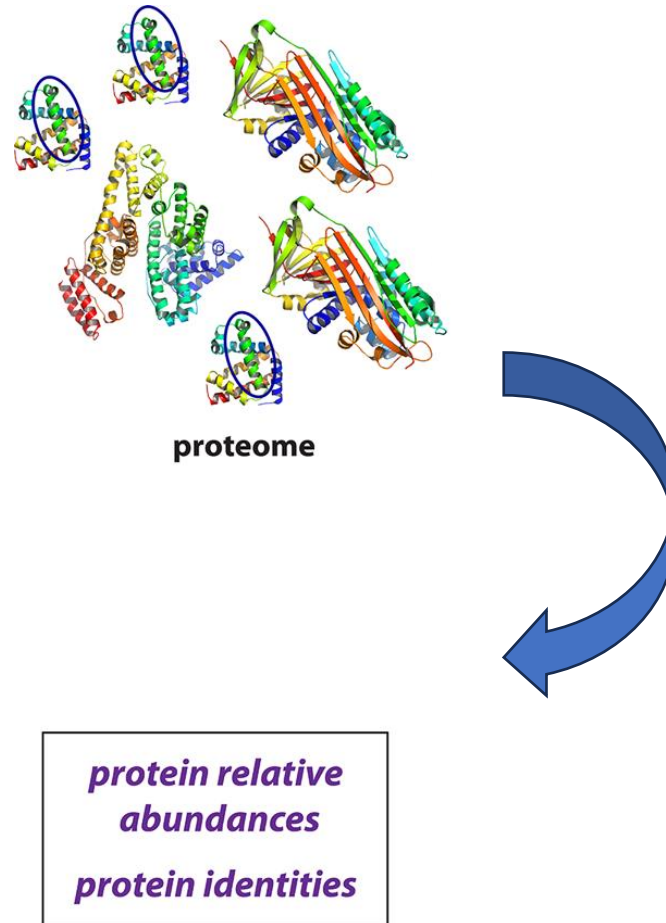


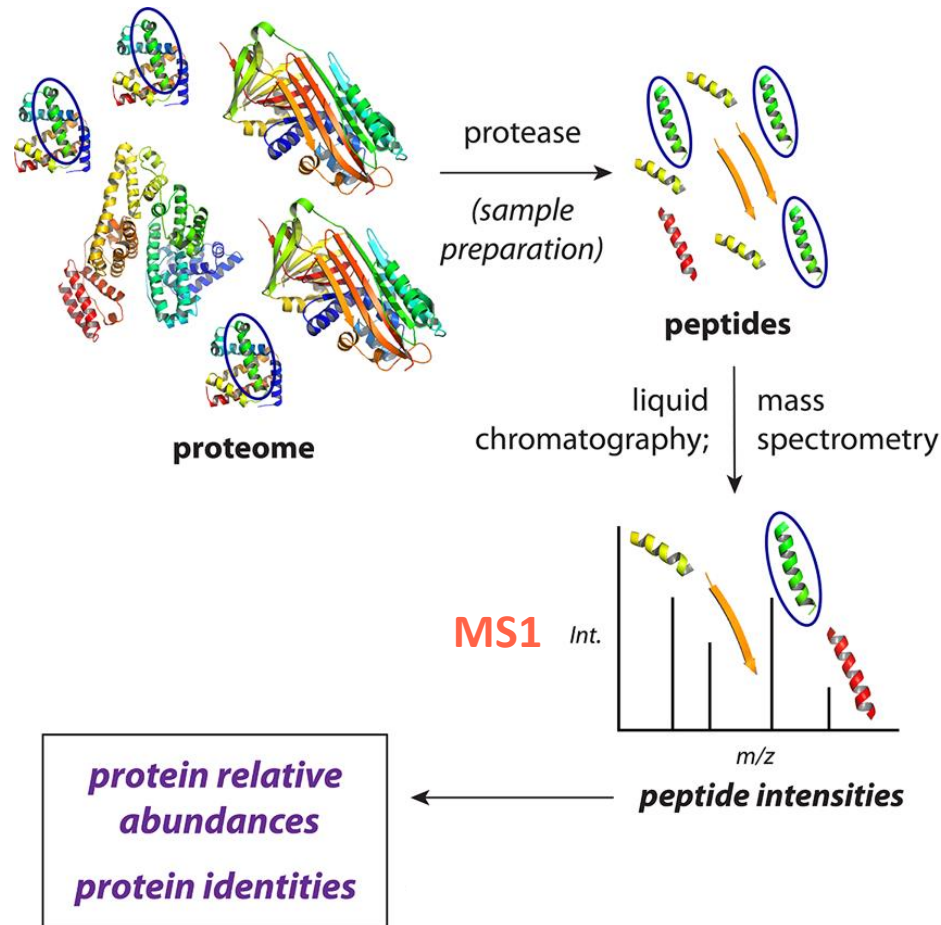
A brief introduction to proteomics mass spectrometry and *de novo* peptide sequencing with Casanovo

Yuhui Hong
2026-09-22

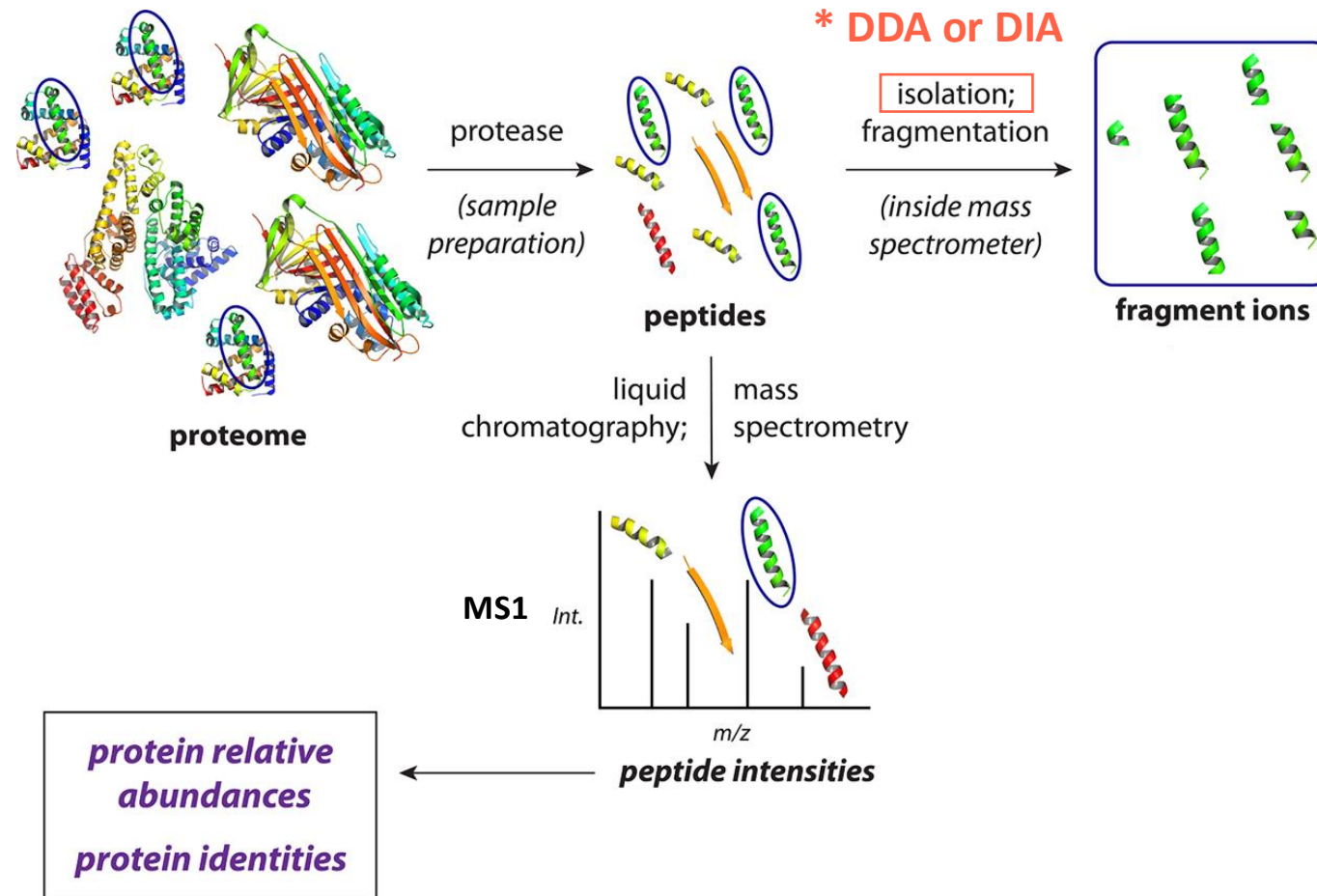
Mass spectrometry identifies and quantifies proteins



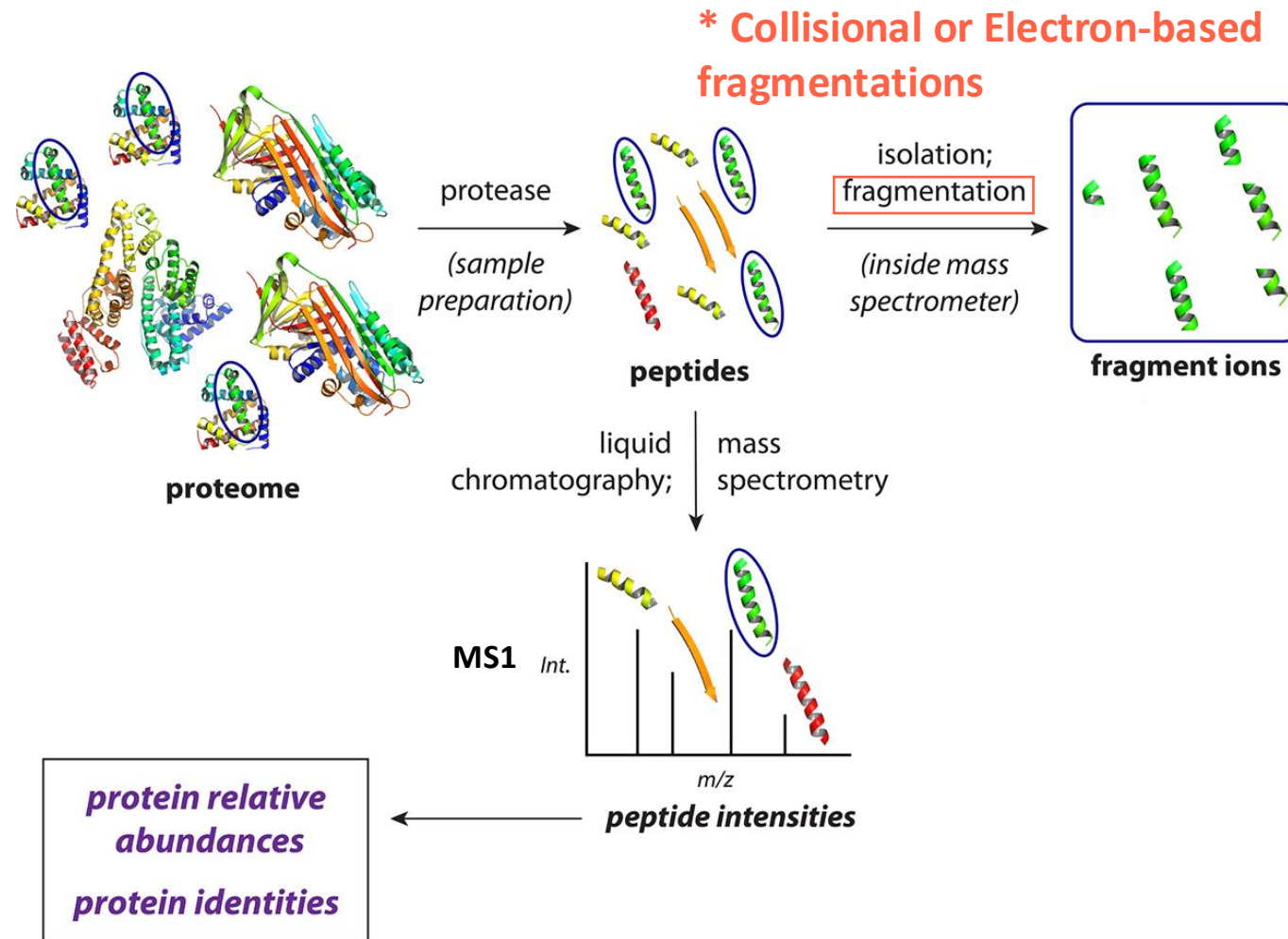
MS1 spectra record the m/z and intensity of intact precursor ions



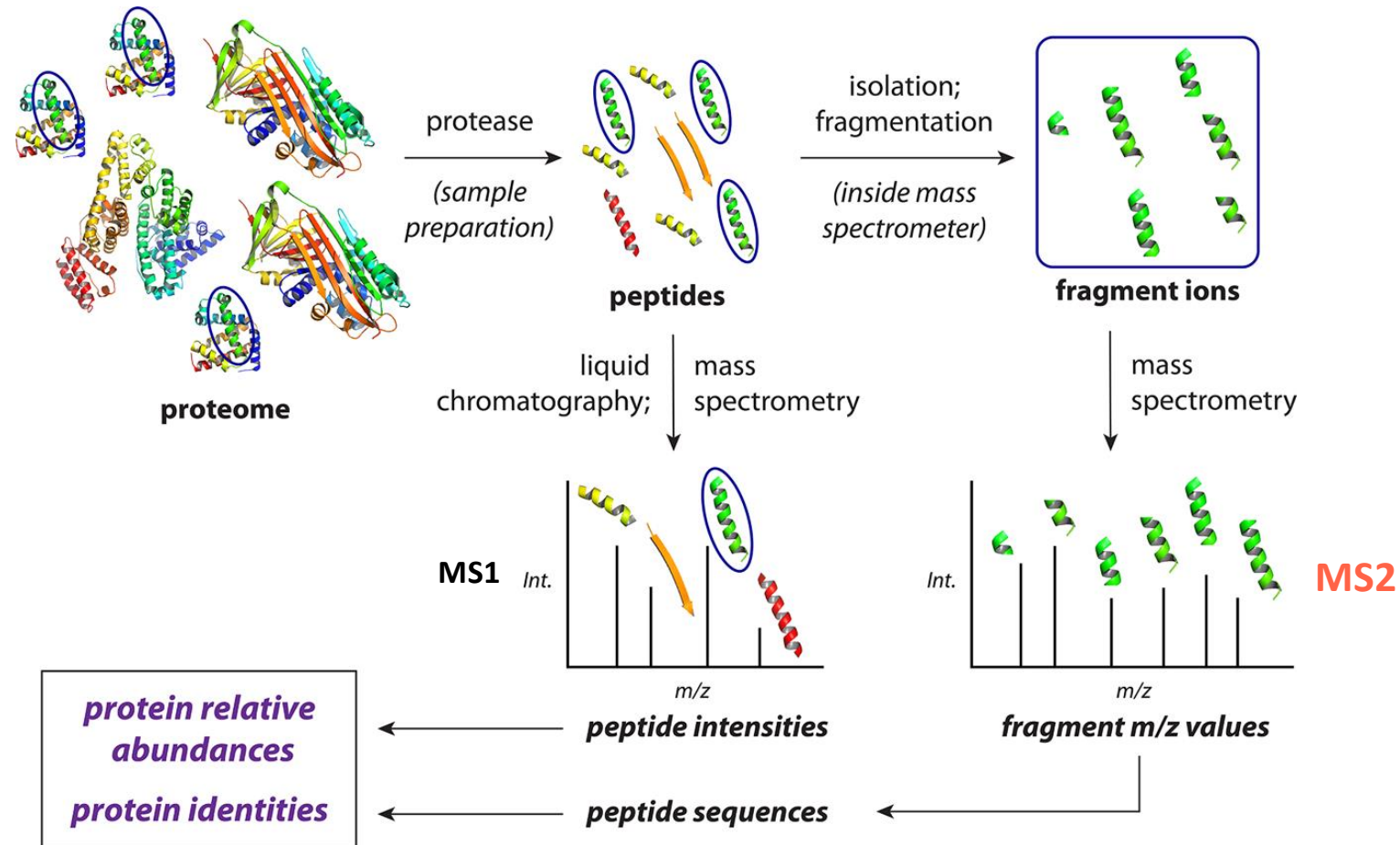
Untargeted proteomics selects precursors using either DDA or DIA



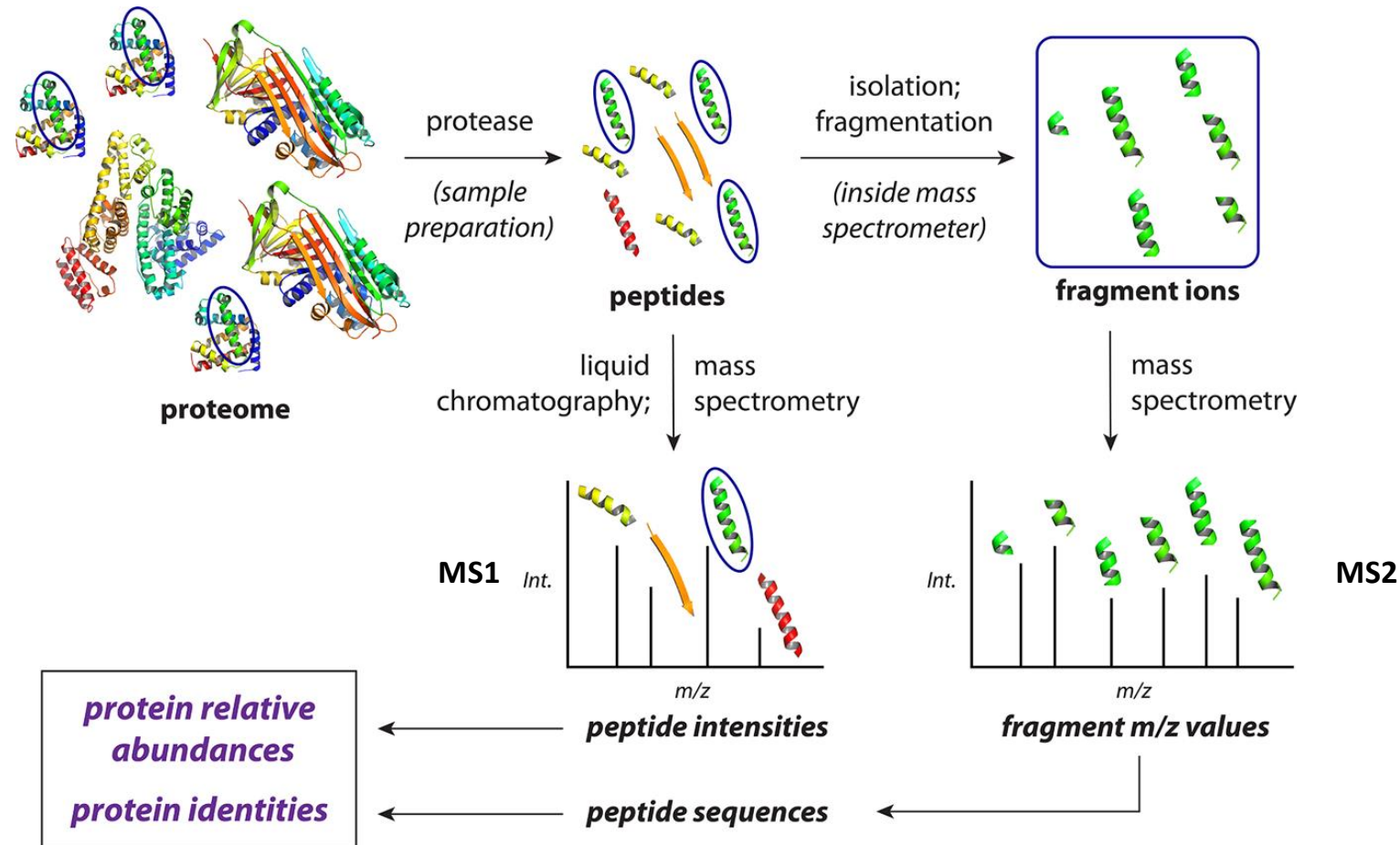
The selected precursors are then fragmented



MS2 (MS/MS) spectra record the intensities of fragment ions from a selected precursor

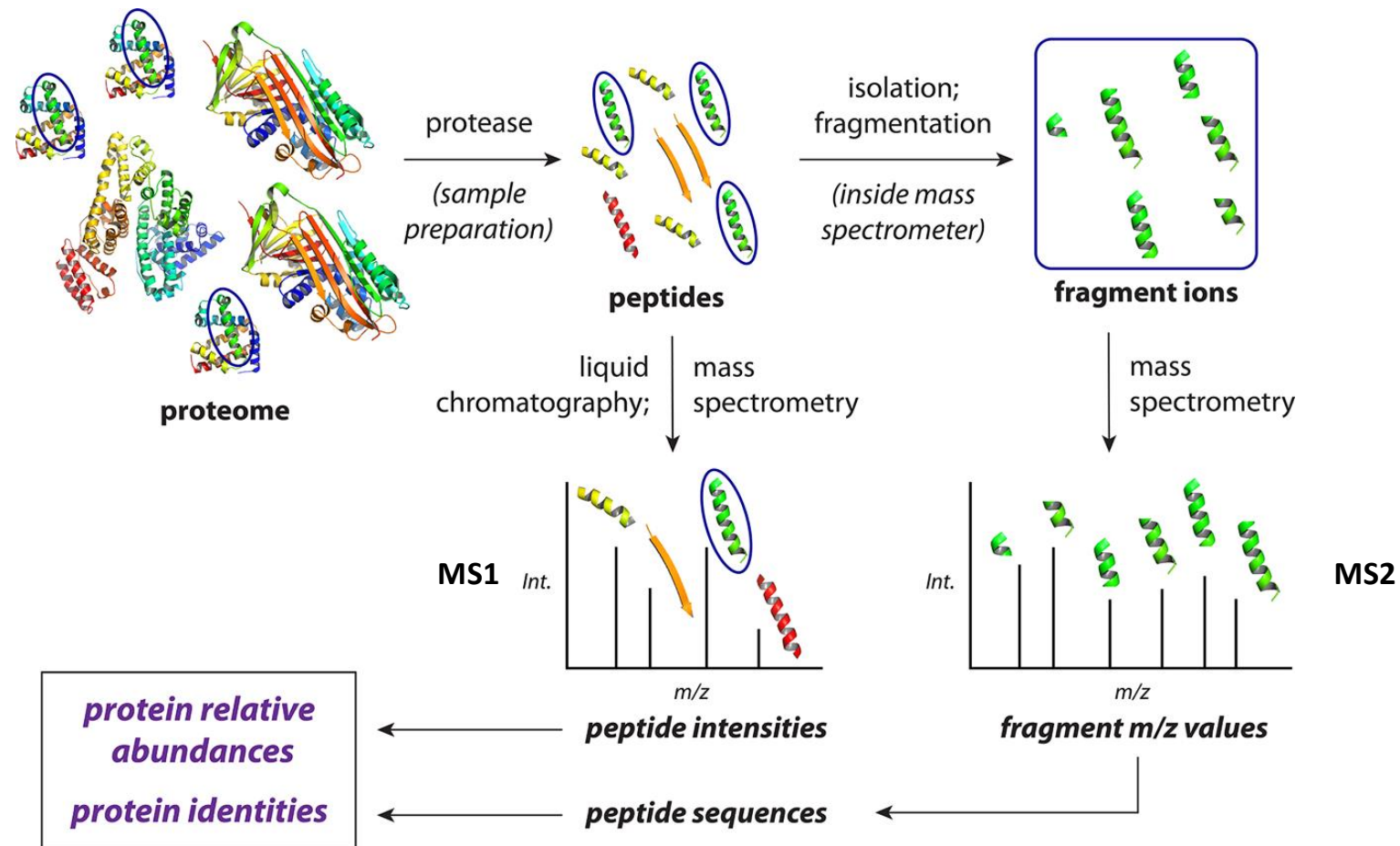


De novo sequencing infers peptide sequences directly from MS/MS spectra

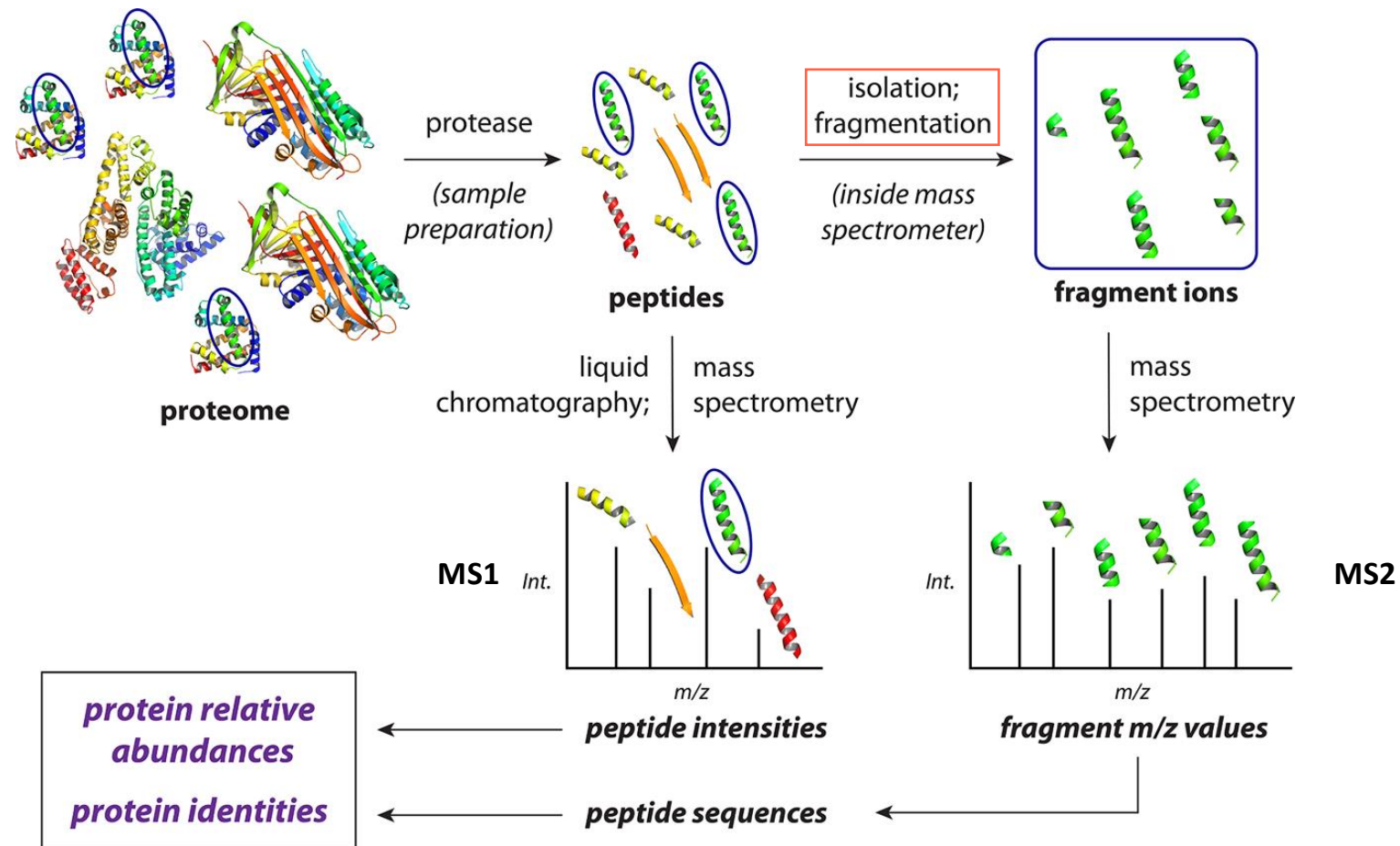


**database searching or
de novo sequencing**

Simplified overview of proteomics mass spectrometry experiments



Simplified overview of proteomics mass spectrometry experiments

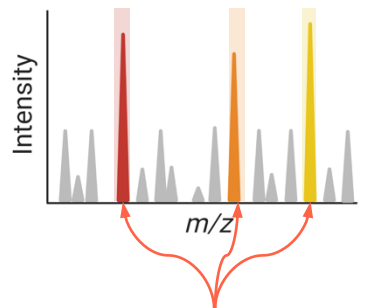


DDA isolates and fragments the most abundant precursors one at a time through a narrow isolation window

DDA (Data-Dependent Acquisition)

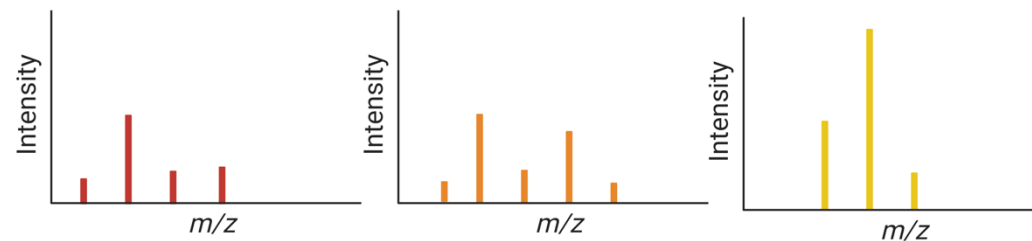
MS1

e.g., 3 most abundant precursors



Isolation windows

MS2

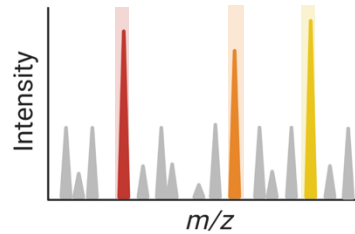


DIA fragments all precursors within wide isolation windows that tile the full MS1 range

DDA

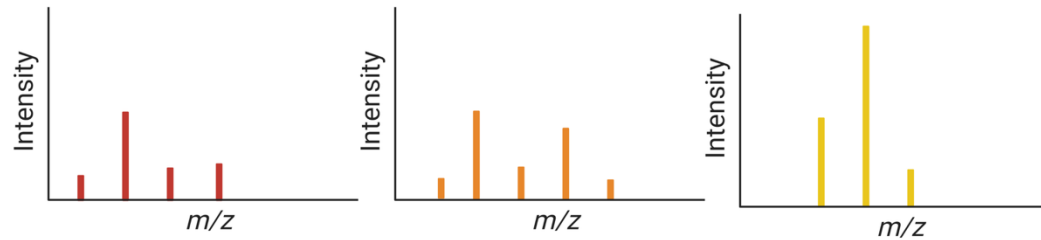
MS1

e.g., 3 most abundant precursors



MS2

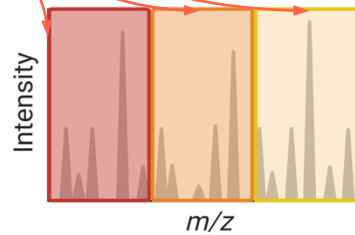
Less complex spectra searched against a protein database



DIA (Data-Independent Acquisition)

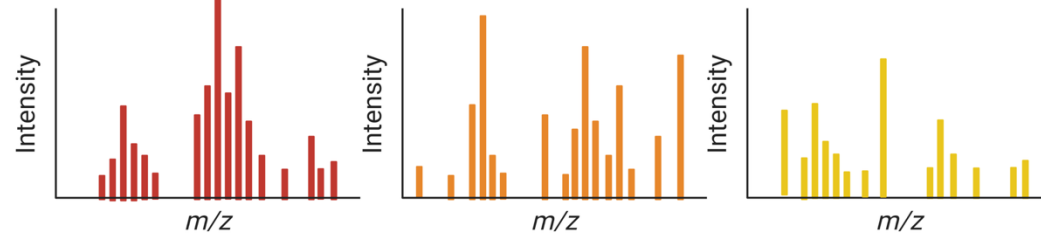
MS1

Isolation windows
e.g., 3 m/z windows

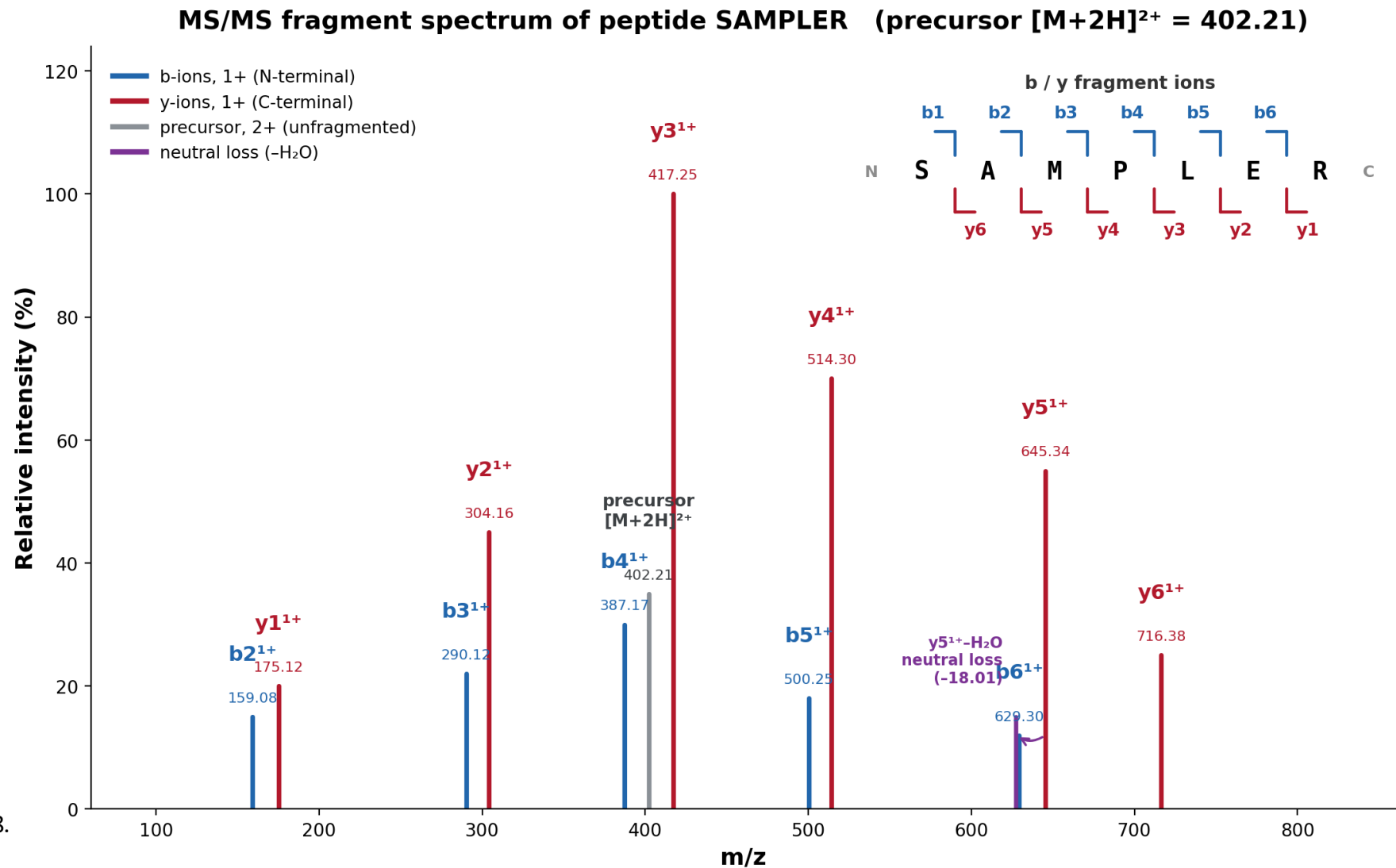


MS2

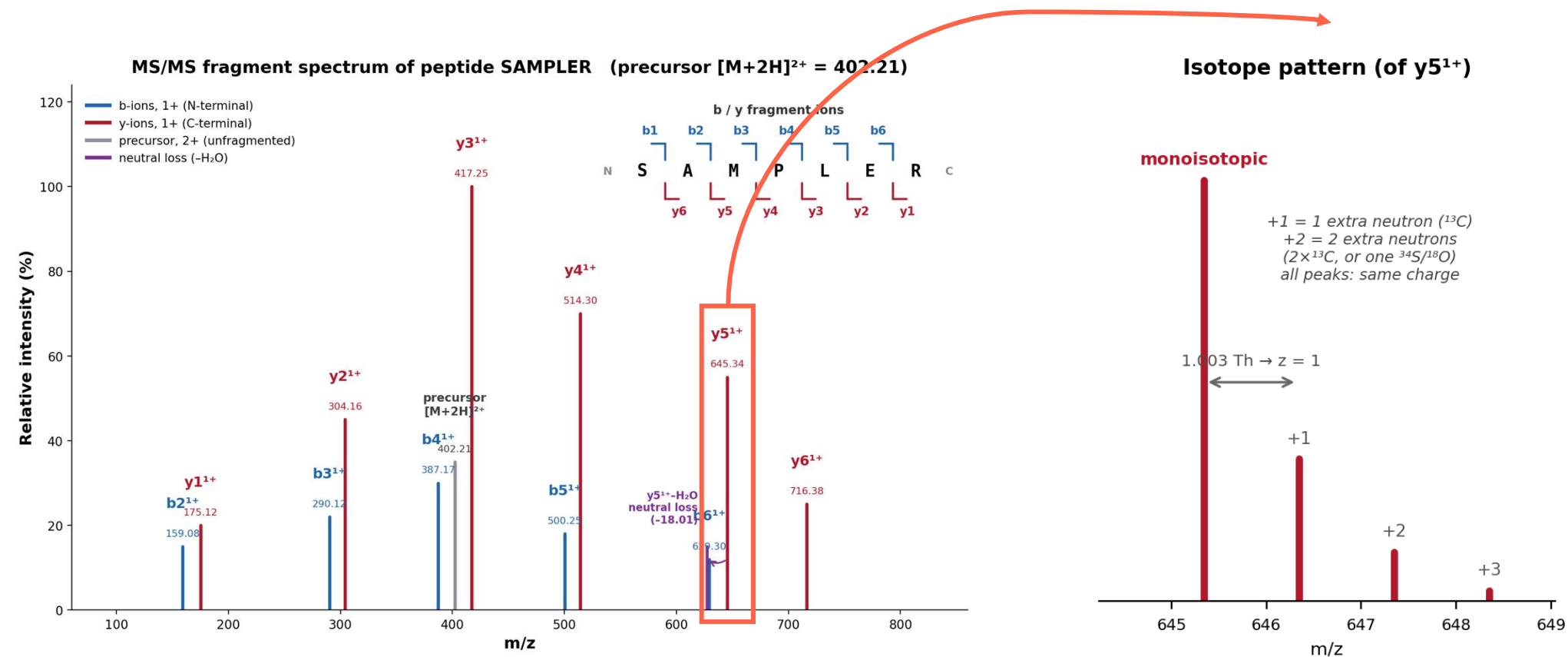
More complex spectra searched against a spectral library



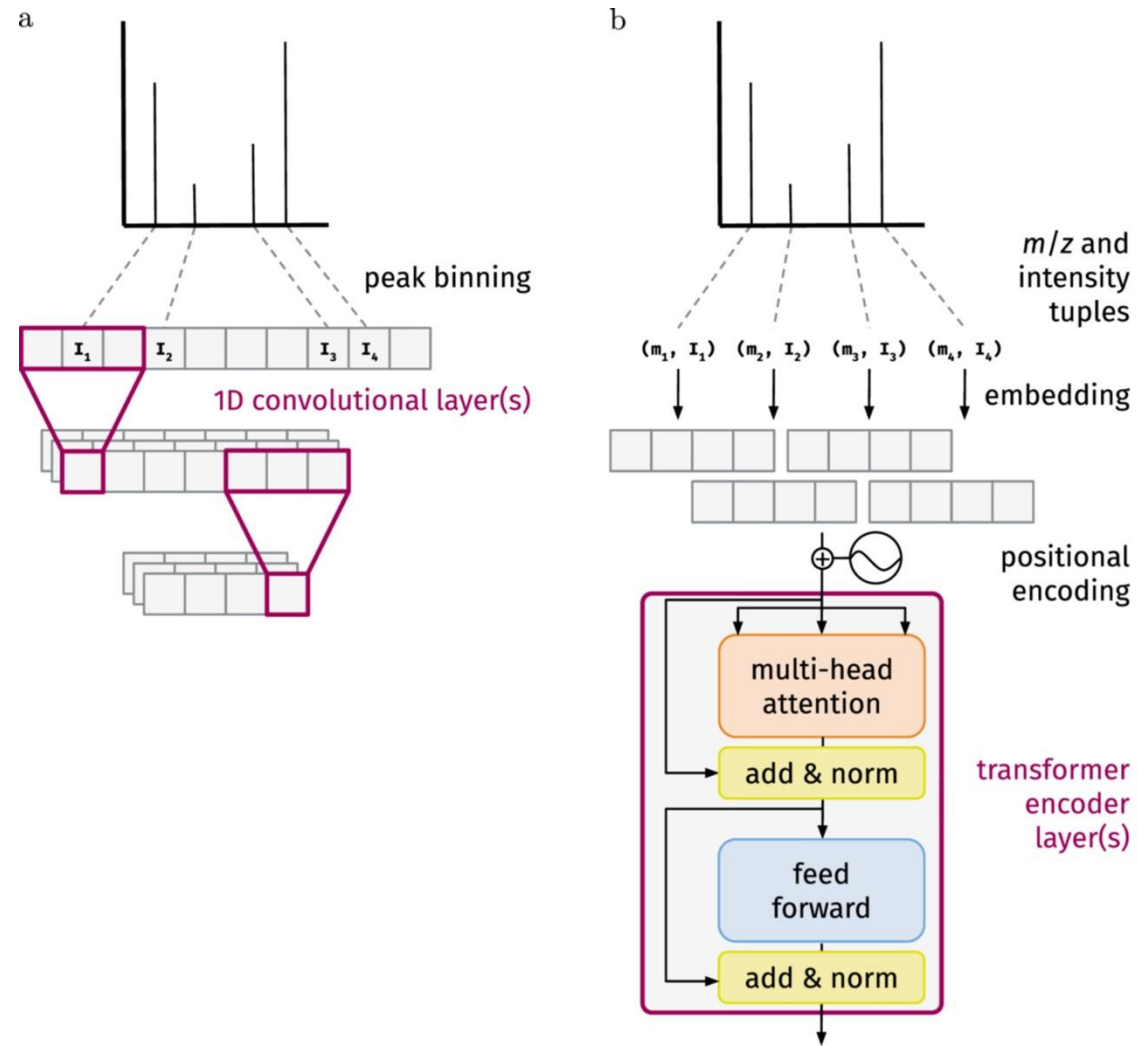
Collisional methods fragment peptides into b- and y-ions



Each fragment appears as a cluster of isotope peaks rather than a single peak

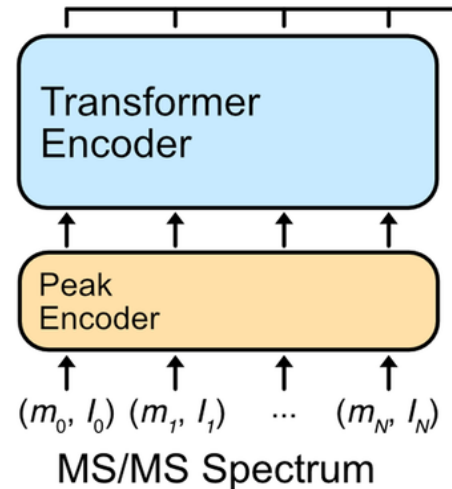


MS/MS spectra can be encoded in two types of representations

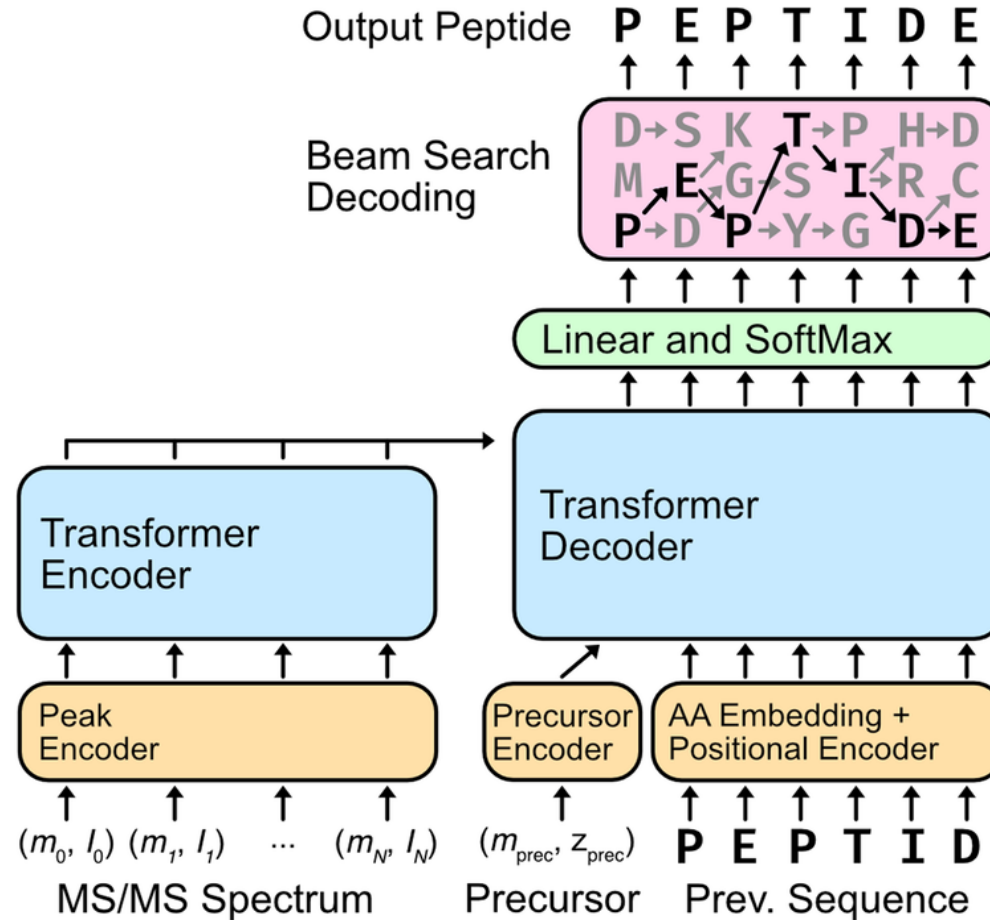


Bittremieux, Wout, et al. "Deep learning methods for de novo peptide sequencing." *Mass Spectrometry Reviews* 45.3 (2026): 507-526.

Casanovo represents peaks as a sequence of (m/z, intensity) tuples and extracts features with a Transformer encoder



Casanovo predicts the amino acid sequence autoregressively



What is next?

- I will apply for the AWS cloud!

- Read the publications:

Casanovo:

Yilmaz, Melih, et al. "De novo mass spectrometry peptide sequencing with a transformer model." *International Conference on Machine Learning*. PMLR, 2022.

Yilmaz, Melih, et al. "Sequence-to-sequence translation from mass spectra to peptides with a transformer model." *Nature communications* 15.1 (2024): 6427.

Cascadia:

Sanders, Justin, et al. "A transformer model for de novo sequencing of data-independent acquisition mass spectrometry data." *Nature Methods* 22.7 (2025): 1447-1453.

Improvement of Casanovo:

Straub, Gwenneth, et al. "Improvements to Casanovo, a deep learning de novo peptide sequencer." *Journal of Proteome Research* 25.2 (2026): 755-764.

Any other related publications

- Hand on the good first issues!