Protein Metrics

Investigating the effects of data-independent acquisition (DIA) strategies on search sensitivity and detection limits

William Resager, Antony Harvey and Ignat Shilov

Protein Metrics, LLC, Boston, MA, USA

Summary

DIA strategies evaluated for host cell protein (HCP) analysis in a complex whole cell lysate matrix

Four isolation methods compared: Narrow, Overlapped, Mixed, and Wide m/z windows

FASTA based spectral library-free searches

USP1 detected at 0.1 fmol spike; match-between-runs (MBR) applied across 84 raw files (21 per strategy)

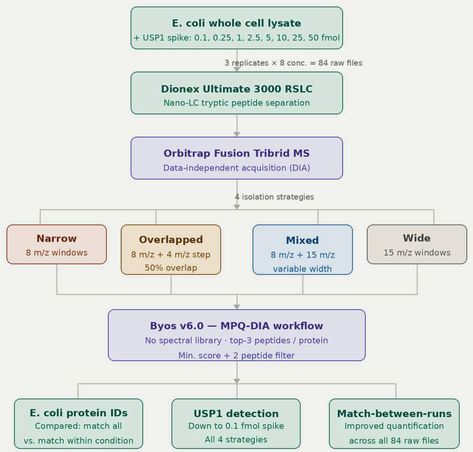
Introduction

Data-independent acquisition (DIA) has continued to gain popularity across many proteomics applications. Of particular interest is its use for host cell protein (HCP) analysis, where the ability to identify low-abundant species is especially valuable.

This study investigates the sensitivity and detection limits of four DIA acquisition strategies using a complex E. coli whole cell lysate spiked with USP1 standard at concentrations ranging from 0.1 to 50 fmol.

Methods

Tryptic peptide datasets of E. coli lysate (Gotti et al.) with USP1 standard (Sigma) dilutions were acquired on an Orbitrap Fusion Tribrid MS (Thermo Fisher Scientific) with a Dionex Ultimate 3000 RSLC nano-LC. Four isolation methods were tested: Narrow (8 m/z), Overlapped (8 m/z, 4 m/z step), Mixed (8+15 m/z), and Wide (15 m/z). Data were analyzed using the MPQ-DIA workflow in Byos v6.0 (Protein Metrics, LLC). Match-between-runs (MBR) was applied automatically, and results were filtered to the top 3 peptides per protein with a minimum score cutoff and at least 2 peptides per protein ID.



Data Review

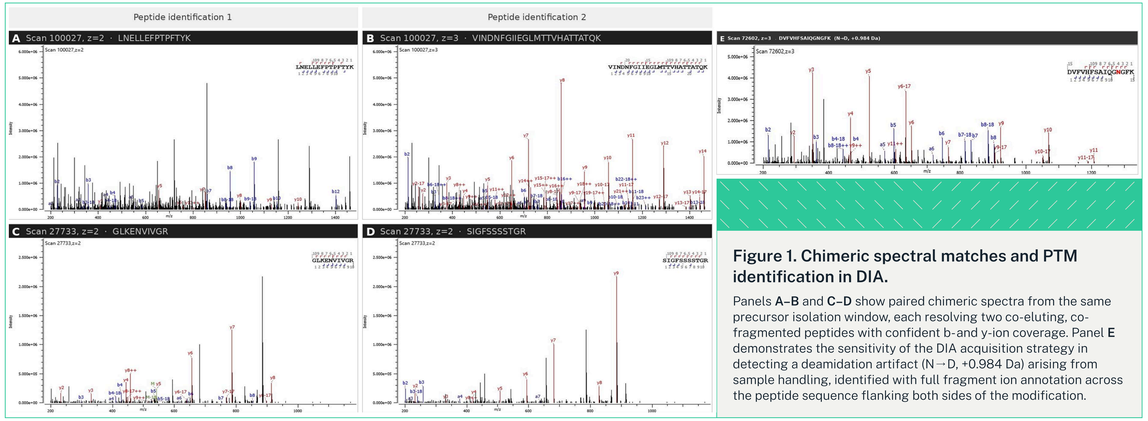


Figure 1. Chimeric spectral matches and PTM identification in DIA.

Panels **A–B** and **C–D** show paired chimeric spectra from the same precursor isolation window, each resolving two co-eluting, co-fragmented peptides with confident b- and y-ion coverage. Panel **E** demonstrates the sensitivity of the DIA acquisition strategy in detecting a deamidation artifact (N→D, +0.984 Da) arising from sample handling, identified with full fragment ion annotation across the peptide sequence flanking both sides of the modification.

PSM filtering	
Minimum peptide length	6
Maximum peptide length	45
Maximum missed cleavage count	2
Minimum peptide score	6.00
Minimum peptide replicate count	2
Keep homologous sequences	— No
Minimum peptide matching a protein count	2
Include modifications	All
Match between runs	
Enable match between runs	✓ Yes
Mode	Between all samples
Reported peptide filtering	
Retain only primary MS2 peptide ID	✓ Yes
Use only top charge state for quantification	✓ Yes
Use the same top N peptides for quantification	✓ Yes
Top N count	3
Peptide prioritization	Unique prioritized over Shared
Apply RSD limit for replicates	✓ Yes
Maximum acceptable %RSD for replicates	30.00

Automatic match-between-runs (MBR) with a **%RSD filter** ensures that only quality peptide IDs are carried forward, guaranteeing robust quantitation across samples

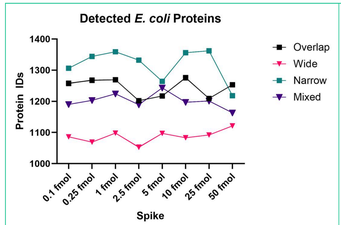
Replicate-based identification thresholds require a peptide to be identified in a minimum of 2 out of 3 replicates, combined with a minimum peptide-per-protein filter, eliminating unreliable single-observation identifications

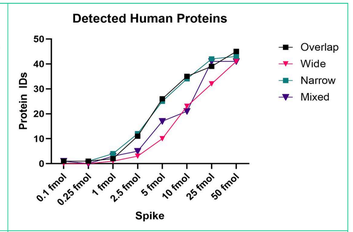
Homology filtering discards homologous peptides by default while preserving the top-N peptide requirement, under a top-3 setting, at least one unique peptide is retained alongside up to two homologous peptides

Unique prioritized over Shared peptide selection preferentially fills the top-N quota with unique peptides first before drawing on shared peptides, in contrast to the "Most Abundant Top-N" setting which selects purely by abundance regardless of uniqueness

Pep Filtering: The algorithm 'mutates' the sequences to create 'decoy' hits. Decoys are then scored with 'true' hits and used in calculating PEP probability of random match for each ID

Results





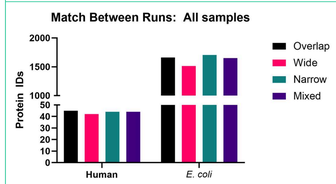


Figure 3. Comparison of match-between-runs conditions on protein identification and sensitivity.

Top: Detected E. coli (left) and human/USP1 (right) protein IDs when match-between-runs is constrained to within each spike condition only. Bottom left: Extending match-between-runs across all samples maintains sensitivity with no appreciable variation observed across spike concentrations, allowing for maintained sensitivity even at lower spike conditions.

Conclusions

All four DIA isolation strategies successfully detected USP1 peptides down to the lowest tested spike concentration of 0.1 fmol, demonstrating the inherent sensitivity of the MPQ-DIA workflow regardless of isolation window configuration. Meaningful differences between strategies began to emerge at 2.5 fmol, where Narrow and Mixed windows consistently yielded higher protein identifications compared to Overlapped and Wide strategies. This divergence became more pronounced at intermediate concentrations, suggesting that tighter isolation windows provide a tangible sensitivity advantage when quantifying low-abundance targets against a complex background proteome.

For the E. coli matrix proteins, all strategies produced stable identification counts across the full concentration range, confirming that background proteome coverage is largely insensitive to USP1 spike level. The application of match-between-runs further reinforced this consistency, aligning identification levels across replicates and reducing run-to-run variability for both the USP1 spike-in and the endogenous E. coli proteins. This was particularly evident at low concentrations where direct identification rates are significantly lower, with match-between-runs effectively rescuing quantitative information that would otherwise be lost.

Taken together, these findings indicate that narrower isolation windows offer superior sensitivity for detecting and quantifying low-abundance host cell proteins in complex matrices, while match-between-runs provides a robust complementary strategy for maximizing quantitative completeness across all acquisition conditions.

Conflict of Interest Statement

All authors are employees of Protein Metrics LLC