diaPASEF Enables High-throughput Proteomic Analysis of Host Cell Proteins for Biopharmaceutical Process Development

Taku Tsukidate[1]*, Alyssa Q. Stiving[1], Shannon Rivera[1], Ansuman Sahoo[2], Sri Madabhushi[2], Xuanwen Li[1]*

1. Analytical Research & Development Mass Spectrometry, Merck & Co., Inc., 126 East Lincoln Avenue, Rahway, New Jersey 07065, United States

2. Biologics Process Research & Development, Merck & Co., Inc., 126 East Lincoln Avenue, Rahway, New Jersey 07065, United States

ABSTRACT: Monitoring and quantifying host cell proteins (HCPs) in biotherapeutics production processes is crucial to ensure product quality, stability, and safety. Liquid chromatography–mass spectrometry (LC–MS) analysis has emerged as an important tool for identifying and quantifying individual HCPs. However, LC-MS–based approaches face challenges due to the wide dynamic range between HCPs and the therapeutic protein, as well as laborious sample preparation and long instrument time. To address these limitations, we evaluated the application of parallel accumulation–serial fragmentation combined with data-independent acquisition (diaPASEF), to HCP analysis for biopharmaceutical process development applications. We evaluated different library generation strategies and LC methods, demonstrating the suitability of these workflows for various HCP analysis needs such as in-depth characterization and high-throughput analysis of process intermediates. Remarkably, the diaPASEF approach enabled the quantification of hundreds of HCPs that were undetectable by a standard data-dependent acquisition mode while considerably improving sample requirement, throughput, coverage, quantitative precision, and data completeness.

Host cell proteins (HCPs) are protein impurities derived from host cells during the biotherapeutics production process. Several HCPs are considered high risk of affecting product quality, stability, and/or safety and require control strategies for removal and monitoring.[1–4] The enzyme-linked immunosorbent assay (ELISA) employing anti-HCP polyclonal antibodies is the most used method to monitor total HCP amount during process development.[5] However, since ELISA relies on immunoreactivity, it cannot detect non-immunogenic proteins or identify individual proteins. In this regard, liquid chromatography–mass spectrometry (LC–MS) analysis has become an important tool to identify and quantify individual HCPs.[6]

One substantial challenge encountered by LC-MS–based approaches is the wide (> 10^6) dynamic range between the sparse HCPs and the abundant therapeutic protein in the drug substance. To enable LC-MS–based HCP analysis, we and others have reported various sample preparation and LC-MS strategies. For example, Protein-A affinity purification[7] and molecular weight cutoff filtration[8] deplete monoclonal antibodies (mAb) while chemical proteomics[9] and ProteoMiner[10] enrich certain HCPs. Alternatively, native digestion selectively digests proteolytically more labile HCPs over mAbs.[11] In addition, offline fractionation[12,13], long gradient LC[14], and ion mobility spectrometry[15,16] improve separation between HCP– and therapeutic protein–derived peptides. However, these strategies typically involve somewhat laborious sample preparation and/or require long instrument time (> 2 h / sample), which considerably limits sample throughput.[13] Furthermore, data-dependent acquisition (DDA), which remains to be the method of choice for HCP analysis, might not sufficiently cover the required dynamic range to allow for adequate detection or quantification of HCPs.[17,18]

Data-independent acquisition (DIA) is an attractive alternative to DDA.[19] Typical DDA methods select the most abundant precursor ions of an MS1 scan for fragmentation and acquisition of MS2 spectra, leading to a loss of information, especially for sparse precursors, and partial data reproducibility due to the stochastic nature of precursor selection. In contrast, DIA methods slide pre-defined precursor isolation windows across the entire MS1 scan range for fragmentation and acquisition of MS2 spectra, enabling high reproducibility, sensitivity, and quantification across a broad dynamic range. Previous studies have applied various DIA methods to HCP analysis and demonstrated their advantages.[16,17,20–23]Among different DIA approaches, the diaPASEF method, which synchronizes ion mobility separation and mass selection to achieve great selectivity and sensitivity, is a transformative technology.[24,25] However, the application of diaPASEF to HCP analysis has not yet been reported in peer-reviewed literature.

In this manuscript, we report the first application of diaPASEF to HCP analysis in the biopharmaceutical industry and describe our optimized workflows (**Figure 1**). We systematically evaluated several spectral library generation strategies and LC methods using our internal pipeline molecules and identified workflows suitable for various HCP analysis needs in biotherapeutics process research and development.

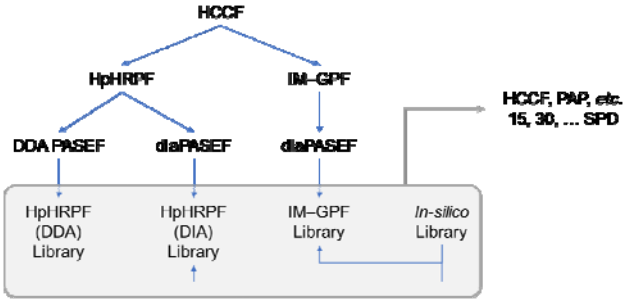


Figure 1. Overview of diaPASEF–based HCP analysis workflows. Spectral libraries were generated from HCCF employing different fractionation strategies and acquisition modes or *in silico*. These spectral libraries were used to process diaPASEF data of biotherapeutics process samples that were acquired with several LC methods. HCCF: harvested cell culture fluid; HpHRPF: high pH reversed-phase fractionation; IM–GPF: ion mobility–coupled gas-phase fractionation; PAP: protein A pool; SPD: samples per day.

Materials and Methods

Protein samples. Pierce Hela Protein Digest Standard was purchased from Thermo Scientific. mAb1 HCCF, mAb2 HCCF, and mAb1 PAP were internally sourced.

Proteomics sample preparation from mAb1 or mAb2 HCCF. mAb1 or mAb2 HCCF (20 µL) was incubated with 4 M urea, 25 mM TCEP-HCl, and 50 mM acrylamide in 50 mM tris-HCl pH 8.0 (total volume: 50 µL) at 60 °C for 30 min. An aliquot (20 µL) was incubated with Sera-Mag SpeedBead (10 µL, 200 µg) and acetonitrile (70 µL) at 20 °C for 10 min with agitation at 500 rpm. The beads were washed three times with 80 % EtOH (200 µL) and then incubated with 50 mM tris-HCl pH 8.0 (20 µL) containing 0.4 µg trypsin and 0.1 µg LysC at 37 °C for 18 h. Peptide concentration was determined with the Nanodrop A280 assay and/or the BCA assay (Thermo Scientific) and adjusted to 0.02 g/L with 0.1 % formic acid before loading 20 µL (400 ng) on Evotip Pure Evotips (Evosep) according to the manufacturer's instructions. The remainder was fractionated with Pierce High pH Reversed-Phase Peptide Fractionation Kit (Thermo Fisher Scientific), concentrated to dryness, reconstituted with 0.1 % formic acid to the final concentration of 0.02 g/L, and loaded (400 ng) on Evotip Pure Evotips according to the manufacturer's instructions.

Proteomics sample preparation from mAb1 PAP. mAb1 PAP were subjected to native digestion as previously described[11]. Peptide concentration was adjusted to 0.02 g/L (based on the initial input amount) with 0.1 % formic acid before loading 20 µL (400 ng) on Evotip Pure Evotips according to the manufacturer's instructions.

LC–MS analysis. Peptides were separated on the Evosep One LC system with Pepsep C18 columns using the manufacturer's pre-defined methods denoted by daily sample throughput e.g., 30 samples per day (SPD) and analysed on the Bruker timsTOF Pro 2 system. The standard DDA PASEF method was used without modification. The diaPASEF method was optimized by adjusting the isolation windows either manually or with the py_diAID software[26]. The IM–GPF schemes were designed with a custom R script.

Data analysis. DDA PASEF data were processed with MaxQuant v2.4.10[27]. Use of .NET Core was disabled. Match between runs (MBR) was disabled for spectral library generation and enabled for all other analyses. All other settings were left as default. MaxQuant evidence and msms files were reformatted with a custom R script and used to generate a spectral library with DIA-NN v1.8.1[28]. All diaPASEF data were processed with DIA-NN v1.8.1. Mass accuracies were set to 10 ppm for spectral library generation and to 15 ppm for all other processing. MBR was disabled for spectral library generation and enabled for all other analyses as MBR tended to have a small but positive effect even when processing data with IM-GPF or HpHRPF libraries (**Supporting Information Figures 1 & 2**). For the generation of the *in silico* spectral library from sequence databases, the precursor charge range and the mass range were restricted to 2−3 and 300–1200 Th, respectively. The global precursor and global protein FDR thresholds were set to 1 % for all analyses. This was achieved by filtering DIA-NN main reports on the Global.Q.Value column and the Global.PG.Q.Value column in MBR-disabled data processing and on the Lib.Q.Value column and the Lib.PG.Q.Value in MBR-enabled data processing. FDR estimation was independently validated by the two-species library analysis as previously described[29] (**Supporting Information Figure 3**). The data post-processing analysis was performed in R version 4.0.3. Protein quantification was performed using the MaxLFQ algorithm as implemented in the iq package[30] based on the Fragment.Quant.Corrected column of DIA-NN main reports. Chromatograms were visualized in Skyline-daily v19.0.9.

Results and Discussion

Optimization of ion mobility–coupled gas phase fractionation scheme. Choice of spectral library has significant impact on DIA MS data quality as well as total experiment time.[31–33] For example, Wen et al. reported that a spectral library generated via HpHRPF and DDA yielded 10–20 % more identifications than an *in silico* spectral library in their diaPASEF analysis of human and mouse cell lysates.[32] On the other hand, HpHRPF results in sample matrices that are different from the target samples, which may affect retention time accuracy. In this regard, gas-phase fractionation (GPF) is an attractive alternative to rapidly generate spectral libraries.[33] For example, Penny et al. demonstrated that the ion mobility–coupled GPF (IM–GPF) workflow performed comparably with the HpHRPF workflow in terms of protein identifications in the HeLa cell protein digest.[34] Similarly, Rice and Belani demonstrated the IM–GPF workflow outperformed the HpHRPF workflow in terms of protein identifications in depleted plasma samples.[35] While both studies employed the isolation window width of 5 Th for their GPF library generation, narrower windows are desirable to improve the selectivity of precursor selection. Thus, we optimized the isolation window width and the number of PASEF scans for the IM–GPF scheme with the commercial HeLa protein digest standard (**Figure 2**).

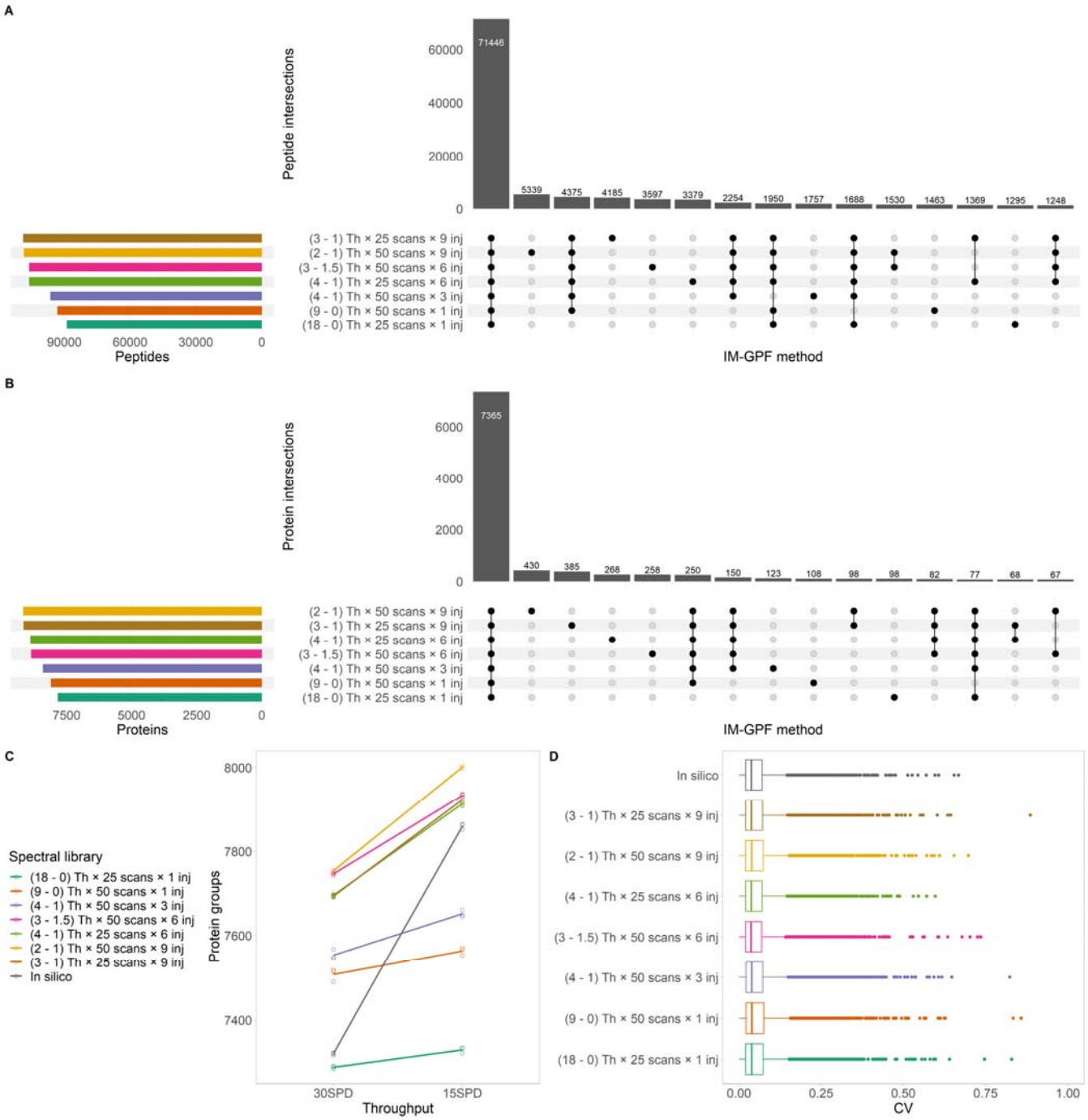


Figure 2. Optimisation of IM–GPF scheme with HeLa protein digest standard. The precursor isolation window width and the number of PASEF scans were varied. Number and overlap of peptides (A) and proteins (B) in various IM–GPF libraries. Horizontal bars indicate the number of peptides or proteins in each library. Vertical bars indicate the number of peptides or proteins shared by several libraries (intersection) as indicated by dotes below. Fifteen largest intersections are shown. Number of quantified protein groups (C) and coefficients of variation from the 30 SPD data (D) in three replicate injections, employing the IM–GPF and *in silico* libraries during data processing. The IM–GPF scheme names and corresponding library names represent (window width – overlap) Th × (number of PASEF scans) × [number of required injections (abbreviated as inj)]. Evosep method: 15 SPD for spectral library data acquisition; 15 or 30 SPD for three replicate injections. Global peptide and global protein FDRs < 0.01.

The IM–GPF schemes with narrower isolation windows yielded larger spectral libraries with more peptide and protein identifications but required more injections to cover the whole mass range (**Figure 2A–B**). The methods with 50 PASEF scans yielded 2.5–3 data points per peak at full width at half maximum (FWHM) on average whereas the methods with 25 PASEF scans yielded ~ 4 data points per peak at FWHM on average, suggesting that these schemes may also be suitable for shorter gradient LC methods (*vide infra*).[36] As expected, the libraries shared most of the library peptides and proteins. On the other hand, the use of different libraries for the analysis of HeLa diaPASEF data resulted in modest difference in the numbers of protein identifications (**Figure 2C**). The larger libraries (6 or 9 injections) consistently outperformed the

smaller libraries (1 or 3 injections) and an *in-silico* library generated by DIA-NN in terms of protein identifications. Protein quantification was highly precise with median coefficients of variation of 3–4 % and unaffected by the choice of spectral library (**Figure 2D**). Interestingly, the "(3 – 1.5) Th × 50 scans × 6 injections" library outperformed the "(3 – 1) Th × 25 scans × 9 injections" library, highlighting the importance of specificity in spectral library[37]. Overall, applying the effective 1-Th isolation windows as in the "(2 – 1) Th × 50 scans × 9 injections" scheme resulted in the best performance (**Supporting Information Figure 4**).

Optimisation of spectral library workflow for biotherapeutics process intermediates. With the optimized IM–GPF scheme in hand, we evaluated several spectral library generation workflows with a harvested cell culture fluid (HCCF) from one of our proprietary monoclonal antibody processes (mAb1, **Figure 3, Supporting Information Figure 5**). We processed mAb1 HCCF with one-step denaturation, reduction and alkylation followed by the SP3 cleanup and on-bead digestion to prepare a protein digest.[38] We then subjected the digest to the offline HpHRPF employing a commercial kit, which generates 10 fractions, or to the optimized IM–GPF scheme to generate spectral libraries. The HpHRP fractions were analysed with a standard DDA PASEF method (HpHRPF–DDA) or a diaPASEF method with non-overlapping 9-Th isolation windows and 50 PASEF scans (HpHRPF–DIA). The cycle time and hence the number of data points per peak of this diaPASEF method matches those of the optimized IM–GPF method under the same LC gradient. The HpHRPF–DIA workflow resulted in the largest spectral library including > 16,000 peptides (**Figure 3A**) and > 1,800 proteins (**Figure 3B**) that are unique to this library and not part of the IM–GPF or HpHRPF–DDA libraries. On the other hand, the HpHRPF–DDA workflow resulted in the smallest spectral library. Accordingly, the use of the HpHRPF–DIA library for HCCF diaPASEF data processing resulted in the largest number of protein identifications (**Figure 3C**). In contrast to the HeLa diaPASEF data, the IM–GPF libraries underperformed an *in-silico* library generated by DIA-NN in terms of protein identifications in the HCCF diaPASEF data acquired with the 15 SPD method. However, the IM-GPF library advantageously outperformed the *in-silico* library in the HCCF diaPASEF data when increasing sample throughput. For example, processing 100SPD diaPASEF data with either the IM-GPF (15 SPD) library or the HpHRPF (15 SPD, DIA) library yielded similar number of protein identifications and abundance distributions (**Figure 3D**). This is likely because many of those peptides that are unique to the HpHRPF (15 SPD, DIA) library were not resolved well enough with the short gradient from more abundant peptides that are also part of the IM-GPF (15 SPD, DIA) library. On the other hand, protein abundance distribution clearly shifted towards more abundant proteins when the same data were processed with the HpHRPF (15 SPD, DDA) library, perhaps reflecting potential abundance bias in the DDA-based library. Quantitative precision was largely unaffected by the choice of spectral library (median CV ~ 6 %).

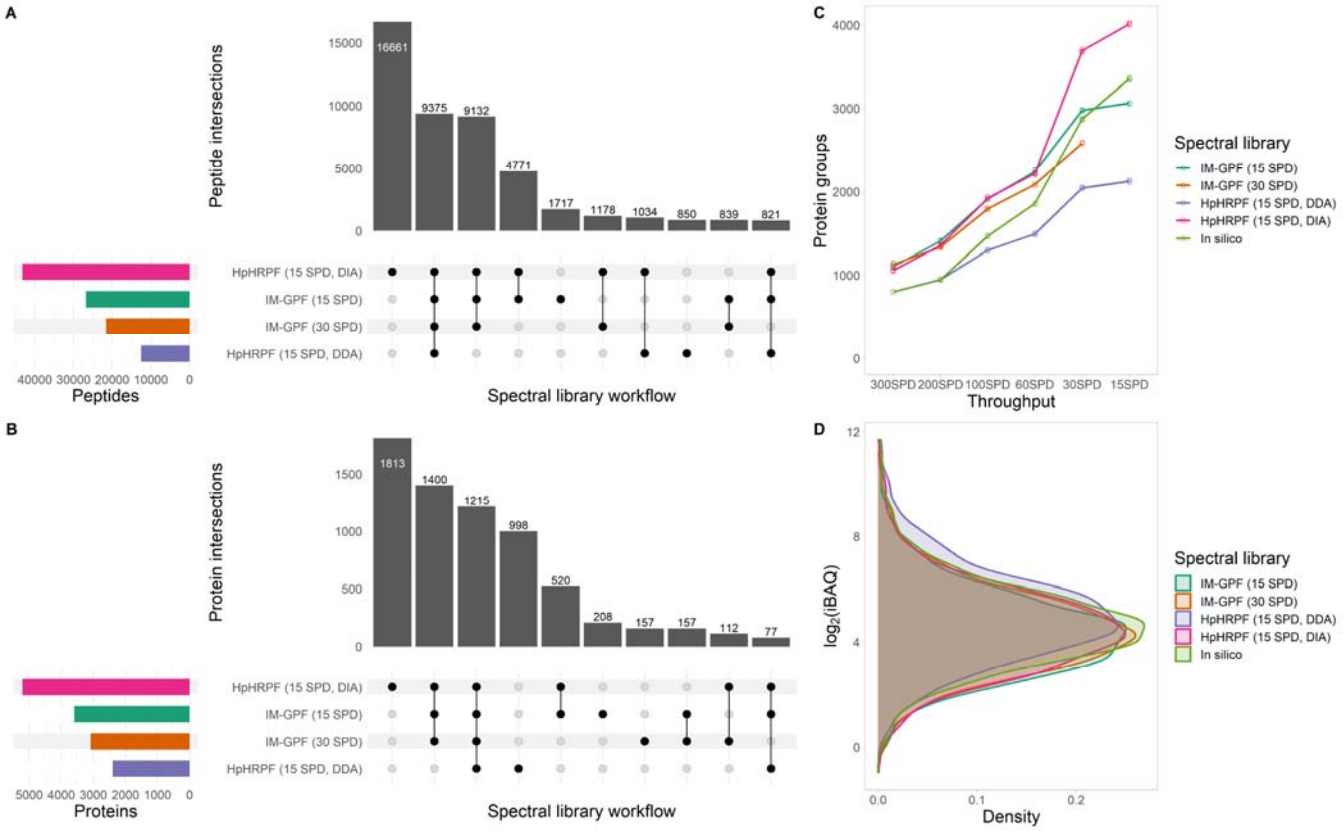


Figure 3. Evaluation of spectral library generation workflows with mAb1 HCCF. Empirical spectral libraries were generated by the optimal IM–GPF scheme or HpHRPF combined with DDA PASEF or diaPASEF. Number and overlap of peptides (A) and proteins (B) in various libraries. Horizontal bars indicate the number of peptides or proteins in each library. Vertical bars indicate the number of peptides or proteins shared by several libraries (intersection) as indicated by dotes below. Ten largest intersections are shown. Number of quantified protein groups in three replicate injections (C) and distribution of average protein abundance score

(iBAQ) (D), employing the empirical and *in-silico* libraries during data processing. D: 100 SPD. Global peptide and global protein FDRs < 0.01.

Similarly, we evaluated the performance of these spectral libraries for the analysis of mAb1 protein A pool (PAP) diaPASEF data (**Figure 4, Supporting Information Figures 6 & 7**). We processed mAb1 PAP according to a native digestion protocol as previously described[11]. The HpHRPF–DIA and the IM–GPF (15 SPD) libraries led to similar protein identifications (**Figures 4A**) and quantitative precisions (median CV ~ 12 %). On the other hand, the use of HpHRPF–DDA (median CV ~ 14 %) or the *in-silico* libraries (median CV ~ 10 %) resulted in significantly smaller numbers of protein identifications. Of note, protein abundance distribution clearly shifted towards more abundant proteins in the case of the *in-silico* library (**Figure 4B**), highlighting the importance of experimental spectral library in the diaPASEF analysis of biotherapeutics process samples to avoid the possibility of missing lower abundance HCPs.

Next, we compared the diaPASEF method and the standard DDA PASEF method in the analysis of the mAb1 PAP sample. The diaPASEF method quantified a majority (50 of 72) of proteins quantified by the DDA PASEF method and 613 additional HCPs (**Figure 4C**). The 22 proteins that were quantified by the DDA PASEF method but not by the diaPASEF were supported only by 24 peptides in total, of which 23 peptides were not part of any of the four experimental spectral libraries. Visual inspection of corresponding spectra revealed many questionable assignments. Moreover, none of these 22 proteins are considered high-risk HCPs.[1] On the other hand, diaPASEF revealed the presence of > 600 additional HCPs that the DDA PASEF method missed (**Figure 4D**). These HCPs included several high-risk HCPs and polysorbate-degradative enzymes such as PLA2G7 and PPT1. In addition, quantification was more precise with the diaPASEF method (median CV ~ 12 %) than with the DDA PASEF method (median CV ~ 16 %).

Furthermore, the diaPASEF method was clearly superior to our platform method that employs a three-hour gradient on a microflow LC and leverages conventional DDA on the Q Exactive HF-X mass spectrometer[2]—The new method required 100-fold less sample, increased throughput by > 10-fold, expanded identification coverage by up to 10-fold, considerably improved quantitative precision, and improved data completeness by five-fold (**Supporting Information Figure 8**).

GO term analysis of mAb1 HCCF and mAb1 PAP suggested that most HCP mass (~ 90 %) was released from plasma membrane, cytoplasm and/or nucleus through cell rapture and that (actively secreted) extracellular proteins accounted for a small fraction of the total HCP mass (~ 10 %) (**Supporting Information Figure 9**). In terms of HCP persistence, the abundance in HCCF and the molecular weight of HCP were strongly correlated with the abundance of HCP in PAP: $\rho_{HCCF} = 0.79$; $\rho_{MW} = -0.81$. On the other hand, the isoelectric point and the hydrophobicity index lacked correlation with the abundance in PAP. In addition, distributions of pI or GRAVY values were similar between HCCF and PAP. These results extend similar findings in previous studies beyond the limited numbers of HCPs to many other diverse HCP populations.[39,40] In addition, the HCP peptides overall eluted earlier than the HeLa protein digest, which may help LC method optimization. Altogether, these results indicate that the diaPASEF method combined with rapid IM–GPF spectral library generation enables wide proteome coverage in biotherapeutics process samples.

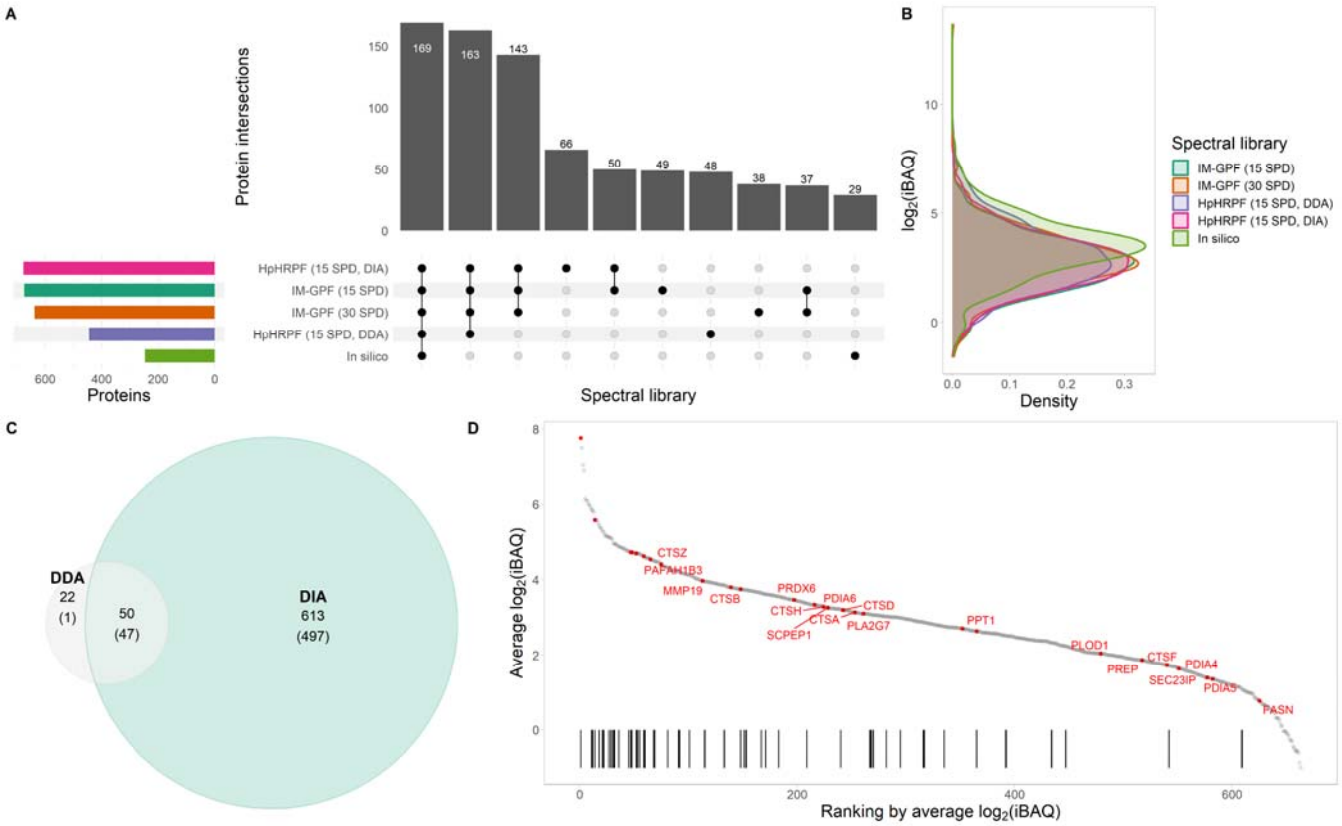


Figure 4. Evaluation of spectral library generation workflows with mAb1 PAP. Number and overlap of protein groups that were quantified in all three replicate injections (A) and distribution of average protein abundance score (iBAQ) (B), employing the empirical HCCF and *in silico* libraries during data processing. In A, horizontal bars indicate the number of proteins; vertical bars indicate the number of shared proteins (intersection); ten largest intersections are shown. (C) Comparison of HCP identifications between DDA PASEF and diaPASEF methods. The numbers in parentheses indicate the numbers of proteins with at least two unique peptides. (D) Waterfall plot of average protein abundance score for the 663 HCPs quantified with the diaPASEF method. High-risk HCPs are highlighted as red points and, if not quantified with the DDA PASEF method, are labelled with their protein symbols. The black bars indicate quantification of the HCP with the DDA PASEF method and the DDA method's significant ($p = 2 \times 10^{-8}$, Kolmogorov-Smirnov test) bias towards abundant HCPs. 30 SPD. Global peptide and global protein FDRs < 0.01.

High-throughput diaPASEF analysis of biotherapeutics production process intermediates. To assess the potential of the optimal diaPASEF workflow to further advance biotherapeutics process understanding, we tested its performance in a high-throughput analysis of mAb2 HCCFs (**Figure 5**). These HCCF samples originated from several mAb2 clones that were cultured under various conditions and harvested at several time points, covering 72 different conditions. We generated a spectral library from 18 repeated IM-GPF measurements of a sample pool with the effective 1-Th isolation window *i.e.*, "(2 – 1) Th × 25 scans × 18 injections" method to balance the optimal isolation window width and cycle time and analysed the 72 single-shot samples in addition to 12 blanks, 12 system suitability samples, and 11 sample pools in a random order at the 100-SPD throughput. Despite the large number of runs required to accomplish this experiment, with the 60 and 100 SPD methods this took < 36 hours from start to finish. Sample pools were highly reproducible with the correlation coefficient (Pearson's r) > 0.97 between any two runs and median CV ~ 6 %, demonstrating the excellent performance of the method. Sample carry-over effect was virtually absent. We quantified ~ 1800 HCPs with few missing values (**Figure 5A**) including ~ 30 high-risk HCPs[1] such as PLA2G7 (**Figure 5B**). Culture condition, time point, and clone had clear impact on overall HCP profiles and high-risk HCP levels. For example, PLA2G7 was more abundant in the HCCFs of cells cultured under the condition A than in the HCCFs of cells cultured under the condition B. On the other hand, we did not find well-connected network of the high-risk HCPs, which is perhaps expected given the diversity of pathway, function, subcellular localization etc. of these HCPs. Detailed analysis of this data will be presented elsewhere (Sahoo *et al.*, manuscript in preparation).

In-silico libraries could be advantageous when analyzing a large set of heterogenous samples. Indeed, we observed modest (< 10 %) increase in the number of quantified HCPs compared to the IM-GPF approach (**Supporting Information Figure 10**). However, data processing time increased considerably from 134 min for the IM-GPF approach including the time for processing the 18 IM-GPF files to 2082 min for the library-free approach. Thus, both approaches are valid options depending on study goals.

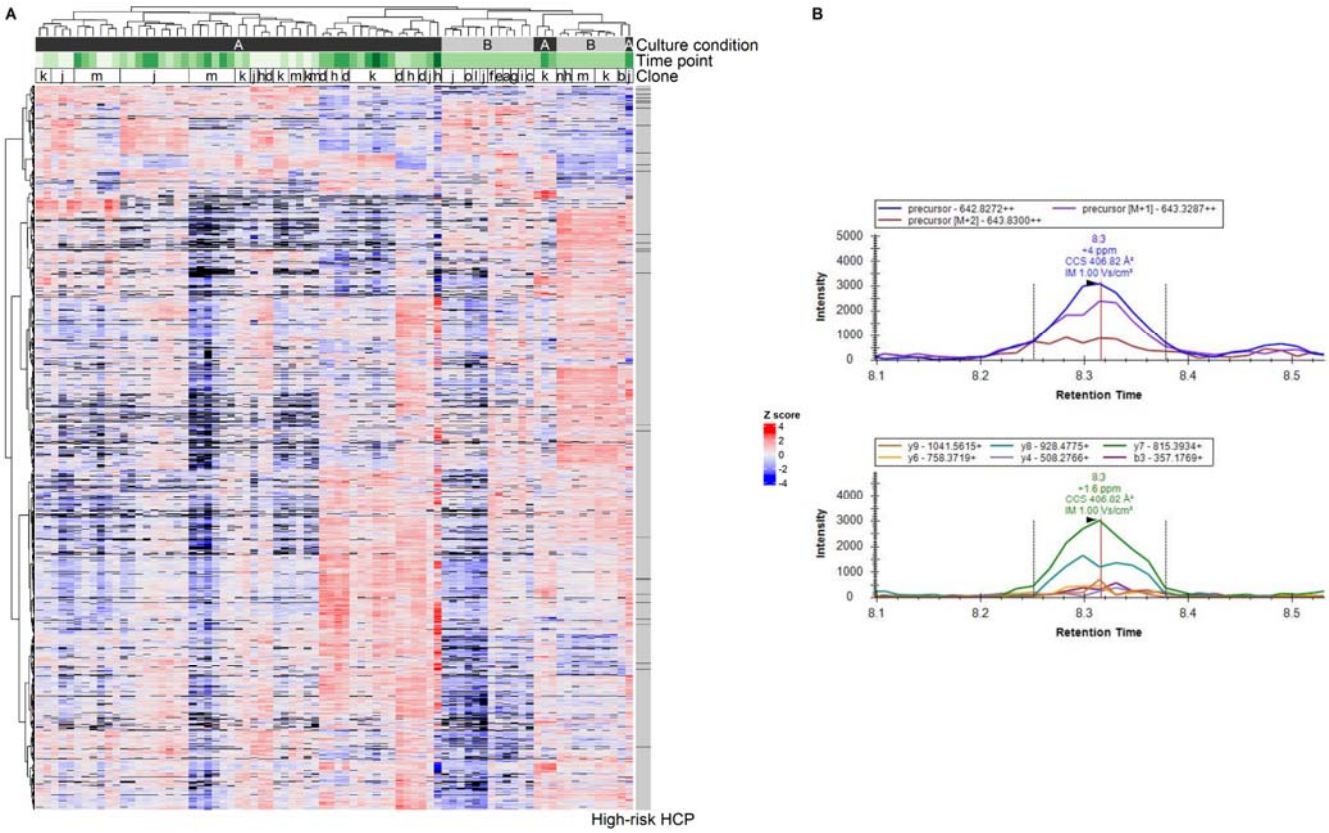


Figure 5. High-throughput diaPASEF analysis of mAb2 HCCFs. (A) Heatmap visualization of protein quantity. Each row corresponds to a protein group and each column corresponds to a sample. Fill colour indicates row-wise Z score. High-risk HCPs are indicated with black bars in the annotation panel on the right. Sample pools were highly reproducible with the correlation coefficient (Pearson's r) > 0.97 between any two runs and median CV ~ 6 %. (B) A representative extracted ion chromatogram for the PLA2G7-derived peptide ENILGSYFDVK and its MS2 fragments. Global peptide and global protein FDRs < 0.01.

Conclusions

In summary, we optimized a diaPASEF-based workflow for HCP analysis and demonstrated its potential to advance biotherapeutics process understanding. We evaluated several spectral library generation strategies and LC methods for biotherapeutics process samples and identified workflows suitable for various HCP analysis needs including high-throughput analysis of harvests and in-depth characterization of purification process intermediates. With an optimized workflow in hand, we demonstrated that the diaPASEF method can reveal the presence of hundreds of HCPs in a PAP sample that are undetectable by the standard DDA PASEF method including several high-risk HCPs and should facilitate investigative studies. We also demonstrated high-throughput analysis of HCCF samples, which should advance our understanding of HCPs prior to purification and inspire new control strategies.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website.

Additional experimental details and results (PDF)

AUTHOR INFORMATION

Corresponding Author

T.T.: taku.tsukidate@merck.com; X.L.: xuanwen.li@merck.com

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Conflict of Interest Disclosure

The authors declare no competing financial interest.

ACKNOWLEDGMENT

T.T. and A.S. thank MRL Postdoctoral Program. The authors thank Douglas D. Richardson and Hillary A. Schuessler for guidance and support.

REFERENCES

(1) Jones, M.; Palackal, N.; Wang, F.; Gaza-Bulseco, G.; Hurkmans, K.; Zhao, Y.; Chitikila, C.; Clavier, S.; Liu, S.; Menesale, E.; Schonenbach, N. S.; Sharma, S.; Valax, P.; Waerner, T.; Zhang, L.; Connolly, T. "High-Risk" Host Cell Proteins (HCPs): A Multi-Company Collaborative View. *Biotechnol. Bioeng.* **2021**, *118* (8), 2870–2885. https://doi.org/10.1002/bit.27808.

(2) Li, X.; An, Y.; Liao, J.; Xiao, L.; Swanson, M.; Martinez-Fonts, K.; Pavon, J. A.; Sherer, E. C.; Jawa, V.; Wang, F.; Gao, X.; Letarte, S.; Richardson, D. D. Identification and Characterization of a Residual Host Cell Protein Hexosaminidase B Associated with N-Glycan Degradation during the Stability Study of a Therapeutic Recombinant Monoclonal Antibody Product. *Biotechnol. Prog.* **2021**, *37* (3), e3128. https://doi.org/10.1002/btpr.3128.

(3) Li, X. Recent Applications of Quantitative Mass Spectrometry in Biopharmaceutical Process Development and Manufacturing.

Journal of Pharmaceutical and Biomedical Analysis. Elsevier B.V. September 20, 2023, p 115581. https://doi.org/10.1016/j.jpba.2023.115581.

(4) Li, X.; Wang, F.; Li, H.; Richardson, D. D.; Roush, D. J. The Measurement and Control of High-Risk Host Cell Proteins for Polysorbate Degradation in Biologics Formulation. *Antib. Ther.* **2022**, *5* (1), 42–54. https://doi.org/10.1093/abt/tbac002.

(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* (12), 2367–2379. https://doi.org/10.1002/bit.25327.

(6) 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* (1), 2213365. https://doi.org/10.1080/19420862.2023.2213365.

(7) Madsen, J. A.; Farutin, V.; Carbeau, T.; Wudyka, S.; Yin, Y.; Smith, S.; Anderson, J.; Capila, I. Toward the Complete Characterization of Host Cell Proteins in Biotherapeutics via Affinity Depletions, LC-MS/MS, and Multivariate Analysis. *MAbs* **2015**, *7* (6), 1128–1137. https://doi.org/10.1080/19420862.2015.1082017.

(8) 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* (5), 3751–3757. https://doi.org/10.1021/acs.analchem.9b05081.

(9) Li, X.; Chandra, D.; Letarte, S.; Adam, G. C.; Welch, J.; Yang, R.-S.; Rivera, S.; Bodea, S.; Dow, A.; Chi, A.; Strulson, C. A.; Richardson, D. D. Profiling Active Enzymes for Polysorbate Degradation in Biotherapeutics by Activity-Based Protein Profiling. *Anal. Chem.* **2021**, *93* (23), 8161–8169. https://doi.org/10.1021/acs.analchem.1c00042.

(10) Zhang, S.; Zhao, B.; Adaniya, S.; Xiao, H.; Li, N. Ultrasensitive Quantification Method for Understanding Biologically Relevant Concentrations of Host Cell Proteins in Therapeutics. *Anal. Chem.* **2023**, *95* (14), 6002–6008. https://doi.org/10.1021/acs.analchem.3c00020.

(11) 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* (10), 5436–5444. https://doi.org/10.1021/acs.analchem.7b00304.

(12) Kufer, R.; Haindl, M.; Wegele, H.; Wohlrab, S. Evaluation of Peptide Fractionation and Native Digestion as Two Novel Sample Preparation Workflows to Improve HCP Characterization by LC-MS/MS. *Anal. Chem.* **2019**, *91* (15), 9716–9723. https://doi.org/10.1021/acs.analchem.9b01259.

(13) Ma, J.; Kilby, G. W. Sensitive, Rapid, Robust, and Reproducible Workflow for Host Cell Protein Profiling in Biopharmaceutical Process Development. *J. Proteome Res.* **2020**, *19* (8), 3396–3404. https://doi.org/10.1021/acs.jproteome.0c00252.

(14) Yang, F.; Li, D.; Kufer, R.; Cadang, L.; Zhang, J.; Dai, L.; Guo, J.; Wohlrab, S.; Greenwood-Goodwin, M.; Shen, A.; Duan, D.; Li, H.; Yuk, I. H. Versatile LC–MS-Based Workflow with Robust 0.1 Ppm Sensitivity for Identifying Residual HCPs in Biotherapeutic Products. *Anal. Chem.* **2022**, *94* (2), 723–731. https://doi.org/10.1021/acs.analchem.1c03095.

(15) 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* (20), 10283–10291. https://doi.org/10.1021/acs.analchem.5b02103.

(16) Beaumal, C.; Beck, A.; Hernandez□Alba, O.; Carapito, C. Advanced Mass Spectrometry Workflows for Accurate Quantification of Trace□level Host Cell Proteins in Drug Products: Benefits of FAIMS Separation and Gas□phase Fractionation DIA. *Proteomics* **2023**, *23* (16), e2300172. https://doi.org/10.1002/pmic.202300172.

(17) Zheng, X.; Nie, S.; Greer, T.; Johnson, R. O. B.; 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* (10), 4383–4390. https://doi.org/10.1021/acs.analchem.0c03931.

(18) Strasser, L.; Oliviero, G.; Jakes, C.; Zaborowska, I.; Floris, P.; Ribeiro da Silva, M.; Füssl, F.; Carillo, S.; Bones, J. Detection and Quantitation of Host Cell Proteins in Monoclonal Antibody Drug Products Using Automated Sample Preparation and Data-Independent Acquisition LC-MS/MS. *J. Pharm. Anal.* **2021**, *11* (6), 726–731. https://doi.org/10.1016/j.jpha.2021.05.002.

(19) Ludwig, C.; Gillet, L.; Rosenberger, G.; Amon, S.; Collins, B. C.; Aebersold, R. Data□independent Acquisition□based SWATH□MS for Quantitative Proteomics: A Tutorial. *Mol. Syst. Biol.* **2018**, *14* (8), 1–23. https://doi.org/10.15252/msb.20178126.

(20) Oh, Y. H.; Mendola, K. M.; Choe, L. H.; Min, L.; Lavoie, A. R.; Sripada, S. A.; Williams, T. I.; Lee, K. H.; Yigzaw, Y.; Seay, A.; Bill, J.; Li, X.; Roush, D. J.; Cramer, S. M.; Menegatti, S.; Lenhoff, A. M. Identification and Characterization of CHO Host□cell Proteins in Monoclonal Antibody Bioprocessing. *Biotechnol. Bioeng.* **2023**, No. August. https://doi.org/10.1002/bit.28568.

(21) Kreimer, S.; Gao, Y.; Ray, S.; Jin, M.; Tan, Z.; Mussa, N. A.; Tao, L.; Li, Z.; Ivanov, A. R.; Karger, B. L. Host Cell Protein Profiling by Targeted and Untargeted Analysis of Data Independent Acquisition Mass Spectrometry Data with Parallel Reaction Monitoring Verification. *Anal. Chem.* **2017**, *89* (10), 5294–5302. https://doi.org/10.1021/acs.analchem.6b04892.

(22) Walker, D. E.; Yang, F.; Carver, J.; Joe, K.; Michels, D. A.; Yu, X. C. A Modular and Adaptive Mass Spectrometry-Based Platform for Support of Bioprocess Development toward Optimal Host Cell Protein Clearance. *MAbs* **2017**, *9* (4), 654–663. https://doi.org/10.1080/19420862.2017.1303023.

(23) Pythoud, N.; Bons, J.; Mijola, G.; Beck, A.; Cianférani, S.; Carapito, C. Optimized Sample Preparation and Data Processing of Data-Independent Acquisition Methods for the Robust Quantification of Trace-Level Host Cell Protein Impurities in Antibody Drug Products. *J. Proteome Res.* **2021**, *20* (1), 923–931. https://doi.org/10.1021/acs.jproteome.0c00664.

(24) Pengelley, S.; Belau, E.; Evers, W.; Albers, C.; Tremintin, G.; Suckau, D. Sub-Ppm HCP Screening in 15 Minutes Using Dia-PASEF. 2021.

(25) Meier, F.; Brunner, A. D.; Frank, M.; Ha, A.; Bludau, I.; Voytik, E.; Kaspar-Schoenefeld, S.; Lubeck, M.; Raether, O.; Bache, N.; Aebersold, R.; Collins, B. C.; Röst, H. L.; Mann, M. DiaPASEF: Parallel Accumulation–Serial Fragmentation Combined with Data-Independent Acquisition. *Nat. Methods* **2020**, *17* (12), 1229–1236. https://doi.org/10.1038/s41592-020-00998-0.

(26) Skowronek, P.; Thielert, M.; Voytik, E.; Tanzer, M. C.; Hansen, F. M.; Willems, S.; Karayel, O.; Brunner, A. D.; Meier, F.; Mann, M. Rapid and In-Depth Coverage of the (Phospho-) Proteome With Deep Libraries and Optimal Window Design for Dia-PASEF. *Mol. Cell. Proteomics* **2022**, *21* (9), 100279. https://doi.org/10.1016/j.mcpro.2022.100279.

(27) Sinitcyn, P.; Hamzeiy, H.; Salinas Soto, F.; Itzhak, D.; McCarthy, F.; Wichmann, C.; Steger, M.; Ohmayer, U.; Distler, U.; Kaspar-Schoenefeld, S.; Prianichnikov, N.; Yılmaz, Ş.; Rudolph, J. D.; Tenzer, S.; Perez-Riverol, Y.; Nagaraj, N.; Humphrey, S. J.; Cox, J. MaxDIA Enables Library-Based and Library-Free Data-Independent Acquisition Proteomics. *Nat. Biotechnol.* **2021**, *39* (12), 1563–1573. https://doi.org/10.1038/s41587-021-00968-7.

(28) Demichev, V.; Messner, C. B.; Vernardis, S. I.; Lilley, K. S.; Ralser, M. DIA-NN: Neural Networks and Interference Correction Enable Deep Proteome Coverage in High Throughput. *Nat. Methods* **2020**, *17* (1), 41–44. https://doi.org/10.1038/s41592-019-0638-x.

(29) Bruderer, R.; Bernhardt, O. M.; Gandhi, T.; Xuan, Y.; Sondermann, J.; Schmidt, M.; Gomez-varela, D.; Reiter, L. Optimization of Experimental Parameters in Data-Independent Mass Spectrometry Significantly Increases Depth and Reproducibility of Results *. *Mol. Cell. Proteomics* **2017**, *16* (12), 2296–2309. https://doi.org/10.1074/mcp.RA117.000314.

(30) Pham, T. V; Henneman, A. A.; Jimenez, C. R. Iq□: An R Package to Estimate Relative Protein Abundances from Ion

Quantification in DIA-MS-Based Proteomics. *Bioinformatics* **2020**, *36* (8), 2611–2613. https://doi.org/10.1093/bioinformatics/btz961.

(31) Barkovits, K.; Pacharra, S.; Pfeiffer, K.; Steinbach, S.; Eisenacher, M.; Marcus, K.; Uszkoreit, J. Reproducibility, Specificity and Accuracy of Relative Quantification Using Spectral Librarybased Data-Independent Acquisition. *Mol. Cell. Proteomics* **2020**, *19* (1), 181–197. https://doi.org/10.1074/mcp.RA119.001714.

(32) Wen, C.; Gan, G.; Xu, X.; Lin, G.; Chen, X.; Wu, Y.; Xu, Z.; Wang, J.; Xie, C.; Wang, H. R.; Zhong, C. Q. Investigation of Effects of the Spectral Library on Analysis of DiaPASEF Data. *J. Proteome Res.* **2022**, *21* (2), 507–518. https://doi.org/10.1021/acs.jproteome.1c00899.

(33) Pino, L. K.; Just, S. C.; MacCoss, M. J.; Searle, B. C. Acquiring and Analyzing Data Independent Acquisition Proteomics Experiments without Spectrum Libraries. *Mol. Cell. Proteomics* **2020**, *19* (7), 1088–1103. https://doi.org/10.1074/mcp.P119.001913.

(34) Penny, J.; Arefian, M.; Schroeder, G. N.; Bengoechea, J. A.; Collins, B. C. A Gas Phase Fractionation Acquisition Scheme Integrating Ion Mobility for Rapid DiaPASEF Library Generation. *Proteomics* **2023**, *23* (7–8), 1–8. https://doi.org/10.1002/pmic.202200038.

(35) Rice, S. J.; Belani, C. P. Optimizing Data-Independent Acquisition (DIA) Spectral Library Workflows for Plasma Proteomics Studies. *Proteomics* **2022**, *22* (17), 1–11. https://doi.org/10.1002/pmic.202200125.

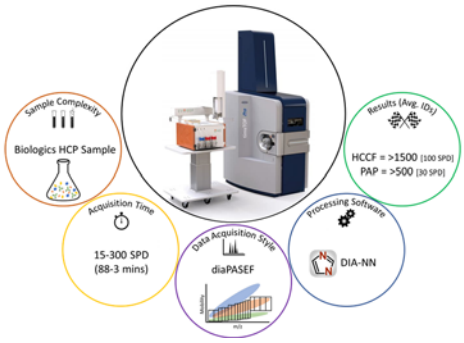
(36) Doellinger, J.; Blumenscheit, C.; Schneider, A.; Lasch, P. Increasing Proteome Depth While Maintaining Quantitative Precision in Short-Gradient Data-Independent Acquisition Proteomics. *J. Proteome Res.* **2023**, *22* (6), 2131–2140. https://doi.org/10.1021/acs.jproteome.3c00078.

(37) Rosenberger, G.; Bludau, I.; Schmitt, U.; Heusel, M.; Hunter, C. L.; Liu, Y.; Maccoss, M. J.; Maclean, B. X.; Nesvizhskii, A. I.; Pedrioli, P. G. A.; Reiter, L.; Röst, H. L.; Tate, S.; Ting, Y. S.; Collins, B. C.; Aebersold, R. Statistical Control of Peptide and Protein Error Rates in Large-Scale Targeted Data-Independent Acquisition Analyses. *Nat. Methods* **2017**, *14* (9), 921–927. https://doi.org/10.1038/nmeth.4398.

(38) Hughes, C. S.; Moggridge, S.; Müller, T.; Sorensen, P. H.; Morin, G. B.; Krijgsveld, J. Single-Pot, Solid-Phase-Enhanced Sample Preparation for Proteomics Experiments. *Nat. Protoc.* **2019**, *14* (1), 68–85. https://doi.org/10.1038/s41596-018-0082-x.

(39) Aboulaich, N.; Chung, W. K.; Thompson, J. H.; Larkin, C.; Robbins, D.; Zhu, M. A Novel Approach to Monitor Clearance of Host Cell Proteins Associated with Monoclonal Antibodies. *Biotechnol. Prog.* **2014**, *30* (5), 1114–1124. https://doi.org/10.1002/btpr.1948.

(40) Zhang, Q.; Goetze, A. M.; Cui, H.; Wylie, J.; Tillotson, B.; Hewig, A.; Hall, M. P.; Flynn, G. C. Characterization of the Co-Elution of Host Cell Proteins with Monoclonal Antibodies during Protein A Purification. *Biotechnol. Prog.* **2016**, *32* (3), 708–717. https://doi.org/10.1002/btpr.2272.



For Table of Contents Only