Comprehensive isoform annotation and quantitation in DIA data

Ignat Shilov, Christian Heckendorf, Abhishek Roushan, Gary Wilson, Claire Bramwell

Protein Metrics LLC, Boston MA

Introduction

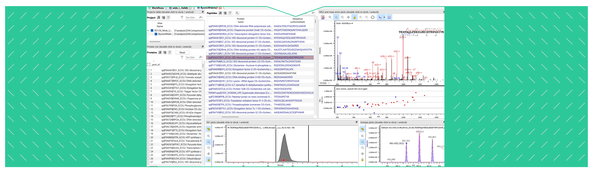
Over the past several years, Data Independent Acquisition (DIA) methods have become increasingly standard in proteome identification and quantification workflows. As DIA adoption has grown, a variety of data processing tools have emerged, showing steady improvements in both performance and sensitivity. However, challenges remain — such as managing coeluting shared fragments, handling large data volumes, and addressing computational complexity due to the combinatorial nature of DIA data.

Protein Metrics, LLC's MS/MS search engine, well-regarded for its role in Data Dependent Acquisition (DDA) applications within the PTM workflow, has been widely adopted by researchers and pharmaceutical companies. Building on this foundation, Protein Metrics recently introduced a new workflow specifically designed to support DIA data. This poster presents the development of a high-performance, scalable algorithm prototype tailored for DIA analysis, representing a significant advancement in our capabilities for handling complex proteomic datasets.

Methods

Dotmatics **BYOS** — **MPQ DIA**: Identification and Quantitation of Protein Mixtures in DIA Samples

Dotmatics **BYONIC** — **NEW SEARCH ENGINE**



The prototype of the new search engine was evaluated within the framework of existing Byos® Host Cell Proteome (HCP) and Multi Protein Quant (MPQ) workflows. In these tests, the new engine was run in place of Protein Metrics' traditional Byonic™ search engine.

Results: Comparing capabilities of detecting/quantifying peptide and protein isoforms

In this work we new approaches to extend the capabilities of Byos® platform for DIA analysis, by introducing a new search engine, designed to suite well the nature of DIA data. Our efforts focus on enhancing three key aspects

1. Increased sensitivity to ion signals by leveraging multiple isotopes, while maintaining robustness against potential interferences affecting individual isotopes.
2. Improved scoring discrimination to better distinguish true matches from interfering signals.
3. More accurate detection of elution boundaries for both precursor and fragment ion signals.

Our approach is centered around preserving flexibility throughout the analysis pipeline:

- We retain all matched fragment signals across multiple charge states and alternative hypotheses until the final scoring stage. This ensures that no potentially valuable match is prematurely discarded.
- The new search engine supports matching all relevant isotopic peaks — whether from centroided or profile-mode data — to observed signals.
- By incorporating profile data, we enhance true signal discrimination, enabling the recovery of matches even in cases of complex isotope interference. Uniquely, this approach allows for direct matching of raw profile points. Elution boundaries for both precursors and fragments are refined using multiple isotope observations.
- Finally, we aggregate signals from multiple adjacent MS2 scans around the precursor apex to further enhance match discrimination and boundary precision.

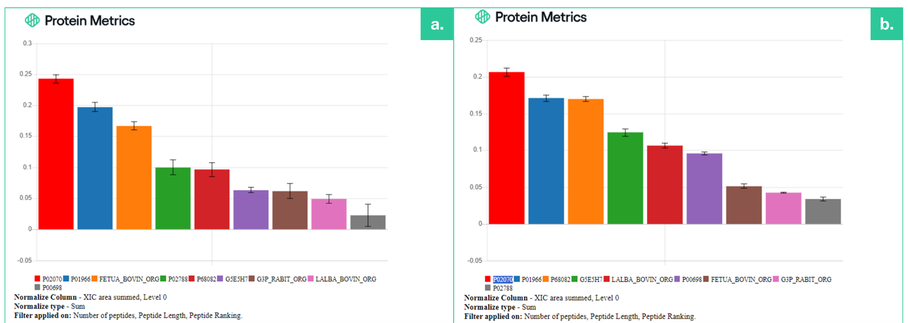


Fig. 1. Both the existing Byos® MPQ workflow using Byonic™ (a) and the new Byos® MPQ DIA workflow with the prototype search engine (b) successfully quantified all nine spiked-in proteins. However, the two approaches exhibit differing levels of quantification variability across replicate runs.

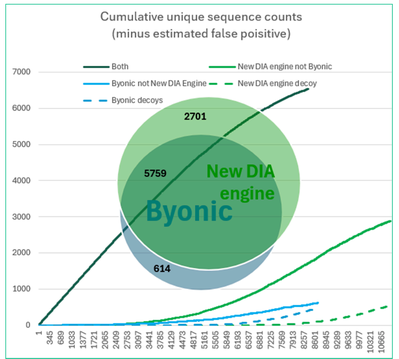


Fig. 2. Comparison of unique sequence identifications between two engines. The plots are limited by 10% False Discovery Rate cutoff and Vienn diagram numbers are calculated based on 1% FDR.

Fig. 5. Example of new ID from complex isotope mixture

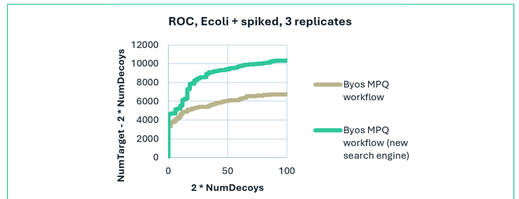


Fig. 3. Higher identification rates demonstrated by decoy analysis

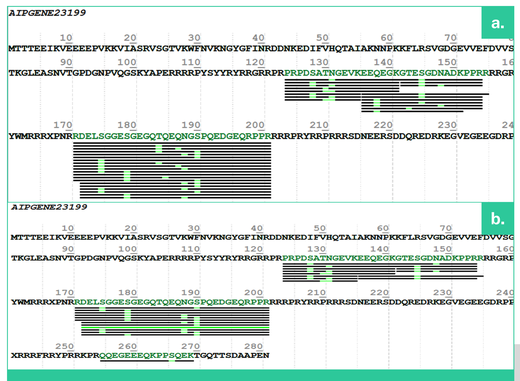


Fig. 4. (a) The Byonic™ engine can produce additional, sometimes conflicting, modification placements due to increased noise in MS2 spectra acquired farther from the elution apex. (b) The new search engine integrates MS2 data from multiple scans surrounding the elution apex, resulting in improved fidelity of modification site localization.

Conclusions

The new search engine prototype demonstrates enhanced discrimination of true positive matches compared to Byonic™. It successfully recovers additional identifications in challenging cases involving complex isotope interference.

We also observed improved quantification precision, reflected in lower coefficient of variation (CV) values in DIA data analysis. Furthermore, the new approach shows a reduced incidence of sporadic isoform reporting, indicating greater consistency and reliability in identification results.