

# A turn-key approach for large-scale identification of complex post-translational modifications

Jian Wang<sup>1</sup>, Veronica G. Anania<sup>2</sup>, Jeff Knott<sup>3</sup>, John Rush<sup>3</sup>, Jennie R Lill<sup>2</sup>,

Philip E. Bourne<sup>4</sup>, Nuno Bandeira<sup>4,5,6</sup>

<sup>1</sup>Bioinformatics Program, University of California, San Diego, La Jolla, USA

<sup>2</sup>Protein Chemistry Department, Genetech Inc., 1 DNA Way, South San Francisco, California

<sup>3</sup>Cell Signaling Technologies, Danvers, MA

<sup>4</sup>Skaggs School of Pharmacy and Pharmaceutical Sciences, UCSD, San Diego, La Jolla, USA

<sup>5</sup>Center for Computational Mass Spectrometry, University of California, San Diego, La Jolla USA

<sup>6</sup>Department of Computer Science and Engineering, University of California, San Diego, La Jolla, USA

Correspondence to:

Nuno Bandeira

Center for Computational Mass Spectrometry

Department of Computer Science and Engineering

University of California, San Diego

9500 Gilman Drive, Mail Code 0404

La Jolla, CA 92093-0404, USA

Email: [bandeira@ucsd.edu](mailto:bandeira@ucsd.edu)

Phone: 1-858-534-8666

Fax: 1-858-534-7029

## Supplementary Material

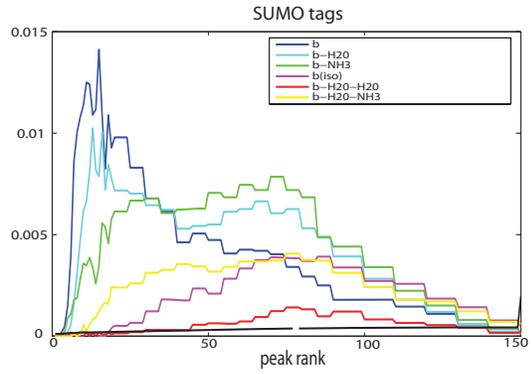
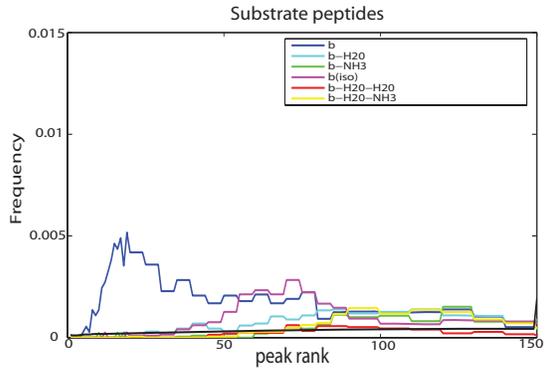
### Figure 1: Comparison of fragmentation patterns of unlinked and SUMOylated peptides

Fragmentation patterns are represented as the distributions of peak intensity ranks for all peaks matched to a particular type of fragment ions; peaks are ranked from most intense to least intense. The fragmentation pattern of substrate peptides and SUMO tag are observed to have different statistics. For example, as shown in a),  $H_2O$  and  $NH_3$  losses from b-ions are more frequently observed in SUMO tag peptides than in substrate peptides. To capture the fragmentation pattern of SUMOylated peptides, fragment ions are divided into two categories: linked-fragments and unlinked fragments (see Figure 2b in the main text). Linked fragments are from peptide fragment ions that are covalently linked to a second peptide. In general, unlinked fragments have fragmentation patterns similar to that of unlinked peptides. As an example, the fragmentation patterns of  $\gamma$ -ion from unlinked peptide and SUMOylated peptide are compared in b). On the other hand, linked-fragments have fragmentation patterns different from those of unlinked peptides. In particular, it was observed that multiply-charged fragments are more prominent as compared to unlinked peptides. The fragmentation pattern of triply charged  $\gamma$ -ion from unlinked peptides and SUMOylated peptides are illustrated in c).

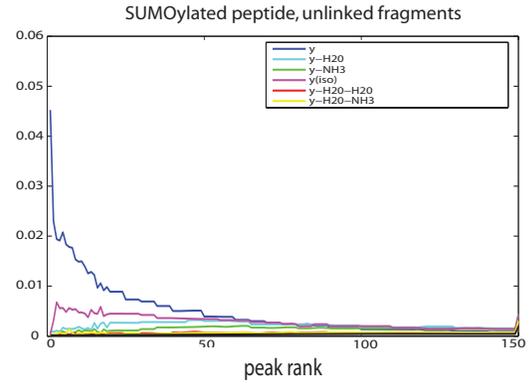
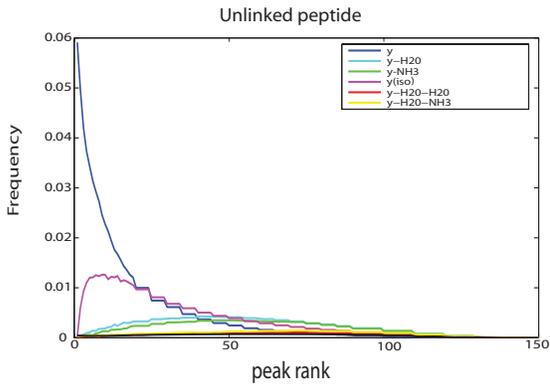
### Figure 2: Features of SUMOylated peptides not identified by Specialize

In the SUMO datasets, Specialize identified 41 out of the 68 SUMOylated peptides found by Mascot. SUMOylated peptides not identified by Specialize were found to have two common features: a) the substrate peptides are of very long length ( $\geq 27a.a.$ ) or b) the SUMO tag does not fragment very well and thus results in peaks with relatively low intensity in the MS/MS spectrum. As shown in a), SUMOylated peptides identified by Specialize (blue line) and peptides in the human NIST spectral library (cyan line) have similar

a) Fragmentation statistics of singly-charged b-ions from charged-3 precursors



b) Fragmentation statistics of singly-charged y-ions from charged-3 precursors



c) Fragmentation statistics for triply-charged y-ions from charged-3 precursors

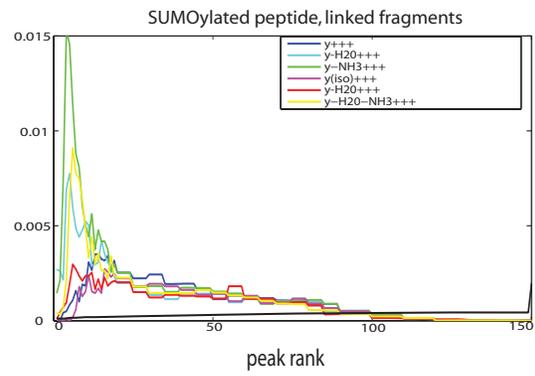
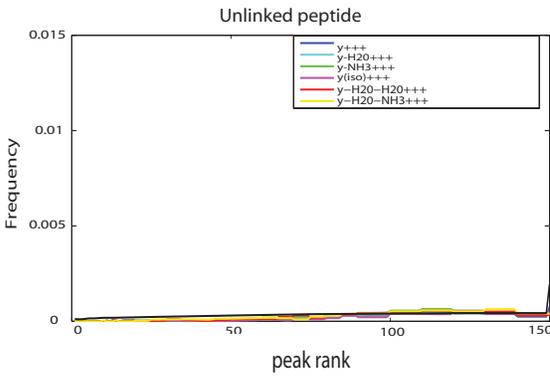


Figure 1:

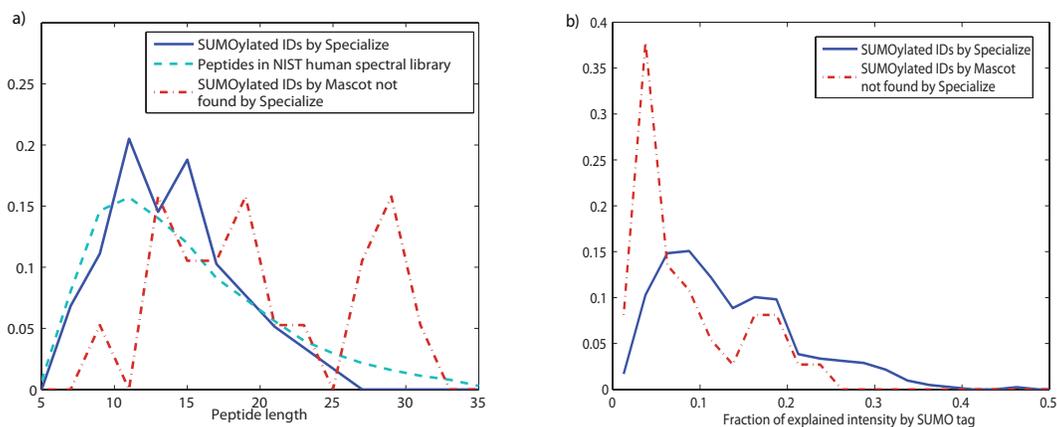


Figure 2:

length distributions indicating that Specialize can identify peptides with a wide range of lengths. However, a large fraction of SUMOylated peptides not identified by Specialize tend to be long ( $\geq 25a.a.$ ), indicating that the current model in Specialize may not generalize well to the class of very long peptides. b) Similarly, in the synthetic peptide libraries, the fragments from the SUMO tag usually contribute 10-20% of the total intensity in the resulting MS/MS spectra (see Figure 2 in the main text). Such trends were also observed for SUMOylated peptides identified by Specialize in the Human SUMO dataset (blue line). However for the SUMOylated peptides not identified by Specialize were observed to have much lower explained intensity from SUMO tags, on average contributing only 5% of the total intensity (red line).