A Modification-Aware Framework for Fast Open Modification Spectral Library Search

Younghee Seo[1], Seungjin Na[2] and Eunok Paek[1,*]

[1]Department of artificial intelligence, Hanyang University
[2]Digital Omics Research Center, Korea Basic Science Institute
*Corresponding author: eunokpaek@gmail.com

The analysis of peptides using mass spectrometry is a fundamental yet complex task in proteomics. Recently, various algorithms have been developed to interpret mass spectrometry data more effectively. Among these, spectral library search through Open Modification Search (OMS) has gained prominence, particularly with the introduction of ANN-SoLo, which enhances the efficiency of OMS by clustering candidate peptides using an Approximate Nearest Neighbor (ANN)-based algorithm. However, ANN-SoLo does not incorporate modification information during clustering, potentially overlooking modified peptides. To address this limitation, we propose a novel approach that integrates complementary spectra to account for modification information during clustering. By ensuring that both modified and unmodified spectrum pairs are grouped within the same cluster, we improve the identification of modified peptides. Furthermore, our modification-aware method refines the selection of candidate peptides, allowing for more accurate identification based on potential modification sites. This approach not only accelerates the search process but also improves the accuracy of OMS in spectral library searches, offering significant benefits to proteomic research.