APPENDIX

A. FURTHER EXPLANATION OF THE CALIBRATION PROCEDURE

To understand why the calibration procedure in Equation 6 is helpful, consider the case when $f_1(E)$ and $f_2(E)$ are both high. The reason can be that this particular spectrum is simply more dense than others. On the other hand, if only one of the two values ($f_1(E)$ or $f_2(E)$) is higher than usual, this cannot be explained by the density of spectrum, we claim that this is real signal. Thus, when we expand $f(A)$, there are four terms

$$\lambda_1\lambda_2 f_1(A)+\bar{\lambda}_1\bar{\lambda}_2 f_2(A)+\lambda_1\bar{\lambda}_2\frac{f_1(E)}{f_2(E)}f_2(A)+\bar{\lambda}_1\lambda_2\frac{f_2(E)}{f_1(E)}f_1(A).$$

The first two terms are non-calibrated so we focus on the last two terms

$$\lambda_1\bar{\lambda}_2\frac{f_1(E)}{f_2(E)}f_2(A) + \bar{\lambda}_1\lambda_2\frac{f_2(E)}{f_1(E)}f_1(A)$$

These can be seen as a function of $\frac{f_2(E)}{f_1(E)}$. We know that a function like $ax+\frac{b}{x}$ will have a higher value if x is close to 0 or 1. So this is exactly what we need. We encourage that $f_1(E)$ and $f_2(E)$ are different, thus improving the calibration.

B. EMPIRICAL WEIGHTS

In this section, we show how we derive our empirical weights that lead to the improved performance demonstrated in Figure 5.

Figure 8 shows the average intensities of observed peaks near b-ions or y-ions in high confidence PSMs ($q = 0.01$) for the worm-01 charge +2 data set. Figure 9 shows the average intensities for *Plasmodium* TMT-10. In the figures, we see strong signals peaks at the b- and y-ions, as well as peaks with offsets of +1 Th peaks (+1 isotope peaks), -17 Th (NH_3-loss), -18 Th (H_2O-loss) and (for b-ions) -28 Th peaks (CO-loss).

The edges are added to E based on these average intensities for b-ion and y-ions separately. The weight of each edge is $w(\{v,u\}) = w'(\{v,u\})x(m_v)$, where m_v is the m/z of v, and x is the preprocessed observed spectrum as introduced in Section 2.1. Recall that in Section 3.1 $w'(\{v,u\})$ is determined by $m_v - m_u$. For low resolution data like yeast and worm, $w'(\{v,u\})$ is read from Table 3. For high resolution data, the $w'(\{v,u\})$ values are then read from Figure 9. In both settings, we calculate $m_v - m_u$ and use the number in the table or intensity in the plot.

For charge +3 data, where the b-ions and y-ions are charge +1 or charge +2, we do similar steps but using the mass-to-charge ratio of the higher charged ion, where $m_{v,c+} = \frac{m_{v,1+}+c-1}{c}$ is the m/z of charge $c+$ ion of v.

C. SELECTION OF SCORE FUNCTION PARAMETERS

Our submodular functions use the hyperparameters λ_1, λ_2 and $\{k_u\}_{u\in U}$. To select values for these hyperparameters, we performed an FDR-based evaluation with cases $k_u \in \{1,2,3,4,5\}$ when $\lambda_1 = \lambda_2 = 1.0$ on worm-01-ch2. The result (Figure C(a)) shows that $k_u = 2$ yields the best performance. Next, we tested cases $\lambda_1 \in \{0.4, 0.6, 0.8, 1.0\}$ on yeast-01-ch2, while fixing $\lambda_2 = 1.0$. We then selected $\lambda_1 = 0.6$ based on these results (Figure C(b)). Next, we tested $\lambda_2 \in \{0.4, 0.6, 0.8, 1.0\}$ while fixing $\lambda_1 = 0.6$. The value $\lambda_2 = 0.8$ had the best performance (Figure C(c)). For α, we calculate the average score of SGM,

$$\alpha_1 = \frac{\sum_{s\in S}\sum_{p\in P_s\cup D_s}\mathfrak{s}(p,s)}{\sum_{s\in S}|P_s\cup D_s|}$$

and the average foreground score of SEQUEST,

$$\alpha_2 = \frac{\sum_{s\in S}\sum_{p\in P_s\cup D_s}\langle p,s\rangle}{\sum_{s\in S}|P_s\cup D_s|},$$

where S is the set of all observed spectra and P_s and D_s are corresponding target and decoy sets of $s \in S$. The value α is then chosen to be $\frac{\alpha_2}{\alpha_1}$. Although the derivation of these parameters was empirical in our study, we hope in future work to develop strategies that can learn these automatically.

Table 3: The empirical weights $w(\{e_v, e_u\})$ used for low resolution data (yeast and worm). For each v and u, we compute the mass-to-charge ratio difference $m_v - m_u$ and read the correspond entry from the table.

b-ion						
$m_v - m_u$	-28	-27	-19	-18	-17	-16
$w'(\{v,u\})$	0.1101	0.0225	0.0121	0.3128	0.2364	0.0784
$m_v - m_u$	-15	-12	-1	0	+1	+2
$w'(\{v,u\})$	0.0112	0.0107	0.0481	0.6122	0.2514	0.0511
y-ion						
$m_v - m_u$	-18	-17	-16	0	+1	+2
$w'(\{v,u\})$	0.1364	0.1179	0.0345	1	0.4253	0.0741

D. HYPERPARAMETER GENERALIZATION

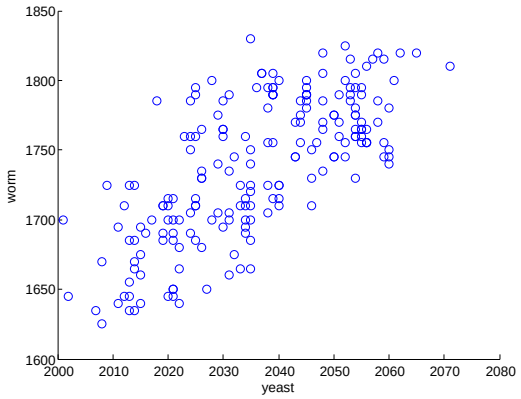
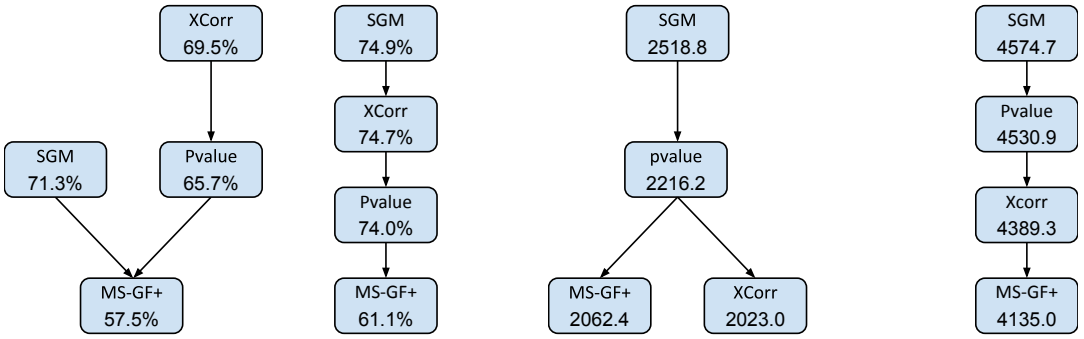


Figure 11: The plot shows the number of high-confidence PSMs ($q \leq 0.01$) obtained by SGM on the yeast data (x-axis) versus the worm data set (y-axis). Every point represents the performance of one combination of parameters.

Our submodular function $f(A)$ contains three hyperparameters. Ideally, these hyperparameters generalize across the different datasets. To test this, we tried all combinations of hyperparameters on one run from the yeast and worm data sets (yeast-01 and worm-01). The results (Figure 11) show that the parameters that work well on one dataset tend also to work well on the other dataset, implying that the parameters generalize well across datasets.



(A) *Plasmodium* TMP (B) human TMP (C) *Plasmodium* accepted PSMs (D) human accepted PSMs

Figure 6: Statistical comparison of methods. Each panel plots, for a single data set, the comparison between four methods in terms of the target match percentage or the number of targets PSMs accepted at $q < 0.01$. A directed edge from A to B means that method A's mean score is significantly larger ($p < 0.05$) than method B's mean score, according to a Wilcoxon signed-rank test. The numbers in the nodes are mean values.

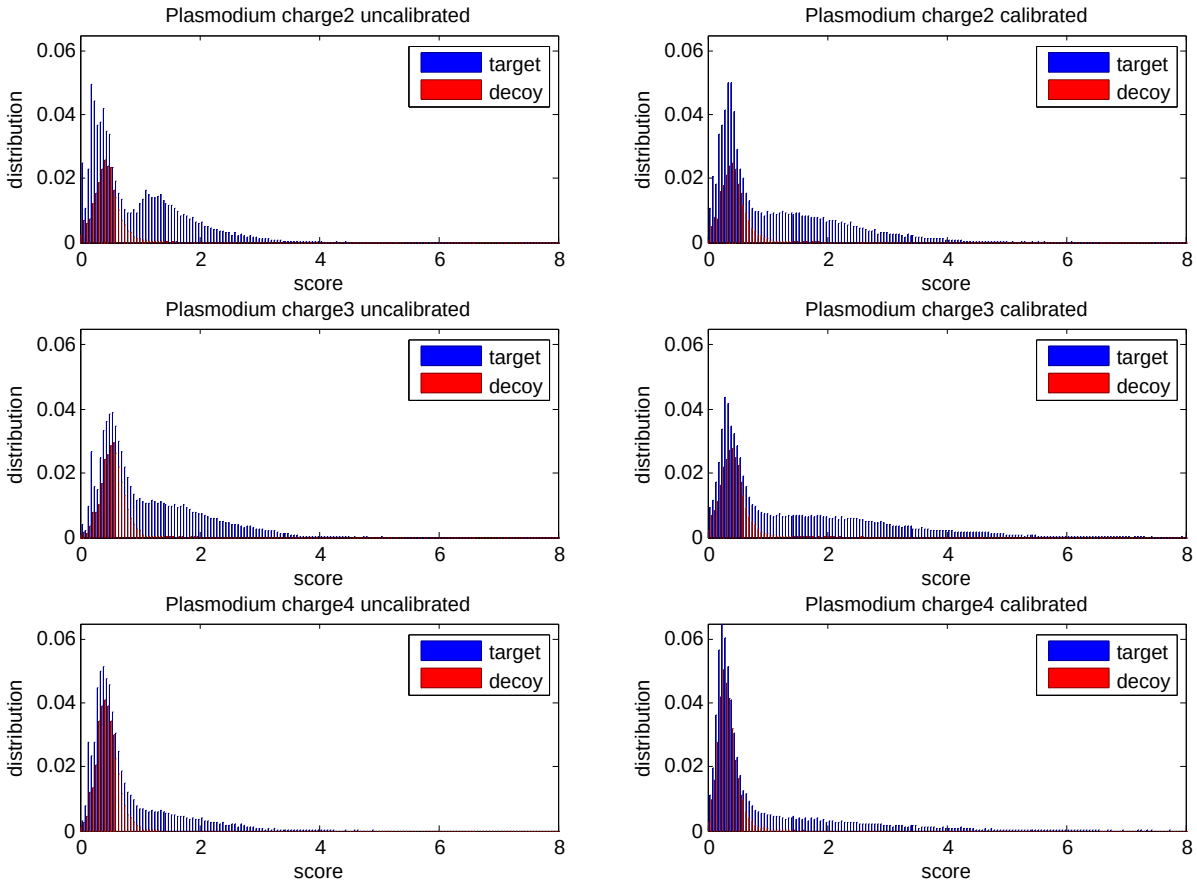


Figure 7: Score calibration of SGM. Each panel plots, for a single charge state, the SGM score distribution of top-scoring PSMs, separated into target and decoy distributions. Panels on the left are uncalibrated scores, and panels on the right are calibrated. In each plot, the x-axis is normalized so that the score threshold at $q = 0.01$ equals 1.

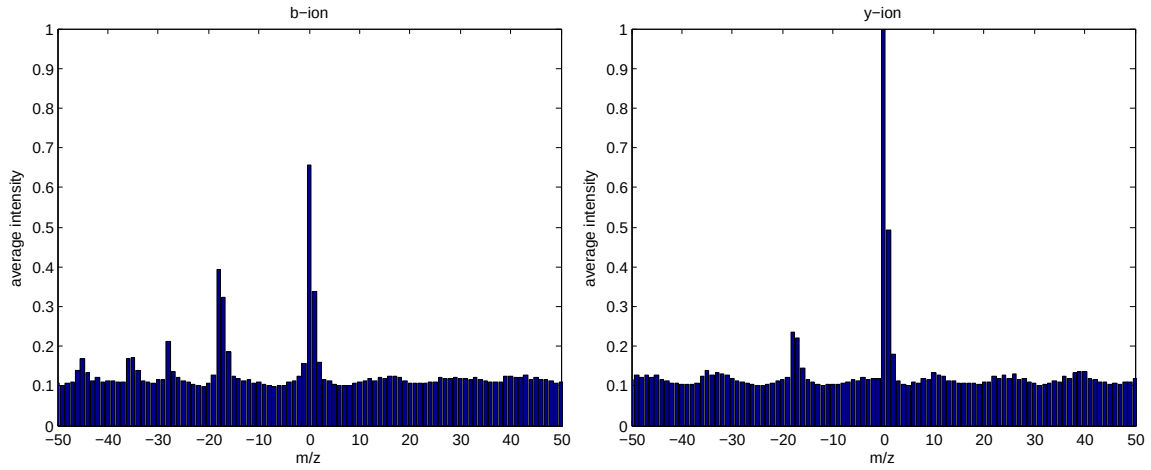


Figure 8: Average intensity near b-ion and y-ion peaks from high-confidence ($q < 0.01$) PSMs in the worm-01 dataset. We see strong signals at m/z=0 (central peak), m/z=+1 (+1 isotope), m/z=-17 (NH_3 loss), m/z=-18 (H_2O loss) and m/z=-28 (CO loss for b-ion only).

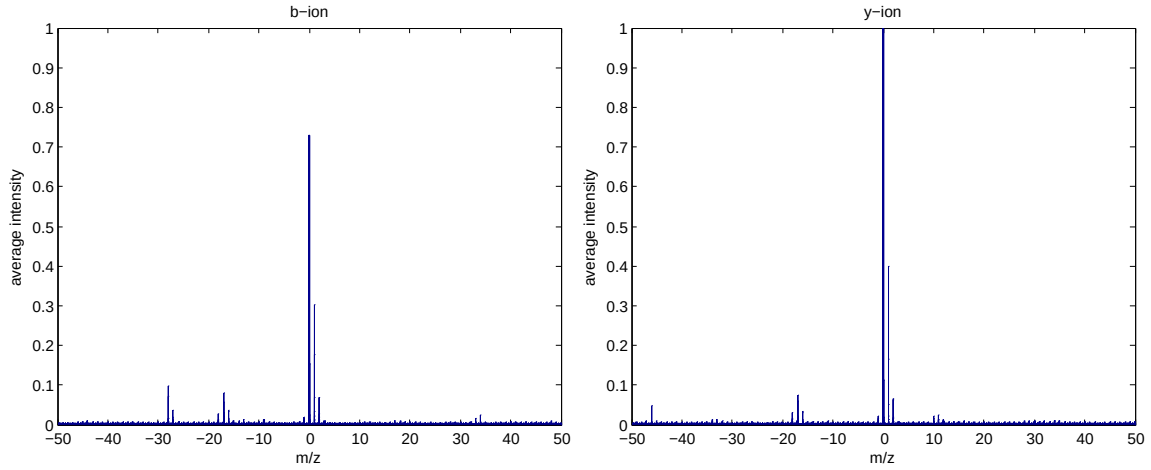
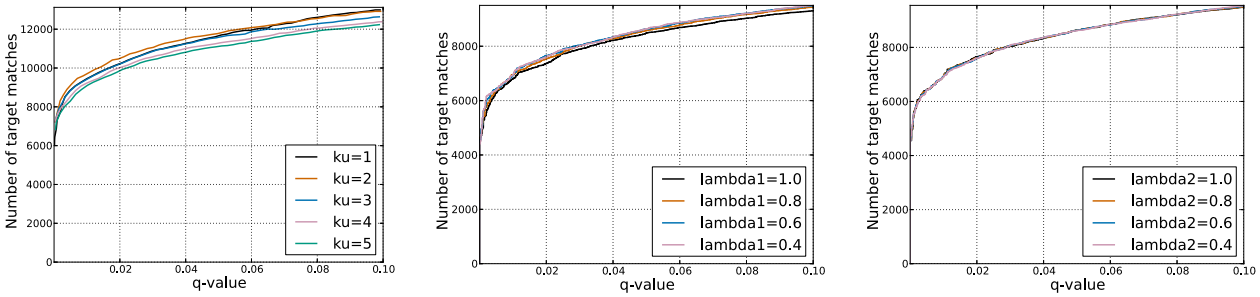


Figure 9: Average intensity near b-ion and y-ion from high-confidence ($q < 0.01$) PSMs in the *Plasmodium* TMT-10 dataset. We see strong signals at m/z=0 (central peak), m/z=+1 (+1 isotope), m/z=-17 (NH_3 loss), m/z=-18 (H_2O loss) and m/z=-28 (CO loss for b-ion only). The intensities in this figure are also used for the empirical weights $w(\{e_v, e_u\})$ for high resolution data for both *Plasmodium* and human. For each v and v, we compute the mass-to-charge ratio difference $m_v - m_u$ and then use the corresponding peak intensity.



k_u λ_1 λ_2

Figure 10: FDR-based evaluation of SGM using different hyperparameters.