Definitive benchmarking of DDA and DIA for host cell protein analysis on the Orbitrap Astral in a regulatory-aligned framework

Somar Khalil,* Michel Plisnier

GSK, Rue de l'Institut, 89, Rixensart, 1330, Belgium

ABSTRACT: Host cell proteins (HCPs) are critical quality attributes in biotherapeutics that require accurate, specific, and comprehensive quantification. Mass spectrometry (MS)-based workflows are increasingly adopted to overcome the coverage and specificity limitations of immunoassays. This study benchmarks the performance of the Orbitrap Astral mass spectrometer for label-free HCP analysis, comparing data-dependent acquisition (DDA, Top80) and data-independent acquisition (DIA, 4 m/z non-staggered windows) modes. We applied a statistically rigorous framework integrating a stable isotope-labeled HCP mixture for traceable quantification, entrapment-based empirical false discovery proportion estimation, deterministic protein inference, and stratified bootstrapping. Both acquisition modes demonstrated exceptional quantitative fidelity ($R^2 \geq 0.99$ for absolute abundance, total error within ±30% acceptance limits). DIA outperformed DDA in identifications, yielding 45% more proteins and 68% more peptides. Hierarchical Bayesian modeling revealed superior differential linearity in DIA (mean slope ≈ 1.0) compared to DDA (slope ≈0.8). Stratified bootstrap analysis confirmed linearity and accuracy across the dynamic range, with DIA achieving lower limits of quantification (0.6 ppm) versus DDA (1.6 ppm). While both workflows reliably quantified most high-risk HCPs, DIA provided expanded proteome coverage and enhanced fold-change precision. These findings validate the Orbitrap Astral as a high-fidelity platform for HCP analysis in both modes and establish a defensible, regulatory-aligned MS-based framework for routine use in quality control environments.

Host cell proteins (HCPs) are process-related impurities originating from expression systems used in recombinant biotherapeutic production[1-2]. As critical quality attributes, residual HCPs must be quantitatively monitored due to their capacity to impair product stability, induce immunogenic responses, or elicit adverse clinical outcomes[3-6]. Traditionally, HCP monitoring has relied on enzyme-linked immunosorbent assays, which offer high-throughput quantification but suffer from limited specificity, narrow coverage, and an inability to resolve individual proteins[7-8]. These constraints have driven the increasing adoption of liquid chromatography-tandem mass spectrometry (LC-MS/MS) as an orthogonal platform capable of direct monitoring of individual HCPs[9-10]. Among LC-MS/MS acquisition strategies, data-dependent acquisition (DDA) remains widely used but is hindered by stochastic precursor selection, leading to variable detection of low-abundance species[11]. In contrast, data-independent acquisition (DIA) employs systematic, window-based precursor fragmentation, yielding highly multiplexed MS2 spectra with improved reproducibility and reduced missingness across replicates[12-14].

The Orbitrap Astral MS integrates high-resolution Orbitrap MS1 detection with an ultra-fast Asymmetric Track Lossless (Astral) MS2 analyzer. This architecture enables true parallel acquisition with scan rates exceeding 200 Hz, narrow DIA isolation windows (2–4 Th), and high ion utilization[15-16], offering substantial gains in proteome coverage, quantitation precision, and acquisition speed over legacy instruments. These capabilities are particularly relevant for HCP analysis, which demands deep dynamic range and high sensitivity against a dominant background matrix[17-19].

Despite these advancements, adoption of DIA in regulated workflows has been limited by concerns over the computational deconvolution of multiplexed MS2 spectra, which introduces an additional layer of inference compared to the direct precursor to fragment pairing observation in DDA[20-21]. While target-decoy approaches are well established for false discovery rate (FDR) control in DDA, some DIA pipelines require careful implementation to achieve equivalent stringency[20-22]. As regulatory expectations increasingly emphasize empirical error characterization, the use of DIA workflows with narrow precursor windows and high-resolution MS1 scans, combined with entrapment-based estimation of the false discovery proportion (FDP), is essential to confirm analytical specificity[23-29].

A further challenge in MS-based HCP analysis is the lack of ground truth for endogenous impurities. While stable isotope-labeled (SIL) peptides are often used, they fail to replicate the complexity and physicochemical behavior of endogenous HCPs[30-32]. In contrast, spiked-in heterologous whole-protein mixtures provide broader coverage, and enable evaluation of linearity, accuracy, and precision across dilution series in alignment with ICH Q2(R2) guidelines[32]. However, such mixtures may not fully represent the biological heterogeneity or matrix interactions of endogenous impurities. In contrast, whole-protein SIL-HCP mixtures derived from the production cell line offer a more representative alternative. When spiked into a purified drug substance matrix, they closely simulate native process-related impurities[33], enabling performance evaluation under conditions the reflect the

complexity of real biologic products[34]. Their structural and physicochemical diversity supports worst-case impurity modeling and enables traceable benchmarking across the analytical range for assessing trueness, precision, and recovery in line with ICH Q2(R2). When used routinely, SIL-HCP standards can also serve as system suitability test materials, supporting verification of recovery and bias against predefined acceptance criteria[35-36]. This represents a biologically relevant alternative to generic reference materials such as HeLa digests or commercial proteomic standards.

As biopharmaceutical development increasingly embraces risk-based impurity control, there is a growing need to move beyond proteome-depth metrics toward rigorous qualification of MS-based workflows. However, most published studies remain exploratory and lack the analytical, computational, and statistical structure required for regulatory-aligned HCP monitoring. To address this gap, we present a comprehensive benchmarking study of the Orbitrap Astral MS for label-free HCP quantification, comparing DDA (Top80) and DIA (4 m/z non-staggered windows). The evaluation was conducted within a statistically rigorous framework integrating: (1) a SIL-HCP mixture derived from Chinese hamster ovary (CHO) cells spiked into a NISTmAb matrix for traceable quantification, (2) empirical estimation of the FDP using two complementary entrapment strategies, (3) deterministic protein inference via greedy parsimony constrained by Occam's razor[37-38], and (4) stratified bootstrap analysis to enable distribution-aware profiling of linearity, accuracy, and data-driven lower limits of quantification (LLOQ) across a seven-point concentration series.

This study delivers a multi-metric performance evaluation of DDA and DIA on the Orbitrap Astral and establishes a scalable, defensible, and regulatory-aligned framework for qualifying MS-based HCP workflows in quality control environments.

EXPERIMENTAL SECTION

Sample preparation. NISTmAb reference material (humanized IgG1κ, NIST RM 8671) was used as the model product matrix. A SIL-HCP standard derived from CHO cells (Sigma) was spiked into the mAb. Seven defined spike-in levels (L1–L7: 4%, 6%, 8%, 10%, 12%, 14%, and 16% w/w relative to the mAb) were prepared. Sample preparation was adapted from our previous study with minor modifications[34]. Briefly, each replicate containing 20 µg of mAb protein was solubilized in 0.1% RapiGest (Waters) and 50 mM triethylammonium bicarbonate (Thermo Fisher Scientific). Proteins were reduced with 8 mM dithiothreitol at 50 °C for 45 minutes and alkylated with 16 mM iodoacetamide in the dark at room temperature for 30 minutes. Digestion was performed overnight at 37 °C using a Trypsin/Lys-C mix (Thermo Fisher Scientific) at a 1:40 enzyme-to-substrate ratio. The reaction was quenched with 0.5% trifluoroacetic acid (Biosolve). Peptides were desalted using C18 spin columns (Thermo Fisher Scientific), dried by vacuum centrifugation, and reconstituted in 0.1% formic acid (Biosolve). Each digest was spiked with four MassPREP protein digest standards (P02769, P00330, P00489, and P00924, Waters) at 0.37, 1.02, 1.37, and 5.14 fmol/µL, respectively. All samples were analyzed in triplicate by DDA and DIA in parallel on the Orbitrap Astral MS.

Liquid Chromatography-Mass Spectrometry. Tryptic peptides corresponding to a 500 ng theoretical mAb load per injection were analyzed using a Vanquish Neo UHPLC system (Thermo Fisher Scientific) operated in direct injection mode. Peptides were separated on an EASY-Spray PepMap Neo C18 column (100 Å, 2 µm, 75 µm ID × 50 cm, Thermo Fisher Scientific) maintained at 55 °C. Mobile phase A consisted of 0.1% formic acid in water, and mobile phase B consisted of 80% acetonitrile, 20% water, and 0.1% formic acid. The LC gradient was as follows: 3–8% B over 3 min at 500 nL/min; 8–32% B over 22 min at 300 nL/min; 32–50% B over 5 min; ramp to 99% B over 1 min at 500 nL/min; and hold at 99% B for 4 min. Electrospray ionization was performed at 1.8 kV. The ion transfer capillary was set to 285 °C, and the ion funnel RF level to 40. For DDA, MS1 scans were acquired in the Orbitrap at 240K resolution (at m/z 200) over an m/z range of 380–1500, with a maximum injection time (maxIT) of 20 ms and automatic gain control (AGC) target of 5 × 106. A Top80 method selected precursors with charge states ≥2 and intensities ≥5 x 103, using a quadrupole isolation width of 0.7 Th. MS2 spectra were acquired in the Astral analyzer using higher-energy collisional dissociation (HCD) at 27% normalized energy, with an m/z range of 120–2000, maxIT of 6 ms, and AGC target of 5 × 103. For DIA, MS1 scans were also acquired at 240K resolution over m/z 380–980, with a 10 ms maxIT and AGC target of 5 × 106. MS2 spectra were acquired in the Astral analyzer using non-overlapping 4 Th precursor isolation windows. Fragment ions were measured over m/z 150–2000 using HCD at 25%, 4 ms maxIT, and AGC target of 5 × 104. The total DIA MS1 cycle time was 0.6 s.

Data processing and protein inference. Raw data were processed using vendor-specific software: SpectroMine (v5.0, Biognosys) for DDA and Spectronaut (v20.0, Biognosys) for DIA, using the DirectDIA library-free workflow. Both searches were conducted with identical parameters against a composite FASTA database containing 83419 *Cricetulus griseus* entries from UniProt (Swiss-Prot and TrEMBL), the NISTmAb heavy and light chain sequences, and the four MassPREP digest standards. Trypsin/P was specified as the proteolytic enzyme, allowing up to one missed cleavage. Mass tolerances were set to auto. Fixed modifications included carbamidomethylation (C) and SIL labels on arginine (+10 Da) and lysine (+8 Da). Variable modifications included methionine oxidation and N-terminal acetylation. For targeted searches of MassPREP standards, only carbamidomethylation was used as a fixed modification. Peptides were filtered for lengths of 7–30 amino acids, with a maximum of two variable modifications and precursor charge states between 2+ and 4+. A uniform q-value threshold of 0.01 was applied at the precursor, peptide, and protein levels. In Spectronaut, confidence filtering was configured using q-value and posterior error probability thresholds of 0.01 at both the run and experiment levels, in line with best-practice recommendations[39]. No normalization or imputation was applied at this stage. Quantification was performed at the MS1 level in DDA (extracted ion chromatograms) and at the fragment-ion level in DIA (peak area summation).

To remove platform-specific bias and standardize reporting, protein inference was decoupled from the search engine and performed via a deterministic pipeline. Following 1% FDR filtering, stripped peptide sequences were re-annotated against an in-silico digest of the C. griseus proteome. Protein inference was implemented using a greedy parsimony algorithm that iteratively

assigns peptides to the minimal protein group explaining the observed evidence. This was followed by strict Occam's razor filtering to retain only groups supported by at least one unique peptide. For each group, the lead protein was defined as the member supported by the highest number of constituent peptides. Groups lacking unique evidence were retained as is.

All computational workflows were executed on a high-performance virtual machine with a 64-core AMD EPYC processor and 256 GB RAM. Downstream statistical analysis and visualization were performed in Python (v3.12) using in-house scripts based on NumPy, pandas, and seaborn. Bayesian modeling was implemented using PyMC, and posterior diagnostics were performed with ArviZ. Final figures were assembled in GraphPad Prism (v10.5).

Empirical FDP estimation. To empirically validate statistical control over false discoveries, we implemented a search-centric entrapment strategy adapted from Wen et al.[22]. In this framework, the reference list of target peptides is derived from a standard search against the unmodified database. Two orthogonal entrapment approaches were employed to probe error rate control under distinct assumptions. First, a shuffled database was constructed by randomly permuting the internal amino acid sequences of all identified C. griseus peptides while preserving the C-terminal residue (K/R). Duplicates matching native peptides were removed, and the final set was down-sampled to maintain a 1:1 target-to-entrapment ratio. Second, a trimmed database was generated from an *in-silico* tryptic digest of the Arabidopsis thaliana proteome, retaining only those foreign proteins whose peptides had no overlap with the native *C. griseus* peptidome. The resulting non-overlapping peptide set was likewise down-sampled to match the size of the target space. Each entrapment FASTA was concatenated with the target database, and the combined databases were used for re-analysis under identical search conditions. FDP was estimated at multiple q-value thresholds (τ) using two distinct methods:

The combined FDP was calculated as:

$$\mathrm{FDP}_{\mathrm{Combined}}(\tau) = \frac{N_E(\tau)\cdot(1+r^{-1})}{N_D(\tau)} \quad (1)$$

where $N_E(\tau)$ is the number of identified entrapment peptides below threshold (τ), $N_D(\tau)$ is the total number of discoveries, and r is that ratio of target to entrapment peptides in the database $(r = 1)$.

The paired FDP was calculated as:

$$\mathrm{FDP}_{\mathrm{Paired}}(\tau) = \frac{N_E(\tau)+\sum_{i=1}^{N_p} I\left(q_{E_i}\leq\tau\wedge\left(q_{T_i}>\tau\vee q_{E_i}<q_{T_i}\right)\right)}{N_T(\tau)+N_E(\tau)} \quad (2)$$

For each of the N_p target-entrapment peptide pairs $(T_i,\ E_i)$, I is an indicator function that counts entrapment hits that are either identified alone or have a better q-value than their target counterparts. $N_T(\tau)$ and $N_E(\tau)$ are the total counts of targets and entrapment peptides identified below (τ), respectively.

Statistical analysis.

Data filtering and normalization: Post-inference, peptide-level data were processed through a sequential quality control pipeline. Outlier peptides were identified at each spike-in level using a robust modified Z-score, where peptides with $|Z_i| > 2.6$ were excluded. The score was calculated as:

$$Z_i = \frac{0.6745(x_i-\tilde{x})}{\mathrm{MAD}} \quad (3)$$

where x_i is the log2-transformed intensity, $\tilde{x}$ is the median intensity, and MAD is the median absolute deviation. Subsequently, intra-protein consistency was enforced by removing any peptide whose median $\log_2$-intensity deviated by more than 30-fold from its parent protein's median. Data were then corrected for loading variations via total sum scaling within each spike level. Finally, a precision filter was applied, discarding any peptide with coefficient of variation (CV%) exceeding 25% across its triplicates. This filtering strategy was validated by comparing $\log_2$ intensity density distributions before and after processing (Supporting Information, Figure S1), confirming the removal of noise and improved symmetry.

Hierarchical Bayesian modeling of differential linearity: To evaluate fold-change linearity at the protein level, we implemented a hierarchical Bayesian regression model using partial pooling. The model estimates individual protein-level linearity parameters (β_j) drawn from a common group-level distribution, regularizing the estimates. Observed fold-change data (y_{ij}) was modeled as a function of the expected fold change (x_i) across spike-in levels, assuming a Student's T-distribution likelihood. The model structure is defined as:

$$y_{ij} \sim \mathrm{StudentT}(\alpha_j + \beta_j x_i, \sigma, \nu)\beta_j \sim \mathrm{Normal}(\mu_\beta, \sigma_\beta)nu \sim \mathrm{Gamma}(2,0.1) \quad (4)$$

Where (μ_β) is the group-level mean linearity, and the degrees-of-freedom parameter (ν) is estimated from the data using a weakly informative Gamma hyperprior. Posterior distributions were inferred using the No-U-Turn Sampler (NUTS)[40]. Model convergence was confirmed by ensuring the Gelman-Rubin diagnostic $(\hat{R})$ was below 1.01 and visual inspection of trace plots.

Global HCP accuracy profile: Quantification accuracy for total HCP content was evaluated using the total error approach that generates an accuracy profile, integrating systematic error (trueness) and random error (precision). Trueness was measured as the mean relative bias from the nominal spike level. Precision was assessed using two-sided parametric \beta-expectation tolerance intervals, constructed to contain 95% of future measurements with 95% confidence. Given the sample size of n=3 technical triplicates, parametric intervals were calculated as $\bar{y} \pm k \cdot s$,where $(\bar{y})$ and (s) are the sample mean and standard deviation, respectively. The statistical tolerance factor was a fixed value of k=9.01, as established by Hahn-Meeker for achieving 95%/95% coverage/confidence at n=3 under the assumption of normality[41]. The predefined acceptance criterion for the total error was set at ±30%.

Stratified bootstrap analysis: To assess quantification fidelity across the full abundance range, we implemented a stratified non-parametric bootstrapping strategy. At each spike-in level, proteins were stratified into four abundance-based bins (P0–15, P15–35, P35–50, and P50–100). Within each stratum, protein abundances (in ppm) were resampled with replacement (n=12,000 iterations) to generate empirical distributions of the mean. From these distributions, we derived two-sided 95% β-expectation tolerance intervals using the 2.5th and 97.5th percentiles. An accuracy criterion of ±35% was applied to define acceptable total error.

Combined with linearity criterion (coefficient of determination (R^2) ≥ 0.98), this framework enabled data-driven estimation of the LLOQ. The LLOQ was defined as the lowest abundance stratum meeting both accuracy and linearity thresholds.

RESULTS AND DISCUSSION

DIA delivers superior CHO proteome coverage and data completeness. A head-to-head comparison on the Orbitrap Astral revealed that the DIA workflow consistently outperformed DDA in proteome coverage, data completeness, and quantitative precision across all spike-in levels. DIA yielded substantially greater depth, identifying more protein groups and peptides than DDA at every concentration point (Figure 1A–B). This enhanced discovery power was further evidenced by a higher number of peptides per protein (Figure 1C), a key metric for confident protein inference. Although DDA yielded a slightly greater absolute number of single-hit proteins (Figure 1D), the proportion of single-hits relative to the total was lower in DIA at all levels. The deterministic sampling of DIA enabled superior data completeness[42-45]. As shown in Figure 1E, DDA exhibited substantial missingness, with over 50% of peptides dropping out at lower spike levels due to stochastic, intensity-driven precursor selection. In contrast, DIA maintained high peptide coverage across all concentrations. Venn diagrams (Figure 1F–G) indicate that 87% of DDA-identified proteins and 96% of peptides were also detected by DIA, while DIA uniquely identified a large number of additional features. Figure 1I illustrates the fundamental difference in acquisition strategies: DDA shows a sparse, banded precursor distribution, whereas DIA achieves uniform, continuous sampling across the retention time and m/z space. This translates into improved precision, as seen in the distribution of peptide CV% values (Figure 1K–L). DIA demonstrated lower and tighter CV% distributions across all levels. Figure 1H further highlights this advantage: DIA exhibited better precision across the full intensity range, especially at the low-abundance end. In contrast, DDA showed a non-monotonic pattern, with improved CV% at mid-range intensities but degraded precision for both low- and high-abundance peptides. This trend was captured by the LOWESS regression, confirming that DIA maintains stable and reliable quantification across the dynamic range. Collectively, these results validate DIA on the Orbitrap Astral as a high-performance acquisition strategy that delivers deeper coverage, higher data completeness, and superior precision, thereby supporting more comprehensive monitoring of residual HCPs.

Both workflows demonstrate well-characterized FDR control. Empirical FDP analysis confirmed that both acquisition workflows achieved effective statistical error control. For DDA, using SpectroMine, all three entrapment-based estimators, paired-shuffled, combined-shuffled, and combined-trimmed, returned conservative FDP values well below the nominal 1% threshold across all q-value cutoffs (Figure 1J, left). The combined-shuffled estimator exhibited a slight non-monotonic trend at higher q-values, where the FDP curve decreased beyond 0.8%. This phenomenon is a distinctive characteristic of a well-calibrated search that successfully models the score distribution of false positives using a target-decoy strategy[46]. It signifies that the SpectroMine search engine achieved excellent separation between high-scoring true identifications and the low-scoring decoy population. As a result, relaxing the q-value cutoff initially accumulates true positives faster than false positives, causing the FDP ratio to temporarily decrease before inevitably rising at more permissive thresholds[47]. In DIA, the FDP curves displayed distinct characteristics (Figure 1J, right) highlighting the importance of using orthogonal validation strategies. The FDP curves are largely invariant to the nominal FDR threshold, a phenomenon indicative of a highly bimodal distribution of precursor q-values where the algorithm confidently separates true and false hits. The analysis demonstrated effective FDR control when using the paired-shuffled and combined-trimmed estimators, both yielding an empirical FDP near or below the nominal 1% target. Crucially, however, the combined-shuffled method resulted in an empirical FDP of approximately 1.5% at a 1% nominal FDR, indicating a slight inflation. This divergence strongly suggests that the inflation observed with the combined-shuffled method is an artifact of its core assumption being violated. The target-decoy strategy fundamentally assumes that decoy hits are a faithful statistical proxy for false-positive target hits. Our data indicates that the simple shuffled peptides were more susceptible to spurious matches in the complex, chimeric DIA spectra than the more realistic foreign peptides from A. thaliana. This led to an overestimation of false positives by the combined-shuffled estimator, a known potential pitfall of certain decoy strategies[48]. The fact that the more robust entrapment methods showed proper error control confirms that the underlying DIA identifications are reliable. While FDP underestimation is a known challenge in some DIA pipelines[22, 49], the stringent, multi-level filtering applied in our Spectronaut workflow[39] provides the necessary confidence to conclude that both workflows, when appropriately configured and validated, deliver reliable identifications suitable for regulated applications.

Global quantitative trends align, but methods are not interchangeable for individual HCPs. Protein-level quantification was performed using the Hi3 label-free approach, aggregating the three most intense quantifiable peptides per spike level through summation[34,50]. DIA outperformed DDA in analytical depth, consistently quantifying more peptides and proteins across all concentrations (Figure 2A–B). The overlap analysis revealed a shared core of approximately 1509 quantifiable proteins, while DIA uniquely quantified an additional 1600 proteins (Figure 2C). Bland–Altman analysis of $\log_2$ fold change (log2FC) between representative spike-in comparisons was performed (Figure 2D). Although the mean bias between DDA and DIA was close to zero, the 95% limits of agreement were wide, extending to approximately ±2 $\log_2$ units, corresponding to a potential four-fold discrepancy for individual proteins. This indicates that while the methods are concordant in aggregate, they are not quantitatively interchangeable for tracking individual HCPs. Differential linearity was further assessed using a hierarchical Bayesian regression model. Posterior distributions of the (μ_β) showed that DIA achieved near-perfect linearity, with values around 1.0, while DDA exhibited a systematic underestimation with $\mu\beta \approx 0.8$ (Figure 2E). Markov chain Monte-Carlo model diagnostics confirmed excellent convergence of the NUTS samples, with trace plots displaying stable sampling and posterior densities appearing unimodal and well-defined (Figure 2F). These results confirm that DIA provides more accurate fold-change estimates across the dynamic range, while DDA tends to underestimate differential abundance.

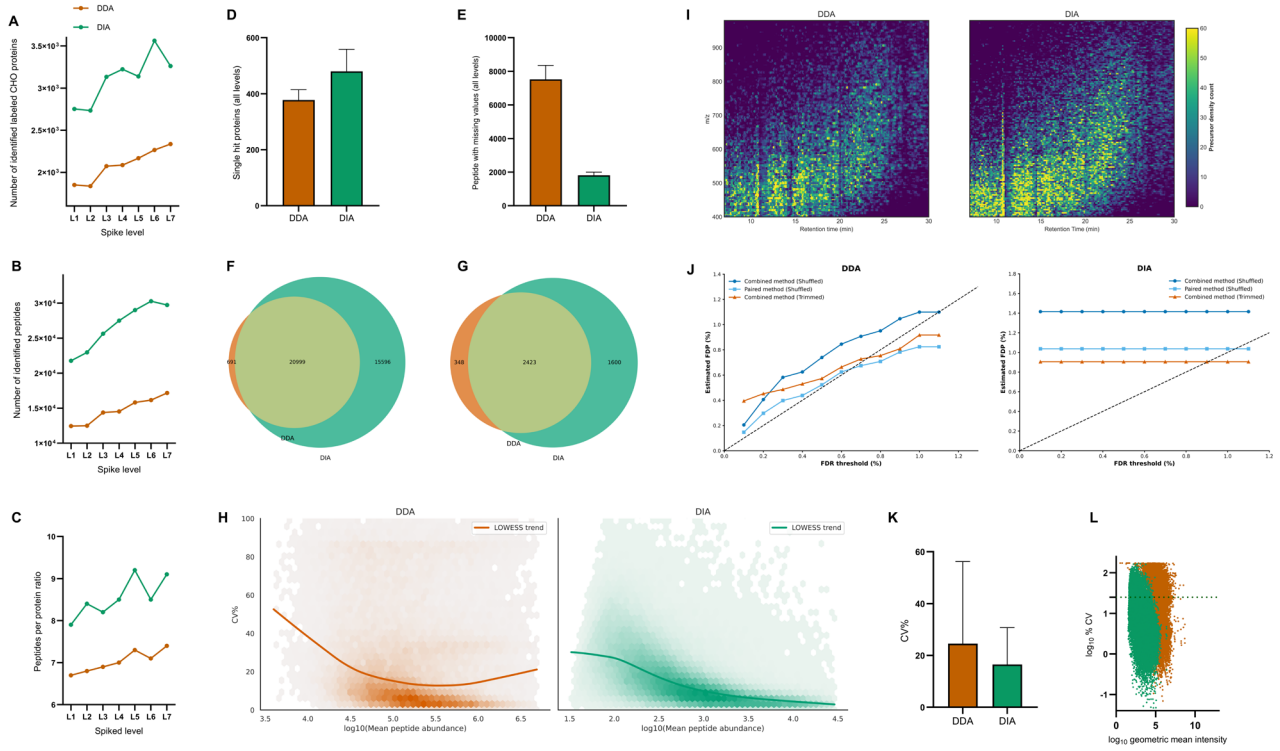


Figure 1. Comparative evaluation of DDA and DIA workflows for proteome coverage, data completeness, and quantitative precision. Line plots show the number of identified proteins (**A**), peptides (**B**), and the average peptides per protein (**C**) across seven spike-in levels. Bar charts compare the total number of single-hit proteins (**D**) and peptides with missing values (**E**) between modes. Venn diagrams illustrate the overlap in identified proteins (**F**) and peptides (**G**). (**H**) Hexbin plots show the relationship between peptide abundance and CV% with a LOWESS trendline. (**I**) 2D density maps visualize precursor distribution in retention time versus m/z space. (**J**) Empirical FDP estimation using paired and combined entrapment strategies.

(**K**) Bar chart comparing the mean peptide-level CV% across all concentrations. (**L**) Scatter plot of log-transformed CV% versus log-transformed geometric mean intensity. Error bars represent the standard deviation across all spike levels.

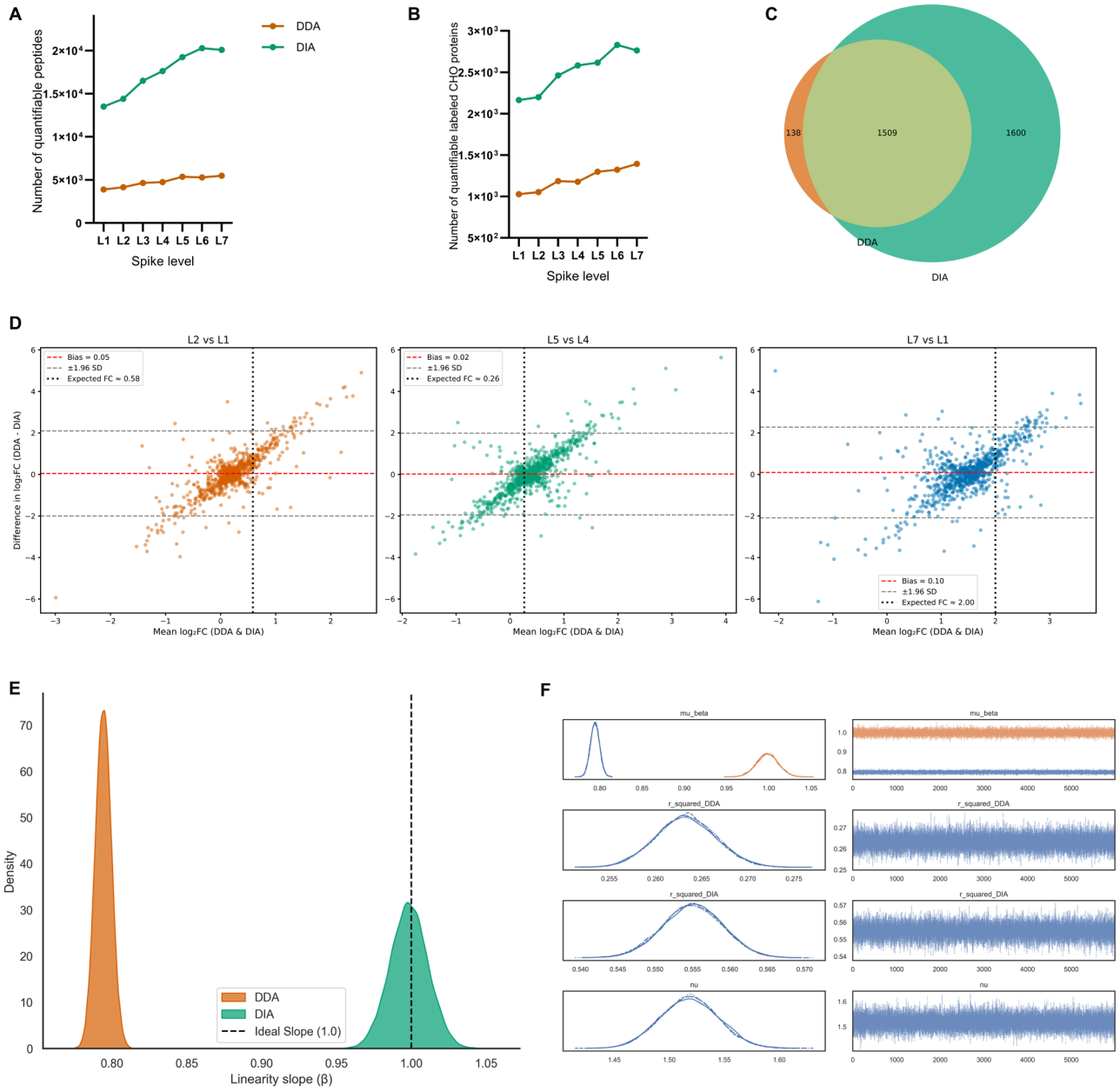


Figure 2. Quantitative performance comparison and assessment of differential linearity. Line plots show the number of quantifiable peptides (**A**) and proteins (**B**) across seven spike-in levels for DDA and DIA. (**C**) Venn diagram illustrating the overlap in quantifiable proteins between the two acquisition modes. (**D**) Bland-Altman plots comparing the $\log_2$FC agreement between DDA and DIA for three representative concentration comparisons; plots show the mean bias (solid line) and 95% limits of agreement (dashed lines). (**E**) Posterior density distributions from the hierarchical Bayesian model for the mean linearity slope parameter (μ_β), comparing DDA and DIA against the ideal slope of 1.0. (**F**) MCMC diagnostic plots from the Bayesian analysis, showing posterior density distributions (left) and trace plots (right) for key model parameters, confirming model convergence.

Both methods accurately quantify total HCP abundance. To enable absolute quantification, summed top-three peptide intensities of the four MassPREP standards were plotted against their known spike-in amounts (fmol) to generate a universal response factor, which was used to convert measured peptide intensities into absolute HCP abundance (in ng), using molecular weights for scaling. Regression of measured versus theoretical HCP abundance demonstrated excellent linearity for global HCP abundance in both workflows, with R^2 values exceeding 0.99 (Figure 3A–B). These results confirm that both DDA and DIA accurately capture proportional changes in total HCP content over the tested concentration range (20–80 ng). Analytical accuracy was further assessed using the total error approach, which integrates systematic error (trueness) and random error (precision) into a unified accuracy profile. For DDA, relative bias was consistently negative across all concentrations, ranging from −10% to −20%. For DIA, the bias was concentration-dependent, starting near +10% at the lowest level and shifting toward −15% at higher concentrations (Figure 3C–D). In both cases, the two-sided 95% β-expectation tolerance intervals remained fully within the predefined ±30% acceptance limits.

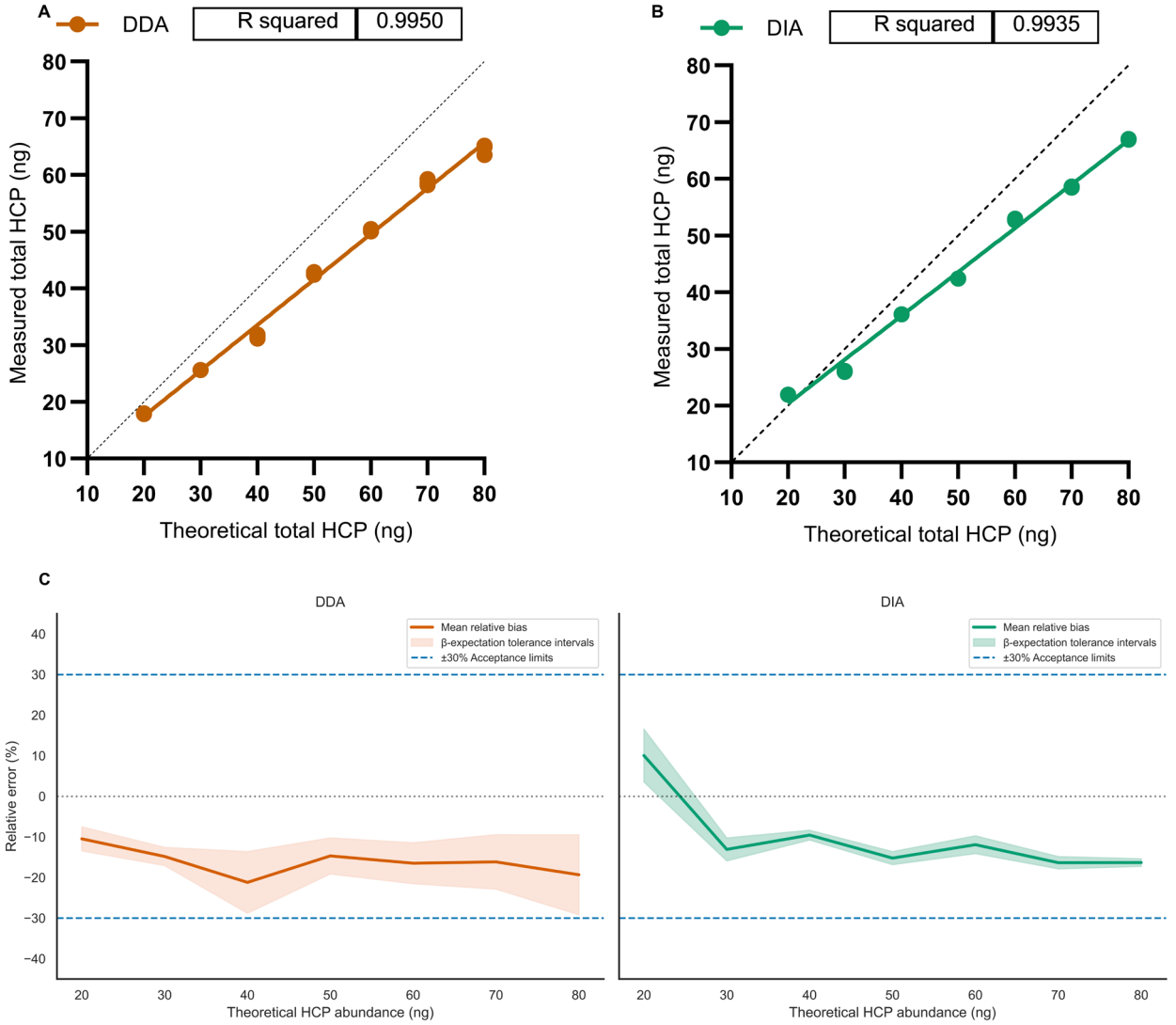


Figure 3. Linearity and accuracy of total HCP quantification. (**A-B**) Regression plots of measured versus theoretical total HCP abundance (ng) for DDA (**A**) and DIA (**B**) workflows, with R^2 values indicating excellent linearity. The dashed line represents the ideal 1:1 relationship. (**C-D**) Accuracy profiles using a total error approach for DDA (**C**) and DIA (**D**). The solid line denotes the mean relative bias (trueness), the shaded area represents the 95% β-expectation tolerance intervals (precision), and the dashed lines indicate the predefined ±30% acceptance limits for total error.

Stratified analysis reveals superior precision and lower LLOQ for DIA. To assess quantification fidelity beyond global performance metrics, we implemented a stratified evaluation framework based on protein abundance percentiles. At each spike-in level, proteins were grouped into four abundance strata (P0–15, P15–35, P35–50, P50–100), and bootstrapped resampling (n = 12,000 iterations) was performed within each bin to generate empirical distributions of the mean abundance. Faceted accuracy profiles (Figure 4A–B) demonstrated that both workflows maintained quantification accuracy within the predefined ±35% acceptance limits across most spike levels and abundance bins. DIA exhibited consistently narrower β-expectation tolerance intervals. The bootstrapped distributions of mean abundance (Figure 4C–D) displayed narrow, symmetric

kernel density estimate (KDE) plots that progressively shifted with increasing spike levels, confirming preserved linearity across the dynamic range. This was further supported by regression of the bootstrapped means against theoretical values within each stratum, which yielded R^2 values ≥ 0.98 for all bins (Figure 4E). Using this framework, the practical LLOQ was defined as the lowest abundance bin that satisfied both linearity and accuracy criteria. Based on these results, the LLOQ was estimated at approximately 1.6 ppm for DDA and 0.6 ppm for DIA, confirming the enhanced sensitivity of DIA for trace-level HCP measurement.

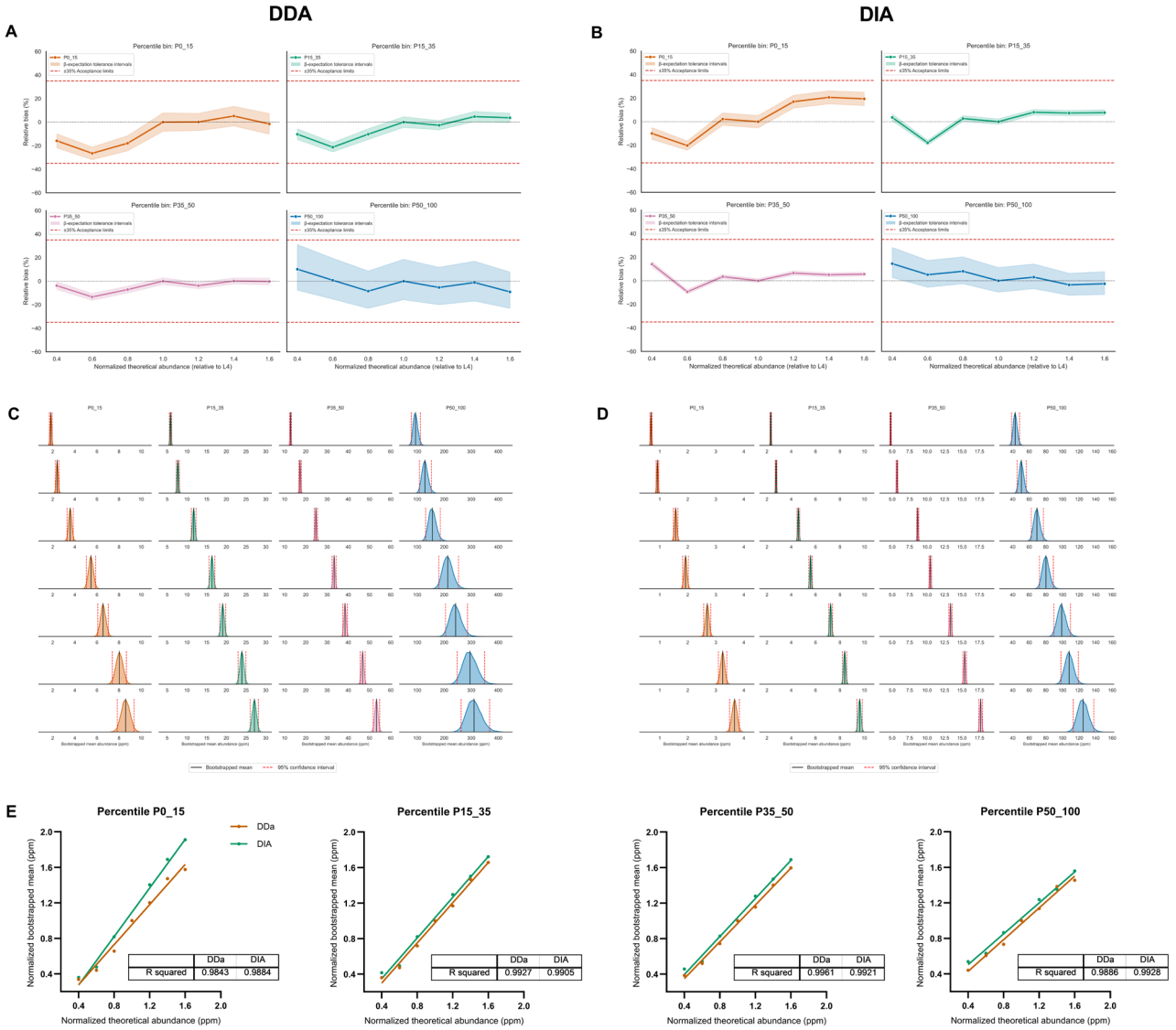


Figure 4. Stratified analysis of quantification fidelity across the HCP abundance range. (**A**-**B**) Faceted accuracy profiles for DDA (**A**) and DIA (**B**) based on bootstrapped abundance percentile bins. Solid lines represent the mean relative bias, shaded regions represent the 95% β-expectation tolerance intervals, and dashed lines indicate the $\pm 35\%$ acceptance limits. (**C**-**D**) KDE plots showing the empirical distribution of bootstrapped mean abundances across all spike levels (rows) and abundance strata (columns) for DDA (**C**) and DIA (**D**). (**E**) Linearity plots of the normalized bootstrapped means versus normalized theoretical abundance within each percentile bin, with R^2 values reported for both DDA and DIA.

High-risk HCPs are reliably monitored by both modes. Another objective of this study was to assess the ability of each workflow to monitor high-risk HCPs known to impact product quality and safety51. At the lowest spike level (L1, 20 ng), both DDA and DIA successfully quantified 19 of the 22 high-risk proteins listed in Table 1. These included well-characterized impurities such as Cathepsin B (90 ppm in DDA vs 79 ppm in DIA) and Clusterin (74 ppm in both modes), which showed excellent quantitative agreement. However, notable discrepancies were observed for a subset of proteins. For example, Peroxiredoxin-1 was quantified at 103 ppm in DDA but only 2 ppm in DIA, while Heat shock cognate 71 kDa protein was measured at 382 ppm in DDA versus 119 ppm in DIA. These differences reflect mode-dependent variability in quantification performance. A direct comparison of quantitative quality revealed that DIA consistently produced higher median per-protein linearity (R^2) values, both within the high-risk panel and across the broader CHO proteome (Supporting Information, Figures S2 and S3). A greater proportion of DIA-quantified proteins exceeded the R2 ≥ 0.95 threshold compared to DDA. This enhanced performance can be attributed to the fundamental difference in acquisition strategy: DIA integrates MS2 fragment-level signals in a systematic and reproducible manner, reducing susceptibility to

precursor-level interference, which can compromise quantification fidelity in DDA.

A framework for regulatory-aligned HCP workflow validation. Despite growing interest in MS-based HCP monitoring, most existing workflows lack the analytical rigor and validation structure required for implementation in regulated environments. In particular, critical elements such as empirical error rate control, reproducible protein inference, and analytical range qualification are rarely addressed in ways that align with regulatory expectations. This study addresses that gap by introducing a comprehensive analytical and computational framework designed to meet the core requirements of ICH Q2(R2). Specificity was reinforced through empirical estimation of FDP using two orthogonal entrapment strategies, assessing error rates at the precursor level. Linearity was demonstrated using a seven-point spike-in series, with R^2 values ≥ 0.99 across the validated range. Trueness and precision were jointly assessed using total error-based accuracy profiles, incorporating β-expectation tolerance intervals at 95% confidence. Both DDA and DIA workflows satisfied predefined acceptance criteria of ±30% for global HCP abundance and ±35% in the stratified percentile-based analysis. The practical LLOQ was determined by integrating both accuracy and linearity criteria across abundance strata, yielding an LLOQ of 1.6 ppm for DDA and 0.6 ppm for DIA.

For applications such as release testing, where conservative error control is required, a slightly lower nominal FDR (e.g., 0.8%) could be applied in cases where certain estimators systematically underestimate FDP. This provides a practical risk-mitigation strategy to ensure empirical error rates remain within a 1% threshold.

In addition to meeting conventional validation metrics, the use of a SIL-HCP standard provided a biologically representative model of worst-case impurity behavior, enabling traceable benchmarking and system suitability verification in line with risk-based impurity control principles. Reproducible protein inference was ensured through deterministic parsimony-based grouping to maintain consistency across datasets. Unlike vendor-based grouping strategies, which may rely on proprietary logic and change between software versions, this fixed and transparent approach improves traceability and supports long-term consistency, an essential consideration for regulatory transparency. While advanced methods such as stratified bootstrapping and Bayesian modeling were applied here to rigorously define method performance and LLOQ during development, routine implementation could rely on a streamlined system suitability test using the SIL-HCP standard to verify recovery and bias against predefined acceptance criteria.

Together, these elements establish a defensible and regulatory-aligned framework for qualifying MS-based HCP workflows. The approach presented here offers a practical blueprint for transitioning discovery proteomics into a validated tool suitable for process development, lot release, and comparability studies.

CONCLUSIONS

This study provides a comprehensive benchmark of DDA and DIA acquisition on the Orbitrap Astral MS, establishing a new reference point for regulatory-aligned HCP analysis. Our multi-dimensional evaluation confirmed that while DIA offers greater proteome coverage, improved precision, and lower LLOQ, both acquisition modes support high-fidelity quantification suitable for residual HCP monitoring. The analytical framework developed here integrates SIL-HCP standards, empirical FDP estimation, deterministic protein inference, and stratified performance evaluation to meet the core requirements of ICH Q2(R2). These components collectively facilitate the implementation of MS-based HCP workflows in quality control and comparability contexts.

Several limitations should be acknowledged. This study was conducted using technical triplicates, a single mAb matrix, and the CHO proteome; while similar performance is anticipated for other organisms, further validation is warranted. Software dependencies may also influence transferability. Additionally, the analytical range and LLOQ values reported here reflect standard tryptic digestion without mAb depletion or enrichment. Application of such enrichment strategies could further expand the quantifiable range and lower LLOQ thresholds by an order of magnitude. Finally, the performance characteristics observed here are specific to the Orbitrap Astral, with its ultra-fast scan speed and high resolution, and may not directly translate to other MS platforms. Future work should extend this framework to diverse biologic matrices, software pipelines, and instrument architectures to support broader adoption of MS-based HCP monitoring in regulated settings. While this study lays the foundation for method validation in discovery proteomics, formal validation exercises with sufficient assay replication will be necessary to fully characterize method performance in alignment with ICH guidelines.

In summary, this work validates the Orbitrap Astral as a premier platform for quantitative HCP analysis and provides the biopharmaceutical industry with a practical blueprint for qualifying untargeted MS-based workflows for routine use in process development, quality control, and comparability studies.

Table 1. List of high-risk HCPs searched in DDA and DIA modes at the lowest spike level (L1: 20 ng)

Accession no.	Protein name	DDA (ppm)	DIA (ppm)
G3H0L9	Cathepsin B[52]	90	79
G3INC5	Cathepsin L[53]	43	41
Q9EPP7	Cathepsin Z[54]	187	53
G3GTT2	C-C motif chemokine[55]	42	7
G3HNJ3	Clusterin[56]	74	74
A4URF0	C-X-C motif chemokine[57]	9	6
G3I8R9	Heat shock cognate 71 kDa protein[58]	382	119
G3HQY6	Lipase[59]	21	9
G3H6V7	Lipoprotein lipase[59]	119	75
A0A061IFE2	Liver carboxylesterase[60]	10	15
G3HRK9	Matrix metalloproteinase-19[57]	17	44
Q9JKY1	Peroxiredoxin-1[61]	103	2
G3I6T1	Phospholipase B-like 2 (PLBL2)[62]	165	85
G3I0W1	phosphopyruvate hydratase[63]	137	80
G3IIE7	Procollagen-lysine[64]	165	91
G3HC31	Protein S100[57]	N/A	0.3
G3H3Q1	Pyruvate kinase[65]	132	82
G3IBF4	Serine protease (HTRA1)[65]	115	72
G3IIB1	Sialate O-acetylesterase[66]	25	82
G3HR95	Carboxypeptidase D[67]	N/A	N/A
G3I3Y6	Glutathione S-transferase[68]	N/A	N/A
G3IA12	Transforming growth factor-β1[55]	N/A	N/A

ASSOCIATED CONTENT

Supporting Information

Additional details and results.

AUTHOR INFORMATION

Corresponding Authors

*E-mail: Somar.x.khalil@gsk.com

Funding Sources

The authors are full-time employees of the GSK group of companies with stock compensation.

Author contribution

The authors contributed to the design of the study, acquisition and analysis of data, and manuscript development.

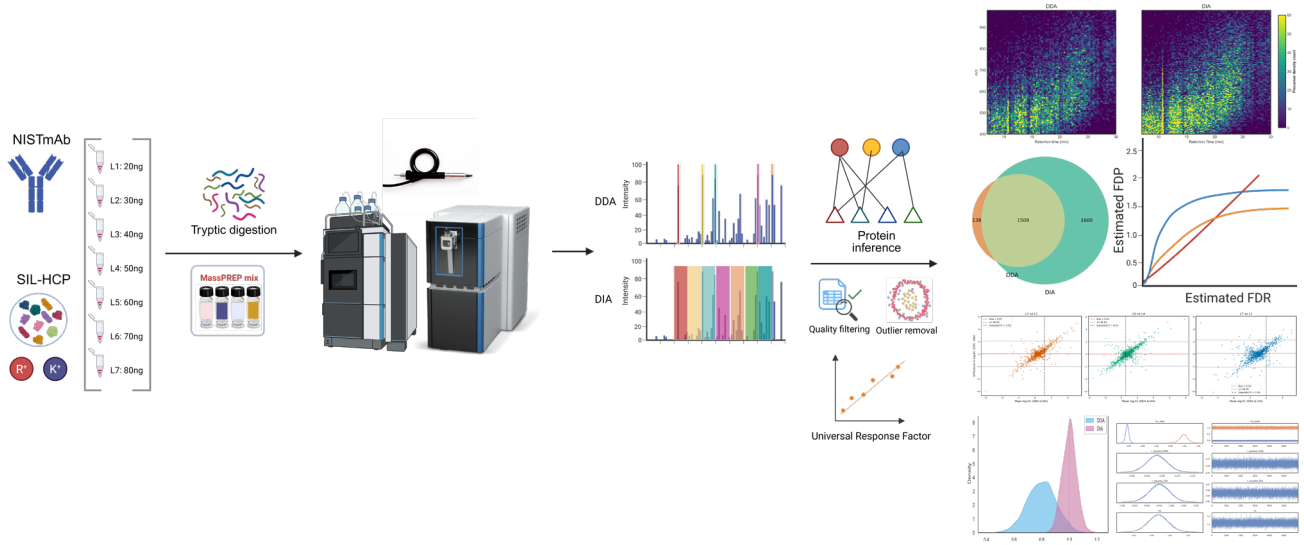
Acknowledgments

The authors gratefully acknowledge Pascal Bourguignon and Nora Zaïm for their contributions to the laboratory work that supported this study. We also thank Jean-François Dierick for his valuable input and critical guidance on topics related to analytical method validation.

REFERENCES

1. Host Cell Protein-Mediated Adjuvanticity and Immunogenicity Risks of Biotherapeutics. Panikulam, Sherin, et al. 2025, Biotechnology Advances, Vol. 81, p. 108575.
2. Host Cell Proteins in Monoclonal Antibody Processing: Control, Detection, and Removal. Ito, Takao, et al. 4, 2024, Biotechnology Progress, Vol. 40, p. e3448.
3. Host Cell Proteins in Biologics Manufacturing: The Good, the Bad, and the Ugly. Kornecki, Martin, et al. 3, 2017, Antibodies, Vol. 6, p. 13.
4. Characterization of Host Cell Proteins in the Downstream Process of Plant-Based Biologics Using LC-MS Profiling. Panapitakkul, Chalisa, et al. 2024, Biotechnology Reports, Vol. 44, p. e00856.
5. Recent Developments in Bioprocessing of Recombinant Proteins: Expression Hosts and Process Development. Tripathi, Nagesh K., and Ambuj Shrivastava. 2019, Frontiers in Bioengineering and Biotechnology, Vol. 7, p. 420.
6. Technical Advancement and Practical Considerations of LC-MS/MS-Based Methods for Host Cell Protein Identification and Quantitation to Support Process Development. Guo, Jia, et al. 1, 2023, mAbs, Vol. 15, p. 2213365.
7. Host Cell Protein Testing by ELISAs and the Use of Orthogonal Methods. Zhu-Shimoni, Judith, et al. 12, 2014, Biotechnology and Bioengineering, Vol. 111, pp. 2367-79.
8. Host Cell Proteins in Biotechnology-derived Products: A Risk Assessment Framework. De Zafra, Christina L. Zuch, et al. 11, 2015, Biotechnology and Bioengineering, Vol. 112, pp. 2284-91.
9. A Highly Sensitive and Robust LC-MS Platform for Host Cell Protein Characterization in Biotherapeutics. Ji, Qinqin, et al. 2023, Biologicals, Vol. 82, p. 101675.
10. Label-Free Shotgun Proteomics: Exploiting a Reliable and Sensitive Method to Monitor Residual Host-Cell Proteins in Monoclonal Antibody Products. Khalil, Somar, et al. 2023, Journal of Pharmaceutical and Biomedical Analysis Open, Vol. 1, p. 100012.
11. Benchmarking Quantitative Performance in Label-Free Proteomics. Dowell, James A., et al. 4, 2021, ACS Omega, Vol. 6, pp. 2494-504.

12. Data-Independent Acquisition Mass Spectrometry (DIA-MS) for Proteomic Applications in Oncology. Krasny, Lukas, and Paul H. Huang. 1, 2021, Molecular Omics, Vol. 17, pp. 29-42.
13. Host Cell Protein Quantification Workflow Using Optimized Standards Combined with Data-Independent Acquisition Mass Spectrometry. Hessmann, Steve, et al. 5, 2023, Journal of Pharmaceutical Analysis, Vol. 13, pp. 494-502.
14. Host Cell Protein Profiling by Targeted and Untargeted Analysis of Data Independent Acquisition Mass Spectrometry Data with Parallel Reaction Monitoring Verification. Kreimer, Simion, et al. 10, 2017, Analytical Chemistry, Vol. 89, pp. 5294-302.
15. Evaluating the Performance of the Astral Mass Analyzer for Quantitative Proteomics Using Data-Independent Acquisition. Heil, Lilian R., et al. 10, 2023, Journal of Proteome Research, Vol. 22, pp. 3290-300.
16. Parallelized Acquisition of Orbitrap and Astral Analyzers Enables High-Throughput Quantitative Analysis. Stewart, Hamish I., et al. 42, 2023, Analytical Chemistry, Vol. 95, pp. 15656-64.
17. Narrow-Window DIA: Ultra-Fast Quantitative Analysis of Comprehensive Proteomes with High Sequencing Depth. Guzman, Ulises H., et al. 2023, Systems Biology.
18. Acquisition and Analysis of DIA-Based Proteomic Data: A Comprehensive Survey in 2023. Lou, Ronghui, and Wenqing Shui. 2, 2024, Molecular & Cellular Proteomics, Vol. 23, p. 100712.
19. Narrow Window Data-Independent Acquisition on the Orbitrap Astral Mass Spectrometer Enables Fast and Deep Coverage of the Plasma Glycoproteome. Jager, Shelley, et al. 2024, Biochemistry.
20. Data-Independent Acquisition: A Milestone and Prospect in Clinical Mass Spectrometry–Based Proteomics. Fröhlich, Klemens, et al. 2024, Molecular & Cellular Proteomics, p. 100800.
21. Quantifying the Impact of Chimera MS/MS Spectra on Peptide Identification in Large-Scale Proteomics Studies. Houel, Stephane, et al. 8, 2010, Journal of Proteome Research, Vol. 9, pp. 4152-60.
22. Assessment of False Discovery Rate Control in Tandem Mass Spectrometry Analysis Using Entrapment. Wen, Bo, et al. 7, 2025, Nature Methods, Vol. 22, pp. 1454-63.
23. Narrow Precursor Mass Range for DIA–MS Enhances Protein Identification and Quantification in Arabidopsis. Zhang, Huoming, and Dalila Bensaddek. 9, 2021, Life, Vol. 11, p. 982.
24. Correlation-Based Deconvolution (CorrDec) To Generate High-Quality MS2 Spectra from Data-Independent Acquisition in Multisample Studies. Tada, Ipputa, et al. 16, 2020, Analytical Chemistry, Vol. 92, pp. 11310-17.
25. Robust, Precise, and Deep Proteome Profiling Using a Small Mass Range and Narrow Window Data-Independent-Acquisition Scheme. Fröhlich, Klemens, Regula Furrer, et al. 2024, Journal of Proteome Research, Vol. 3, p. 23.
26. Ultra-Fast Label-Free Quantification and Comprehensive Proteome Coverage with Narrow-Window Data-Independent Acquisition. Guzman, Ulises H., et al. 12, 2024, Nature Biotechnology, Vol. 42, pp. 1855-66.
27. Improving Proteomic Identification Using Narrow Isolation Windows with Zeno SWATH Data-Independent Acquisition. Gu, Kongxin, et al. 8, 2024, Journal of Proteome Research, Vol. 23, pp. 3484-95.
28. Statistical Control of Peptide and Protein Error Rates in Large-Scale Targeted Data-Independent Acquisition Analyses. Rosenberger, George, et al. 9, 2017, Nature Methods, Vol. 14, pp. 921-27.
29. Target-Decoy Approach and False Discovery Rate: When Things May Go Wrong. Gupta, Nitin, et al. 7, 2011, Journal of the American Society for Mass Spectrometry, Vol. 22, pp. 1111-20.
30. Versatile and Multiplexed Mass Spectrometry-Based Absolute Quantification with Cell-Free-Synthesized Internal Standard Peptides. Masuda, Keiko, et al. 2022, Journal of Proteomics, Vol. 251, p. 104393.
31. Large-Scale Targeted Proteomics Using Internal Standard Triggered-Parallel Reaction Monitoring (IS-PRM). Gallien, Sebastien, et al. 6, 2015, Molecular & Cellular Proteomics, Vol. 14, pp. 1630-44.
32. Host Cell Protein Quantitation by LC-MS. Experimental Demonstration, Qualification, and Comparison of Methods in USP 1132.1. Chrone, Victor G., et al. 2025, Journal of Pharmaceutical and Biomedical Analysis, Vol. 265, p. 117051.
33. Whole Proteomes as Internal Standards in Quantitative Proteomics. Ong, Shao-En. 7, 2010, Genome Medicine, Vol. 2, p. 49.
34. Comparative Analysis of MS/MS Search Algorithms in Label-Free Shotgun Proteomics for Monitoring Host-Cell Proteins Using Trapped Ion Mobility and ddaPASEF. Khalil, Somar, and Michel Plisnier. 2025, Journal of Pharmaceutical and Biomedical Analysis Open, p. 100082.
35. A Practical Approach to Estimate Analytical Method Variability from Routine Testing. Bonifas, Andrew P., and Yi Li. 2024, Journal of Pharmaceutical and Biomedical Analysis, Vol. 249, p. 116344.
36. Guidelines and Considerations for the Use of System Suitability and Quality Control Samples in Mass Spectrometry Assays Applied in Untargeted Clinical Metabolomic Studies. Broadhurst, David, et al. 6, 2018, Metabolomics, Vol. 14, p. 72.
37. Interpretation of Shotgun Proteomic Data. Nesvizhskii, Alexey I., and Ruedi Aebersold. 10, 2005, Molecular & Cellular Proteomics, Vol. 4, pp. 1419-40.
38. IDPicker 2.0: Improved Protein Assembly with High Discrimination Peptide Identification Filtering. Ma, Ze-Qiang, et al. 8, 2009, Journal of Proteome Research, Vol. 8, pp. 3872-81.
39. Optimizing Spectronaut Search Parameters to Improve Data Quality with Minimal Proteome Coverage Reductions in DIA Analyses of Heterogeneous Samples. Baker, Christa P., et al. 6, 2024, Journal of Proteome Research, Vol. 23, pp. 1926-36.
40. The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. Hoffman, Matthew D., and Andrew Gelman. 2011, arXiv.
41. Hahn, Gerald J., and William Q. Meeker. Statistical Intervals: A Guide for Practitioners. s.l. : Wiley, 1991.
42. Analysis of DIA Proteomics Data Using MSFragger-DIA and FragPipe Computational Platform. Yu, Fengchao, et al. 1, 2023, Nature Communications, Vol. 14, p. 4154.
43. Dual Data-Independent Acquisition Approach Combining Global HCP Profiling and Absolute Quantification of Key Impurities during Bioprocess Development. Husson, Gauthier, et al. 2, 2018, Analytical Chemistry, Vol. 90, pp. 1241-47.
44. Benchmarking of Analysis Strategies for Data-Independent Acquisition Proteomics Using a Large-Scale Dataset Comprising Inter-Patient Heterogeneity. Fröhlich, Klemens, et al. 1, 2022, Nature Communications, Vol. 13, p. 2622.
45. An Inflection Point in High-Throughput Proteomics with Orbitrap Astral: Analysis of Biofluids, Cells, and Tissues. Hendricks, Nathan G., et al. 9, 2024, Journal of Proteome Research, Vol. 23, pp. 4163-69.
46. Target-Decoy Search Strategy for Increased Confidence in Large-Scale Protein Identifications by Mass Spectrometry. Elias, Joshua E., and Steven P. Gygi. 3, 2007, Nature Methods, Vol. 4, pp. 207-14.
47. Semi-Supervised Learning for Peptide Identification from Shotgun Proteomics Datasets. Käll, Lukas, et al. 1, 2007, Nature Methods, Vol. 4, pp. 923-25.
48. False Discovery Rates in Spectral Identification. Jeong, Kyowon, et al. 16, 2012, BMC Bioinformatics, Vol. 13.
49. Chromatogram Libraries Improve Peptide Detection and Quantification by Data Independent Acquisition Mass

Spectrometry. Searle, Brian C., et al. 1, 2018, Nature Communications, Vol. 9, p. 5128.
50. Absolute Quantification of Proteins by LCMSE. Silva, Jeffrey C., et al. 1, 2006, Molecular & Cellular Proteomics, Vol. 5, pp. 144-56.
51. High-risk' Host Cell Proteins (HCPs): A Multi-company Collaborative View. Jones, Marisa, et al. 8, 2021, Biotechnology and Bioengineering, Vol. 118, pp. 2870-85.
52. CHO Cathepsin B Identified as the Protease Responsible for a Target Bispecific Antibody Fragmentation. Hu, Lixia, et al. 2022, Protein Expression and Purification, Vol. 199, p. 106144.
53. Cathepsin L Causes Proteolytic Cleavage of Chinese-Hamster-Ovary Cell Expressed Proteins During Processing and Storage: Identification, Characterization, and Mitigation. Luo, Haibin, et al. 1, 2019, Biotechnology Progress, Vol. 35, p. e2732.
54. Proteomic Analysis of Host Cell Protein Dynamics in the Culture Supernatants of Antibody-Producing CHO Cells. Park, Jin Hyoung, et al. 1, 2017, Scientific Reports, Vol. 7, p. 44246.
55. Experience with Host Cell Protein Impurities in Biopharmaceuticals. Vanderlaan, Martin, et al. 4, 2018, Biotechnology Progress, Vol. 34, pp. 828-37.
56. Clusterin Binds by a Multivalent Mechanism to the Fc and Fab Regions of IgG. Wilson, Mark R., and Simon B. Easterbrook-Smith. 3, 1992, Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology, Vol. 1159, pp. 319-26.
57. Identification and Tracking of Problematic Host Cell Proteins Removed by a Synthetic, Highly Functionalized Nonwoven Media in Downstream Bioprocessing of Monoclonal Antibodies. Gilgunn, S., et al. 2019, Journal of Chromatography A, Vol. 1595, pp. 28-38.
58. Effects of Lysosomal Biotherapeutic Recombinant Protein Expression on Cell Stress and Protease and General Host Cell Protein Release in C Hinese Hamster Ovary Cells. Migani, Damiano, et al. 3, 2017, Biotechnology Progress, Vol. 33, pp. 666-76.
59. Knockout of a Difficult-to-remove CHO Host Cell Protein, Lipoprotein Lipase, for Improved Polysorbate Stability in Monoclonal Antibody Formulations. Chiu, Josephine, et al. 5, 2017, Biotechnology and Bioengineering, Vol. 114, pp. 1006-15.
60. Rapid Polysorbate 80 Degradation by Liver Carboxylesterase in a Monoclonal Antibody Formulated Drug Substance at Early Stage Development. Zhang, Sisi, et al. 11, 2020, Journal of Pharmaceutical Sciences, Vol. 109, pp. 3300-07.
61. Mass Spectrometric Evaluation of Upstream and Downstream Process Influences on Host Cell Protein Patterns in Biopharmaceutical Products. Falkenberg, Heiner, et al. 3, 2019, Biotechnology Progress, Vol. 35, p. e2788.
62. Specific Immune Response to Phospholipase B-Like 2 Protein, a Host Cell Impurity in Lebrikizumab Clinical Material. Fischer, Saloumeh Kadkhodayan, et al. 1, 2017, The AAPS Journal, Vol. 19, pp. 254-63.
63. Identification and Characterization of Co-Purifying CHO Host Cell Proteins in Monoclonal Antibody Purification Process. Liu, Xinrong, et al. 2019, Journal of Pharmaceutical and Biomedical Analysis, Vol. 174, pp. 500-08.
64. Evaluating Immunogenicity Risk Due to Host Cell Protein Impurities in Antibody-Based Biotherapeutics. Jawa, Vibha, et al. 6, 2016, The AAPS Journal, Vol. 18, pp. 1439-52.
65. Host Cell Protein Removal from Biopharmaceutical Preparations: Towards the Implementation of Quality by Design. Goey, Cher Hui, et al. 4, 2018, Biotechnology Advances, Vol. 36, pp. 1223-37.
66. Subcellular Site of the Biosynthesis ofO-Acetylated Sialic Acids in Bovine Submandibular Gland. Schauer, Roland, et al. 3, 1988, Glycoconjugate Journal, Vol. 5, pp. 257-70.
67. Carboxypeptidase D Is the Only Enzyme Responsible for Antibody C-terminal Lysine Cleavage in Chinese Hamster Ovary (CHO) Cells. Hu, Zhilan, et al. 10, 2016, Biotechnology and Bioengineering, Vol. 113, pp. 2100-06.
68. Proteomics in Biomanufacturing Control: Protein Dynamics of CHO-K1 Cells and Conditioned Media during Apoptosis and Necrosis. Albrecht, Simone, et al. 6, 2018, Biotechnology and Bioengineering, Vol. 115, pp. 1509-20.

Graphical abstract