Host cell protein quantification workflow using optimized standards combined with data-independent acquisition mass spectrometry

Steve Hessmann, Cyrille Chery, Anne-Sophie Sikora, Annick Gervais, Christine Carapito

Journal of Pharmaceutical Analysis 13 (2023) 494–502

Contents lists available at ScienceDirect

Journal of Pharmaceutical Analysis

journal homepage: www.elsevier.com/locate/jpa

Original article

Host cell protein quantification workflow using optimized standards combined with data-independent acquisition mass spectrometry

Steve Hessmann [a], Cyrille Chery [b], Anne-Sophie Sikora [b], Annick Gervais [b], Christine Carapito [a,*]

[a] BioOrganic Mass Spectrometry Laboratory (LSMBO), IPHC UMR 7178, University of Strasbourg, CNRS, ProFI – FR2048, Strasbourg, 67000, France
[b] Department of Analytical Development Sciences for Biologicals, UCB Pharma S.A., Braine l'Alleud, 1420, Belgium

ARTICLE INFO

Article history:
Received 14 March 2022
Received in revised form
20 February 2023
Accepted 25 March 2023
Available online 31 March 2023

Keywords:
Host cell proteins
Absolute quantification standards
Data-independent acquisition

ABSTRACT

Monitoring of host cell proteins (HCPs) during the manufacturing of monoclonal antibodies (mAb) has become a critical requirement to provide effective and safe drug products. Enzyme-linked immunosorbent assays are still the gold standard methods for the quantification of protein impurities. However, this technique has several limitations and does, among others, not enable the precise identification of proteins. In this context, mass spectrometry (MS) became an alternative and orthogonal method that delivers qualitative and quantitative information on all identified HCPs. However, in order to be routinely implemented in biopharmaceutical companies, liquid chromatography-MS based methods still need to be standardized to provide highest sensitivity and robust and accurate quantification. Here, we present a promising MS-based analytical workflow coupling the use of an innovative quantification standard, the HCP Profiler solution, with a spectral library-based data-independent acquisition (DIA) method and strict data validation criteria. The performances of the HCP Profiler solution were compared to more conventional standard protein spikes and the DIA approach was benchmarked against a classical data-dependent acquisition on a series of samples produced at various stages of the manufacturing process. While we also explored spectral library-free DIA interpretation, the spectral library-based approach still showed highest accuracy and reproducibility (coefficients of variation < 10%) with a sensitivity down to the sub-ng/mg mAb level. Thus, this workflow is today mature to be used as a robust and straightforward method to support mAb manufacturing process developments and drug products quality control.

1. Introduction

For 30 years now, the monoclonal antibody (mAb) market has remarkably grown up with a plethora of approved antibodies by the U.S. Food and Drug Administration and European Medicines Agency and a current sales market of over $100 billion [1,2]. The high specificity of mAbs to target molecules or antigens and their various mechanisms of action enable their use as pharmaceuticals for a wide range of applications [3]. The high demands of mAbs require the production of well-characterized drug products in terms of the mAb structure and its impurities, namely host cell proteins (HCPs) remaining from the production process. These impurities are included in the critical quality attributes risk assessment as they can affect the product efficacy and the patient's safety by inducing immunogenic reactions [4,5]. Guidelines state classically that the acceptable HCP amount in the final drug product should be below 100 ng/mg mAb [6]. Ultimately, the level of impurities should be as low as possible as issues related to HCPs may arise from specific proteins rather than from overall impurities amounts [7–9]. Of note is that the HCP profile can be affected by numerous upstream process decisions [10] (cell culture duration, feeding strategies or culture temperature) or by the production upscale for commercialisation [11], which highlights the need to be able to finely monitor HCPs throughout all steps of the manufacturing process. Indeed, specific and sensitive analytical methods allowing reaching five to six orders of magnitude dynamics are needed to detect trace level HCPs in the presence of the mAb [12]. Enzyme-linked immunosorbent assays (ELISA) are commonly used for this purpose as they provide the sensitivity and throughput requested [13]. However, ELISA has several limitations as it provides a global amount as an output without individual identification of the HCPs present and its coverage is incomplete [14]. Since immunogenic risk or mAb degradation are related to

Peer review under responsibility of Xi'an Jiaotong University.
* Corresponding author.
E-mail address: ccarapito@unistra.fr (C. Carapito).

https://doi.org/10.1016/j.jpha.2023.03.009

specific HCPs unrelated to their amounts, these drawbacks raise an urgent need for alternative methods.

In this context, mass spectrometry (MS) became the most promising alternative to monitor HCPs allowing risk assessment with individual HCP identification and unbiased quantification. In recent years, liquid chromatography-tandem MS (LC-MS/MS)-based studies have been conducted. On the one hand, data-dependent acquisition (DDA) strategies were successfully applied allowing global HCP profiling and reliable individual HCP quantification down to the sub ng/mg mAb level [15–17]. However, DDA analysis still suffers from stochasticity, the presence of missing values and a discrimination towards the quantification of most abundant proteins, which become significant issues when the HCP impurities are present at trace levels compared to the biotherapeutic. On the other hand, targeted strategies (selected reaction monitoring or paralleled reaction monitoring) were applied for robust and accurate quantification of targeted HCPs and allowed quantification down to the sub ng/mg mAb level [7]. However, the development of a targeted quantification assay is time consuming, compared to the implementation of a global DDA method, and it is still limited to the selection of about hundred targets.

In parallel, advances in MS have highlighted the potential of data-independent acquisition (DIA) on high-resolution/accurate mass instruments. DIA is based on the co-isolation and co-fragmentation of all ions contained in predefined *m/z* windows of variable widths to cover the entire mass range. The acquisition of MS2 signals from all detectable species allows recording complete digital proteome maps while aiming at sensitivity, quantification accuracy and robustness equivalent to pure targeted methods [18]. These advantages make DIA approaches attractive and particularly appropriate for HCP monitoring. However, the bottleneck of DIA MS today still resides in the data processing step. Indeed, each MS2 scan contains the fragments' information of all co-isolated precursors, rendering peptide identifications and further quantitative signals extraction difficult. The use of sample-specific spectral libraries generated from DDA runs to extract quantitative information from DIA data is still the mostly used route for DIA data interpretation, but the generation of spectral libraries requires time and ideally the implementation of prior fractionation of the studied proteome. The recent development of spectral library-free algorithms certainly holds promises to further increase the interest and applicability of DIA strategies for the monitoring of HCPs [15,19].

In addition, if not coupled with isotope dilution, the method needs to allow the estimation of absolute amounts of all individual HCPs. In this regard, the Top3 strategy introduced by Silva et al. [20] in 2006 has been successfully applied in a few studies and the use of three, four, five or seven standard proteins has been reported [15,19,21,22]. Some methods have been developed using a single reference protein while others are based on an average amount calculated from the estimation of each standard protein.

In this context, we developed an original MS-based HCP quantification workflow with improved quantification performances thanks to the use of an internal calibration curve, the HCP Profiler standard [23], and an optimized DIA method on a fast-scanning Quadrupole (Q)-Orbitrap instrument. We applied this workflow to a sample series collected at various stages of the manufacturing process [19].

2. Experimental

2.1. Reagents and material

Crude harvest and post protein A affinity chromatography (PPA) samples of an immunoglobulin-G4 mAb A33 were obtained from a Chinese hamster ovary DG44 cell culture, as described by Husson et al. [19] and in the Supplementary data (Fig. S1). A CHO-DG44 mock cell line sample was provided by UCB Pharma S.A. (Braine l'Alleud, Belgium) to generate the spectral library further used for DIA data extraction. HCP Profiler beads (Anaquant, Villeurbanne, France), first introduced by Trauchessec et al. [23], were spiked in all samples to derive absolute HCPs quantities. In summary, from 18 *Escherichia coli* (*E. coli*) proteins digested and analyzed by LC-MS and tryptic peptides reporting the best MS response were selected. After the confirmation of their specificity to the *E. coli* proteome, 54 selected peptides were adsorbed at known amounts to a water-soluble polymer bead via the READYBEADS™ technology of Anaquant. All chemicals were acquired from Sigma-Aldrich (Sigma-Aldrich, St. Louis, MO, USA).

2.2. MAb quantification

The mAb titer was determined using an Agilent 1100 series high performance liquid chromatography (HPLC) system (Agilent Technologies, Santa Clara, CA, USA) and a 1 mL HiTrap protein G HP column (GE Healthcare Life Sciences, Chicago, IL, USA). The flow rate was set at 1 mL/min and 100 μL of sample were injected. A wash solution composed of 20 mM sodium phosphate at pH 7 was used to clean the column and mAb elution was performed with 20 mM glycine (pH 2.8). The mAb concentration was determined after peaks integration using a standard curve of purified mAb (data not shown).

2.3. Protein quantification

The resuspension of the protein pellets was done in gel loading buffer (10 mM Tris, 1 mM ethylenediaminetetraacetic acid, 5% β-mercaptoethanol, 5% sodium dodecyl sulfate (SDS), 10% glycerol, pH 6.8) and total protein concentration was measured using a RC DC™ Protein Assay kit (Bio-Rad laboratories, Hercules, CA, USA) following manufacturer's protocol.

2.4. Sample preparation

CHO-DG44 mock cell line sample was fractionated onto 12% acrylamide SDS-polyacrylamide gel electrophoresis (SDS-PAGE) for spectral library generation. Harvest and PPA samples were stacked in a single band for HCP quantification. The 24 gel bands of the fractionated mock cell line sample and stacked bands were cut into small pieces. Proteins were in-gel reduced with 10 mM dithiothreitol (Sigma-Aldrich) for 30 min at 60 °C. Alkylation was performed with 55 mM iodoacetamide (Sigma-Aldrich) for 30 min in the dark. Then trypsin (Promega, Madison, WI, USA) was added to a 1:50 enzyme:substrate ratio (we estimated 1 μg of proteins in each band of the mock cell line fractionation). Samples were incubated overnight at 37 °C (14 h). Peptides were extracted from gel bands using 60% acetonitrile (ACN; Sigma-Aldrich) and 0.1% formic acid (FA; Sigma-Aldrich) for 1 h under agitation and a second step with 100% ACN for 1 h. After vacuum drying, samples were resuspended in 2% ACN and 0.1% FA to a final protein concentration of 0.4 μg/μL. In all samples, retention time standards (indexed retention time (iRT) kit, Biognosys, Schlieren, Switzerland) were spiked. For the HCP Profiler quantification (Anaquant), one bead was spiked in 150 μL of 0.2 ng/μL protein solution. For the mix of standard proteins, four accurately quantified standard proteins (on column 10 fmol of yeast alcohol dehydrogenase (ADH, P00330), 2 fmol of rabit phosphorylase b (PYGM, P00489), 0.5 fmol of bovine serum albumin (BSA, P02769) and 0.2 fmol of yeast enolase (ENL, P00924)) from the MassPREP Digestion Standard Kit (Waters, Milford, CT, USA) were spiked.

2.5. NanoLC-MS/MS acquisitions

DDA and DIA acquisitions were performed on a NanoAcquity ultra-high performance liquid chromatography (UPLC) device (Waters) coupled to a Q-Exactive HF-X mass spectrometer (Thermo Fisher Scientific Inc., Bremen, Germany). Mobile phase A was 0.1% (*V/V*) FA in water and mobile phase B was 0.1% (*V/V*) FA in ACN. The equivalent of 400 ng of proteins was trapped onto a Symmetry C_{18} precolumn (20 mm × 180 μm, 5 μm; Waters) and eluted on an Acquity UPLC BEH130 C_{18} column (250 mm × 75 μm, 1.7 μm; Waters). A 115 min chromatographic gradient (2%–35% B in 95 min, 35%–80% B in 1 min, 80% B for 5 min, 80%–2% B in 1 min and maintained 2% B for 13 min) was applied at 400 nL/min, with a column temperature set at 60 °C. The Q-Exactive HF-X source temperature was set at 250 °C and spray voltage to 2 kV. The system was fully controlled by XCalibur software v4.0.27.19, 2013 (Thermo Fisher Scientific Inc.) and NanoAcquity UPLC console v1.51.3347 (Waters). The three injection replicates of DIA and DDA were performed in a randomized injection sequence. The MS proteomics data have been deposited to the ProteomeXchange consortium via the PRIDE partner repository with the dataset identifier PXD029305 [24].

2.6. DDA acquisition

Full scan MS spectra (*m/z* 375–1500) were acquired in positive mode at a resolution of 120,000 at *m/z* 200, a maximum injection time of 60 ms and an automatic gain control (AGC) target value of 3×10^6. The 10 most intense multiply charged peptides per full scan (charge states ≥ 2) were isolated using a *m/z* 2 window and fragmented using higher energy collisional dissociation (normalized collision energy set at 27). MS/MS spectra were acquired with a resolution of 15,000 at *m/z* 200, a maximum injection time of 60 ms and an AGC target value of 1×10^5, and dynamic exclusion was set to 40 s.

2.7. DIA acquisition

Full-scan MS spectra were collected from *m/z* 350–1500 at a resolution of 60,000 at *m/z* 200 with an AGC target fixed at 3×10^6 and a maximum injection time of 60 ms. Fragments analysis (MS/MS) was subdivided into 40 windows of variable widths. Two acquisition methods were developed for harvest and PPA samples (Tables S1 and S2). Resolution was set to 30,000 at *m/z* 200 and AGC target was fixed at 1×10^6 with an automatic maximum injection time.

2.8. DDA data treatment

Raw DDA files were converted to .mgf peaklists using MsConvert and were submitted to Mascot database search on a local server (version 2.5.1, MatrixScience, London, UK) against a FASTA database including all *Critecutulus griseus* entries extracted from UniProtKB/TrEMBL (56,566 protein entries, February 15, 2021) together with their reversed sequences, as well as the iRT retention time standards, the four standard proteins of the MassPREP Digestion Standard Kit, HCP Profiler kit proteins, the mAb heavy and light chains and common contaminants. Spectra were searched with a mass tolerance of 5 ppm in MS mode and 0.05 Da in MS/MS mode. One trypsin missed cleavage was tolerated. Carbamidomethylation of cysteine residues was set as fixed modification. Oxidation of methionine residues and acetylation of proteins N-termini were set as variable modifications. Identification results were imported into Proline software version 1.6 (http://proline.profiproteomics.fr) for validation [25]. A false discovery rate (FDR) of 1% was set at the peptide level using adjusted e-value and the protein level using Mascot modified mudpit scores. Peptide abundances were extracted with Proline software using an extraction *m/z* tolerance and peptide spectrum matches/peak matching *m/z* tolerance of 5 ppm. Alignment of the LC-MS runs was performed using loess smoothing, peptide identity method and with a time tolerance of 300 s. Cross assignment of peptide ions abundances was performed among harvest or PPA samples using a *m/z* tolerance of 5 ppm and a retention time tolerance of 40 s.

2.9. Cricetulus griseus, CHO spectral library generation

A reference spectral library combining a series of analyses conducted on different samples in DDA mode was generated using the Spectronaut and Pulsar algorithms (v.14.5; Biognosys). This series of analyses comprised the 24 gel bands obtained by SDS-PAGE fractionation of the CHO DG44 mock cell line and all DDA analyses of harvest and PPA samples, including iRT retention time standards and the 18 proteins from the HCP Profiler kit. Raw DDA files were uploaded into Spectronaut and searched with the Pulsar algorithm against a FASTA database containing all *Critecutulus griseus* entries extracted from UniProtKB/TrEMBL (56,566 protein entries, February 15, 2021), as well as the iRT retention time standards, the 18 proteins from HCP Profiler kit, the reference sequence of the mAb and common contaminants. Trypsin/P enzyme was used and one missed cleavage was tolerated. Carbamidomethylation of cysteine residues was set as fixed modification. Oxidation of methionine residues and acetylation of proteins N-termini were set as variable modifications. MS and MS/MS mass tolerances were set in dynamic mode. The spectral library was validated as follows: a FDR of 0.01 was set at peptide spectrum matches, peptides and proteins levels. Fragment ions window was set between *m/z* 300 and 1800 with four to six fragments per precursor.

2.10. DIA data treatment

DIA data was analyzed with a peptide-centric approach using the Spectronaut algorithm and the upper described in house generated spectral library (v.14.5; Biognosys). Trypsin/P was used as digestion enzyme with one missed cleavage allowed. Carbamidomethylation of cysteine residues was set as a fixed modification. Oxidation of methionine residues and acetylation of proteins' N-termini were set as variable modifications. For quantitative data extraction, MS and MS/MS mass tolerances, extracted ion chromatogram (XIC) and retention time windows were all set as dynamic. iRT regression type was set to local (non-linear) regression. A FDR of 1% was set at precursors and proteins levels. At this extraction stage, a sparse Q-value filter was applied. Peptide quantities corresponding to the sum of four to six fragments XIC areas (interference correction parameter was turned on) were calculated. Precursors with a Q-value below 0.01 were used for iRT profiling.

2.11. HCP Top3 quantification

After data extraction, a list of identified peptides with their corresponding intensities was exported in Excel format for both DDA and DIA data. Prior the Top3 quantification, filters were applied to remove oxidized and acetylated peptides alongside with their non-modified counterparts. Precursors inferred to host organism proteins, standard proteins and precursors with charge states 2 and 3 were kept. For DDA data, a maximum of one precursor validated by cross-assignment was allowed. Precursors with more than one Q-value > 0.01 or profiled were removed for DIA data. For both acquisition methods, quantity estimation was performed using precursors' intensities showing a coefficient of variation (CV) below 20% within injection

triplicates. Finally, HCP peptides showing 100% sequence identity with a semi-tryptic or non-tryptic mAb peptide were removed. After applying these stringent validation filters in DDA and DIA modes, peptide intensities were obtained by summing all precursor intensities and protein intensities by summing the three most intense peptides intensities. For HCP Profiler quantification, a calibration curve of the log2 (Top3 standard peptides abundance) in function of log2 (standard proteins quantity) is obtained and allowed estimating protein mol quantities. For the four standard proteins mix, the universal signal response factor (MS signal/mol of protein) was calculated using PYGM as a reference, and allowed estimating protein mol quantities. Finally, for both quantification methods, protein molecular weights and injected mAb quantity were used to estimate individual HCP ng/mg mAb amounts.

3. Results and discussion

3.1. A multi-stage HCP profiling workflow optimization

A sample set described in Husson et al. [19] including two levels of HCP complexity was used to investigate bioprocess developments by MS. Two cell culture durations (seven and ten days), three harvest procedures (no shear, low shear or high shear) and two protein A purification protocols (one standard and one including intermediate column washes with 25 mM Tris, 10% isopropanol, 1 M urea, pH 9) were investigated resulting in four HCP-rich harvest samples and seven purified PPA fractions (Fig. S1). Fig. 1 summarizes the different levels of optimization/benchmarking that were conducted. First, two quantification methods were benchmarked. On one hand, samples were spiked with a mixture of four standard proteins

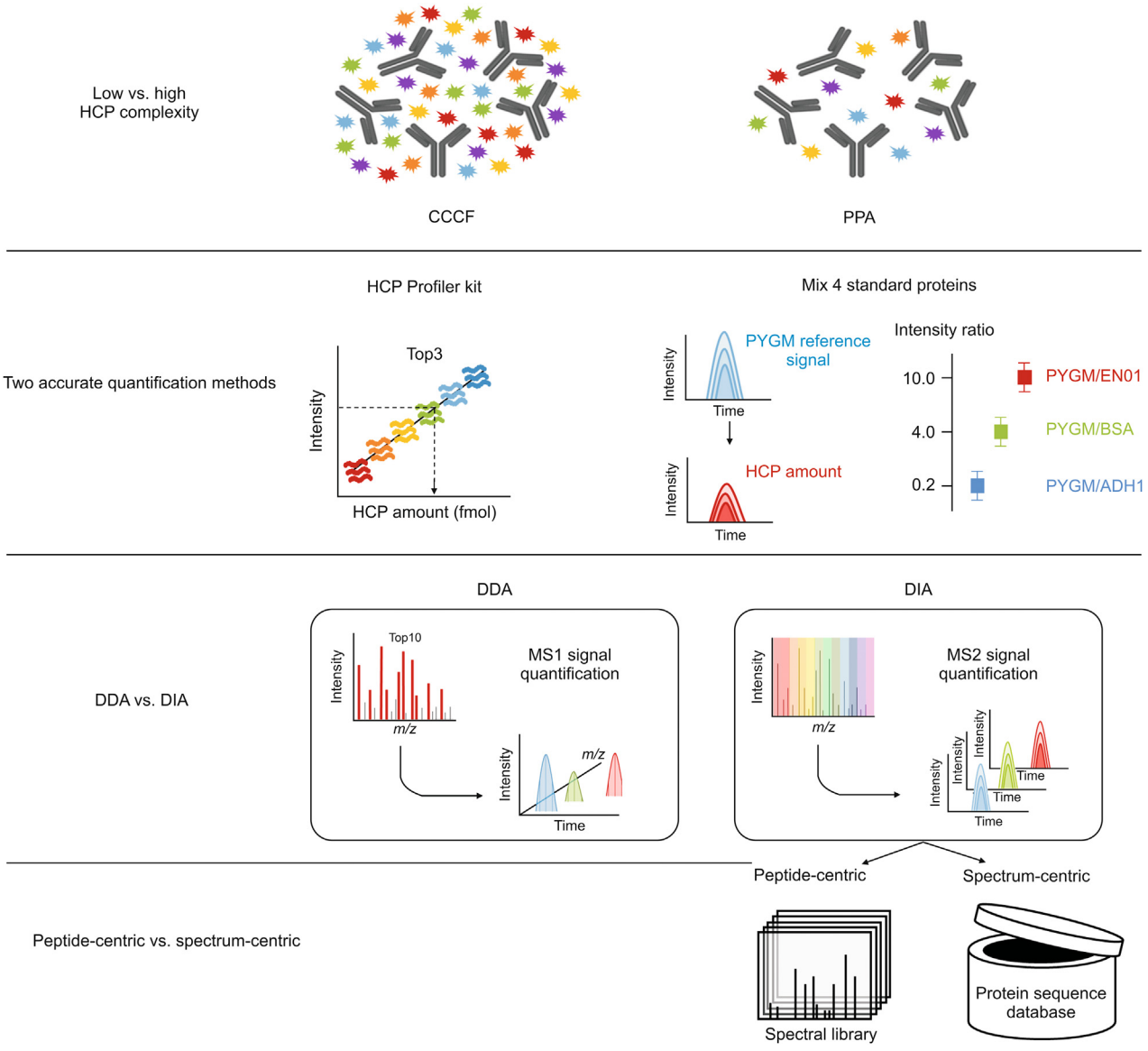

Fig. 1. Experimental design for the optimization of a robust mass spectrometry (MS)-based quantification strategy for host cell protein (HCP) monitoring. CCCF: clarified cell culture fluid; PPA: post protein A; PYGM: rabit phosphorylase b; ENO1: yeast enolase 1; BSA: bovine serum albumin; ADH1: yeast alcohol dehydrogenase; DDA: data-dependent acquisition; DIA: data-independent acquisition.

PYGM, ADH, BSA and ENL from Waters (Mix 4P), that was previously applied to derive HCP amounts with MS methods [21]. On the other hand, samples were spiked with an original mixture of peptides coated on a water-soluble bead releasing controlled amounts of a 2.5 log peptide range after solubilization (READYBEADS™ technology). Second, DIA methods were finely tuned on a fast-scanning Q-Orbitrap instrument to thoroughly compare the performances achieved with DIA methods against more classical DDA approaches on the same instrument. Finally, two DIA data extraction and interpretation strategies, a peptide-centric approach requiring the prior acquisition of a reference spectral library and a library-free spectrum-centric approach, were evaluated.

3.2. Implementation of the original HCP profiler standard for more accurate HCP quantification

Top3 quantification strategies assume that the sum of the MS response of the three best responding peptides per mole of protein is constant within a CV of less than 10%. Starting from this assumption, an internal standard can be used to calculate a signal response factor (Top3 peptides signal/mol) and to estimate an absolute amount of each individual HCP (HCP Top3 peptides signal/signal response factor). The use of the Mix 4P has been previously reported by others and used for HCP quantification, considering the PYGM protein as a reference and the three other proteins (ADH, BSA, and ENL) to calculate ratios, as internal controls [15,19,21]. However, using a single standard protein to derive absolute amounts of HCPs covering a large range of abundances is not ideal. Indeed, the standard protein used to derive the absolute amount of a given protein should be close to a ratio of 1 in abundance with the given protein to be quantified. Therefore, the development of finely tuned standards for accurate HCP quantification using MS methods is a valuable challenge for the field. In this context, we have implemented and evaluated an original standard enabling the inclusion of an internal calibration curve in each sample. This standard based on the READYBEADS™ technology developed by Anaquant [23] is composed of a water-soluble polymer bead, which releases unlabeled peptides at known amounts. Eighteen tripeptides distributed over six concentration points ranging from 1 to 500 fmol; and thus a total of 54 peptides ranging over 2.5 orders of magnitude are adsorbed on the bead. The extracted ion chromatograms of those 54 peptides allow building an internal calibration curve that can then be further used to derive each individual HCP amount. The robustness and reproducibility of this standard were assessed by CVs calculated on the slopes, intercepts and R^2 of the 33 calibration curves obtained on the 11 samples, and all CVs were all below 2.2% (Table S3 and Fig. S2). On average, 1464 HCPs were quantified in harvest fractions with global quantities between 222,646 and 365,145 ng/mg mAb, and 115 HCPs in PPA fractions representing 569 to 19,153 ng/mg mAb (Fig. 2). Overall, the quantification results obtained with the HCP Profiler and the Mix 4P are consistent in the sense that similar conclusions can be drawn regarding the manufacturing process impact on the HCP Profiles. However, the global HCP amounts are in general higher using the HCP Profiler standard, except for PPA 5 and harvest 1 samples, while the numbers of HCPs quantified were lower (on average 34% and 13% less HCPs quantified for PPA and harvest fractions, respectively). In order to understand this overall difference in derived HCP amount, we compared the individual amounts obtained for all HCPs quantified with both methods (Fig. S3). The ratios between both strategies were consistent with a median ratio of 1.3% and 78% of the 5305 ratios spanning within a factor 2. A closer look into individual and known-to-be problematic HCPs is illustrated in Fig. S4 and supports comparable individual quantities estimated for: serine protease HTRA1 (G3IBF4_CRIGR) known for its protease activity [26,27], putative phospholipase B-like 2 (G3I6T1_CRIGR) known to be immunogenic [28] and clusterin (G3HNJ3_CRIGR) known to be difficult to remove [29], like the other two. Since individual peptides' ionization efficiencies and response factors vary, taking into account the MS response of 54 peptides spiked over a large concentration range drawing an abundance-related calibration curve rather than of only 3 peptides from a single protein ultimately leads to a more accurate amount estimation. This explains the overall differences and more reliable amounts derived from the HCP Profiler standard. Conversely, the slightly reduced numbers of quantified HCPs with HCP Profiler, more noticeable in the less complex PPA fractions, may be imputed to competition/suppression effects due to the larger concentration range of spiked standards. However, robustness and reliability of quantification may prevail over coverage and the HCP Profiler presents further advantages. Indeed, the 54 peptides can also be used as LC-MS/MS quality controls serving as retention time and intensity

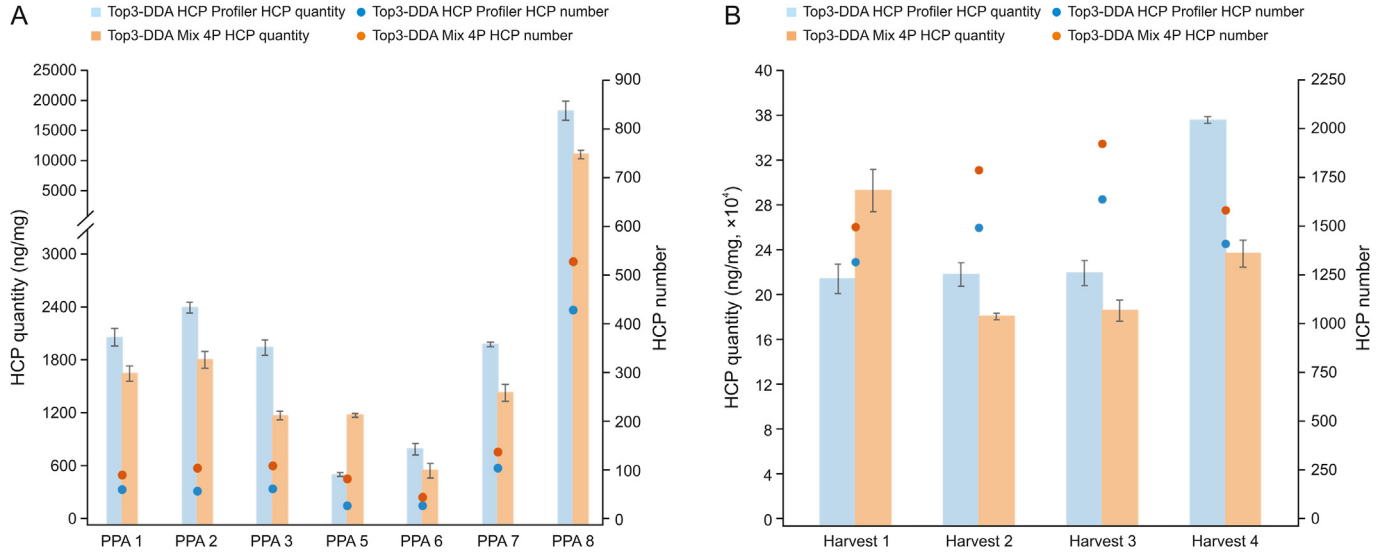

Fig. 2. Benchmarking of the host cell protein (HCP) Profiler standard against the Mix 4P using a MS1-XIC data-dependent acquisition (DDA) approach. Comparison of global HCPs numbers and amounts obtained for (A) post protein A affinity chromatography (PPA) and (B) harvest fractions on a Q-Exactive HF-X using both methods. Bar heights represent the means of the global HCP amounts in injection triplicates. Error bars represent the standard deviation. Dots indicate the numbers of quantified HCPs.

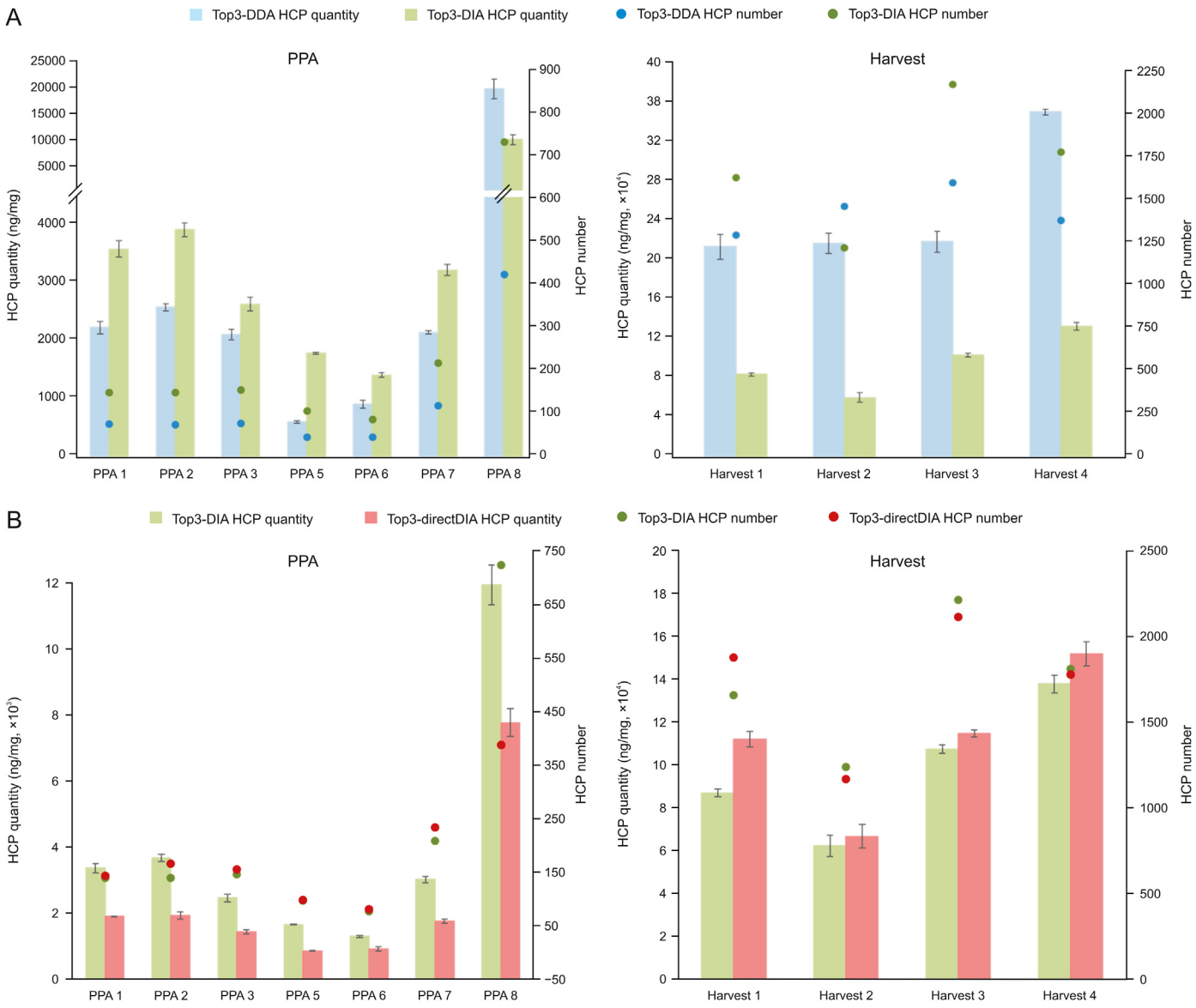

Fig. 3. Evaluation of data-independent acquisition (DIA) against data-dependent acquisition (DDA) strategies for global host cell protein (HCP) profiling using the HCP Profiler standard. (A) Quantification results obtained for post protein A affinity chromatography (PPA) and harvest fractions. Bar heights represent the means of the global HCP amounts in injection triplicates. Error bars represent the standard deviation. (B) Numbers of HCPs quantified per sample for PPA and harvest fractions.

peptide anchors throughout injection series. Moreover, the demonstrated high reproducibility of the calibration curves, unrelated to the sample complexity, ensures a broad applicability of the method all over the mAb manufacturing process. Finally, due to its easy and ready-to-use characteristics, this original standard will allow avoiding user-induced analytical biases that may occur while preparing the mixtures of standard proteins at known amounts. For all those reasons and while considering the use of isotope dilution with highly purified heavy labeled standards as the gold standard method for absolute quantification of key HCPs [7,19], the HCP Profiler standard offers a valuable compromise to estimate absolute amounts for all detectable HCPs while providing the best overview of the overall HCPs content.

3.3. Implementation of a DIA method for improved HCP profiling

In order to combine the original quantification standard previously described with highest performing MS acquisition methods, we developed a dedicated MS2-based DIA approach and benchmarked it against DDA data acquired in parallel on the same samples and instrument. We first generated the most comprehensive spectral library from a fractionated CHO DG44 mock cell line sample combined with DDA runs acquired on all harvest and PPA fractions in order to get full advantages of a peptide-centric DIA approach. The combination of these DDA data allowed us to generate a project-specific spectral library containing 40,281 peptides derived from 3978 protein groups. Then, DIA variable isolation windows methods were developed for both sample types based on the distribution of precursors over the *m/z* acquisition range observed in DDA, the MS and MS/MS scan times of the instrument and the theoretical number of MS cycles per chromatographic peak. Thus, two methods composed of 40 variable isolation windows, one dedicated to HCP-rich harvest samples and a second one dedicated to PPA samples, were developed (Fig. S5 and Tables S1 and S2). After data acquisition and peptides signals extractions, we applied stringent validation filters. The first filter applied acts as a signal quality filter: precursors with more than one Q-value > 0.01 and/or profiled were removed from DIA results and those with more than

one cross-assigned attribution were removed from DDA data. The second quality filter applied refers to the reproducibility of the signals as all precursors with CVs above 20% were excluded. Then, a sequence homology filter was applied using the BLASTP [30] (v.2.10.0+) algorithm run against the mAb heavy and light chain sequences. HCP peptides showing 100% sequence identity with a semi-tryptic or non-tryptic mAb peptide were removed. While this last filter does not have a major impact, it is nonetheless important. Indeed, a mAb peptide resulting from an unspecific trypsin cleavage, wrongly attributed to an HCP sequence, would lead to a significant overestimation of this HCP amount and eventually even of the overall HCP amount. Finally, the selection of the three most intense peptides per protein was performed and absolute HCP amounts were estimated using the HCP Profiler standard.

An average of 1737 HCPs with a global estimation between 62,792 and 138,297 ng/mg mAb for harvest samples and 221 HCPs in PPA fractions with a quantity between 1339 and 11,992 ng/mg mAb were obtained by DIA (Fig. 3). In comparison to DDA results, the overall estimated HCP amounts in DIA were higher for PPA samples, while the HCP-rich harvest fractions showed lower overall amounts in DIA compared to DDA. This later observation potentially highlights the presence of highly interfered MS1 signals in DDA that may result in an overestimation of the HCP amounts in highly complex harvest samples. We observed a significant benefit of DIA on the HCPs coverage, as approximately 19% more HCPs were quantified for harvest and 111% more for PPA samples. An increased intra-HCP dynamic was also noted: while DDA was able to achieve 2.4 to 5.5 orders of magnitude within the least and most abundant HCP, MS2-based DIA allowed to reach a dynamic between 3.5 and 6.1 (Table S4). Furthermore, the precision of the DIA data extraction was assessed by the CVs calculated on all HCP peptides' intensities within injection replicates. A median of 9.0% was obtained for DIA compared to 9.8% for DDA, highlighting a slight improvement of the data extraction when taking into account that three times more peptides were quantified by DIA, i.e., 21,403 and 6661 peptides respectively (Fig. S6). When focusing on the 3779 HCPs commonly quantified in the eleven samples by both methods, only 52% of the ratios of the quantity obtained using DIA versus DDA are within a factor of 2 with a median of 0.79 (Fig. S7A). Among these common proteins, 1735 HCPs were quantified with a Top3 in DIA while this number drops to 1648 in DDA (Fig. S7B). As we sum the areas of the Top3 peptides, the increased number of peptides per HCP obtained in DIA has a direct and positive impact on the HCPs quantification. Overall, our HCP Profiler-DIA method has demonstrated its ability to extract signals close to the background noise, which is a significant benefit when HCP impurities of interest are present at trace levels. Compared to results obtained with a standard ELISA assay, our HCP Profiler-DIA method shows global quantities higher by a factor 4 on average for PPA 1 to 7 and reached a factor 32 for PPA 8 obtained with a modified protocol (Fig. S8). Husson et al. [19] attributed this increased amount of HCPs in PPA 8 to the longer culture duration (10 days vs. 7 days) and a modified PPA affinity chromatography protocol (including intermediate column washes with 25 mM Tris, 10% isopropanol, 1 M urea, pH 9) that led to the drop of some HCPs, such as the immunogenic phospholipase B-like 2 protein (Fig. S4B), and concomitantly to a higher diversity/number of HCPs quantified. These results have been previously highlighted by others and imputed to the limitations of ELISA assays, including the inability to detect non-immunogenic or degraded HCPs [31,32]. Altogether, the increased HCP map coverage and the accurate and reproducible quantification capabilities of our HCP-Profiler DIA method make it a valuable approach to achieve a reliable overview of the HCP content of samples from various steps of the mAb manufacturing process, with a sensitivity down to sub-ng/mg mAb.

3.4. Evaluation of a spectral library-free DIA data interpretation strategy

The generation of a spectral library requests sample preparation and instrument acquisition time. Even though this takes less time than the development of a specific ELISA kit, a new dedicated library must be generated for each drug product. In addition, HCPs not covered by the library will not be searched for, even if they are present in detectable amounts. Promising alternative approaches for DIA data extraction have been introduced to avoid the use of spectral libraries. With the aim to make our DIA method even more straightforward, we evaluated the spectrum-centric approach called directDIA in Spectronaut. This approach relies on the generation of pseudo MS/MS spectra from the DIA runs which will then be used to query a search against the organism's database as it is done in classical DDA search engines. We assessed the performance of the directDIA strategy against our previously described peptide-centric approach. At the level of the HCP Profiler standard, comparable reproducibility of the signal extraction was achieved as CV values calculated on the slopes, intercepts and R^2 of the 33 calibration curves were all below 2% (Table S3). The results were also very consistent regarding HCPs contents as comparable quantification results and close numbers of quantified HCPs were obtained (Fig. 3). In addition, equivalent sensitivity down to the sub-ng/mg mAb level was achieved (Table S4). Both extraction strategies were able to achieve 3 to 6 orders of magnitude between the least and most abundant HCP. Looking further at individual HCP amounts, we compared the quantities obtained using DIA over directDIA for the 5833 HCPs commonly detected with both methods. A median of 0.96 was obtained and the amounts estimated were in accordance with 82% of the common HCPs within a factor 2 (Fig. S9A). Similarly, close numbers of peptides were used to quantify those common HCPs, respectively 11,436 and 11,090 for DIA and directDIA and proteins were identified with comparable numbers of peptides with both methods (Fig. S9B). Furthermore, when overlapping the lists of individual HCPs quantified, Venn diagrams show good correlation of both DIA strategies (Fig. S10). However, HCPs quantified with both DIA and directDIA methods only represent 52% of all quantified HCPs. On the one hand, 800 HCPs quantified using the spectral library could not be extracted by directDIA, which highlights the important room of improvement that still exists for spectrum-centric search algorithms. On the other hand, 974 HCPs were missed using the spectral library-based approach. On average, 260 HCPs for each harvest sample were not identified in the spectral library (Fig. S11). These HCPs cover about 5 orders of magnitude between the least and most abundant HCP with a maximum reaching thousands of ng/mg mAb (Table S5). The numbers and quantities of these HCPs are not negligible and thus also enlighten the limitations of the spectral library extraction approach as HCPs not present in the library will neither be identified nor quantified. However, spectral library free results should still be taken with caution, as one could seriously argue that HCPs with an estimated amount around thousands of ng/mg mAb should be identified in the spectral library with at least a few peptides. The number of false positives must still be significant in directDIA although difficult to properly estimate yet. In addition, as for the generation of the reference spectral library, the spectral library free-based approach also strongly relies on the use of the CHO protein sequence database, which is yet poorly annotated and curated (56,565 entries in UniProtKB/TrEMBL). As a result, the high redundancy of the database likely significantly hinders the extraction of specific peptides. As a conclusion, while already showing promising results, the current limitations of the spectral library-free approach make it premature to be readily implemented in a regulated environment working with CHO cell-based bioproducts, but this may well change in a near future.

4. Conclusion

We have demonstrated that the original HCP Profiler standard offers a ready-to-use solution for accurate HCP quantification with its internal six-points calibration curves representing a move towards the standardization ultimately requested for the implementation of MS-based methods in a biopharmaceutical environment. We prove here that the standard undergoes no loss of performance when combined with a DIA workflow on a Q-Orbitrap instrument. Our study also demonstrates again the advantages of MS methods over ELISA assays. On average, our MS-based approach identifies each HCP with 5 peptides containing 7 to 28 amino acids, which could be considered as the equivalent of 5–14 epitopes. Therefore, a finer granularity of results is obviously achieved with the main difference being the detection principle. MS allows to avoid the inherent gaps in the ELISA tests such as the absence of specific polyclonal antibodies or impaired binding due to loss of conformational epitopes. Similarly, we could argue on the gap related to the spectral library-based DIA approach, which is restricted to identifying HCPs that are present in the library. However, the development of a new spectral library takes less time than the months requested to generate a specific ELISA kit. In addition, anti-HCP antibodies are perishable and have to be reproduced by a new immunization campaign whenever needed. By contrast, within a week, it is possible to generate a comprehensive spectral library specific to the mAb produced, namely a library generated from the analysis of a mock cell line and/or HCP-rich harvest samples. This library can be updated endlessly with new analyses in case of manufacturing process changes suspected to lead to the presence of new HCPs. Thus, once generated, the spectral library can be unlimitedly used to extract signals from any DIA analysis. Altogether, the combination of the HCP Profiler with an optimized spectral library-based DIA method presents a sufficient robustness to consider its implementation within a biopharmaceutical environment to support process development or batch-to-batch consistency. In a short-term perspective, spectral library-free approaches, which are more straightforward and non-dependent on any prior information, will likely become the best suited way to support the release of safer bioproducts.

CRediT author statement

Steve Hessmann: Conceptualization, Methodology, Software, Formal analysis, Investigation, Writing - Original draft preparation, Visualization; **Cyrille Chery:** Writing - Reviewing and Editing, Supervision; **Anne-Sophie Sikora:** Writing - Reviewing and Editing, Supervision; **Annick Gervais:** Writing - Reviewing and Editing, Supervision, Project Administration; **Christine Carapito:** Conceptualization, Methodology, Validation, Writing - Original draft preparation, Reviewing and Editing, Project Administration.

Declaration of competing interest

The authors declare that there are no conflicts of interest.

Acknowledgments

This project was supported by the “Association Nationale de la Recherche et de la Technologie” and UCB Pharma S.A. (Belgium and France) via the CIFRE fellowship of Steve Hessmann. This work was supported by the “Agence Nationale de la Recherche” via the French Proteomic Infrastructure ProFI FR2048 (ANR-10-INBS-08-03). The authors thank Laura Herment and Tanguy Fortin from Anaquant for their support in the use of the HCP Profiler standard.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jpha.2023.03.009.

References

[1] A.L. Grilo, A. Mantalaris, The increasingly human and profitable monoclonal antibody market, Trends Biotechnol. 37 (2019) 9–16.
[2] H. Kaplon, A. Chenoweth, S. Crescioli, et al., Antibodies to watch in 2022. mAbs 14 (2022), 2014296.
[3] M. Suzuki, C. Kato, A. Kato, Therapeutic antibodies: Their mechanisms of action and the pathological findings they induce in toxicity studies, J. Toxicol. Pathol. 28 (2015) 133–139.
[4] T. Cui, B. Chi, J. Heidbrink Thompson, et al., Cathepsin D: Removal strategy on protein A chromatography, near real time monitoring and characterisation during monoclonal antibody production, J. Biotechnol. 305 (2019) 51–60.
[5] B. Yang, W. Li, H. Zhao, et al., Discovery and characterization of CHO host cell protease-induced fragmentation of a recombinant monoclonal antibody during production process development, J. Chromatogr. B 1112 (2019) 1–10.
[6] R. Molden, M. Hu, E.S. Yen, et al., Host cell protein profiling of commercial therapeutic protein drugs as a benchmark for monoclonal antibody-based therapeutic protein development, mAbs 13 (2021), 1955811.
[7] X. Gao, B. Rawal, Y. Wang, et al., Targeted host cell protein quantification by LC-MRM enables biologics processing and product characterization, Anal. Chem. 92 (2020) 1007–1015.
[8] X. Li, Y. An, J. Liao, et al., 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. 37 (2021), e3128.
[9] S.S. Rane, R.J. Dearman, I. Kimber, et al., Impact of a heat shock protein impurity on the immunogenicity of biotherapeutic monoclonal antibodies, Pharm. Res. 36 (2019), 51.
[10] H. Falkenberg, D.M. Waldera-Lupa, M. Vanderlaan, et al., Mass spectrometric evaluation of upstream and downstream process influences on host cell protein patterns in biopharmaceutical products, Biotechnol. Prog. 35 (2019), e2788.
[11] D.S. Chahar, S. Ravindran, S.S. Pisal, Monoclonal antibody purification and its progression to commercial scale, Biologicals 63 (2020) 1–13.
[12] K.N. Valente, N.E. Levy, K.H. Lee, et al., Applications of proteomic methods for CHO host cell protein characterization in biopharmaceutical manufacturing, Curr. Opin. Biotechnol. 53 (2018) 144–150.
[13] J. Zhu-Shimoni, C. Yu, J. Nishihara, et al., Host cell protein testing by ELISAs and the use of orthogonal methods, Biotechnol. Bioeng. 111 (2014) 2367–2379.
[14] M. Vanderlaan, J. Zhu-Shimoni, S. Lin, et al., Experience with host cell protein impurities in biopharmaceuticals, Biotechnol. Prog. 34 (2018) 828–837.
[15] N. Pythoud, J. Bons, G. Mijola, et al., 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. 20 (2021) 923–931.
[16] D.E. Walker, F. Yang, J. Carver, et al., A modular and adaptive mass spectrometry-based platform for support of bioprocess development toward optimal host cell protein clearance, mAbs 9 (2017) 654–663.
[17] F. Yang, D.E. Walker, J. Schoenfelder, et al., A 2D LC-MS/MS strategy for reliable detection of 10-ppm level residual host cell proteins in therapeutic antibodies, Anal. Chem. 90 (2018) 13365–13372.
[18] F. Zhang, W. Ge, G. Ruan, et al., Data-independent acquisition mass spectrometry-based proteomics and software tools: A glimpse in 2020, Proteomics 20 (2020), e1900276.
[19] G. Husson, A. Delangle, J. O'Hara, et al., Dual data-independent acquisition approach combining global HCP profiling and absolute quantification of key impurities during bioprocess development, Anal. Chem. 90 (2018) 1241–1247.
[20] J.C. Silva, M.V. Gorenstein, G.-Z. Li, et al., Absolute quantification of proteins by LCMSE: A virtue of parallel MS acquisition, Mol. Cell. Proteomics 5 (2006) 144–156.
[21] C.E. Doneanu, A. Xenopoulos, K. Fadgen, et al., Analysis of host-cell proteins in biotherapeutic proteins by comprehensive online two-dimensional liquid chromatography/mass spectrometry, mAbs 4 (2012) 24–44.
[22] D. Li, A. Farchone, Q. Zhu, et al., Fast, robust, and sensitive identification of residual host cell proteins in recombinant monoclonal antibodies using sodium deoxycholate assisted digestion, Anal. Chem. 92 (2020) 11888–11894.
[23] M. Trauchessec, A.M. Hesse, A. Kraut, et al., An innovative standard for LC-MS-based HCP profiling and accurate quantity assessment: Application to batch consistency in viral vaccine samples, Proteomics 21 (2021), e2000152.
[24] E.W. Deutsch, A. Csordas, Z. Sun, et al., The ProteomeXchange consortium in 2017: Supporting the cultural change in proteomics public data deposition, Nucleic Acids Res. 45 (2017) D1100–D1106.
[25] D. Bouyssié, A.M. Hesse, E. Mouton-Barbosa, et al., Proline: An efficient and user-friendly software suite for large-scale proteomics, Bioinformatics 36 (2020) 3148–3155.
[26] N. Aboulaich, W.K. Chung, J.H. Thompson, et al., A novel approach to monitor clearance of host cell proteins associated with monoclonal antibodies, Biotechnol. Prog. 30 (2014) 1114–1124.

[27] J. Pezzini, G. Joucla, R. Gantier, et al., Antibody capture by mixed-mode chromatography: A comprehensive study from determination of optimal purification conditions to identification of contaminating host cell proteins, J. Chromatogr. A 1218 (2011) 8197–8208.
[28] B. Tran, V. Grosskopf, X. Wang, et al., Investigating interactions between phospholipase B-like 2 and antibodies during Protein A chromatography, J. Chromatogr. A 1438 (2016) 31–38.
[29] N.E. Levy, K.N. Valente, L.H. Choe, et al., Identification and characterization of host cell protein product-associated impurities in monoclonal antibody bioprocessing, Biotechnol. Bioeng. 111 (2014) 904–912.
[30] C. Camacho, G. Coulouris, V. Avagyan, et al., BLAST+: Architecture and applications, BMC Bioinformatics 10 (2009), 421.
[31] A. Farrell, S. Mittermayr, B. Morrissey, et al., Quantitative host cell protein analysis using two dimensional data independent LC-MS^{E}, Anal. Chem. 87 (2015) 9186–9193.
[32] S. Kreimer, Y. Gao, S. Ray, et al., Host cell protein profiling by targeted and untargeted analysis of data independent acquisition mass spectrometry data with parallel reaction monitoring verification, Anal. Chem. 89 (2017) 5294–5302.