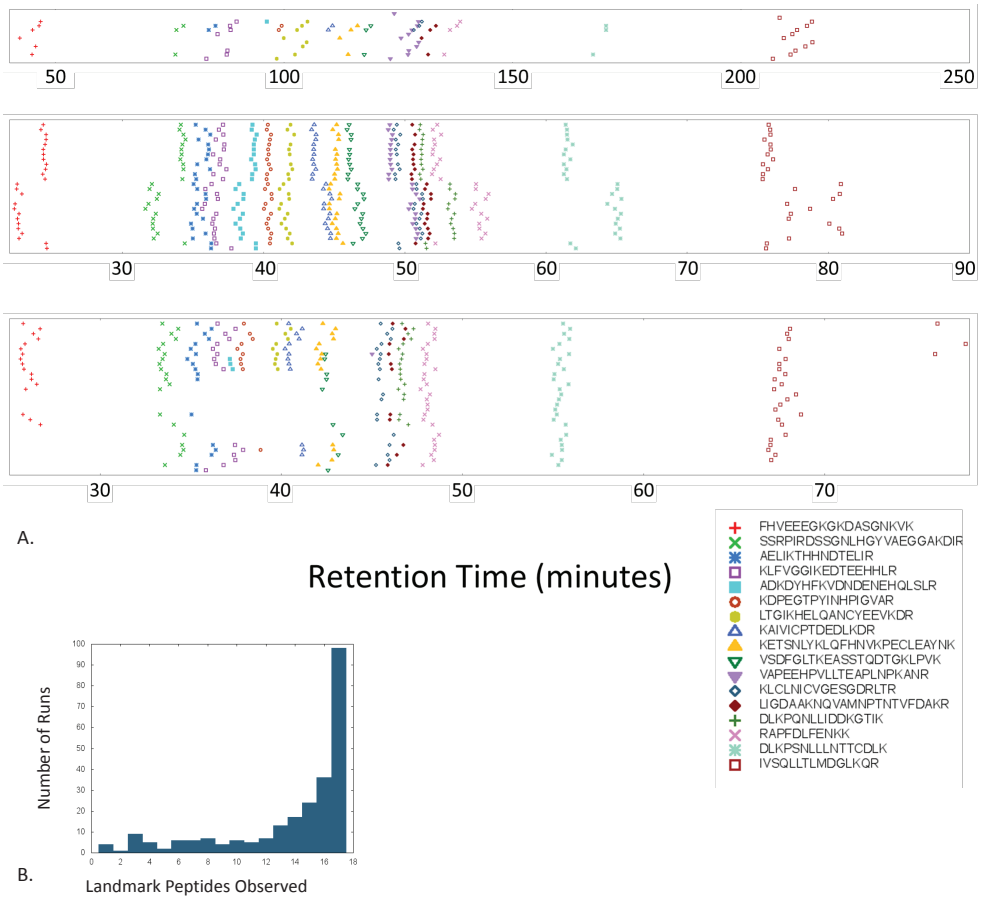

### Endogenous Peptides for Retention Time Landmarks

Starting with the 50 most commonly seen peptides, we winnowed down the list to find a set of peptides that both covered the entire elution profile and consistently eluted in the same order relative to each other. An in-house script automated the process. First we record the relative order of the 50 peptides in all 250 runs, for each pair of peptides A and B, keeping track of how often A came before B. Next we use a greedy algorithm to select a consistent set.

- Start the set with one peptide.
- For each remaining peptide, try adding it to the set in the appropriate order.
- If it cannot be placed unambiguously relative to the existing peptides in the set, eliminate this peptide.

We found seventeen that eluted in a consistent order. They are plotted at their observed retention times in several library runs in Figure 2.

**FIGURE 2. A. Retention times of landmark peptides in library data. The observed retention times of the seventeen peptides selected to act as landmarks were plotted for 68 runs in the library. Runs from each of three gradients are plotted together. The rank order of the peptides is the same in all runs, but the absolute times differ even for runs with the same gradient. Peptides are distributed across the entire gradient, with a higher density in the early-to-middle times. B. Histogram of number of landmark peptides in each run. Not every peptide was observed in every run, but there are enough in most cases to cover the whole gradient.**

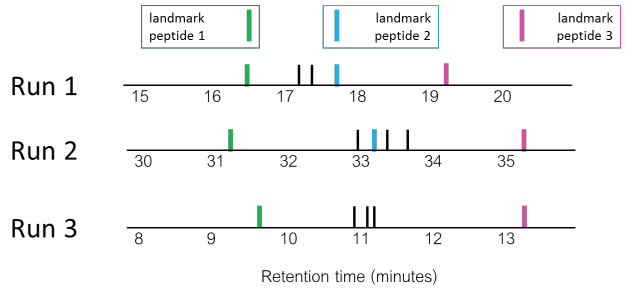



## Use Relative Retention Times to Estimate RT on New Gradient

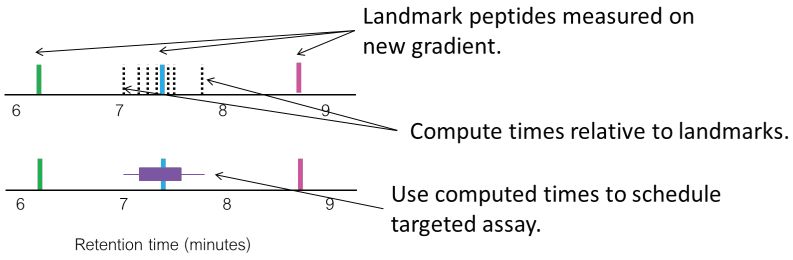
The Crystal library computes a relative retention time for each peptide stored in the library as a distance between the two nearest landmark peptides. (Figure 3a) These are used to estimate the retention times on a new gradient (Figure 3b). First, the RT of the landmarks must be measured on the new gradients. Then the relative RTs can be projected on to this new gradient and the average time is taken as the estimate.

We estimated the times of 1750 peptides on a 4 hour gradient. In addition, we compare our estimates to estimates based on peptide hydrophobicity (Krohkin, 2009). The accuracy of the estimate is measured as the difference between the estimated and observed times. Figure 4 plots the accuracy of the two estimation methods as well as accuracy of the library predictions as a function of the observed time. Library predictions are much closer than the hydrophobicity predictions to the observed retention times with most falling with in +/- 10 minutes of the observed time. Predictions are not consistently earlier or later than observed, but there is a slight trend for the prediction to be too early at the beginning of the run and too late at the end of the run. This may be due to having fewer landmarks at the ends of the run.

**FIGURE 3. A. Observed retention times of target peptides are stored as the distance between the two nearest landmark peptides. B. Retention time predictions are made by projecting the relative times on to the known times of the landmarks on a new gradient.**

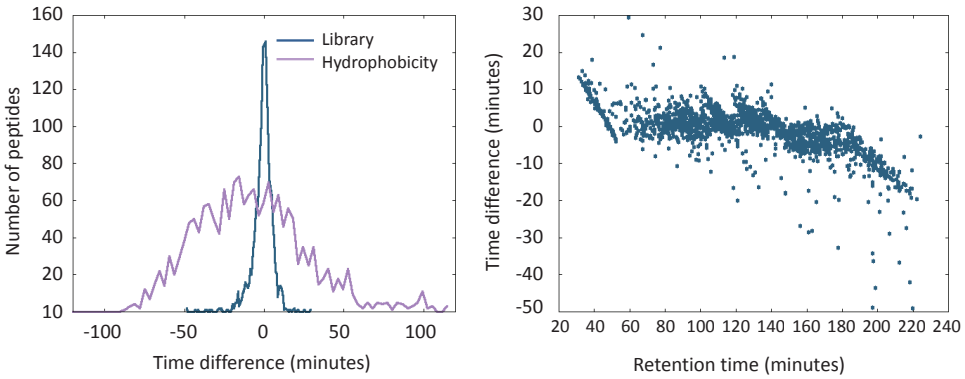



A.




B.

**FIGURE 4. Comparison of estimated and observed retention times of 1750 peptides. A. Histogram of predicted minus observed retention times for both prediction methods. B. Library-predicted minus observed retention time vs. the observed retention time.**




A B

## Conclusion

Endogenous peptides can successfully act as retention time landmarks and accurately estimate RT in new gradients.

- Spectrum libraries capture valuable retention time information.
- Our algorithm finds endogenous peptides with consistent elution behavior to act as standards.
- We can accurately predict the retention time of any library peptide by estimating it relative to the standard peptides. Therefore, comparisons can more easily be made across datasets with accurate mass and retention time measurements (AMT). This capability also enables method transfer to scheduled LC-MRM.
- Library-based estimated retention times are closer to the observed times than predictions made based on hydrophobicity.

## References


1. Krokhin OV, Spicer V. (2009) Peptide retention standards and hydrophobicity indexes in reversed-phase high-performance liquid chromatography of peptides. *Anal Chem* 81(22):9522-30.


## Acknowledgments


This research has been funded by the National Cancer Institute (P30 CA076292, P50-CA119997 to EH, R21 CA169979 to EH, R21 CA169980 to JK) and the American Lung Association (LCD-257857-N to JK) as well as the Moffitt Foundation.





PN-MSACL-2014-Spectral-Libraries-Frewen_E_03/14S