**Title (20 words max):**

Rapid and Accurate Global PTM Discovery (G-PTM-D) Using Defined Mass Windows

**Introduction (120 words max):**

Correctly identifying protein posttranslational-modifications (PTMs) is crucial to understanding many aspects of protein function in biological processes. Methods to identify and localize PTMs from complex biological samples have been limited, but a few techniques are emerging. G-PTM-D [Journal of Proteome Research, DOI: 10.1021/acs.jproteome.6b00034] is a recently developed tool for global identification of PTMs. We enhanced this method by using non-standard mass-window searches, and demonstrate its effectiveness in identification of numerous types of PTMs, including high-mass modifications such as glycosylations. The number of identified modifications increased by 20% and the search time decreased by an order of magnitude. We also show the advantages of using mass windows in a standard database search, when PTM discovery is not the primary goal.

**Methods (120 words max):**

We developed a modified version of the Morpheus software for bottom-up spectral database searching that we call MetaMorpheus, which integrates a spectral calibration procedure with the G-PTM-D workflow. Two multi-fraction mammalian datasets with deep proteome coverage were used to evaluate the performance of the methods. Uniprot XML protein databases containing only reviewed human/mouse proteins were employed. A 10ppm precursor mass tolerance was used for the initial calibration step, and then reduced to 5ppm for subsequent searches in defined narrow mass windows (*notch* searches). For the Global PTM Discovery (G-PTM-D) process, we allowed 120 modifications including PTMs, adducts, and chemical modifications. For the discovery of other modifications, we used a novel wide mass-window search (*interval* search*)*.

**Preliminary Data (300 words max):**

By employing the novel notch search strategy instead of the wide-mass search in the G-PTM-D process, using only the modifications listed in Uniprot, the number of confidently identified PTMs increased by 8%. The overall search time dropped significantly, from 30 hours to 3.5 hours for all datasets. Using an expanded modification list that includes chemical modifications such as adducts and large mass PTMs such as glycosylations, the overall identification rate increased from 164,697 to 179,345 peptide-spectral matches (PSMs), which in turn increased the number of modified peptides by an additional 20%. We identified hundreds of glycosylated peptides in these unenriched samples, with many of these modifications exceeding 1000 Daltons. Algorithmic improvements in the G-PTM-D procedure itself, such as searching for combinations of modifications and for modifications with neutral losses make it an excellent tool for PTM discovery.

Previously described wide mass search strategies for discovery and localization of unknown modifications benefit from using alternative, carefully designed search modes based on interval and notch searches. We discovered that limiting mass shifts to a lower bound of -187 Da (corresponding to the largest mass difference that could be attributed to loss of a single residue, Tryptophan), and no upper bound, is an important step in eliminating spurious PSMs. Furthermore, treating highly suspect mass shifts corresponding to residue additions/substitutions/deletions in a separate notch search with individualized false discovery rate estimates is beneficial. An automated tool built into MetaMorpheus allows confidently identifying novel modifications based on these search results. These strategies for increasing confidence in identifications with unusual mass shifts can be used in any database search.

**Novel Aspect (20 words max):**

Defined mass windows enable rapid database searches and significantly improve identification and discovery rates of PTMs including high-mass glycosylations.