proteomes

MDPI

Article

Prospective ICH Q2(R2)-Aligned Total-Error Validation of Label-Free Untargeted Proteomics for Host Cell Protein Quantification in Biotherapeutics

Somar Khalil *, Jean-François Dierick, Pascal Bourguignon and Michel Plisnier

Technical Research & Development, GSK, 1330 Rixensart, Belgium
* Correspondence: somar.x.khalil@gsk.com

Abstract

Background: Untargeted proteomics enables quantitative host cell protein (HCP) determination in biotherapeutics, yet no workflow has been validated under ICH Q2(R2) for regulated quality control. **Methods**: A prospective total-error (TE) validation of label-free ddaPASEF proteomics was performed. A stable isotope-labeled whole-proteome standard was spiked into NISTmAb at seven levels (20–80 ng) and analyzed in four independent assays (198 injections), supporting one-way random-effects ANOVA with Welch–Satterthwaite adjustment. Peptide-level identification error was evaluated by dual entrapment. **Results**: Empirical false-discovery proportions were below 1% at q = 0.01. Weighted least-squares regression (R^2 = 0.993) confirmed stable proportional compression with 81–85% recovery. Repeatability dominated the variance structure (median CV 2.7%); intermediate precision SD ranged from 0.69% to 3.81%. Both 95% β-expectation and 95/95 content tolerance intervals were contained within ±30% at all levels, defining a validated range of 20–80 ng. Abundance-stratified TE profiling revealed concentration-dependent calibration heterogeneity, with stratum-specific intervals within ±35% defining an abundance-aware LLOQ of 3.6 ppm (P95 = 3.87 ppm). Robustness under independent search software (FragPipe v24.0, CCC = 0.998) and cross-platform acquisition (Astral, CCC = 0.980) remained within ±30% limits. **Conclusions**: This constitutes the first prospective ICH Q2(R2)-aligned validation of untargeted proteomics for HCP quantification, with a transferable statistical framework for high-dimensional analytical methods.

Keywords: host cell proteins; label-free quantification; LC-MS/MS; ddaPASEF; ICH Q2(R2); total error; accuracy profile; analytical validation

Academic Editors: Jens R. Coorssen and Matthew P. Padula

Received: 16 March 2026
Revised: 18 April 2026
Accepted: 21 April 2026
Published: 23 April 2026

1. Introduction

Residual host cell proteins (HCPs) are a heterogeneous class of process-related impurities characterized by broad abundance distributions and protein-specific clearance kinetics during purification. Clinically relevant species with enzymatic or immunogenic activity may contribute minimally to total mass yet disproportionately influence risk [1–4]. Polyclonal enzyme-linked immunosorbent assays (ELISAs) generate antibody-dependent composite signals with incomplete and variable proteome coverage. A single ELISA readout is consequently poorly suited for mechanistic interpretation at the individual protein level, and orthogonal methods are now expected for complex biologics [5].

Liquid chromatography–tandem mass spectrometry (LC–MS/MS) enables direct identification and quantification of individual HCPs [6–9]. In purified monoclonal antibody

(mAb) matrices, optimized digestion and multidimensional separations can reach detection limits in the low-ppm range [10–15]. Trapped ion mobility with parallel accumulation–serial fragmentation (TIMS-PASEF) improves both sensitivity and selectivity for low-abundance impurities in product-dominated backgrounds by increasing precursor sampling density and mitigating co-elution in the gas phase [16–22]. Label-free quantification based on top-peptide summation allows estimation of total HCP burden without isotopic internal standards, with quantitative performance driven by peptide selection, ionization behavior, and data-processing algorithms [23–27].

Untargeted HCP proteomics still lacks precedent as a formally validated quality control release assay under International Council for Harmonisation (ICH) Q2(R2). Label-free data-dependent acquisition (DDA) introduces abundance-dependent precursor sampling stochasticity; protein-level estimates aggregate peptides with non-random missingness; and shared sequences create inference ambiguity governed by protein-grouping rules. Database search configuration and false discovery rate (FDR) thresholds modify identification depth and downstream quantitative stability in ways not addressed by classical univariate validation constructs [28–30]. These features define a high-dimensional measurement system whose uncertainty structure differs from that of single-analyte assays addressed by ICH guidance.

DIA acquisition modes eliminate stochastic precursor sampling but impose additional constraints that currently preclude direct regulatory deployment for untargeted HCP quantification. Library-dependent DIA operates within a closed analyte space requiring a pre-constructed and version-locked spectral library, which constitutes a functional component of the measurement system subject to its own revalidation upon modification. Library-free DIA removes this constraint but introduces a computational deconvolution layer without direct precursor-to-fragment traceability, and no 21 CFR Part 11-compliant software currently supports this workflow for GMP deployment. DDA was therefore selected as the acquisition mode compatible with the open-analyte, traceable identification chain required for a prospective regulatory validation.

ICH Q2(R2) defines validation characteristics including trueness, precision, linearity, and quantitation limit over a specified reporting range, and permits combined evaluation to support reliable routine performance [31]. Application to untargeted proteomics requires formal modeling of identification error and hierarchical variance components covering sample preparation, assay, analyst, and instrument platform. ICH Q14 further emphasizes lifecycle management, predefined acceptance criteria, and ongoing performance verification [32]. A partial regulatory analogue exists in new peak detection within multi-attribute methods, where untargeted signals are qualified as limit tests using predefined suitability criteria [33]. That framework, however, operates on a fixed product digest with an established reference profile and does not generalize to HCP proteomics, where the set of reportable proteins is database-derived and inference-driven.

The Current United States Pharmacopeia (USP) ⟨1132.1⟩ provides operational guidance for MS-based HCP analysis and describes three quantitative strategies (relative to product protein, spiked intact proteins, and stable isotope-labeled peptides) [34]. The chapter distinguishes product-specific ICH-aligned method validation from broader system qualification but does not define the measurand for untargeted outputs in metrological terms, nor establish a statistical framework linking database composition, FDR control, and protein grouping to validated performance characteristics. Quantitation limits remain defined under ICH Q2(R1) terminology and are not extended to integrated bias–variance evaluation under Q2(R2). The chapter itself acknowledges that MS-based quantitation accuracy is "often reported as plus/minus two-fold" and presents stochastic non-detection of lower-abundance HCPs as an intrinsic measurement property without requiring formal

characterization of this variability. USP ⟨1132.1⟩ therefore codifies operational best practices without articulating a validation architecture connecting identification error, inference rules, and quantitative performance to predefined acceptance criteria under Q2(R2).

Recent work has demonstrated that the three USP ⟨1132.1⟩ strategies can satisfy ICH Q2(R2) criteria when applied to two predefined HCPs (Clusterin and Lipoprotein Lipase) spiked into a mAb matrix [35]. In that design, measurands are specified ex ante, and validation characteristics are estimated for fixed analytes under acceptance criteria permitting two-fold recovery (50–200%) and 30% CV, without variance-component decomposition or total-error (TE) integration of bias and precision. Protein inference is not evaluated because the analyte space is predefined. That framework does not extend to untargeted HCP proteomics, where protein identities are determined by database search and grouping rules at the point of analysis and where the reportable measurand aggregates hundreds of individually inferred protein-group quantities. Detectability, grouping stability, and quantitative behavior within the endogenous HCP population remain uncharacterized under any formal validation framework. Extrapolation from two spike-defined proteins to hundreds of database-derived species presumes quantitative homogeneity that has not been empirically demonstrated, and no mechanism within the targeted design accounts for the stochastic identification variability, abundance-dependent calibration heterogeneity, or inference instability inherent to untargeted workflows.

For untargeted HCP workflows, validation should be conducted at the measurement-system level: identification error must be controlled empirically, performance characterized over the abundance range, and protein groupings maintained under fixed database and inference parameters. Under ICH Q2(R2)/Q14, system-level validation integrates TE-based accuracy profiling as described by the Société Française des Sciences et Techniques Pharmaceutiques (SFSTP) harmonization initiative [36,37], combining predictive intervals with prespecified coverage probabilities and hierarchical variance-component estimation to define the reportable range, lower limit of quantitation (LLOQ), and system suitability test (SST) criteria linked to continued performance verification.

We present a prospective, hierarchical validation of label-free ddaPASEF proteomics for quantitative determination of HCPs in mAb matrices, aligned with ICH Q2(R2) and implemented within the TE approach [37]. The design uses an absolute nanogram load scale and a stable isotope-labeled whole-proteome standard as a compositionally representative calibrant [38]. Empirical false discovery proportion (FDP) control via dual entrapment, deterministic parsimony for stable protein-group assignment, and random-effects ANOVA with Welch–Satterthwaite adjustment are integrated to derive β-expectation and 95/95 content tolerance intervals (TIs) that combine trueness and precision into unified acceptance criteria. Abundance-stratified TE profiling defines an abundance-aware validated range and LLOQs. Cross-software and cross-platform comparisons evaluate the robustness of the integrated analytical–computational system [39]. All tested levels (20–80 ng) satisfied ±30% β-expectation containment, with an aggregate LLOQ of 20 ng and an abundance-aware LLOQ of 3.6 ppm (P95 = 3.87 ppm). To our knowledge, this is the first prospective ICH Q2(R2)-aligned validation of untargeted proteomics for HCP quantification. The statistical framework is transferable to other high-dimensional analytical methods requiring regulatory qualification.

2. Materials and Methods

2.1. Validation Design

Validation employed a prospective hierarchical design aligned with ICH Q2(R2) for quantitative impurity assays and adapted to a high-dimensional LC–MS/MS measurement

system comprising sample preparation, acquisition, database search, protein inference, and quantitative summarization (Figure 1; Table 1) [31,36,37].

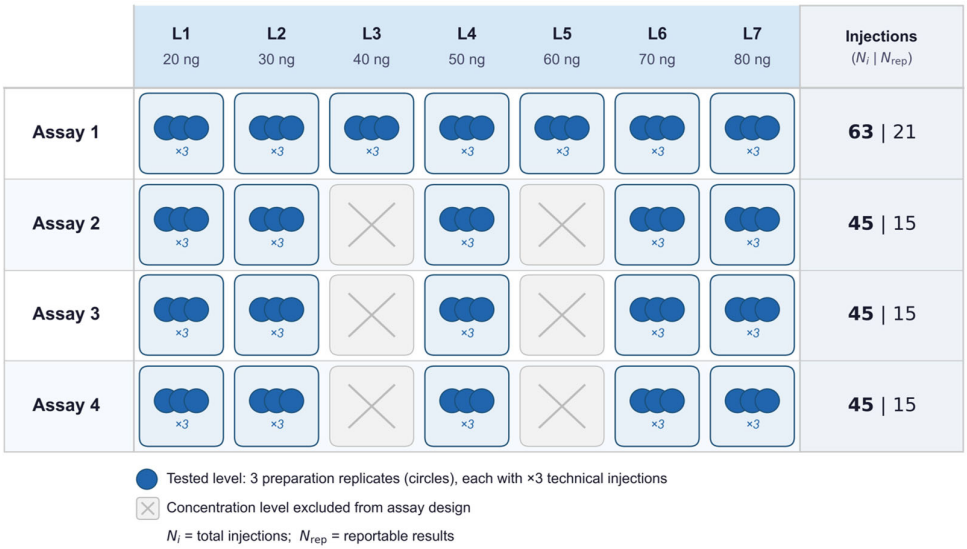

Figure 1. Hierarchical validation design for label-free ddaPASEF HCP quantification. Seven spike levels (20–80 ng) were evaluated across four independent assays. At each included level, three independent preparations were analyzed in technical triplicate (3 × 3). Grey tiles indicate levels not included in a given assay. The right column reports total injections (N_i) and replicate-block reportable results ($N_{(rep)}$), each defined as the arithmetic mean of three technical injections. Levels L3 (40 ng) and L5 (60 ng) were included in Assay 1 only, to support preliminary linearity characterization. Their exclusion from Assays 2–4 was defined a priori in the validation design to concentrate between-assay degrees of freedom at the remaining five levels. Precision and TE estimates at L3 and L5 therefore reflect within-assay repeatability only.

Table 1. Sources of analytical variation incorporated into the validation and robustness design.

Factor	Levels	Role
Assay run	4 independent assays	Between-assay variance
Analyst	3 analysts	Operational robustness (absorbed within-assay variance)
Independent preparation	3 per level per assay	Sample preparation variance
Technical injections	3 per independent preparation	Instrument repeatability
LC column	3 columns	Chromatographic robustness
LC–MS system (validated system)	Evosep One + timsTOF Pro	Primary validated measurement system
LC–MS system (robustness comparison)	Vanquish Neo + Orbitrap Astral [39]	Cross-platform robustness
Software pipeline (validated system)	SpectroMine	Primary processing workflow
Software pipeline (robustness comparison)	FragPipe	Algorithmic robustness

The validated measurand is the aggregate total HCP mass estimated by Hi3 label-free quantification and reported as nanograms of HCP per injection under the defined SIL-HCP calibration and deterministic protein-group inference rules. The measurement system generating this quantity comprised sample preparation, LC–MS/MS acquisition Evosep One (Evosep Biosystems, Odense, Denmark)–timsTOF Pro (Bruker Daltonics, Bremen, Germany) operated in ddaPASEF mode, database search with FDR control, deterministic protein inference, Hi3-based quantification, and post-processing required to compute the reportable total HCP value. All analytical and computational parameters were version-locked before the study began.

The design was defined a priori to support variance decomposition and TE estimation at the replicate-block level. Per spike level, replicate-block relative error was modeled using a one-way random-effects model with assay run as the grouping factor to estimate between-assay and within-assay components. The within-assay component aggregates variation from preparation, analyst, LC column, and residual effects under the locked configuration. Four independent assay runs were performed. Assay 1 included a complete seven-level spike series (20–80 ng). Assays 2–4 included five levels (L1, L2, L4, L6, L7). At each level, three independent preparations were analyzed in technical triplicate (3×3). The arithmetic mean of each triplicate defined the reportable result ($N_{(}rep_{)}$); total injections (N_i) denote all raw LC–MS/MS acquisitions. Levels present in all four assays yielded twelve replicate-block observations spanning assay runs, analysts, and LC columns.

Cross-platform robustness was evaluated by comparing total HCP values for a subset of samples generated on an Evosep One–timsTOF Pro system with those obtained on a Vanquish Neo–Orbitrap Astral (Thermo Fisher Scientific, Waltham, MA, USA) platform under matched spike levels and harmonized processing parameters [38]. Software robustness was assessed by reprocessing identical raw files using an independent search engine, FragPipe.

2.2. Analytical Procedure

The locked protocol is provided in Supplementary Method S1. NISTmAb (RM 8671, National Institute of Standards and Technology, Gaithersburg, MD, USA) served as the matrix. A stable isotope-labeled Chinese Hamster ovary (CHO) whole-proteome standard (SIL-HCP; SILu™CHOP, MSQC12, Sigma-Aldrich, St. Louis, MO, USA) was spiked at seven levels (20–80 ng total HCP injection load). Samples underwent denaturation, reduction, alkylation, and overnight digestion with trypsin/Lys-C (1:40, *w*/*w*; Pierce™ Trypsin/Lys-C Protease Mix, MS Grade, Cat. No. A41007, Thermo Fisher Scientific, Waltham, MA, USA), followed by C18 solid-phase extraction (Pierce™ Peptide Desalting Spin Columns, Cat. No. 89852, Thermo Fisher Scientific, Waltham, MA, USA). Four MassPREP protein digest standards (Waters Corporation, Milford, MA, USA) were added post-digestion for response-factor determination. Peptides (500 ng) were separated on an Evosep One (30 samples/day method) coupled to a timsTOF Pro operated in ddaPASEF mode. Database searching was performed in SpectroMine (v5.2; Biognosys AG, Schlieren, Switzerland) against the Cricetulus griseus UniProt reference proteome (UP000001075, accessed 2 February 2026) (trypsin/P specificity, $\leq$1 missed cleavages) with 1% FDR control at PSM, peptide, and protein levels.

Protein inference was performed using a deterministic parsimony algorithm retaining groups supported by $\geq$1 unique peptide; lead accessions were defined by maximal peptide evidence. The full algorithmic specification is provided in Supplementary Method S2; the inference script is archived at DOI: 10.5281/zenodo.18826281.

Peptide-level filtering included single-hit exclusion, modified Z-score outlier removal [40], intensity deviation screening, total ion current normalization, and replicate CV filtering. Protein abundance was estimated using the Hi3 approach [27]. Absolute HCP mass was calculated as

$$m_{HCP}(ng) = \left(\frac{I_{HCP}}{RF}\right) MW_{HCP} \times 10^{-6}, \quad (1)$$

where I_{HCP} is the summed Hi3 intensity of the protein and $RF = \text{median}_k\left(\frac{I_k}{n_k}\right)$ is the response factor derived from the MassPREP standards (I_k Hi3 intensity; n_k injected amount in fmol). MW_{HCP} is the molecular weight of the lead protein. Total HCP is defined as the sum of all inferred protein-group masses and constitutes the reportable value used for validation analyses [23,27,28].

2.3. Empirical Estimation of False Discovery Proportion

Peptide-level specificity was evaluated independently of target–decoy competition using a search-embedded entrapment strategy [41]. Two orthogonal entrapment spaces were constructed. The first comprised shuffled C. griseus tryptic peptides generated by permuting internal residues while preserving proteolytic termini. The second comprised a trimmed foreign-proteome space derived from in silico digestion of Arabidopsis thaliana and additional non-CHO proteomes, with peptides overlapping the CHO peptidome removed. Each entrapment database was constructed to approximate a 1:1 ratio (r) of unique entrapments to unique target peptides. Entrapment FASTAs were concatenated with the target FASTA and searched under identical SpectroMine parameters.

PSM-level results were collapsed to unique stripped peptide sequences by retaining the minimum q-value per sequence. For a peptide-level threshold τ, the empirical FDP was estimated as

$$FDP(\tau) = \frac{N_E(\tau) \times \left(1 + r^{-1}\right)}{N_D(\tau)}, \quad (2)$$

where $N_{E(\tau)}$ is the number of unique entrapment peptides at or below τ, and $N_{D(\tau)}$ is the total number of unique peptides (target + entrapment) at or below τ. For $r = 1$, the estimator reduces to $2N_E(\tau)/N_D(\tau)$.

FDP was evaluated over the reporting range of τ. Uncertainty was quantified by nonparametric bootstrap at the peptide level: stripped sequences were resampled with replacement, FDP recomputed per resample, and pointwise 95% percentile bands derived. At $\tau = 0.01$, uncertainty in the entrapment proportion N_E/N_D was additionally summarized using a two-sided 95% Wilson score interval and scaled by $(1 + r^{-1})$.

2.4. Statistical Analysis

The complete statistical analysis pipeline is archived at Zenodo (DOI: 10.5281/zenodo.18826281). Data processing and numerical operations used NumPy (v2.0.1) and pandas (v2.3.3). Statistical procedures, including random-effects ANOVA, weighted least-squares (WLS) regression, hierarchical bootstrap resampling, kernel density estimation (KDE), and tolerance interval construction, were implemented in SciPy (v1.17.1) and statsmodels (v0.14.6). Figures were generated in Matplotlib (v3.10.8). Bootstrap procedures employed a fixed NumPy random seed (0). Final figure layout refinement was performed in GraphPad Prism (v11.0; GraphPad Software, Boston, MA, USA).

2.4.1. Linearity

Linearity of total HCP quantification was evaluated by regressing measured total HCP (ng) on nominal spike amount over L1–L7 (20–80 ng) using replicate-block reportable results pooled across assays. The WLS model was specified as

$$Y_{ajk} = \beta_0 + \beta_1 \mu_j + \varepsilon_{ajk} \quad (3)$$

where Y_{ajk} denotes the replicate-block mean at assay a, spike level j, and replicate block k, and μ_j is the nominal spike amount. Weights were defined as

$$w_j = \frac{1}{\widehat{\mathrm{Var}}\left(Y \mid x = \mu_j\right)} \quad (4)$$

where $\widehat{\mathrm{Var}}(\cdot)$ is the empirical variance of replicate-block reportables at level j. Coefficient uncertainty was estimated using HC3 heteroscedasticity-consistent standard errors [42].

Linearity was evaluated by testing the null hypotheses $H_0 : \beta_0 = 0$ and $H_0 : \beta_1 = 1$. Model adequacy was assessed by examination of residual patterns and variance behavior over the concentration range.

2.4.2. Accuracy Profile

Accuracy of total HCP was evaluated on the relative-error scale within a TE framework aligned with SFSTP guidance [36]. Reportable results were defined at the replicate-block level as the mean total HCP (ng) across technical injections within each assay and spike level. Relative error was defined as

$$RE_{ajk} = 100 \times \frac{\left(Y_{ajk} - \mu_j\right)}{\mu_j} \tag{5}$$

where $Y_{\{ajk\}}$ is the replicate-block mean and μ_j the nominal spike amount at level j. Level-specific bias was estimated as

$$Bias_j = \text{mean}\left(RE_{ajk}\right). \tag{6}$$

Bias was modeled as a level-wise step function without parametric smoothing.

2.4.3. Variance Decomposition and Modeling

At each spike level, variance on the relative-error scale was estimated using one-way random-effects ANOVA with assay as a grouping factor. The within-assay variance component σ_W^2 (repeatability) was taken as MS_{within}. The between-assay variance component σ_B^2 was estimated by the method of moments using harmonic-mean replication m_h:

$$\sigma_B^2 = \max\left(\frac{(MS_{between} - MS_{within})}{m_h}, 0\right). \tag{7}$$

Total SD (intermediate precision under varied conditions) was

$$\sigma_T = \sqrt{(\sigma_W^2 + \sigma_B^2)}. \tag{8}$$

Level-specific total SDs were modeled by a log–log relationship:

$$\log(\sigma_{T,j}) = a + b \log\left(\mu_j\right), \tag{9}$$

fitted by weighted least squares with weights proportional to $n_j - 1$, where n_j is the number of replicate-block observations at level j. Predicted SDs $\widehat{\sigma_{T(pred,j)}}$ were used for TI construction.

2.4.4. Tolerance Interval Construction and Acceptance Criteria

The TE approach evaluates whether future routine measurements fall within predefined acceptability limits by integrating systematic bias and random variability into a single decision criterion. At each spike level, two TIs were estimated. The 95% β-expectation interval describes the range within which a future replicate-block relative error is expected to fall with probability β given the observed bias and variance structure. In parallel, a 95/95 content TI was estimated to provide 95% confidence that at least 95% of future errors were contained within the reported bounds. A level was considered validated when the β-expectation interval was fully contained within the predefined ±30% acceptance limits.

For each level,

$$TI_{j(\beta)} = Bias_j \pm K_{\beta,j}\, \widehat{\sigma}_{T(pred,j)}, \tag{10}$$

With $\beta = 0.95$. The coverage factor $K_{\beta,j}$ was derived from the Student t distribution using Welch–Satterthwaite effective degrees of freedom applied to the ANOVA mean-square representation of the total variance estimator: [43]

$$\widehat{\sigma}_T^2 = \widehat{\sigma}_w^2 + \widehat{\sigma}_B^2,\ \nu_{eff,j} = \frac{(c_W MS_W + c_B MS_B)^2}{\frac{(c_W MS_W)^2}{df_W} + \frac{(c_B MS_B)^2}{df_B}}, \tag{11}$$

where $c_W = 1 - 1/m_h$, $c_B = 1/m_h$, $df_W = N_j - A_j$, $df_B = A_j - 1$, A_j is the number of assays and N_j the number of replicate-block observations. Effective degrees of freedom were bounded below by 3. The coverage factor was

$$K_{\beta,j} = t_{\left(\frac{1+\beta}{2}, \nu_{eff,j}\right)}. \tag{12}$$

In parallel, 95/95 content TIs were estimated using a hierarchical cluster bootstrap with parametric simulation of future errors [44]. For $B = 4000$ iterations, assays were resampled with replacement; replicate blocks were resampled within assay at each level. Bias and variance models were refitted per bootstrap sample. Simulated future errors were generated as

$$e^*_{mj} = Bias^*_j + Z_{mj}\,\widehat{\sigma_{T(pred,j)}},\ Z_{mj} \sim \mathcal{N}(0,1). \tag{13}$$

At each level, the inner half-width was defined as the 95th percentile of $\left|e^*_{mj} - Bias^*_j\right|$ over simulated values. The 95/95 half-width $HW_{j(95/95)}$ was the 95th percentile of these inner half-widths across bootstrap replicates. The final interval was

$$TI_{j(95/95)} = Bias_j \pm HW_{j(95/95)}. \tag{14}$$

By construction,

$$\Pr_{boot}\left(\Pr\left(\left|E_j\right| \le HW_{j(95/95)} \mid \widehat{\theta^*}\right) \ge 0.95\right) \ge 0.95, \tag{15}$$

where E_j denotes a future replicate-block relative error and $\widehat{\theta^*}$ the bootstrap-refitted bias and variance parameters.

A level was considered validated when

$$TI_{j(\beta)} \subseteq [-30\%, +30\%]. \tag{16}$$

The ±30% aggregate acceptance limit was derived from prior performance characterization and reflects the combined uncertainty expected for a label-free untargeted LC–MS/MS workflow operating on a multi-protein calibrant across independent assay runs (Supplementary Note S4) [36,37,44]. This limit accommodates known sources of systematic compression in Hi3 quantification, including response-factor dispersion across the SIL-HCP proteome, incomplete digestion equivalence, and ionization-dependent biases, while preserving analytical relevance for aggregate HCP burden estimation [27,45]. The wider ±35% limit applied to abundance-stratified analysis accounts for the additional variance introduced by bootstrap-based stratum-level estimation and the smaller effective protein populations per stratum. Both limits were predefined prior to data analysis and align with TE acceptance criteria commonly applied in quantitative bioanalytical validation.

2.4.5. Abundance-Stratified Total-Error Analysis

Protein abundances were reported in ppm (ng HCP per mg mAb). Stratification was defined once using the L4 spike level. For each protein p, the L4 reference abundance $\overline{Y_{p,L4}}$

was computed as the mean ppm at L4 pooled over assays and technical replicates. Fixed strata (Q1–Q4) were assigned using empirical quantiles of $\left\{\overline{Y_{p,L4}}\right\}$ with cutpoints at the 5th, 25th, 50th, 75th, and 100th percentiles. The 5th-percentile lower bound limited leverage from extremely low-abundance proteins with unstable estimates. Stratum membership was fixed for all subsequent analyses.

For each assay a, spike level j, replicate block r, and stratum b, a stratum-level reportable was obtained by nonparametric bootstrap over proteins within cell (a, j, r, b). Let $\{Y_{pajrb}\}_{p=1}^{n_{ajrb}}$ denote ppm values in that cell. For B bootstrap iterations, proteins were resampled with replacement and the mean was computed. The reportable was the bootstrap expectation:

$$\widehat{\mu_{ajrb}} = \frac{1}{B}\sum_{t=1}^{B}\left(\frac{1}{n_{ajrb}}\sum_{p\in S_{ajrb}^{(t)}} Y_{pajrb}\right). \tag{17}$$

Cells with fewer than 30 proteins were excluded.

A stratum-specific normalization anchor was defined as the mean L4 reportable pooled across assays and replicate blocks:

$$\widehat{A_b} = \mathrm{mean}_{a,r}\left(\widehat{\mu_{a,L4,r,b}}\right). \tag{18}$$

Observed and theoretical spike ratios were

$$R_{\mathrm{obs},ajrb} = \frac{\widehat{\mu_{ajrb}}}{\widehat{A_b}}, \qquad R_{\mathrm{theo},j} = \frac{S_j}{S_{L4}}. \tag{19}$$

Relative error on the ratio scale was

$$RE_{ajrb} = 100 \times \frac{R_{\mathrm{obs},ajrb} - R_{\mathrm{theo},j}}{R_{\mathrm{theo},j}}. \tag{20}$$

Within each stratum, TE accuracy profiles were constructed on RE_{ajrb}. Level-specific bias was estimated as the empirical mean at each spike ratio. Variance was decomposed using one-way random-effects ANOVA with assay as grouping factor. Total SD was modeled by a log–log relationship in the theoretical spike ratio $x_{norm} = S_j/S_{L4}$.

The 95% β-expectation TI was

$$TI_{j,b}(\beta) = Bias_{j,b} \pm K_{\beta,j,b}\sigma_{T,j,b}, \tag{21}$$

where $K_{\beta,j,b} = t\left(\frac{1+\beta}{2}, \nu_{j,b}\right)$.

With $\beta = 0.95$ and $\nu_{j,b}$ obtained by Welch–Satterthwaite (minimum df = 3). A spike ratio was accepted within stratum b when

$$TI_{j,b}(\beta) \subseteq [-35\%, +35\%]. \tag{22}$$

The validated domain was defined as the intersection of acceptable spike ratios across strata. Abundance-aware quantification limits were derived from validated levels. The LLOQ was defined as the 95th percentile of stratum-level reportables at the lowest validated spike ratio within Q1. The ULOQ was defined as the 5th percentile of reportables at the highest validated spike ratio within Q4.

All TI calculations were performed on dimensionless ratios; ppm units are reported for interpretability only and do not affect the statistical evaluation.

2.5. System Suitability Testing

SST was incorporated into the analytical procedure to verify continued quantitative and identification performance during routine operation. SST evaluates performance at the lower boundary of the validated range and differentiates quantitative degradation from chromatographic or instrument-state perturbation.

The primary SST component comprised bracketed acquisitions of SIL-HCP at the LLOQ-proximal level (L1; 20 ng total HCP). One replicate-block reportable result was generated at sequence start and one at sequence end, each defined as the mean of three technical injections. Acceptance limits were derived from L1 validation performance. SIL-HCP recovery for each bracket was required to fall within 70–130% of nominal. Failure of the opening bracket precluded sequence initiation; failure of the closing bracket following a passing opening bracket triggered deviation investigation before results were released.

An orthogonal SST component employed a retention time calibration (RTC) peptide mixture injected at sequence start and end. Two metrics were monitored: median retention time deviation (ΔRT) relative to a locked deployment reference and median RTC MS1 signal intensity relative to the Phase I baseline median. Acceptance thresholds of $\pm$0.50 min (ΔRT) and $\pm$35% (MS1 intensity) were implemented during initial operational deployment and remain subject to refinement based on Phase I control-chart performance.

Longitudinal monitoring of run-level SST metrics was performed using Phase I Individuals–Moving Range (I–MR) control charts. Control-limit derivation and monitoring procedures are described in Supplementary Method S3. Quantitative SIL-HCP limits are directly traceable to the validated TE performance envelope.

3. Results

Validation outcomes for specificity, linearity, trueness, precision, range, and limits of quantification are presented under the predefined TE decision criteria.

3.1. Protein Inference and Identification Depth

Protein inference was performed using a deterministic greedy parsimony algorithm applied to the theoretical tryptic digest of the reference FASTA. Vendor-reported and parsimony-derived group sizes showed substantial concordance along the identity line (Figure 2A). Discordance was concentrated among vendor-defined singleton and small groups (<10 proteins), which expanded into modest multi-protein equivalence classes under parsimony rules. Agreement increased with group size, reflecting rule-dependent resolution of shared-peptide networks.

Parsimony inference yielded a broad distribution of group sizes (Figure 2B), extending to approximately 25–30 proteins per group. Group frequency decreased monotonically with increasing size. Singleton groups predominated, but multi-member equivalence classes constituted a structured fraction of the inferred proteome, consistent with non-unique peptide connectivity under theoretical digestion.

In Assay 1, unique peptides increased from 9534 (L1) to 15,979 (L7), and inferred lead proteins from 1655 to 2446 (Figure 2C,D). Assays 2–4 produced closely matched counts at shared levels. At 50 ng (L4), peptide identifications ranged from 13,563 to 14,047 and lead proteins from 2187 to 2297 (interassay CV < 2% and <3%, respectively). At 20 ng, counts spanned 9358–9410 peptides and 1654–1672 proteins.

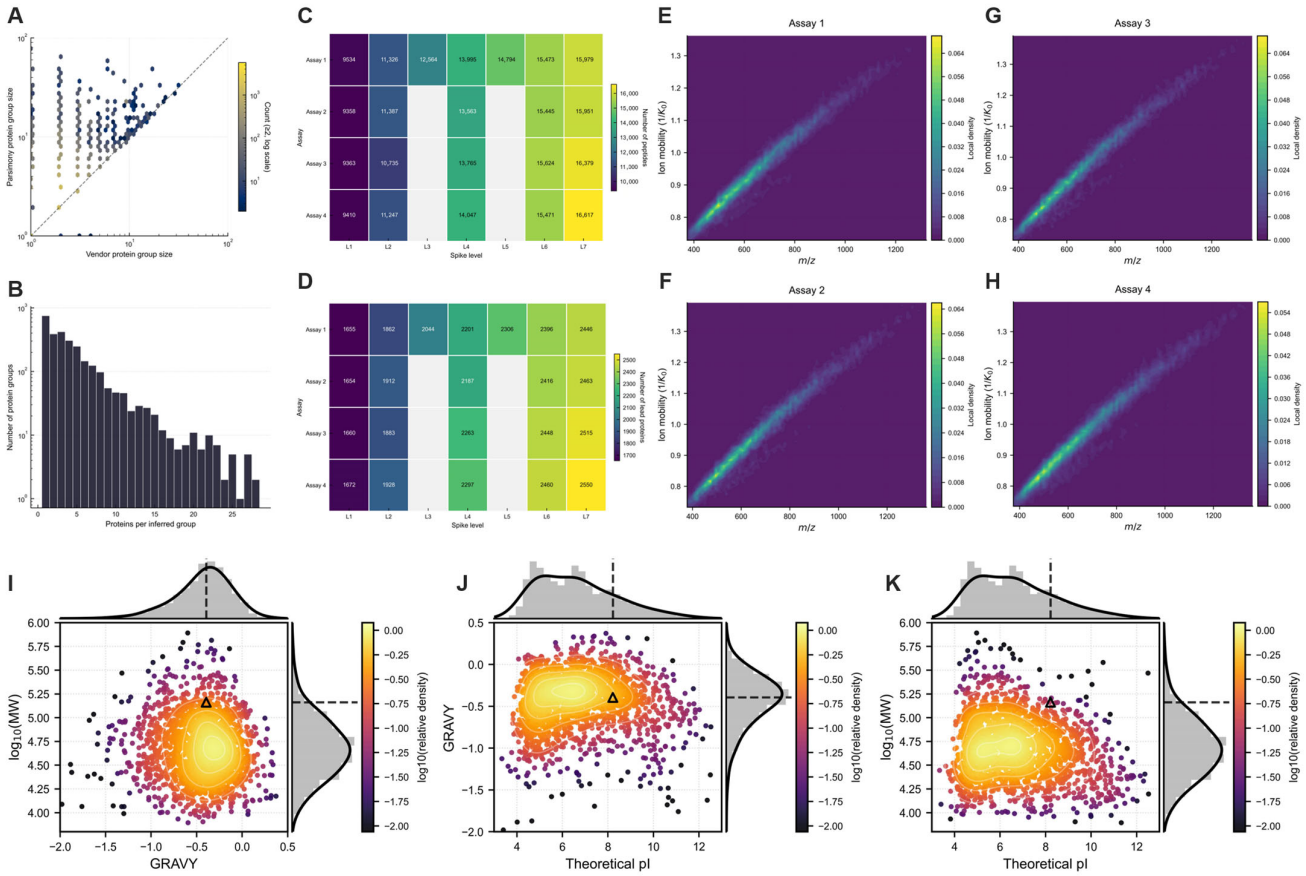

Figure 2. (**A**) Concordance between vendor-reported and parsimony-derived protein-group sizes (Hexbin, log–log scale; dashed line = identity). (**B**) Distribution of parsimony-derived group sizes. (**C**,**D**) Peptide and lead-protein counts across assays and spike levels. (**E**–**H**) Precursor occupancy in the m/z–1/K_0 plane for each assay, shown as density-weighted scatter plots. (**I**–**K**) Physicochemical distribution of inferred CHO proteins (GRAVY, pI, $\log_{10}$MW); the NISTmAb reference is indicated by an open triangle, and dashed lines denote the marginal medians of the CHO protein distribution on each axis.

Precursor-space reproducibility was examined in the m/z–1/K_0 plane (Figure 2E–H). All assays exhibited the characteristic diagonal density distribution defined by mass–collisional cross-section coupling. Kernel density maxima overlapped without measurable displacement, confirming invariant precursor occupancy across runs. Variation in identification depth was attributable to analyte load, not acquisition instability.

Physicochemical coverage of inferred CHO proteins was characterized by theoretical pI, GRAVY index, and $\log_{10}$(MW) (Figure 2I–K). pI values ranged from <4 to >12 with maximal density between 5 and 7. GRAVY values spanned −2.0 to +0.5 (centered ~−0.6). MWs extended from <10 kDa to >500 kDa with a modal density between 30 and 100 kDa. The NIST mAb occupied a peripheral region of this space. No systematic truncation of physicochemical classes was observed.

Spearman protein–protein correlation matrices were computed from $\log_{10}$-transformed intensities. Average-linkage clustering of $(1 - \rho)$ dissimilarities was applied to the 1250 proteins with the highest inter-sample variance at each spike level (Figure 3A). Silhouette maximization over k = 2–12 selected k = 2 at all levels; silhouette widths ranged from 0.237 to 0.285, below conventional thresholds for meaningful cluster separation. Cophenetic correlation coefficients ranged from 0.526 to 0.600 (Supplementary Table S1, Panel A) with no systematic dependence on HCP load. The binary partition lacks compositional coherence and does not define a clearly interpretable grouping.

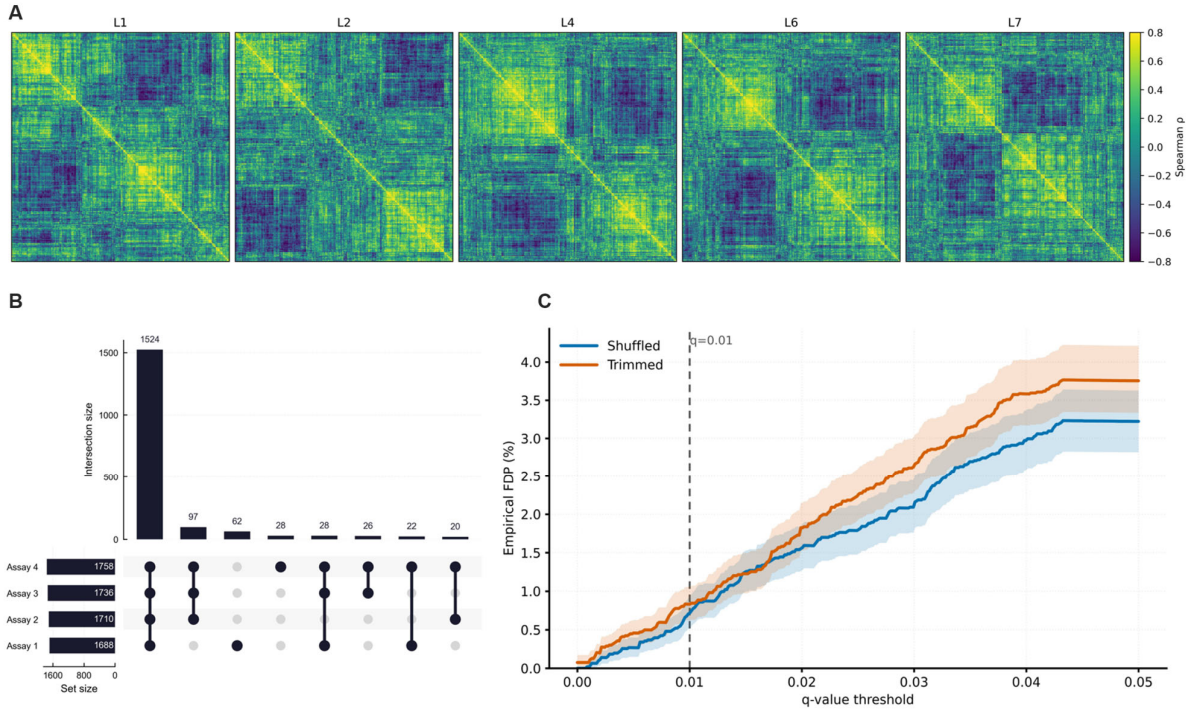

Figure 3. (**A**) Spearman protein–protein correlation matrices over common spike levels (L1, L2, L4, L6, L7), hierarchically clustered on a fixed correlation scale. (**B**) UpSet plot of protein identification overlap across the four assays. (**C**) Empirical peptide-level FDP as a function of q-value threshold (τ) estimated by search-embedded entrapment; shaded regions denote 95% bootstrap bands and the dashed line marks q = 0.01.

Matrix similarity was evaluated on 648 proteins common to all spike levels using Mantel permutation tests (Supplementary Table S1, Panel B). All ten pairwise comparisons rejected matrix independence ($p < 0.001$). Mantel r values were small (0.062–0.298) and decreased with increasing spike-level separation. The protein–protein correlation structure lacks stability across concentration levels and does not justify structured covariance modeling. The absence of a discrete covariance structure supports the exchangeability assumption underlying the one-way random-effects restricted maximum likelihood (REML) variance decomposition. More complex covariance specifications, including cluster-indexed random-effects or block-diagonal residual covariance models, were not required for this dataset.

Consistency of quantifiable proteins was assessed by UpSet analysis (Figure 3B). A total of 1524 proteins were quantified in all four assays (87–90% of assay-specific inventories: 1688–1758 proteins). Three-assay intersections comprised 97 and 62 proteins; assay-unique identifications were minimal.

3.2. Specificity and False Discovery Control

No SIL-HCP heavy peptides were detected in the unspiked mAb control (L0). Endogenous light CHO peptides do not constitute a blank failure mode because quantification is restricted to the isotopically resolved SIL-HCP population.

Peptide-level specificity was evaluated using search-embedded entrapment with concatenated target and entrapment databases constructed to approximate a 1:1 ratio. Empirical FDP curves with pointwise 95% bootstrap percentile bands are shown in Figure 3C. FDP increased as the reported q-value threshold was relaxed. At the operating threshold (q = 0.01), empirical FDP was below 1% for both entrapment constructions: 0.7–0.8% (shuffled) and 0.85–0.9% (trimmed).

The trimmed entrapment space produced marginally higher FDP at stringent thresholds, consistent with increased score competitiveness from preserved amino acid composition. At q = 0.01, both constructions converged to comparable empirical control. Bootstrap-derived FDP provides the primary calibration of identification error. The two-sided 95% Wilson score interval for the entrapment proportion at q = 0.01 (Supplementary Table S2) was consistent with binomial sampling variability.

Hi3 quantification requires three concordant peptides per protein group. Assuming approximate independence and peptide-level FDP ≈ 0.009, the probability of three false-positive peptides mapping to the same protein group is on the order of 10^{-7}. Protein-level false identification is therefore negligible relative to other variance components in the validation; a formal derivation is provided in Supplementary Note S1.

Blank-matrix evaluation excludes spike carryover and matrix-driven false positives. Entrapment analysis gives an empirical upper bound on peptide-level identification error, independent of target–decoy competition. Specificity at the reporting threshold is therefore supported.

3.3. Linearity

WLS regression yielded the fitted relationship $y = 1.25 + 0.798x, R^2 = 0.9928$ (Figure 4A; Table 2). The slope was below unity (HC3 95% CI [0.79, 0.81]; $p < 0.001$), corresponding to proportional compression of approximately 20% over 20–80 ng. The intercept was positive (HC3 95% CI [0.83, 1.67] ng; $p < 0.001$), introducing an additive offset whose relative contribution decreased with concentration (~7% at 20 ng; <2% at 80 ng). The implied relative bias, $\frac{y-x}{x} = \frac{1.25}{x} + (0.798 - 1)$, approaches −20.2% as $x \to \infty$, with partial attenuation at lower concentrations due to the intercept term.

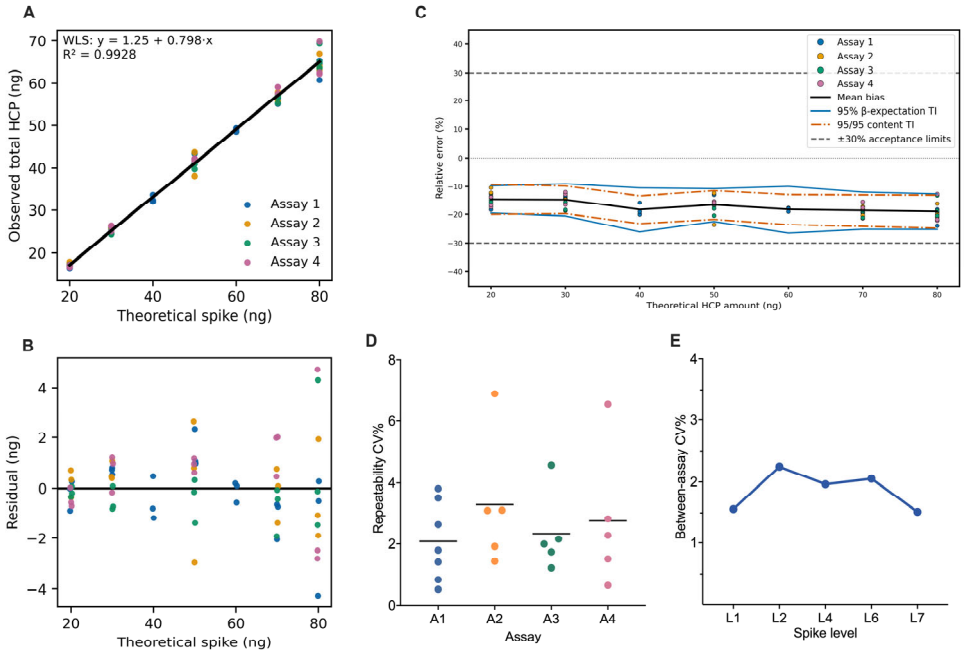

Figure 4. (**A**) WLS regression of observed versus nominal HCP amounts (20–80 ng). (**B**) Regression residuals across spike levels. (**C**) Accuracy profile on the relative-error scale showing mean bias (solid black line), 95% β-expectation tolerance limits (solid blue lines), and 95/95 content tolerance limits (vermillion dot-dash lines); dashed grey horizontal lines denote ±30% acceptance criteria. (**D**) Within-assay repeatability (CV%) by spike level. (**E**) Between-assay dispersion (CV%) by spike level.

Table 2. Linearity summary for total HCP quantification.

Metric	Value
Intercept (ng)	1.25
Intercept SE (HC3)	0.21
Intercept 95% CI (HC3)	[0.83–1.67]
Slope	0.7978
Slope SE (HC3)	0.01
Slope 95% CI (HC3)	[0.79–0.81]
R^2	0.9928
P (Intercept = 0)	<0.001
P (Slope = 1)	<0.001

Residuals were centered without curvature over the tested range (Figure 4B). Absolute dispersion increased with concentration, consistent with intensity-dependent variance in label-free quantification. Inverse-variance weighting stabilized residual variance. Replicate-block responses from all assays followed a common regression function without assay-specific displacement. Compression and offset therefore represent stable properties of the measurement system and were carried in subsequent trueness and TE analyses.

3.4. Accuracy

Trueness was evaluated at the replicate-block level as the signed deviation of the mean total HCP from the nominal spike amount (Table 3). Bias was negative across 20–80 ng, ranging from −14.67% at 20 ng to −19.01% at 80 ng, corresponding to mean recoveries of 85.33% to 80.99%. Bias magnitude increased between 20 and 40 ng and stabilized between −16.7% and −19.0% from 50 to 80 ng. The P05–P95 distribution of replicate-block relative errors showed consistent negative centering at all levels. The 95% confidence interval (CI) for recovery remained below 100% throughout the range.

Table 3. Trueness summary for total HCP quantification.

Level	Spike (ng)	Mean Result (ng)	|Bias| (ng)	Bias (%)	Mean Recovery (%)	P05–P95 Relative Error (%) [1]	95% CI of Recovery (%)
1	20	17.07	2.93	−14.67	85.33	[−17.88, −11.46]	[83.87, 86.79]
2	30	25.56	4.44	−14.79	85.21	[−18.59, −12.30]	[83.80, 86.61]
3	40	32.66	7.34	−18.35	81.65	[−19.94, −16.23]	[76.29, 87.02]
4	50	41.67	8.33	−16.66	83.34	[−21.91, −12.79]	[81.41, 85.26]
5	60	49.03	10.97	−18.29	81.71	[−18.97, −17.82]	[79.99, 83.42]
6	70	56.94	13.06	−18.66	81.34	[−21.26, −15.58]	[80.13, 82.54]
7	80	64.79	15.21	−19.01	80.99	[−22.99, −13.05]	[78.81, 83.17]

[1] The relative error is the pointwise error for one replicate-block reportable, and bias (%) is the systematic component at that spike level.

The bias trajectory aligns with the fitted calibration function. Relative bias becomes increasingly negative with concentration and approaches the asymptotic compression implied by the slope deficit. At 80 ng, the model-implied bias (−18.6%) closely matched the empirical estimate (−19.01%). The trueness profile therefore reflects stable proportional compression and was incorporated into the TE evaluation.

3.5. Precision

Precision was evaluated on the replicate-block relative-error scale using a one-way random-effects ANOVA with assay as the grouping factor (Table 4). Variance components were estimated using method-of-moments estimators. Repeatability (within-assay SD, $\widehat{\sigma_W}$) dominated at all spike levels and ranged from 0.69% (60 ng) to 3.81% (80 ng), with most values between 1% and 3%.

Table 4. Precision summary for total HCP quantification.

Level	Spike (ng)	Mean Result (ng)	Within-Assay SD (%)	Between-Assay SD (%)	Total SD (%)	Within-Assay df	Between-Assay df
1	20	17.07	2.29	0.14	2.30	8	3
2	30	25.56	1.61	1.67	2.32	8	3
3	40	32.66	2.16	0	2.16	2	0 [1]
4	50	41.67	3.09	0	3.09	8	3
5	60	49.03	0.69	0	0.69	2	0 [1]
6	70	56.94	1.35	1.48	2.00	8	3
7	80	64.79	3.81	0	3.81	8	3

[1] Variance components were estimated by one-way random-effects ANOVA using method-of-moments estimators. For levels present in a single assay (40 and 60 ng), between-assay variance was not estimable. Total SD corresponds to intermediate precision under ICH Q2(R2).

The between-assay component ($\widehat{\sigma_B}$) was small and was truncated to zero when $MS_{between} \leq MS_{within}$. Non-zero estimates were observed at 30 ng (1.67%) and 70 ng (1.48%), without systematic concentration dependence.

Total SD ($\widehat{\sigma_{T,j}} = \sqrt{\widehat{\sigma_W^2} + \widehat{\sigma_B^2}}$), corresponding to intermediate precision, ranged from 0.69% to 3.81%, with the largest values at 50 ng (3.09%) and 80 ng (3.81%) (Table 4). Over the validated range, within-assay dispersion drove most of the total variability.

Median repeatability CVs across assays ranged from 2.1% to 3.2% (Figure 4D). Between-assay CV ranged from 1.5% to 2.3% (Figure 4E).

The levels 40 ng and 60 ng were present in Assay 1 only; between-assay variance was therefore non-estimable (df = 0). Precision estimates at these levels reflect repeatability only and contribute to the variance model through the within-assay component.

For levels present in all assays (between-assay df = 3), the hierarchical design limits detection of small between-assay components. The minimum detectable between-assay SD at 80% power is approximately 1.41 $\widehat{\sigma_W}$ (Supplementary Note S2; Table S3). At median $\widehat{\sigma_W} \approx 2.3$, between-assay SD below ~3.2% is statistically indistinguishable from zero. Method-of-moments truncation to zero should therefore be interpreted as compatibility with small between-assay variance and not absence of such variance. REML estimates (Supplementary Table S4) yielded small positive components where truncation occurred, supporting robustness of the variance decomposition.

3.6. Total-Error Accuracy Profile

Accuracy was evaluated on the relative-error scale within the predefined TE approach. Across 20–80 ng, bias was negative at all levels (−14.67% to −19.01%) and exceeded the random component in magnitude. Model-predicted total SD ranged from 2.22% to 2.72%; systematic under-recovery dominated.

Level-specific 95% β-expectation TIs, incorporating Satterthwaite-adjusted effective degrees of freedom (3.0–10.78), ranged from [−19.57, −9.76]% at 20 ng to [−25.28, −12.74]% at 80 ng (Table 5; Figure 4C). All β-expectation TIs lay entirely below 0%. Interval width increased modestly with concentration, consistent with the log–log variance model.

Parallel 95/95 content TIs obtained by hierarchical cluster bootstrap were comparable in location and width, spanning [−20.03, −9.30]% at 20 ng to [−24.84, −13.18]% at 80 ng.

Both TI types were fully contained within the predefined ±30% acceptance limits at all spike levels. The most conservative case occurred at 80 ng, where the lower bound of the 95/95 content TI (−24.84%) remained 5.16 percentage points above the −30% criterion.

Table 5. Accuracy summary for total HCP quantification.

Level	Spike (ng)	Mean Result (ng)	Bias (%)	Pred. Total SD (%) [1]	Satterthwaite df	95 β-Expectation TI (%)	95/95 Content TI (%)
1	20	17.07	−14.67	2.22	10.78	[−19.57, −9.76]	[−20.03, −9.30]
2	30	25.56	−14.79	2.36	6.02	[−20.56, −9.03]	[−19.76, −9.83]
3	40	32.66	−18.35	2.46	3	[−26.17, −10.52]	[−23.26, −13.43]
4	50	41.67	−16.66	2.54	8	[−22.52, −10.81]	[−21.77, −11.56]
5	60	49.03	−18.29	2.61	3	[−26.59, −9.99]	[−23.63, −12.96]
6	70	56.94	−18.66	2.67	5.77	[−25.25, −12.07]	[−24.25, −13.08]
7	80	64.79	−19.01	2.72	8	[−25.28, −12.74]	[−24.84, −13.18]

[1] Predicted total SD denotes the smoothed estimate on the relative-error scale. Effective degrees of freedom were obtained via Welch–Satterthwaite adjustment.

3.7. Analytical Range and LLOQ

The analytical range was defined using the TE criterion requiring the 95% β-expectation TI for replicate-block relative error to lie within ±30%. This condition was satisfied at all tested levels from 20 to 80 ng (Table 5; Figure 4C). At 80 ng, the lower β-expectation bound (−25.28%) remained 4.72 percentage points within the −30% acceptance limit. The validated analytical range is therefore 20–80 ng total HCP per injection.

Following ICH Q2(R2) Section 6, the LLOQ was defined as the lowest experimentally tested level meeting the β-expectation criterion: 20 ng. The corresponding 95/95 content TI provided confirmatory coverage. Levels below 20 ng were not evaluated, and no extrapolation beyond the validated domain was performed.

A signal-to-noise-based LOD is not applicable to this estimator. Total HCP is derived from aggregated protein-group signals under a calibration-compression model; detectability is governed by population-level quantitative performance rather than single-analyte intensity. Reporting is restricted to the validated TE range.

3.8. Abundance-Stratified Total-Error Performance

Total HCP is an aggregate estimator over a heterogeneous abundance distribution. Aggregation reduces variance through averaging but can obscure abundance-dependent bias. TE performance was therefore re-estimated within fixed abundance strata (Q1–Q4) defined at L4. Bootstrap KDEs were well separated by spike level within each stratum (Figure 5A); with 10,000 resamples per cell, bootstrap variability was negligible relative to between-level separation.

Stratified WLS regressions (Table 6; Figure 5B–E) demonstrated abundance-dependent calibration behavior. Slopes decreased with abundance rank: Q1 = 1.06 (95% CI [1.02, 1.10]), Q2 = 0.99 [0.95, 1.04], Q3 = 0.90 [0.89, 0.92], Q4 = 0.80 [0.78, 0.81]. Low-abundance proteins exhibited mild expansion, mid-abundance proteins approximated proportionality, and higher-abundance strata exhibited compression. Intercepts increased with abundance rank (0.26 → 29.57 ppm), introducing positive offsets at low expected abundance. Residual diagnostics showed no curvature (Figure 5F–I).

Trueness patterns followed from these calibration structures (Supplementary Table S6). Q1 and Q2 displayed positive bias at most levels, whereas Q3 and Q4 transitioned from positive bias at low spike ratios to negative bias at higher ratios. In Q4, bias ranged from +23.1% (L1) to −10.9% (L7), consistent with slope compression below unity combined with positive intercept offset.

Stratified bias is defined relative to the L4 anchor. Positive bias in Q1 reflects proportional deviation from the L4 systematic component and does not imply absolute over-recovery relative to nominal concentration. Absolute bias equals the aggregate L4 bias plus the stratum-specific deviation (Supplementary Note S3).

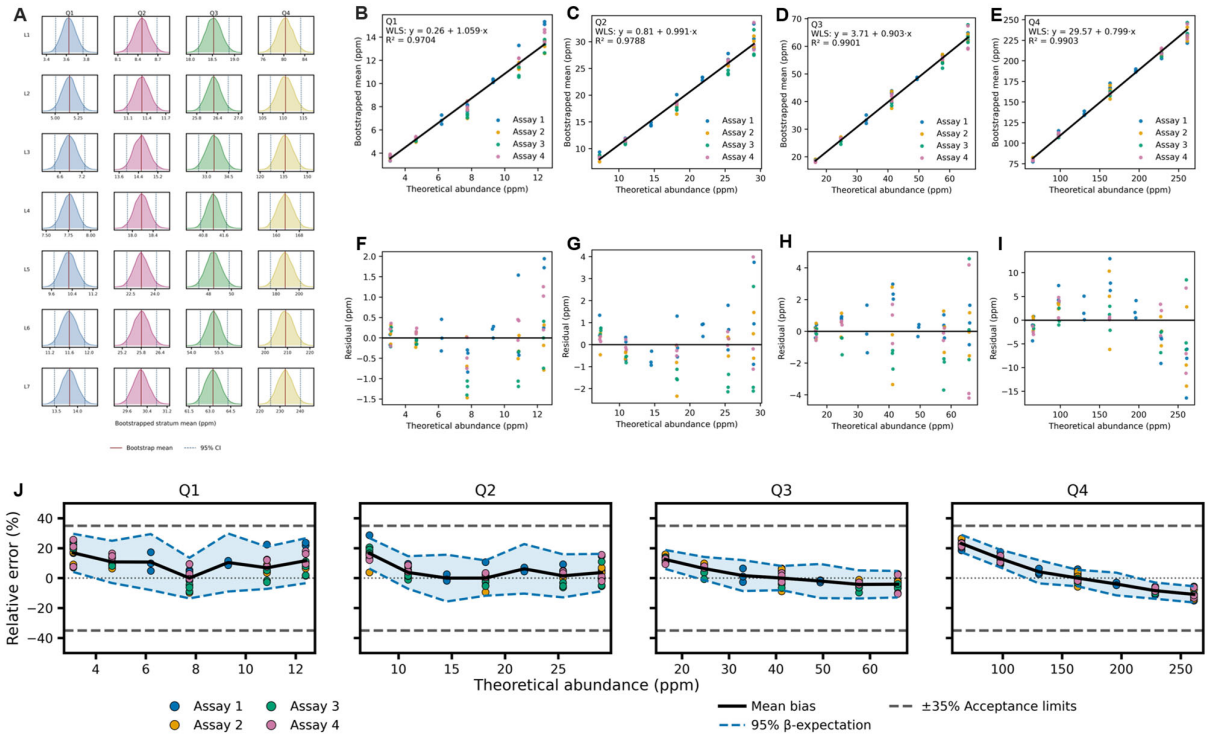

Figure 5. (**A**) Bootstrap density distributions of stratum-level mean abundances (ppm; 10,000 resamples) across spike levels for Q1–Q4; solid lines denote bootstrap means and dashed lines the 2.5th–97.5th percentiles. (**B**–**I**) Weighted least-squares calibration of bootstrapped mean abundance versus theoretical abundance for strata Q1–Q4; upper panels show fitted relationships and lower panels show residuals. (**J**) Stratified accuracy profiles showing relative error of observed-to-anchor response ratios versus theoretical spike ratios; solid lines indicate mean bias, and dashed lines denote 95% β-expectation TIs with ±35% acceptance criteria.

Table 6. Linearity summary for abundance-stratified HCP quantification.

Metric	Value Q1	Value Q2	Value Q3	Value Q4
Intercept (ppm)	0.26	0.81	3.71	29.57
Intercept SE (HC3)	0.09	0.27	0.18	0.92
Intercept 95% CI (HC3)	[0.07–0.44]	[0.27–1.36]	[3.35–4.07]	[27.73–31.41]
Slope	1.0588	0.9906	0.9029	0.7986
Slope SE (HC3)	0.02	0.02	0.01	0.01
Slope 95% CI (HC3)	[1.02–1.10]	[0.95–1.04]	[0.89–0.92]	[0.78–0.81]
R^2	0.9704	0.9788	0.9901	0.9903
p (Intercept = 0)	<0.01	<0.001	<0.001	<0.001
p (Slope = 1)	<0.001	<0.001	<0.001	<0.001

Variance decomposition (Supplementary Table S7) attributed most dispersion to repeatability. Within-assay SDs were largest in Q1 (up to 8.06%) and decreased with abundance rank. Between-assay components were smaller and frequently truncated to zero under the non-negativity constraint. Total SD ranged from 0.80% (Q3, L5) to 8.06% (Q1, L1). Elevated dispersion in Q1 reflects higher leverage of individual proteins within sparse strata.

Stratum-specific TE accuracy profiles (Supplementary Table S8; Figure 5J) were fully contained within ±35% acceptance limits at all spike levels. Bias drove most of the interval displacement. The widest β-expectation interval occurred in Q1 at L1 [6.0, 27.7]%, and the most negative in Q4 at L7 [−15.5, −6.3]%.

The abundance-aware validated range was defined as the intersection of acceptable spike levels across Q1–Q4. The lowest passing level in Q1 defined the abundance-aware

LLOQ (mean 3.62 ppm; conservative P95 = 3.87 ppm). The highest passing level in Q4 defined the abundance-aware ULOQ (mean 232.57 ppm; conservative P05 = 222.97 ppm).

The aggregate LLOQ (20 ng per injection) differs from the abundance-aware LLOQ (3.6 ppm; conservative P95 = 3.87 ppm). The former defines the minimum validated total impurity burden at the sample level, whereas the latter defines the lowest abundance class with demonstrated population-level quantitative performance. These limits address different measurement constructs and are not interchangeable (Supplementary Table S5).

REML estimates (Supplementary Table S9; Figure S1) yielded small positive between-assay components where method-of-moments truncation occurred, supporting the robustness of the variance decomposition.

The aggregate total HCP remains the validated reportable quantity for impurity burden assessment. Abundance-stratified analysis characterizes performance heterogeneity without redefining the primary release measurand.

3.9. Robustness

Robustness was evaluated by varying two analytical factors: data-processing workflow and MS platform. Total HCP (ng) was the reportable. Agreement was assessed using Deming regression (2000 bootstrap resamples for 95% CIs), Lin's concordance correlation coefficient (CCC), and Bland–Altman analysis on the relative-difference scale. Pairwise 95% limits of agreement (LoA) were evaluated against the predefined ±30% acceptance limits.

3.9.1. Software Comparison: SpectroMine vs. FragPipe

A subset of Assay 1 timsTOF Pro data (L1–L7; one preparation per level; technical triplicates retained) was reprocessed using FragPipe (v24.0; MSFragger v4.4.1; IonQuant v1.11.20). Database composition, FDR thresholds, and Hi3 quantification parameters were held constant. Deming regression of FragPipe versus SpectroMine yielded slope 0.959 (95% CI [0.943, 0.984]) and intercept +1.80 ng (Figure 6A; Table 7). Lin's CCC was 0.9982. Bland–Altman analysis showed mean relative bias +1.27% with 95% LoA [−5.8, +8.3]% (Figure 6B). All individual differences and corresponding LoA were within ±30%.

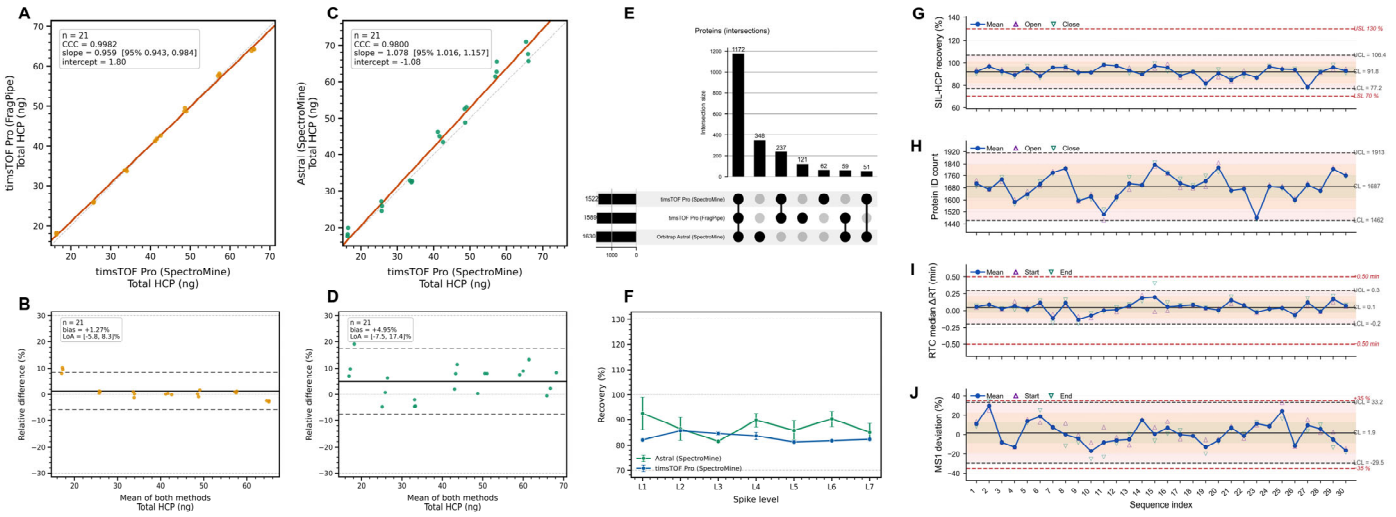

Figure 6. The software comparison (**A**,**B**) is based on a subset of Assay 1 (one independent preparation per level; technical triplicates retained). The platform comparison (**C**,**D**) is based on matched spike levels under harmonized processing parameters and constitutes a robustness assessment, not a full bridging validation; formal platform transfer requires demonstration of preserved variance structure under ICH Q14. (**A**) Deming regression of total HCP quantified by FragPipe versus SpectroMine on identical timsTOF Pro data; dashed line indicates identity. (**B**) Bland–Altman analysis for software

comparison. (**C**) Deming regression of total HCP between timsTOF Pro and Orbitrap Astral (SpectroMine processing). (**D**) Bland–Altman analysis for cross-platform comparison. (**E**) UpSet plot showing overlap of quantified proteins across analytical configurations. (**F**) Level-wise mean recovery (%) for timsTOF Pro and Orbitrap Astral; dashed lines denote 70% and 130% acceptance limits. (**G–J**) Phase I Individuals control charts for run-level SST metrics: SIL-HCP recovery, protein identification count, RTC median retention time deviation, and RTC MS1 intensity deviation; solid lines indicate run-level observations, red dashed lines indicate specification limits, and shaded bands denote ±1 SD process control zones from Phase I I–MR analysis. Corresponding moving-range charts showing short-term variability.

Table 7. Agreement statistics for robustness comparisons.

Category	Comparison	Deming Slope [95% CI]	Deming Intercept (ng)	Lin's CCC	BA Bias (%)	BA LoA (%)	Within ±30%
Software	SpectroMine vs. FragPipe	0.9592 [0.9433, 0.9843]	+1.805	0.9982	1.27	[−5.8, +8.33]	Yes
Platform	timsTOF Pro vs. Astral	1.0785 [1.0158, 1.1569]	−1.081	0.9800	4.95	[−7.51, +17.4]	Yes

3.9.2. Platform Comparison: timsTOF Pro vs. Orbitrap Astral

The timsTOF Pro and Orbitrap Astral were selected for platform comparison because they represent two architecturally different high-performance acquisition systems currently deployed in regulated biopharma analytical laboratories: trapped ion mobility separation with PASEF multiplexing on the one hand, and Orbitrap mass analysis with high-speed data-independent-compatible acquisition on the other. The comparison tests robustness over a meaningful analytical boundary (ion mobility versus non-mobility separation, and different precursor sampling mechanisms), and not between incremental generations of the same architecture. Other contemporary platforms (e.g., quadrupole time-of-flight and lower-resolution ion trap systems) were not included; extension to these platforms requires independent bridging evaluation under ICH Q14.

Total HCP values from timsTOF Pro and Orbitrap Astral were compared following identical SpectroMine processing and matched 500 ng injection load (L1–L7; 20–80 ng). Deming regression yielded slope 1.078 (95% CI [1.016, 1.157]) and intercept −1.08 ng (Figure 6C; Table 7). Lin's CCC was 0.9800. Bland–Altman analysis showed mean relative bias +4.95% with 95% LoA [−7.5, +17.4]% (Figure 6D). The upper LoA bound remained below the +30% acceptance limit. All individual differences were contained within ±30%.

Quantifiable protein overlap across SpectroMine (timsTOF Pro), FragPipe (timsTOF Pro), and SpectroMine (Orbitrap Astral) is shown in Figure 6E. A core set of 1172 proteins was shared across configurations, with smaller configuration-specific subsets.

Level-wise mean recovery (Figure 6F) ranged from 81 to 85% on timsTOF Pro and 85–99% on Orbitrap Astral. LoAs were narrower for software reprocessing than for platform comparison. Observed pairwise differences were of similar magnitude to the estimated intermediate precision from hierarchical validation.

Under ICH Q14 lifecycle principles, platform transfer requires bridging validation demonstrating preservation of variance structure, not solely aggregate agreement.

3.10. System Suitability Testing and Process Stability

SST performance was evaluated across 30 pilot analytical sequences using Phase I Individuals–Moving Range (I–MR) charts (Figure 6G–J; Supplementary Figure S2). Run-level SIL-HCP recovery remained within the 70–130% specification limits for all sequences and was centered near the baseline mean without sustained 3σ violations. Moving-range charts showed no persistent inflation of short-term variability.

Protein identification counts exhibited expected stochastic dispersion associated with DDA. Isolated downward excursions were observed but were not sustained and did not coincide with recovery failure or chromatographic deviation. Moving-range values showed intermittent spikes without evidence of process drift.

RTC median retention time deviation remained within the ±0.50 min specification limits and centered on the deployment reference. RTC MS1 deviation remained within the ±35% tolerance. A single elevated MS1 observation was not accompanied by concurrent recovery or ΔRT shifts and was not followed by sustained deviation. Moving-range charts indicated stable short-term variability for recovery and ΔRT, with intermittent variability in identification count and MS1 deviation consistent with operational noise.

No monitored metric showed sustained loss of statistical control or specification-limit exceedance.

4. Discussion

The working hypothesis of this study was that a label-free untargeted proteomic workflow, evaluated within the TE framework, could satisfy ICH Q2(R2) performance criteria for aggregate HCP quantification despite operating over an inference-dependent analyte space with database-derived protein identities and stochastic DDA sampling. The results confirm this hypothesis within defined boundary conditions. Label-free ddaPASEF proteomics satisfies prespecified performance criteria aligned with ICH Q2(R2) for quantitative HCP determination within 20–80 ng. To our knowledge, this is the first prospective validation of an untargeted proteomic workflow for HCP quantification conducted under ICH Q2(R2) guidelines. The approach validates the measurement system itself, covering identification, inference, and quantitative aggregation steps that jointly determine the reportable measurand, in contrast to targeted MS assays that define analytes a priori.

Systematic and random error components were modeled separately and incorporated into predictive TI constructions. All tested levels satisfied containment under both 95% β-expectation and 95/95 content TI definitions. The WLS calibration model ($\hat{y}$ = 1.25 + 0.798x, R^2 = 0.993) comprises a fixed additive offset contributing a bias term that decays as 1/x, and a proportional compression term (slope deficit ≈ 0.20) inducing concentration-independent multiplicative attenuation. HC3-robust inference confirmed deviation from $\beta_0 = 0$ and $\beta_1 = 1$; residual diagnostics supported first-order specification without curvature. The response structure was reproducible within and between four independent assays. The ±30% aggregate acceptance limit was derived from the USP ⟨1132⟩ immunoassay performance envelope and SFSTP β-expectation guidance [36,37].

The TE framework models trueness and precision jointly. The β-expectation tolerance interval is centered on the estimated bias and expanded by a precision-derived coverage factor applied to the total variance estimator. A structurally stable, proportional compression (WLS slope 0.798; HC3 95% CI [0.79, 0.81]; slope reproducible over four independent assays) contributes a concentration-independent multiplicative bias. This shifts the interval center but does not inflate the interval width beyond that of a zero-bias method with equivalent precision. TE containment holds because the compression is structurally stable, fully characterized by the calibration model, and incorporated into the interval center with negligible estimation uncertainty at the available degrees of freedom.

Stability was verified by replicate-block regression diagnostics (Figure 4B) and by slope reproducibility over the four-assay design. Precision was low relative to bias magnitude. Median repeatability was 2.7% CV; intermediate precision ranged from 1.5% to 2.3% CV. Between-assay variance was minor and frequently truncated to zero under the method-of-moments estimator. Two features explain this: the locked acquisition method yielded stable precursor-space occupancy and reproducible identification depth, and total

HCP, as a composite estimator, attenuates independent protein-level fluctuations upon summation in proportion to population size. With df = 3 for the between-assay component, between-assay SD below approximately 3.2% is statistically indistinguishable from zero (Supplementary Note S2). Zero estimates therefore reflect limited resolving power at the available degrees of freedom, not the absence of between-assay variability. Log–log variance modeling confirmed near-constant relative-error dispersion within the validated domain.

The method exhibits a stable negative bias of approximately −15% to −19%, with mean recoveries between 80.99% and 85.33%. This under-recovery arises from three multiplicative effects: incomplete proteolytic conversion for structurally resistant proteins at the fixed enzyme-to-substrate ratio; ionization suppression within the mAb-dominated matrix; and response-factor mismatch between the MassPREP Hi3 standards and CHO peptides, compounded by residual response-factor dispersion in the SIL-HCP calibrant proteome (Supplementary Table S11). The missed-cleavage distribution (95.81% zero, 4.19% one; Supplementary Table S10) falls within standard bottom-up benchmarks, implying that digestion efficiency is not the dominant contributor. Conservative peptide filtering introduces an additional bias–precision trade-off: exclusion of single-hit, high-variance, and incompletely detected peptides restricts Hi3 quantification to reproducible precursors per protein group, suppressing stochastic DDA sampling noise but decreasing the number of contributing proteins and attenuating recovered total HCP abundance. This attenuation is absorbed into the calibration model as part of the slope deficit and is evaluated jointly with all other systematic components within the TE construction. The ~20% aggregate compression falls within published Hi3 accuracy benchmarks in complex biological matrices, where individual protein-level deviations of 30–50% are common [27,45].

The between-assay CV of bias was <3% and systematic bias dominated the TE budget, exceeding total SD by 5–7-fold. Under ICH Q2(R2), zero bias is not required when systematic deviation is stable and incorporated into predictive limits. The 95% β-expectation and 95/95 content TIs incorporate the empirical bias at each level; at all validated levels the most conservative TI bound (−24.84% at 80 ng) remained more than five percentage points inside the −30% limit. The two constructions address different inferential targets: the former provides expected predictive coverage for a future replicate-block result; the latter provides confidence-guaranteed containment of at least 95% of future results. Concurrent containment under both criteria establishes predictive control under complementary definitions of coverage. Bias reduction would require correction factors derived from orthogonal quantification (Supplementary Notes S3, S5). By comparison, the targeted qualification reported by Chrone et al. [35] evaluated two predefined HCPs under acceptance criteria permitting 50–200% recovery and ≤30% CV without integrating bias and variance into a unified TE metric. The present validation achieves intermediate precision of 1.5–2.3% CV with systematic bias fully propagated into predictive TIs contained within ±30% at all levels, giving a tighter performance envelope for the aggregate measurand.

Peptide-level response-factor equivalence between the SIL-HCP calibrant and endogenous CHO HCPs has not been independently demonstrated. Because the validated measurand is the aggregate total HCP mass derived from Hi3 summation over all quantifiable protein groups, individual protein-level response-factor deviations are partially averaged, and residual calibration mismatch is absorbed into the empirical bias term propagated through the TI construction. Published evaluations of MS-based HCP workflows confirm that heterogeneous protein mixtures can provide reliable aggregate HCP estimates despite imperfect proteome matching, provided systematic bias is empirically characterized within the validation framework [28,38]. This validation applies specifically to this SIL-HCP lot and processing configuration; a change in calibrant source or composition would require reassessment under the revalidation trigger matrix (Supplementary Table S12). A bridg-

ing study between alternative SIL proteome preparations remains a necessary extension. Absolute values are therefore calibrant-relative estimates (see Limitations).

Aggregate TE performance masks abundance-dependent heterogeneity. Stratified WLS regression produced slopes decreasing with abundance rank (Q1 ≈ 1.06; Q2 ≈ 0.99; Q3 ≈ 0.90; Q4 ≈ 0.80), with intercepts increasing correspondingly. The aggregate metric, dominated by Q4 species, reflects high-abundance protein behavior; low-abundance strata exhibit different calibration profiles hidden in the summed result. L4 normalization forces zero bias at the anchor level by construction; stratified bias represents deviation from L4-specific systematic error, not from ground truth (Supplementary Note S3). All strata satisfied ±35% β-expectation containment simultaneously, defining abundance-aware LLOQ and ULOQ reported in ppm. The abundance-stratified LLOQ of 3.6 ppm (P95 = 3.87 ppm) adds resolution at the protein-population level. For individual high-risk protein tracking at sub-ppm concentrations, targeted LC–MS/MS methods with protein-specific internal standards remain the appropriate strategy.

Empirical entrapment analysis established peptide-level FDP below 1% at q = 0.01 under both shuffled and trimmed constructions, with bootstrap-derived FDP of 0.7–0.9% constituting the primary calibration evidence. Propagation to the protein level is attenuated geometrically by the Hi3 requirement of three concordant peptides: at FDP = 0.009, the expected protein-level false-identification rate scales as $FDP^3 \approx 10^{-7}$, negligible relative to other uncertainty components. Deterministic parsimony inference ensures denominator stability for longitudinal monitoring by preventing drift attributable to software-specific grouping heuristics.

Perturbation of the search engine (FragPipe) and MS platform (Orbitrap Astral) did not produce deviations exceeding ±30%. The wider platform limits of agreement reflect additional acquisition-layer variance from ion sampling architecture and duty cycle differences. Under ICH Q14, platform transfer is a lifecycle change and requires bridging validation showing preserved calibration behavior and variance structure, not only aggregate agreement. The two-component SST framework addresses the gap between one-time validation and continued performance verification. LLOQ-proximal bracket injections subject the system to its most demanding quantitative condition at each sequence, while orthogonal retention time monitoring isolates chromatographic and instrument-state perturbations. Revalidation triggers under ICH Q14 lifecycle management are defined prospectively and include sustained SST excursions, database or software version updates altering protein inference by >5%, and extension to matrices outside the validated domain (Supplementary Method S3; Table S12).

HCP risk depends on species-specific biological activity, not on total impurity mass [1–4]. Somatotropin (Omnitrope) clinical material initially quantified at approximately 20 ppm HCP by a generic immunoassay was subsequently shown to contain approximately 1400 ppm by a process-specific assay; the associated anti-therapeutic antibody rate reached 60%, and a reduction in the HCP level to below 50 ppm decreased this rate to below 3% [4]. A CHO-derived Factor IX trial was halted when 26% of subjects developed anti-HCP antibodies, and a Phase II CHO-derived oncology product trial was suspended owing to HCP-related adverse events [4]. Industry practice holds total HCP below 100 ppm [1–4]; species-specific thresholds may be lower. The 3.6 ppm abundance-aware LLOQ (P95 = 3.87 ppm) covers species whose safety thresholds exceed this value under typical mAb HCP control practice. For species with thresholds below 3.6 ppm, targeted LC–MS/MS methods with protein-specific stable isotope-labeled peptide standards remain the appropriate analytical strategy, providing analyte-specific sensitivity and traceability that aggregate workflows cannot provide.

Enrichment or depletion strategies extend the abundance range by reducing the product-dominated background, but the associated trade-offs are rarely formalized. Differential protein recovery during enrichment is matrix- and protocol-dependent, non-uniform over the HCP population, and covaries with physicochemical properties. When recovery is not determined per protein class and propagated into the uncertainty budget, reported concentrations reflect post-enrichment yield, not true burden. Enrichment bias therefore requires separate specificity and recovery studies and cannot be absorbed into an unenriched validation. Extension to enriched workflows is a separate validation exercise under ICH Q14 lifecycle principles [32]. Enrichment and depletion protocols are also modality-specific: strategies optimized for mAbs do not transfer to recombinant vaccines or other biotherapeutic formats. A platform method intended for use beyond mAbs therefore favors an unenriched global digestion workflow.

DIA acquisition eliminates DDA sampling stochasticity and improves per-protein quantitative consistency [39], but three constraints limit immediate regulatory deployment. First, library-dependent DIA operates within a closed analyte space contingent on prior observation, incompatible with the open analyte definition required for untargeted HCP monitoring. Under regulatory validation, the spectral library itself constitutes a functional component of the measurement system: its composition, construction parameters, and source data define the reportable analyte space and must be version-locked, qualified, and subject to revalidation upon modification, imposing a validation burden with no direct analogue in sequence-database-driven DDA workflows. Second, co-fragmentation of all precursors within each isolation window is inherent to all DIA modes and abolishes the direct precursor-to-fragment relationship preserved in DDA. In library-dependent DIA, the spectral library partially resolves this ambiguity through expected fragment patterns and retention time constraints; in library-free DIA, computational deconvolution must perform this reassignment without prior empirical reference, exposing an additional inference layer between raw data and reported protein identity that complicates regulatory traceability of the identification chain. Third, no 21 CFR Part 11-compliant software currently supports fully library-free DIA for GMP deployment. Library-free DIA algorithms remove both the analyte-space constraint and the library dependency and align with the measurement-system validation framework described here, but the available processing software ecosystem has not matured to support GMP implementation. The TE framework is transferable to DIA once these infrastructure and inferential constraints are resolved. Extension to non-CHO expression systems, alternative product formats, or alternative processing pipelines requires bridging validation.

In the defined applicability domain, the method satisfied all predefined regulatory criteria. The statistical complexity of the framework is confined to the validation stage. During routine operation, the procedure reduces to a locked sample preparation protocol, a fixed acquisition method, a version-controlled informatics pipeline, and evaluation of reportable results against predefined acceptance criteria. The SST bracket design provides binary pass/fail decisions at the point of use; the underlying distributional modeling is embedded in the acceptance limits and does not require re-execution.

Untargeted MS-based HCP analysis is unlikely to replace high-throughput immunoassays for routine lot-release screening in the near term. Its principal value lies in providing a platform-level quantitative method with explicit characterization of measurement uncertainty for the impurity population. The framework described here may extend to other analytical domains where analyte identity is determined by computational inference. Glycoproteomics workflows that assign glycopeptide compositions from combinatorial search spaces, lipidomics methods where lipid species are inferred from isobaric precursor fragmentation, and metaproteomics pipelines that resolve taxonomic assignments

from shared peptide evidence each face analogous challenges of inference-conditioned measurand definition and would benefit from system-level TE validation with empirical identification-error calibration.

The bottom-up proteomic workflow employed here infers canonical protein identities from tryptic peptide sequences and does not preserve intact proteoform structure. The protein identification space is bounded by the reference database; HCPs absent from the C. griseus UniProt proteome are not reportable regardless of abundance. Database version updates altering protein entries modify the reportable analyte space and constitute a revalidation trigger (Supplementary Method S3).

The validation was conducted using a SIL-HCP standard spiked into a purified mAb matrix (NISTmAb RM 8671) and not on a process-derived sample containing endogenous HCPs from an actual upstream manufacturing process. Process-derived matrices introduce HCP populations with variable modification states, product adsorption effects, and abundance distributions dependent on purification history; performance in such matrices requires bridging evaluation.

Post-translational modifications, splice variants, single-nucleotide polymorphisms, and other sources of primary-sequence diversity generate a proteoform landscape of considerably greater complexity than canonical accession-level inference captures. Because peptide-to-protein mapping is performed computationally following enzymatic digestion, the measurement system reports inferred protein-group abundances, not direct observations of full-length protein species. Proteoforms sharing identical tryptic peptides collapse into single protein groups, and their individual abundances are not resolved. The aggregate total HCP measurand is therefore a lower-bound estimate of proteoform-level impurity complexity.

Beyond proteoform resolution, additional constraints apply. The validation was performed on a single CHO-derived mAb matrix (NISTmAb RM 8671); generalizability to alternative host-cell expression systems, non-mAb product formats, or matrices with different HCP composition remains undemonstrated. The SIL-HCP calibrant was assumed to approximate aggregate CHO proteome composition, but formal demonstration of peptide-level response-factor equivalence between calibrant and endogenous HCP populations was not performed. The sample size comprised four independent assays, limiting the degrees of freedom available for between-assay variance estimation; between-assay SD below approximately 3.2% was statistically indistinguishable from zero at the available power. Precision and TE estimates at 40 ng and 60 ng, present in a single assay, reflect within-assay repeatability only. The validation was conducted within a single laboratory; multi-site reproducibility was not assessed.

5. Conclusions

The working hypothesis that a label-free untargeted proteomic workflow could satisfy ICH Q2(R2) performance criteria within the TE framework is confirmed under the boundary conditions defined in this study. Label-free ddaPASEF proteomics satisfies all predefined performance criteria for quantitative HCP determination within the validated range of 20–80 ng total HCP per injection. The 95% β-expectation and 95/95 content TIs were contained within ±30% at every tested level, with the most conservative bound (−24.84% at 80 ng) remaining more than five percentage points inside the acceptance limit. The aggregate LLOQ is 20 ng.

Abundance-stratified evaluation identified structured calibration heterogeneity concealed in the aggregate metric. Stratum-specific TIs were contained within ±35% at all spike levels, defining an abundance-aware LLOQ of 3.6 ppm (P95 = 3.87 ppm). The aggregate total HCP remains the validated reportable measurand.

Robustness under variation of search engine (CCC = 0.998) and MS platform (CCC = 0.980) was maintained within ±30% limits. Platform transfer constitutes a lifecycle change requiring bridging validation under ICH Q14.

The SST framework, comprising LLOQ-proximal bracket injections and orthogonal chromatographic monitoring, links validation evidence to continued routine performance verification.

The validated applicability domain is defined as CHO-derived mAb matrices quantified using SIL-HCP calibration on Evosep One/timsTOF Pro in ddaPASEF mode with version-locked data processing (SpectroMine, deterministic parsimony inference, Hi3 quantification). The TE framework and its statistical architecture are transferable to DIA-based workflows and to other high-dimensional analytical methods where analyte identity is determined by computational inference, pending resolution of infrastructure and regulatory traceability constraints. Extension to alternative expression systems, product formats, MS platforms, or informatics pipelines requires bridging validation.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/proteomes14020021/s1: Supplementary Method S1: Locked Analytical Protocol; Supplementary Method S2: Deterministic Parsimony Inference Algorithm; Supplementary Method S3: SST and Statistical Process Control; Supplementary Table S1: Cophenetic correlation coefficients by spike level; Supplementary Table S2: Entrapment fraction among accepted peptides at q = 0.01; Supplementary Table S3: Power analysis for between-assay variance detection; Supplementary Table S4: REML variance estimates (Aggregate); Supplementary Table S5: Dual LLOQ summary; Supplementary Table S6: Trueness summary for abundance-stratified HCP quantification; Supplementary Table S7: Precision summary for abundance-stratified HCP quantification; Supplementary Table S8: Accuracy summary for abundance-stratified HCP quantification; Supplementary Table S9: REML variance estimates (Stratified); Supplementary Table S10: Missed cleavage distribution; Supplementary Table S11: MassPREP response factor dispersion; Supplementary Table S12: Revalidation trigger matrix; Supplementary Table S13: SIL-HCP physicochemical coverage summary; Supplementary Figure S1: REML vs. Method-of-Moments Variance Comparison; Supplementary Figure S2: Moving Range charts for pilot SST metrics; Supplementary Figure S3: Distribution of missed cleavages under trypsin/P digestion; Supplementary Figure S4: MassPREP RF Distribution; Supplementary Figure S5: Coverage of SIL-HCP detected proteins vs. CHO proteome; Supplementary Note S1: Peptide-to-Protein FDP Propagation Bound; Supplementary Note S2: Sensitivity Analysis for Hierarchical Design; Supplementary Note S3: L4 Anchor Normalization and Absolute Bias Decomposition; Supplementary Note S4: Acceptance Limit Derivation; Supplementary Note S5: SIL-HCP Calibrant Representativeness Assessment.

Author Contributions: Conceptualization, S.K. and M.P.; methodology, S.K. and J.-F.D.; software, S.K.; validation, S.K. and J.-F.D.; formal analysis, S.K. and P.B.; investigation, S.K. and J.-F.D.; resources, P.B. and M.P.; data curation, S.K.; writing—original draft preparation, S.K.; writing—review and editing, J.-F.D., P.B. and M.P.; visualization, S.K.; supervision, M.P.; project administration, S.K. and M.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding. The APC was funded by GSK.

Informed Consent Statement: Not applicable.

Data Availability Statement: The complete statistical analysis pipeline (Python 3.12), including random-effects ANOVA, TI construction, abundance-stratified TE profiling, and entrapment FDP estimation, is archived at Zenodo (DOI: 10.5281/zenodo.18826281). The deterministic parsimony inference script and processed quantitative output tables, including peptide-level identifications, protein-group Hi3 abundances, replicate-block reportable results, and entrapment search outputs, are archived at the same repository. Raw LC–MS/MS acquisition files and search results have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with dataset identifier PXD075694 [46].

Acknowledgments: The authors gratefully acknowledge Nora Zaïm for her contributions to the laboratory work that supported this study.

Conflicts of Interest: S.K., J.-F.D., P.B., and M.P. are employees of GSK. The research presented here was conducted as part of analytical method development within GSK and was not influenced by commercial considerations. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

ANOVA	Analysis of variance
CCC	Lin's concordance correlation coefficient
CHO	Chinese hamster ovary
CI	Confidence interval
CV	Coefficient of variation
DDA	Data-dependent acquisition
ddaPASEF	Data-dependent acquisition with parallel accumulation–serial fragmentation
DIA	Data-independent acquisition
ELISA	Enzyme-linked immunosorbent assay
FDP	False discovery proportion
FDR	False discovery rate
GMP	Good manufacturing practice
HC3	Heteroscedasticity-consistent standard errors (type 3)
HCP	Host cell protein
Hi3	Top-three peptide intensity summation
ICH	International Council for Harmonisation
I–MR	Individuals–Moving Range
KDE	Kernel density estimation
LC–MS/MS	Liquid chromatography–tandem mass spectrometry
LLOQ	Lower limit of quantification
LoA	Limits of agreement
mAb	Monoclonal antibody
MW	Molecular weight
ppm	Parts per million (ng HCP per mg product protein)
PSM	Peptide-spectrum match
REML	Restricted maximum likelihood
RTC	Retention time calibration
SD	Standard deviation
SFSTP	Société Française des Sciences et Techniques Pharmaceutiques
SIL-HCP	Stable isotope-labeled host cell protein (standard)
SST	System suitability test
TE	Total error
TI	Tolerance interval
TIMS-PASEF	Trapped ion mobility spectrometry with parallel accumulation–serial fragmentation
ULOQ	Upper limit of quantification
USP	United States Pharmacopeia
WLS	Weighted least squares

References

1. Bracewell, D.G.; Francis, R.; Smales, C.M. The future of host cell protein (HCP) identification during process development and manufacturing linked to a risk-based management for their control. *Biotechnol. Bioeng.* **2015**, *112*, 1727–1737. [CrossRef]

2. Jones, M.; Palackal, N.; Wang, F.; Gaza-Bulseco, G.; Hurkmans, K.; Zhao, Y.; Chitikila, C.; Clavier, S.; Liu, S.; Menesale, E.; et al. "High-risk" host cell proteins (HCPs): A multi-company collaborative view. *Biotechnol. Bioeng.* **2021**, *118*, 2870–2885. [CrossRef] [PubMed]
3. Vanderlaan, M.; Zhu-Shimoni, J.; Lin, S.; Gunawan, F.; Waerner, T.; Van Cott, K.E. Experience with host cell protein impurities in biopharmaceuticals. *Biotechnol. Prog.* **2018**, *34*, 828–837. [CrossRef] [PubMed]
4. De Zafra, C.L.Z.; Quarmby, V.; Francissen, K.; Vanderlaan, M.; Zhu-Shimoni, J. Host cell proteins in biotechnology-derived products: A risk assessment framework. *Biotechnol. Bioeng.* **2015**, *112*, 2284–2291. [CrossRef]
5. Zhu-Shimoni, J.; Yu, C.; Nishihara, J.; Wong, R.M.; Gunawan, F.; Lin, M.; Krawitz, D.; Liu, P.; Sandoval, W.; Vanderlaan, M. Host cell protein testing by ELISAs and the use of orthogonal methods. *Biotechnol. Bioeng.* **2014**, *111*, 2367–2379. [CrossRef]
6. Molden, R.; Hu, M.; E, S.Y.; Saggese, D.; Reilly, J.; Mattila, J.; Qiu, H.; Chen, G.; Bak, H.; Li, N. Host cell protein profiling of commercial therapeutic protein drugs as a benchmark for monoclonal antibody-based therapeutic protein development. *mAbs* **2021**, *13*, 1955811. [CrossRef]
7. Hessmann, S.; Chery, C.; Sikora, A.-S.; Gervais, A.; Carapito, C. Host cell protein quantification workflow using optimized standards combined with data-independent acquisition mass spectrometry. *J. Pharm. Anal.* **2023**, *13*, 494–502. [CrossRef]
8. Guo, J.; Kufer, R.; Li, D.; Wohlrab, S.; Greenwood-Goodwin, M.; Yang, F. Technical advancement and practical considerations of LC-MS/MS-based methods for host cell protein identification and quantitation to support process development. *mAbs* **2023**, *15*, 2213365. [CrossRef]
9. Tank, P.; Vora, S.; D'Souza, F.; Kamble, S.S.; Rameshkumar, N. Advanced mass spectrometry techniques for monitoring biopharmaceutical host cell proteins. *Trends Biotechnol.* **2025**, *44*, 991–1004. [CrossRef]
10. Doneanu, C.E.; Anderson, M.; Williams, B.J.; Lauber, M.A.; Chakraborty, A.; Chen, W. Enhanced Detection of Low-Abundance Host Cell Protein Impurities in High-Purity Monoclonal Antibodies Down to 1 ppm Using Ion Mobility Mass Spectrometry Coupled with Multidimensional Liquid Chromatography. *Anal. Chem.* **2015**, *87*, 10283–10291. [CrossRef]
11. Khalil, S.; Wychowski, A.; Chéry, C.; Gervais, A. Label-free shotgun proteomics: Exploiting a reliable and sensitive method to monitor residual host-cell proteins in monoclonal antibody products. *J. Pharm. Biomed. Anal. Open* **2023**, *1*, 100012. [CrossRef]
12. Huang, L.; Wang, N.; Mitchell, C.E.; Brownlee, T.; Maple, S.R.; De Felippis, M.R. A Novel Sample Preparation for Shotgun Proteomics Characterization of HCPs in Antibodies. *Anal. Chem.* **2017**, *89*, 5436–5444. [CrossRef] [PubMed]
13. Chen, I.-H.; Xiao, H.; Daly, T.; Li, N. Improved Host Cell Protein Analysis in Monoclonal Antibody Products through Molecular Weight Cutoff Enrichment. *Anal. Chem.* **2020**, *92*, 3751–3757. [CrossRef] [PubMed]
14. Nie, S.; Greer, T.; O'Brien Johnson, R.; Zheng, X.; Torri, A.; Li, N. Simple and Sensitive Method for Deep Profiling of Host Cell Proteins in Therapeutic Antibodies by Combining Ultra-Low Trypsin Concentration Digestion, Long Chromatographic Gradients, and BoxCar Mass Spectrometry Acquisition. *Anal. Chem.* **2021**, *93*, 4383–4390. [CrossRef]
15. Chen, I.-H.; Xiao, H.; Li, N. Improved host cell protein analysis in monoclonal antibody products through ProteoMiner. *Anal. Biochem.* **2020**, *610*, 113972. [CrossRef]
16. Meier, F.; Park, M.A.; Mann, M. Trapped Ion Mobility Spectrometry and Parallel Accumulation–Serial Fragmentation in Proteomics. *Mol. Cell. Proteom.* **2021**, *20*, 100138. [CrossRef]
17. Steigenberger, B.; Van Den Toorn, H.W.P.; Bijl, E.; Greisch, J.-F.; Räther, O.; Lubeck, M.; Pieters, R.J.; Heck, A.J.R.; Scheltema, R.A. Benefits of Collisional Cross Section Assisted Precursor Selection (caps-PASEF) for Cross-linking Mass Spectrometry. *Mol. Cell. Proteom.* **2020**, *19*, 1677–1687. [CrossRef]
18. Tsukidate, T.; Stiving, A.Q.; Rivera, S.; Sahoo, A.; Madabhushi, S.; Li, X. diaPASEF Enables High-Throughput Proteomic Analysis of Host Cell Proteins for Biopharmaceutical Process Development. *Anal. Chem.* **2024**, *96*, 12999–13006. [CrossRef]
19. Meier, F.; Brunner, A.-D.; Koch, S.; Koch, H.; Lubeck, M.; Krause, M.; Goedecke, N.; Decker, J.; Kosinski, T.; Park, M.A.; et al. Online Parallel Accumulation–Serial Fragmentation (PASEF) with a Novel Trapped Ion Mobility Mass Spectrometer. *Mol. Cell. Proteom.* **2018**, *17*, 2534–2545. [CrossRef]
20. Meier, F.; Beck, S.; Grassl, N.; Lubeck, M.; Park, M.A.; Raether, O.; Mann, M. Parallel Accumulation–Serial Fragmentation (PASEF): Multiplying Sequencing Speed and Sensitivity by Synchronized Scans in a Trapped Ion Mobility Device. *J. Proteome Res.* **2015**, *14*, 5378–5387. [CrossRef]
21. Charkow, J.; Röst, H.L. Trapped Ion Mobility Spectrometry Reduces Spectral Complexity in Mass Spectrometry-Based Proteomics. *Anal. Chem.* **2021**, *93*, 16751–16758. [CrossRef]
22. Hernandez, D.R.; DeBord, J.D.; Ridgeway, M.E.; Kaplan, D.A.; Park, M.A.; Fernandez-Lima, F. Ion dynamics in a trapped ion mobility spectrometer. *Analyst* **2014**, *139*, 1913–1921. [CrossRef]
23. Zhao, L.; Cong, X.; Zhai, L.; Hu, H.; Xu, J.-Y.; Zhao, W.; Zhu, M.; Tan, M.; Ye, B.-C. Comparative evaluation of label-free quantification strategies. *J. Proteom.* **2020**, *215*, 103669. [CrossRef]
24. Millán-Oropeza, A.; Blein-Nicolas, M.; Monnet, V.; Zivy, M.; Henry, C. Comparison of Different Label-Free Techniques for the Semi-Absolute Quantification of Protein Abundance. *Proteomes* **2022**, *10*, 2. [CrossRef] [PubMed]

25. Ahrné, E.; Molzahn, L.; Glatter, T.; Schmidt, A. Critical assessment of proteome-wide label-free absolute abundance estimation strategies. *Proteomics* **2013**, *13*, 2567–2578. [CrossRef] [PubMed]
26. Wang, X.; Shen, S.; Rasam, S.S.; Qu, J. MS1 ion current-based quantitative proteomics: A promising solution for reliable analysis of large biological cohorts. *Mass Spectrom. Rev.* **2019**, *38*, 461–482. [CrossRef]
27. Silva, J.C.; Gorenstein, M.V.; Li, G.-Z.; Vissers, J.P.C.; Geromanos, S.J. Absolute Quantification of Proteins by LCMSE. *Mol. Cell. Proteom.* **2006**, *5*, 144–156. [CrossRef]
28. Khalil, S.; Plisnier, M. Comparative analysis of MS/MS search algorithms in label-free shotgun proteomics for monitoring host-cell proteins using trapped ion mobility and ddaPASEF. *J. Pharm. Biomed. Anal. Open* **2025**, *6*, 100082. [CrossRef]
29. Nesvizhskii, A.I.; Aebersold, R. Interpretation of Shotgun Proteomic Data. *Mol. Cell. Proteom.* **2005**, *4*, 1419–1440. [CrossRef]
30. Claassen, M. Inference and Validation of Protein Identifications. *Mol. Cell. Proteom.* **2012**, *11*, 1097–1104. [CrossRef]
31. ICH Q2(R2) Validation of Analytical Procedures—Scientific Guideline | European Medicines Agency (EMA). Available online: https://www.ema.europa.eu/en/ich-q2r2-validation-analytical-procedures-scientific-guideline (accessed on 19 February 2026).
32. ICH Q14 Analytical Procedure Development—Scientific Guideline | European Medicines Agency (EMA). Available online: https://www.ema.europa.eu/en/ich-q14-analytical-procedure-development-scientific-guideline (accessed on 2 March 2026).
33. Rogstad, S.; Yan, H.; Wang, X.; Powers, D.; Brorson, K.; Damdinsuren, B.; Lee, S. Multi-Attribute Method for Quality Control of Therapeutic Proteins. *Anal. Chem.* **2019**, *91*, 14170–14177. [CrossRef] [PubMed]
34. United States Pharmacopeial Convention. *⟨1132.1⟩ Residual Host Cell Protein Measurement in Biopharmaceuticals by Liquid Chromatography-Mass Spectrometry*; United States Pharmacopeia: Rockville, MD, USA, 2024. [CrossRef]
35. Chrone, V.G.; Blaszczyk, A.J.; Zhang, D.H.; Nielsen, S.B.; Crawford, J.; Schwämmle, V.; Højrup, P.; Kofoed, T.; Peckham, N.; Mørtz, E. Host cell protein quantitation by LC-MS. Experimental demonstration, qualification, and comparison of methods in USP 1132.1. *J. Pharm. Biomed. Anal.* **2025**, *265*, 117051. [CrossRef] [PubMed]
36. Feinberg, M. Validation of analytical methods based on accuracy profiles. *J. Chromatogr. A.* **2007**, *1158*, 174–183. [CrossRef]
37. Hubert, P.; Nguyen-Huu, J.-J.; Boulanger, B.; Chapuzet, E.; Cohen, N.; Compagnon, P.-A.; Dewé, W.; Feinberg, M.; Laurentie, M.; Mercier, N.; et al. Harmonization of strategies for the validation of quantitative analytical procedures. *J. Pharm. Biomed. Anal.* **2007**, *45*, 82–96. [CrossRef]
38. Henrion, A.; Arsene, C.-G.; Liebl, M.; O'Connor, G. Label-free quantification of host cell protein impurity in recombinant hemoglobin materials. *Anal. Bioanal. Chem.* **2024**, *416*, 387–396. [CrossRef]
39. Khalil, S.; Ho, J.T.C.; Plisnier, M. Definitive benchmarking of DDA and DIA for host cell protein analysis on the Orbitrap Astral in a regulatory-aligned framework 2025. *bioRxiv* **2025**, *preprint*. [CrossRef]
40. Yaro, A.S.; Maly, F.; Prazak, P.; Malý, K. Outlier Detection Performance of a Modified Z-Score Method in Time-Series RSS Observation with Hybrid Scale Estimators. *IEEE Access* **2024**, *12*, 12785–12796. [CrossRef]
41. Wen, B.; Freestone, J.; Riffle, M.; MacCoss, M.J.; Noble, W.S.; Keich, U. Assessment of false discovery rate control in tandem mass spectrometry analysis using entrapment. *Nat. Methods* **2025**, *22*, 1454–1463. [CrossRef]
42. Long, J.S.; Ervin, L.H. Using Heteroscedasticity Consistent Standard Errors in the Linear Regression Model. *Am. Stat.* **2000**, *54*, 217–224. [CrossRef]
43. Satterthwaite, F.E. Synthesis of Variance. *Psychometrika* **1941**, *6*, 309–316. [CrossRef]
44. Mee, R.W. β-Expectation and β-Content Tolerance Limits for Balanced One-Way ANOVA Random Model. *Technometrics* **1984**, *26*, 251–254. [CrossRef]
45. Chevreux, G.; Tilly, N.; Bihoreau, N. Quantification of proteins by data independent acquisition: Performance assessment of the Hi3 methodology. *Anal. Biochem.* **2018**, *549*, 184–187. [CrossRef]
46. Perez-Riverol, Y.; Bandla, C.; Kundu, D.J.; Kamatchinathan, S.; Bai, J.; Hewapathirana, S.; John, N.S.; Prakash, A.; Walzer, M.; Wang, S.; et al. The PRIDE database at 20 years: 2025 update. *Nucleic Acids Res.* **2025**, *53*, D543–D553. [CrossRef]