Supplementary Materials
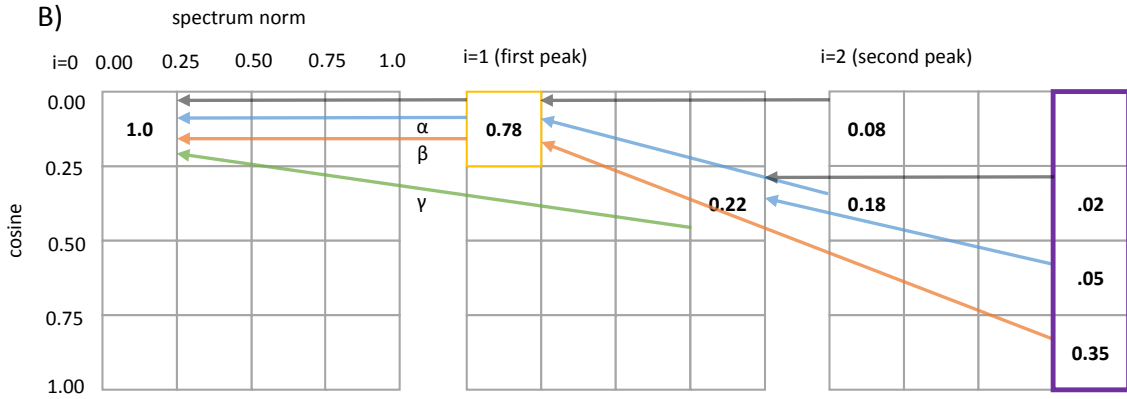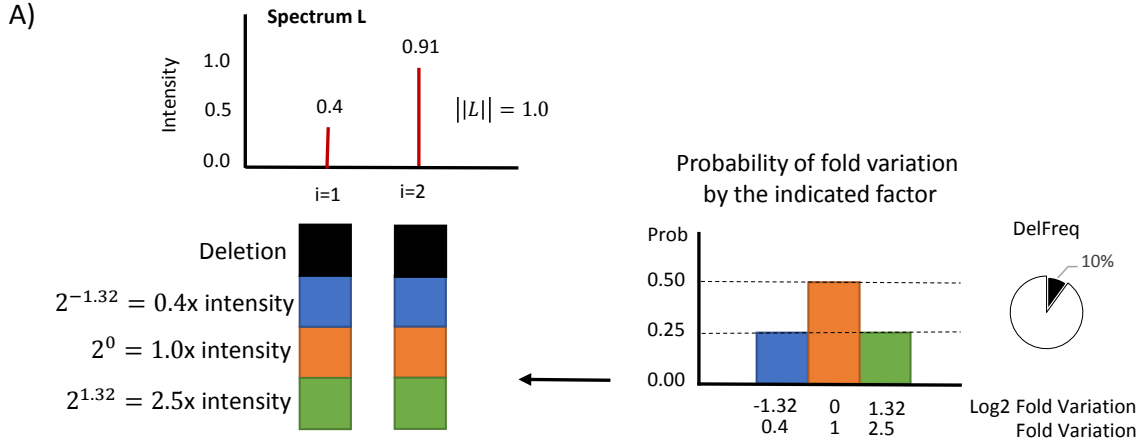
$$LibDP(i, c, p) = \sum_{y=0 \to p} LibDP(i - 1, c - \sqrt{y} \times L_i, p - y) \times Prob(\sqrt{y}|L_i)$$

In the equation for the LibDP recurrence, $c - \sqrt{y} * L_i$ is used because we are iterating over all possible increments in spectrum norm. For each increment y (third term in the LibDP recurrence, $p - y$), this translates into a $\sqrt{y}$ intensity for the replicate spectrum peak. Thus if we consider $\sqrt{y}$ intensity in a replicate spectrum peak, such a peak will contribute $\sqrt{y} \times L_i$ to the accumulated cosine c between the replicate and the library spectrum L, which explains the second term in the LibDP recurrence: $(c - \sqrt{y} \times L_i)$.

CPTAC Test Dataset Search with Sigma UPS1 search

In the initial analysis, the search library did not include Sigma48UPS1 spectra. To determine whether this affected the search results, the search was re-performed with the combined NIST Yeast and Sigma libraries. Supplementary Figure S2 includes ROC curves for this search of the Test dataset against this combined library. At 1% Peptide FDR, Tremolo had 3902 IDs, and SpectraST had 3315, representing a 17% gain. The results show the relative difference between Tremolo and SpectraST to be comparable to that of the search without the Sigma library.

A)

Spectrum L

Intensity

1.0

0.91

0.5   0.4

$||L|| = 1.0$

0.0

i=1   i=2

Deletion

$2^{-1.32} = 0.4x$ intensity

$2^{0} = 1.0x$ intensity

$2^{1.32} = 2.5x$ intensity

Probability of fold variation by the indicated factor

Prob

0.50

0.25

0.00

| | -1.32 | 0 | 1.32 |
| | 0.4 | 1 | 2.5 |

Log2 Fold Variation
Fold Variation

DelFreq

10%

B)

spectrum norm

i=0 0.00   0.25   0.50   0.75   1.0   i=1 (first peak)   i=2 (second peak)

cosine

0.00

1.0   α   0.78   0.08

β

0.25

γ   0.22   0.18   .02

0.50   .05

0.75   0.35

1.00

C)

(α) if replicate peak intensity $\Delta cos = (0.16) \times 0.4 = 0.06$
is $0.4 \times L_1 = 0.16$: $\Delta norm = (0.16)^2 = 0.03$
(β) if replicate peak intensity $\Delta cos = (0.4) \times 0.4 = 0.16$
is $1.0 \times L_1 = 0.40$: $\Delta norm = (0.4)^2 = 0.16$
(γ) if replicate peak intensity $\Delta cos = (1.0) \times 0.4 = 0.4$
is $2.5 \times L_1 = 1.00$: $\Delta norm = (1.0)^2 = 1.0$

D) $(1.0 \times 0.1) + (1.0 \times 0.25 \times 0.9) + (1.0 \times 0.5 \times 0.9) = 0.775$

Supplementary Figure S1: SLGF calculation example using the LibDP dynamic programming recursion. In (A) a very simple library spectrum with only two peaks is considered (with intensity 0.4 and 0.91). Each of these peaks has identical deletion frequency and intensity variation distributions that are discretized to 0.4x, 1.0x, and 2.5x intensity, with 0.25, 0.5, and 0.25 probability respectively. In (B) the dynamic programming recurrence is illustrated with the probability of occurrence shown in each cell (0.0 if empty). The arrows represent the propagation of probability and the color of each arrow represents which portion of the intensity variation distribution yielded the probability update. The last column of the table for the last peak (i=2 in this case) corresponds to the distribution of cosine scores at Euclidean norm 1.0 (shown in purple). This last column is re-normalized to total probability 1.0 and used as the SLGF theoretical score distribution for the library spectrum. The calculation of the $\Delta Cosine$ and $\Delta Norm$ is shown in (C), which explains the differing indices for the dynamic programming for the dynamic programming recurrence between i=1 and i=0. Additionally, the calculation of the probability of the top left cell (shown in yellow) for the $i = 1$ peak is shown in (D). Note that while it is shown that the LibDP matrix has $n_p = 4$ and $n_c = 4$, the actual implementation uses $n_p = 400$ and $n_c = 800$.

2
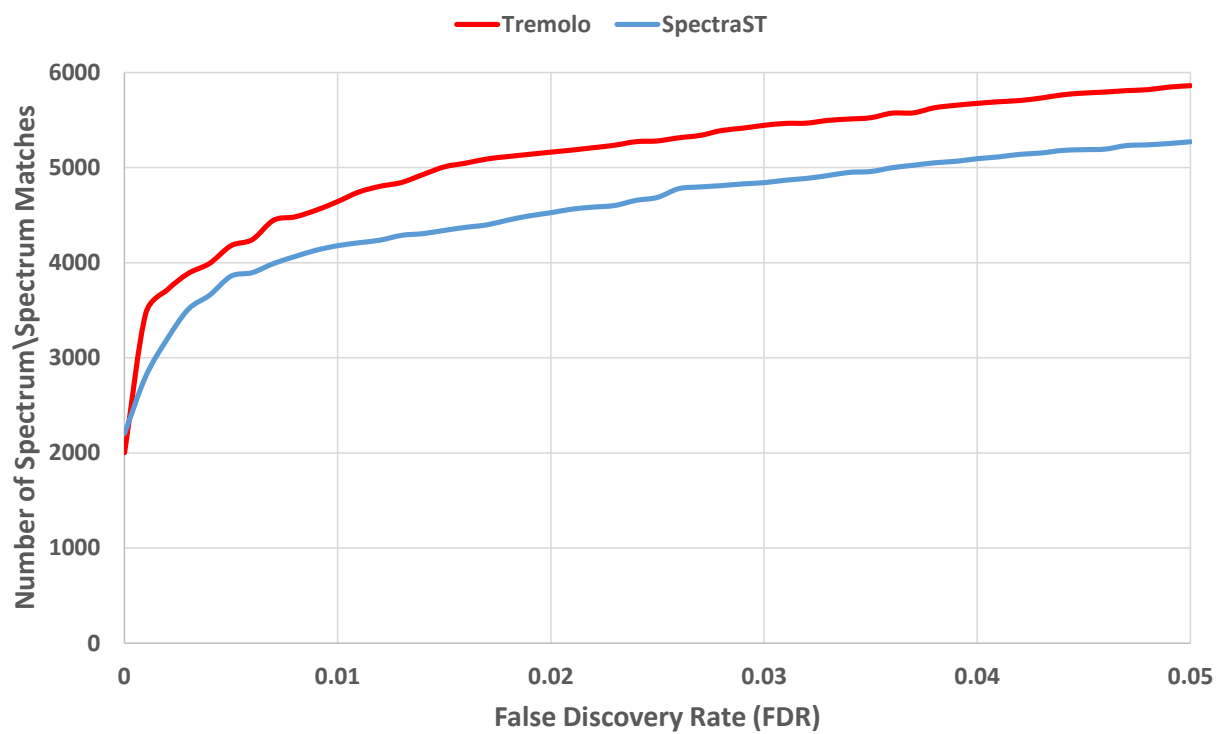
Supplementary Figure S2: Peptide spectral library search sensitivity by FDR comparison between Tremolo and SpectraST on the CPTAC Test dataset with the combined NIST libraries for Sigma and Yeast. At 1% spectrum level FDR, the gain in IDs for Tremolo is approximately 11%.