

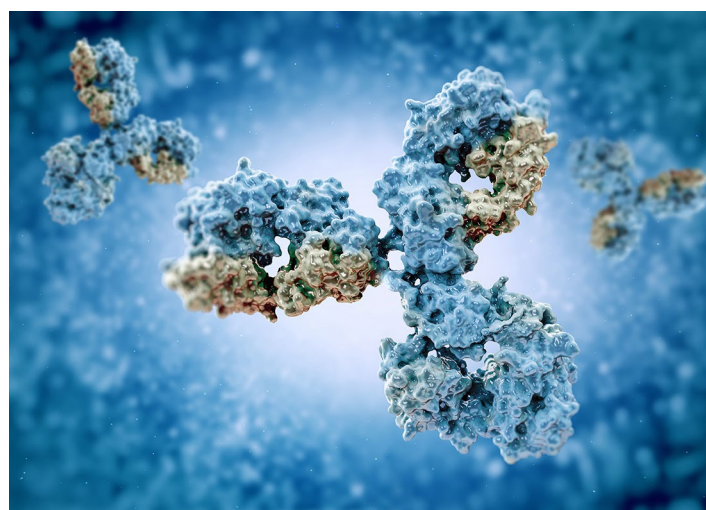
Comprehensive identification and label-free quantitation of host cell protein contaminants using BioPharma Finder 4.1 software

Authors: Silvia Millán-Martín,¹ Sara Carillo,¹ Jonathan Bones¹, Jennifer Sutton²

¹National Institute for Bioprocessing Research and Training (NIBRT), Dublin, Ireland

²Thermo Fisher Scientific

Keywords: NIBRT, biopharmaceutical, biotherapeutic, monoclonal antibody (mAb), host cell protein (HCP), downstream process (DSP), critical quality attributes, IgG, peptide mapping



Key benefits

- Peptide analysis by liquid chromatography–high-resolution accurate mass spectrometry (LC-HRAM-MS) offers orthogonal solutions for detection and monitoring of residual host cell proteins (HCPs) compared to immunological methods with the unbiased discovery of HCP impurities and subsequent quantitation.
- HRAM MS data combined with Thermo Scientific™ BioPharma Finder™ 4.1 software provides comprehensive HCP qualitative and quantitative analysis.
- Using label-free MS quantification, the removal of HCPs during a monoclonal antibody (mAb) downstream process was evaluated after different Protein A purification cycles on the same column.

Goal

To detect, identify, and quantify potential host cell proteins (HCPs) present in an investigational IgG1 mAb and to monitor HCP profile changes during the Protein A purification process by using BioPharma Finder 4.1 software through a peptide mapping analysis experiment that incorporates the host cell protein analysis feature. The study demonstrates the integrated data analysis workflow provides confident and simplified host cell protein detection and quantitation.

Introduction

Host cell proteins are low level process-related impurities derived from the host expression systems during biopharmaceutical manufacturing, which can impact quality or safety, or compromise product stability¹. The International Conference on Harmonisation (ICH) guideline Q11 establishes HCPs as a Critical Quality Attribute (CQA)², and regulatory guidelines demand that HCP levels must be monitored and managed to acceptable levels³. Although exact levels are not specified, a target limit of less than 100 ppm in the final product is commonly employed within the industry. Accordingly, the downstream purification process (DSP) must be designed to achieve robust removal of these impurities.

Most purification processes for monoclonal antibodies (mAbs) involve Protein A affinity chromatography following cell culture harvest. Subsequently, two or three steps, such as anion exchange, cation exchange, and hydrophobic interaction chromatography, are included as polishing steps to remove problematic, co-purifying HCPs⁴.

Label-free quantification is a method in mass spectrometry that aims to determine the relative amount of proteins in two or more biological samples. It allows the unbiased discovery of HCP impurities and subsequent quantitation.

Data analysis requires the availability of software platforms and databases that help in the interpretation of MS/MS data and facilitate confident identification of proteins present. There are several commercially available applications that use workflows to process and report mass spectrometry data. They compare the raw data taken from mass spectrometry or spectral libraries to the information from a selected database and identify proteins from the mass spectra of digested fragments.

BioPharma Finder 4.1 software incorporates a host cell protein analysis workflow as part of the peptide mapping analysis feature, which improves data handling for protein sequence assessment, PTM evaluation, and the detection of HCPs for biopharmaceuticals characterization. For HCP analysis, it allows for database search directly within the application by selecting any of the three protein FASTA databases available or by adding another database of choice.

In the present study, HRAM MS-based HCP identification and label-free quantitation was performed on an in-house produced IgG1 mAb at different stages of the protein A purification process to evaluate the efficiency of the purification within downstream processing.

Experimental

Recommended consumables

- Ultrapure water, 18.2 MΩ·cm resistivity
- Water, Optima™ LC/MS grade (Fisher Chemical™, P/N 10505904)
- Acetonitrile with 0.1% formic acid (v/v), Optima™ LC/MS grade (Fisher Chemical™, P/N 10118464)
- Formic acid, LC-MS grade (>99%, Pierce™, P/N 28905)
- Thermo Scientific™ Acclaim™ VANQUISH™ C18, 2.2 μm, 2.1 × 250 mm column (P/N 074812-V)
- Trypsin Protease, Pierce™ MS Grade (P/N 90058)
- HiTrap™ Protein A (Cytiva)
- Thermo Scientific™ Virtuoso™ vial, clear 2 mL kit with septa and cap (P/N 60180-VT405)
- Thermo Scientific™ Virtuoso™ vial identification system (P/N 60180-VT100)

Sample handling equipment

- Thermo Scientific™ Vanquish™ Flex UHPLC system consisting of:
 - Thermo Scientific™ Vanquish™ System Base (P/N VF-S01-A)
 - Thermo Scientific™ Vanquish™ Binary Pump F (P/N VF-P10-A)
 - Thermo Scientific™ Vanquish™ Split Sampler FT (P/N VF-A10-A)
 - Thermo Scientific™ Vanquish™ Column Compartment H (P/N VH-C10-A)
 - MS Connection Kit Vanquish (P/N 6720.0405)
- Thermo Scientific™ Q Exactive™ Plus hybrid quadrupole-Orbitrap mass spectrometer (P/N IQLAAEGAAPFALGMBDK)
- Thermo Scientific™ Nanodrop™ 2000 Spectrophotometer (P/N ND-2000)

Software packages

- BioPharma Finder 4.1 software
- Thermo Scientific™ Chromeleon™ Enterprise Chromatography Data System (CDS) 7.2.10

Sample preparation

Ig1 mAb expression and purification

Recombinant IgG1 monoclonal antibody was expressed by mammalian cell culture in a CHO cell line. Cells were harvested at day 14, clarified, and sterile filtered. Samples of clarified media were passed through a HiTrap Protein A column (Cytiva) using an ÄKTA™ Avant system (Cytiva), then washed with phosphate buffered saline before elution of mAbs from the Protein A column using 0.1 M sodium citrate, pH 3.2. The elution peak was automatically collected (when the UV 280 nm signal rose above 50 mAU) into 15 mL tubes containing 300 µL of neutralizing buffer (1 M Tris-HCl, pH 9).

Tryptic digest

Protein A eluates were concentrated and buffer exchanged into 1X PBS using 3 K Vivaspin™ 500 concentrators (Sartorius Stedum Biotech, Gottingen, Germany). Quantification of the concentrated protein was carried out using a NanoDrop 2000 spectrophotometer at 280 nm and a BCA assay kit (Pierce Biotechnology, Rockford, IL, USA).

Sample aliquots containing 1 mg of concentrated protein were reduced and alkylated. Proteins were digested using 20 µg sequencing grade trypsin (Promega, Madison, WI, USA) for 18 h at 37 °C at 400 rpm mixing. Formic acid (10% v/v) was added at a 1:10 ratio in volume to halt digestion. The supernatant was vacuum dried using a Thermo Scientific™ SpeedVac™ concentrator. Samples were stored at -30 °C. Peptides were cleaned up using C18 column chromatography⁵.

LC-MS conditions

Peptide samples were dissolved in 0.1% formic acid at a concentration of 1 mg/100 µL. A total of 10 µL sample was injected onto an Acclaim VANQUISH C18 column (2.2 µm, 2.1 mm x 250 mm) for separation over a 65 min linear gradient (Table 1) using buffer A, 0.1% formic acid in water, and buffer B, 0.1% formic acid in acetonitrile.

Table 1. Chromatographic separation - LC gradient conditions

Time (min)	Flow (mL/min)	% Mobile phase B	Curve
0.0	0.300	2.0	5
45.0	0.300	40.0	5
45.5	0.300	80.0	5
50.0	0.300	80.0	5
50.5	0.300	2.0	5
65.0	0.300	2.0	5

Data-dependent (DDA) LC-MS/MS analysis of the tryptic digests was performed using a Vanquish Flex Binary UHPLC system coupled to a Q Exactive Plus hybrid quadrupole-Orbitrap mass spectrometer. The mass spectrometer was operated in positive ion mode at a spray voltage of 3.8 kV and capillary temperature of 320 °C. MS1 spectra were collected in the range of 200–2000 *m/z*. The five most intense precursors were selected for MS/MS, collected in the range of 50–2000 *m/z* with a maximum ion injection time of 200 ms. Tables 2 and 3 detail MS source/analyzer conditions and MS method parameters, respectively.

Table 2. MS source and analyzer conditions

MS source parameters	Setting
Source	Thermo Scientific™ Ion Max™ source with HESI II probe
Sheath gas pressure	25 arbitrary units
Auxiliary gas flow	10 arbitrary units
Probe heater temperature	150 °C
Source voltage	3.8 kV
Capillary temperature	320 °C
S-lens RF level	60

Table 3. MS method parameters utilized for peptide mapping analysis

General	Setting
Run time	0 to 65 min
Polarity	Positive
Full MS parameters	Setting
Mass range	200–2000 <i>m/z</i>
Resolution	70,000
AGC target value	3.0 × 10 ⁶
Max. injection time	100 ms
Default charge state	2
In-source CID	0 eV
Microscans	1
MS ² parameters	Setting
Resolution	17,500
AGC target value	1.0 × 10 ⁵
Isolation width	2.0 <i>m/z</i>
Signal threshold	1.0 × 10 ⁴
Normalized collision energy (NCE)	28
TopN MS ²	5
Max. injection time	200 ms
Fixed first mass	-
Dynamic exclusion	7.0 s
Loop count	5

MS data processing

The identification of protein impurities was performed using BioPharma Finder software version 4.1 via the host cell protein analysis feature. The quantitative analysis of those impurities was performed through peptide mapping analysis using the parameter settings summarized in Tables 4 and 5.

Table 4. BioPharma Finder 4.1 software parameter settings for peptide mapping data analysis

Component detection	Setting
Absolute MS signal threshold	1.2 × 10 ⁴ counts
Typical chromatographic peak width	0.29
Relative MS signal threshold (% base peak)	1.00
Relative analog threshold (% of highest peak)	1.00
Width of Gaussian filter (represented as 1/n of chromatographic peak width)	3
Minimum valley to be considered as two chromatographic peaks	80.00
Minimum MS peak width (Da)	1.20
Maximum MS peak width (Da)	4.20
Mass tolerance (ppm for high-res or Da for low-res)	6.00
Maximum retention time shift (min)	1.76
Maximum mass (Da)	30,000
Mass centroiding cutoff (% from base)	15.00
Identification	Setting
Maximum peptide mass	11,000
Mass accuracy (ppm)	6
Minimum confidence	0.80
Maximum number of modifications for a peptide	1
Unspecified modification	---
N-glycosylation	CHO
Protease specificity	High
Static modifications	Setting
Side chain	Carbamidomethylation (C)
N-term	Gln→Pyro-Glu
Variable modifications	Setting
C-term	Lys
Side chain	Deamidation (N), Double oxidation, Glycation, H ₂ O loss, Hydroxylation, Mannosylation (S), NH ₃ loss, Oxidation (MW)

Table 5. BioPharma Finder 4.1 software parameter settings for HCP analysis

Basic parameters	Setting
Protein database	Chinese hamster <i>Cricetulus griseus</i> UniProt-proteome UP000001075
Acquisition type	High-High (MS ¹ and MS ²)
Precursor mass tolerance	20 ppm
Ions to search	b ion, y ion, NL ion
Mass range (MH ⁺ peptide mass)	350 to 5,000
E-value cutoff	0.100
Protease parameters	Setting
Protease termini	Fully digested
Max num internal miscleavages	2
Enable decoy search	Yes
Modifications	Setting
Static side chain	Carbamidomethylation (C)
Max # of variable modification per peptide	1
Variable side chain	Deamidation (N) Oxidation (MW) Deamidation (Q)
Protein terminal modification	Acetylation (N-term)
Advanced parameters	Setting
Enable methionine protein N-term clip	Yes

Results and discussion

Identification of HCPs that are present during expression at every downstream processing step is of great importance for mitigating potential risk in therapeutic protein manufacturing. During downstream process development, the cell from the harvest is clarified and purified using a series of chromatographic and filtration steps to purify the therapeutic protein from potentially deleterious proteins that can impact the efficacy and safety of the drug. To satisfy larger market demands, the protein load on the chromatography steps needs to be increased. This can either be achieved by increasing the number of cycles or by investing in larger chromatography columns, which will increase considerably the downstream process cost.

The present work shows the evaluation of HCP levels in two protein A eluates that underwent different numbers of purification cycles with the aim to maximize the Protein A purification capacity. To identify and quantify the HCPs present within an in-house produced IgG1 mAb, a peptide mapping approach using HRAM MS detection and HCP data analysis by label-free quantitation with BioPharma Finder software (Figure 1) was applied. By comparing the intensity of the three most abundant identified peptides of low abundant protein impurities with a protein of known concentration (drug substance), a quantitative estimation of the relative abundance can be obtained.

The raw files from the mass spectrometer were incorporated into BioPharma Finder software as part of the peptide mapping experiment and searched against a Chinese hamster database (Chinese hamster *Cricetulus griseus*_uniprot-proteome_UP000001075)⁶. The parameter for precursor mass tolerance was set to 20 ppm. The maximum number of missed cleavages was set to two. Carbamidomethylation of cysteine (C) residues was specified as a fixed modification; methionine (M) oxidation and asparagine (N) and glutamine (Q) deamidation were specified as variable modifications and N-terminal acetylation was also stated. The application uses an E-value cutoff to calculate the expected number of hits in the database with a quality score greater than or equal to the cutoff. About 124 proteins were identified in total in each LC-MS/MS run with scores ranging from 10 to 206 and confidence over 90% using Comet⁷ as the source MS/MS sequence database search tool.

As an example, Figure 2 shows the host cell protein quantitation data analysis workflow in BioPharma Finder software represented by three panels: HCP quantitation results on the protein (a) and peptide (b) levels and the HCP Quant Plot (c) representing the HCP level trends for the studied samples, allowing end-users to directly evaluate DSP experiment. Additionally, other graphical details can be shown such as the chromatogram, trend ratio (when comparing different samples), trend MS area, peptide sequence coverage, protein sequence, full scan spectra, and MS² spectra.

With the aim to provide highly confident identification of the detected HCPs, custom filters were applied to the list of identified peptides and proteins using default filter settings. Initially, only proteins identified based on at least two peptides were considered (filter 1), which resulted in 37 identified HCPs with unique and non-unique peptides from a total of 124 identified proteins by default. The next filter step consisted of selecting those proteins containing only unique peptides (filter 2). Subsequently, only those peptides with good overall structural resolution (≤ 1.5) were included for confident and reliable MS² data (filter 3). The average structural resolution (ASR) is a score that evaluates the completeness of fragmentation coverage of a peptide. An ASR of 1.0 indicates a full fragment coverage of a peptide sequence and is displayed on the fragment coverage map. Finally, only those proteins with at least three identified peptides (filter 4) were used for quantitative purposes. The Venn diagram⁸ shown in Figure 3 summarizes the number of identified HCPs after applying the combination of custom filters mentioned above.

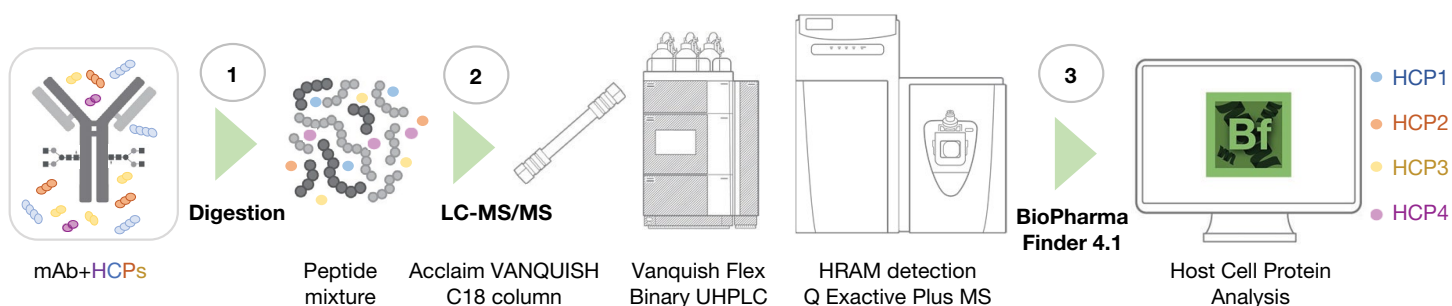


Figure 1. Overview of experimental setup for host cell protein analysis and quantitation. Purified samples were digested with trypsin (1) and the generated peptides subsequently analyzed by reversed-phase chromatography coupled to HRAM MS detection (2). Data analysis was performed by a peptide mapping approach with the host cell protein quantitation feature within BioPharma Finder software (3).

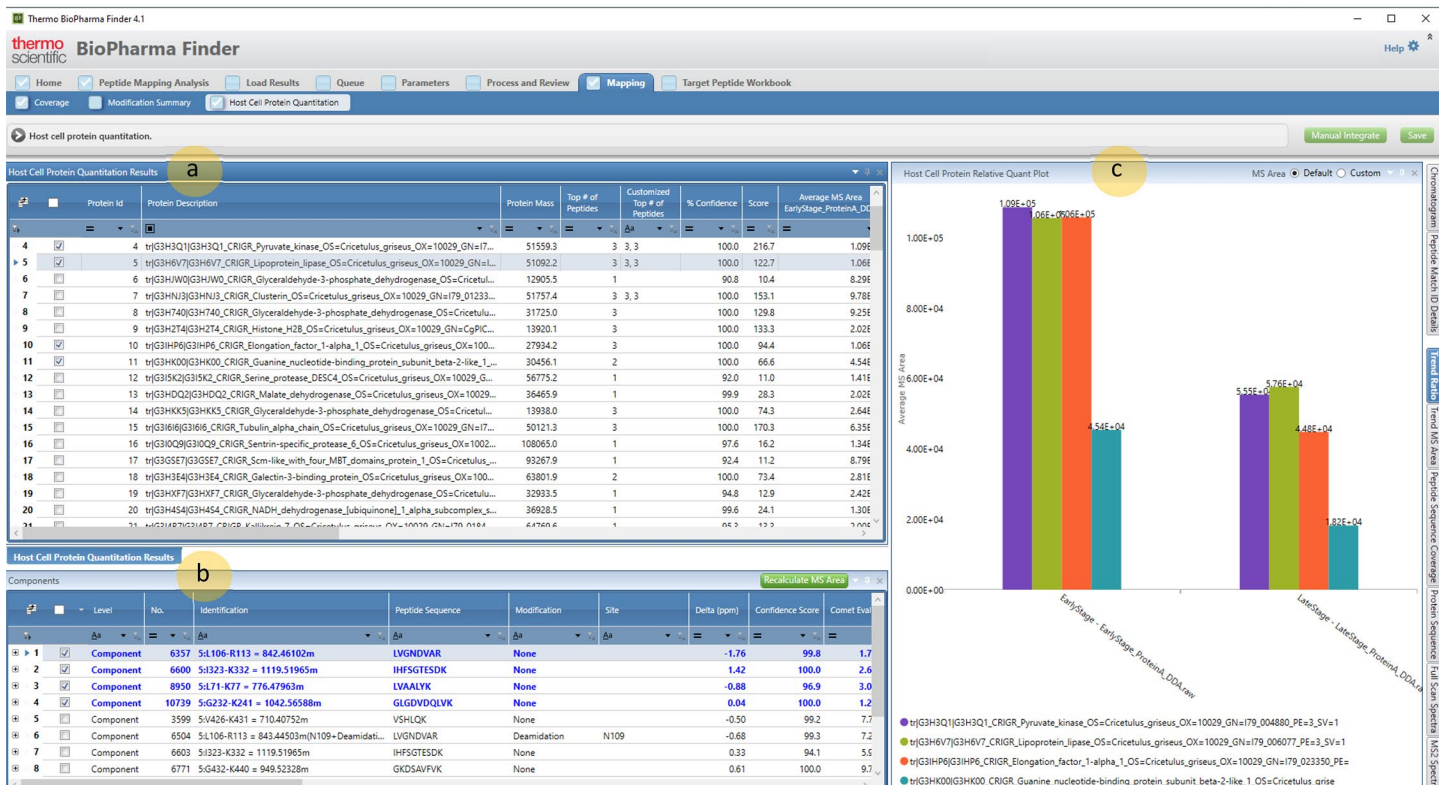


Figure 2. BioPharma Finder data analysis workflow for host cell protein quantitation. a) Host cell protein quantitation results; b) Components table with selected peptides for label-free quantitation highlighted in blue (three most abundant peptides); c) Host cell protein relative quant plot.

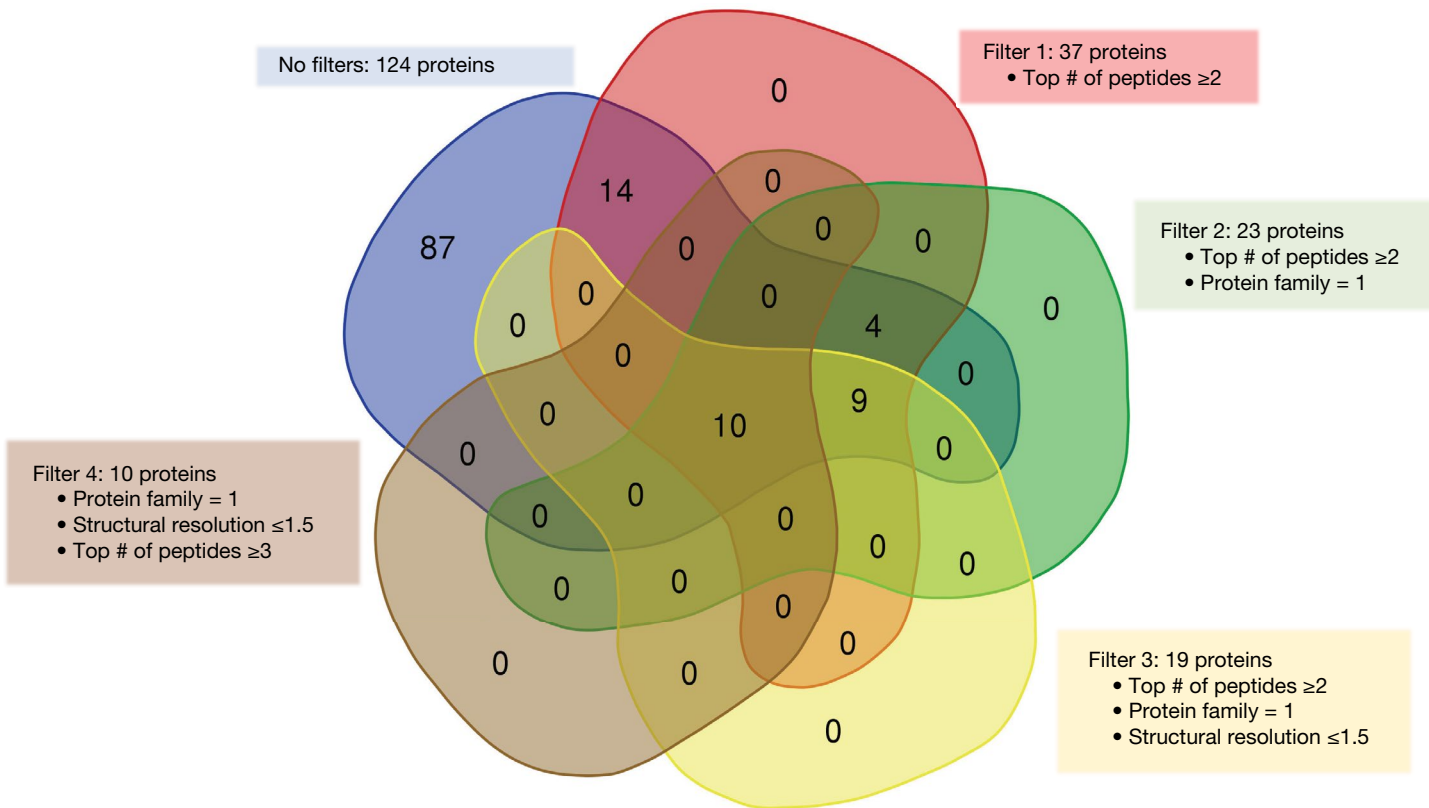


Figure 3. Venn diagram of the HCP proteins identified from the BioPharma Finder software host cell protein quantitation feature when applying or not a combination of custom filters using Chinese hamster *Cricetus griseus_uniprot-proteome_UP000001075* database

The accurate mass of precursor ions (predominantly less than 4 ppm) plus excellent MS² spectral quality provided by the Q Exactive Plus mass spectrometer further increases the confidence in HCP identifications, even for proteins present at very low abundances. As an example, Figure 4 shows the total ion chromatogram (TIC) for a 100 µg sample load used for HCP analysis (gray trace, a), combined with the extracted ion chromatograms representing the three unique peptides of lipoprotein lipase used for label-free quantitation (highlighted by red traces, b).

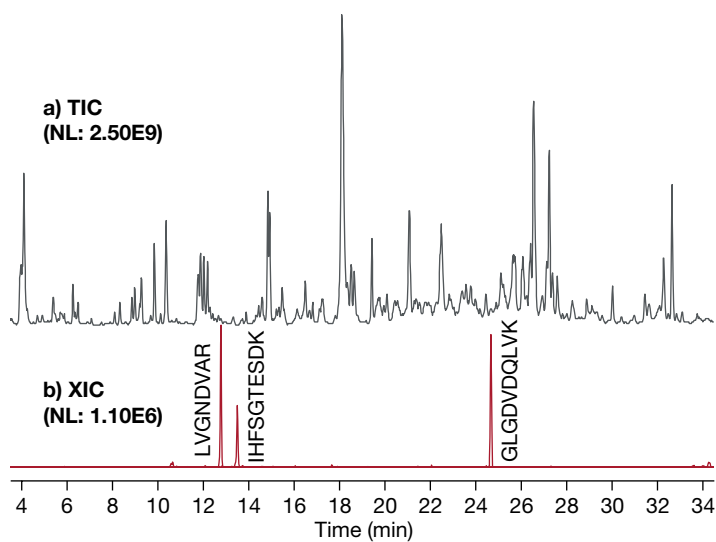


Figure 4. Overlay of the TIC for DSP-Stage1 sample (100 µg sample load) used for HCP analysis (a) and extracted ion chromatograms (XICs) for the three most abundant peptides for lipoprotein lipase (b) (NL: normalized largest intensity)

Their corresponding MS/MS spectra and fragment ion assignments are summarized in Figure 5 for DSP-Stage1 and DSP-Stage2 samples, respectively. Resulted spectra showed similar quality despite the fact that they were slightly different in intensity. Fragment mass errors observed for the ion series of the peptides were <5 ppm of the delta mass values. This allowed high confidence ID in both cases.

The list of identified HCPs following the three filter rules is summarized in Table 6, where a comparison of the average MS area from the two studied samples was evaluated with the aim to compare the levels of detected HCPs at different stages of the purification step (Figure 6). The sample referred to as DSP-Stage1 resulted from early cycles of Protein A chromatography, while the sample referred to as DSP-Stage2 was obtained from the end of Protein A purification life cycles. As observed in Figure 6, DSP-Stage2/Stage1 average MS area ratio (from top 3 or top 2 peptides) shows values close to 0.5 for most of the detected HCPs, which means over 120 cycles for the Protein A purification step reduced by half the levels of detected HCPs when compared to the eluate obtained from early cycles of Protein A purification. For Nidogen-1 protein, further purification cycles did not have any impact in the levels of this HCP. Some problematic HCPs including lipoprotein lipase and nidogen continue to penetrate and are particularly difficult to remove even after polishing steps such as anion/cation exchange or hydrophobic interaction chromatography through resin association or co-elution with mAbs⁹.

Protein MS areas from the top three peptides for each HCP and the areas for the top three peptides from the IgG1 drug substance were used to calculate individual HCP concentration (ppm or ng HCP/mg mAb) based on label-free quantitation¹⁰. The average MS area for the top three peptides from the studied IgG1 mAb was consistent for the two studied eluates (1.24E+08 average value) with CV <2.9%. Results showed Protein A chromatography was able to reduce HCP levels even after 120 cycles where they were detected from very low abundance (<10 ppm) up to over 500 ppm, except for elongation factor 2 of which levels were over 1000 ppm, indicating further purification steps would be needed.

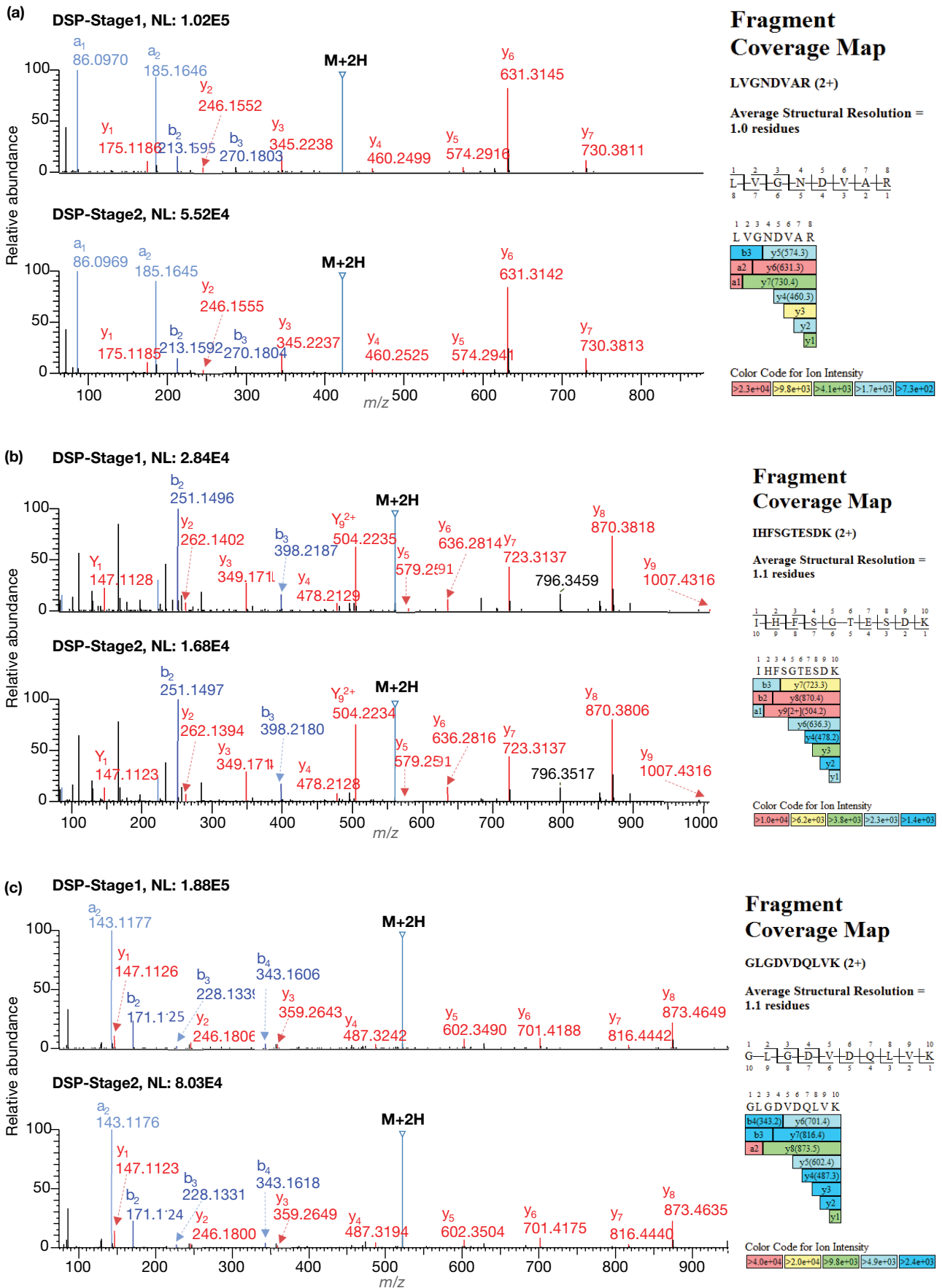


Figure 5. Representative MS² spectra and corresponding fragment coverage map for the top three lipoprotein lipase detected peptides for the two studied samples. Fragment ion assignments in blue (a-, b-ions) and red (y-ions) for the peptides a) LVGNDVAR, b) IHFGTESDK, and c) GLGDVDQLVK that were used to quantify lipoprotein lipase in the drug substance. Even at such low abundance, nine unique peptides were identified in total, providing additional confidence (Table 6).

Table 6. List of CHO proteins identified with at least two unique peptides in a peptide mapping based database search. 19 HCPs (out of a total of 124 identified HCPs) containing unique peptides and with overall structural resolution ≤ 1.5 were identified, and only those containing at least three unique peptides were quantified for the studied IgG1 mAb by using the average MS area of the top three peptides. Results are expressed as ng HCP/mg mAb (ppm).

Accession	Protein description	Protein mass	Top # of peptides	% Confidence	Score	Average MS area DSP-Stage1	Average MS area DSP-Stage2	DSP1-Stage1 HCP amount (ppm)	DSP1-Stage2 HCP amount (ppm)
tr G3HKG8	40S ribosomal protein S3a	29945	2	100.0	59.9	3.08E+04	1.36E+04	---	---
tr G3IE21	Aldose reductase-related protein 2	36340	3	100.0	89.5	5.89E+04	2.90E+04	47	24
tr G3HJ3	Clusterin	51757	3	100.0	153.1	9.78E+05	6.72E+05	773	554
tr G3HSL4	Elongation factor 2	97669	3	100.0	113.9	5.32E+06	2.21E+06	4207	1825
tr G3HQM6	Endoplasmic reticulum chaperone protein BiP	92622	2	100.0	60.5	5.59E+04	3.54E+04	---	---
tr G3H3E4	Galectin-3-binding protein	63802	2	100.0	73.4	2.81E+04	1.36E+04	---	---
tr G3I3Y6	Glutathione S-transferase P	25226	3	100.0	148.0	7.29E+04	3.27E+04	58	27
tr G3HK00	Guanine nucleotide-binding protein subunit beta-2-like 1	30456	2	100.0	66.6	4.54E+04	1.82E+04	---	---
tr G3HC84	Heat shock protein HSP 90-beta	47808	3	100.0	102.6	7.27E+04	3.63E+04	57	30
tr G3H6V7	Lipoprotein lipase	51092	3	100.0	122.7	1.06E+05	5.76E+04	84	48
tr G3HRK9	Matrix metalloproteinase-19	58942	2	100.0	57.0	2.27E+04	8.90E+03	---	---
tr G3I3U5	Nidogen-1	30091	2	100.0	112.1	2.91E+04	2.81E+04	---	---
tr G3H8V4	Phospholipid transfer protein	54374	3	100.0	178.1	1.04E+05	6.86E+04	83	57
tr G3H3Q1	Pyruvate kinase	51559	3	100.0	216.7	5.89E+04	2.98E+04	47	25
tr G3IBF4	Serine protease	28718	3	100.0	151.6	7.58E+04	3.27E+04	60	27
tr G3HDR3	T-complex protein 1 subunit delta	42136	2	100.0	89.0	3.86E+04	1.65E+04	---	---
tr G3HYB7	T-complex protein 1 subunit eta	54906	2	100.0	99.6	2.38E+04	1.08E+04	---	---
tr G3HG83	T-complex protein 1 subunit gamma	60620	2	100.0	86.8	2.19E+04	1.12E+04	---	---
tr G3I2M1	T-complex protein 1 subunit zeta	57975	3	100.0	122.6	2.97E+04	1.09E+04	15	8

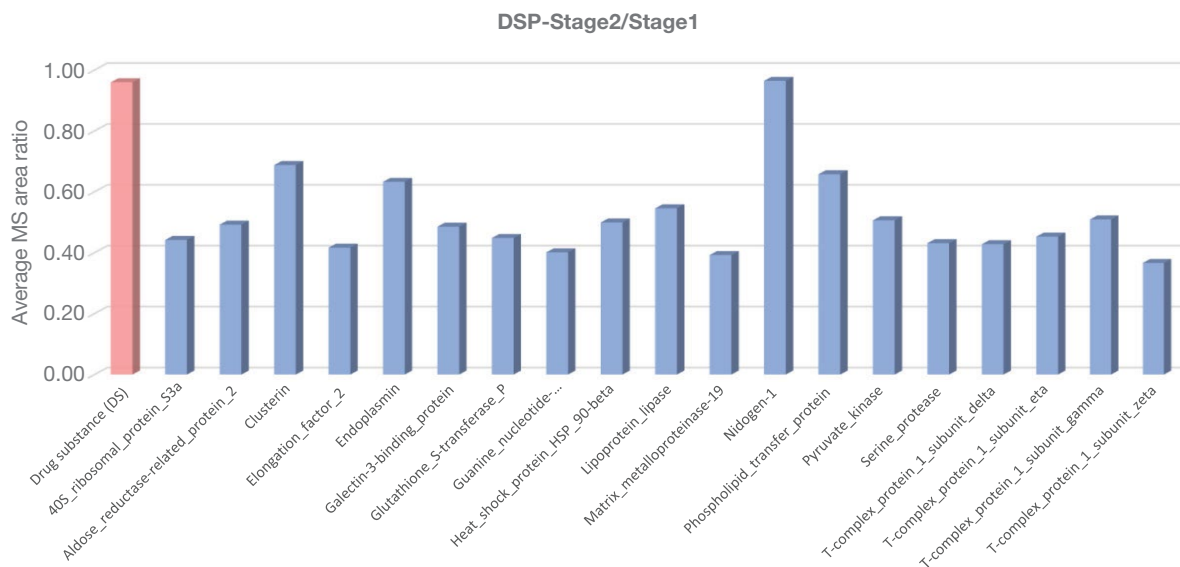


Figure 6. Comparison bar graph between DSP-Stage2 and DSP-Stage1 samples. Average MS area ratios of the top three or top two unique peptides for the 19 identified CHO cell proteins within BioPharma Finder software are shown (see Table 6 for details).

Conclusions

- We have developed a host cell protein data analysis workflow to assess protein clearance during DSP.
- The developed workflow demonstrates the suitability of the data analysis platform for reliable identification and quantitative analysis of trace impurities in monoclonal antibodies, which supports biopharmaceutical manufacturing process.
- Label-free quantitation utilizing a protein of known concentration (drug substance) offers a good alternative to absolute HCP quantitation using an internal standard, and a good estimation of the relative abundance can be obtained at very low concentration.
- High-quality MS/MS data obtained with the Q Exactive Plus hybrid quadrupole-Orbitrap mass spectrometer in combination with the use of custom filters for the obtained protein list allowed confident identification of a total of 19 HCPs. For quantitative purposes, only those proteins with at least three identified peptides were considered.
- The results obtained in this study demonstrate the applicability of BioPharma Finder software's new host cell protein quantitation feature for the analysis of low-level impurities present in an investigational IgG1 mAb.

References

1. Zuch de Zafra, C.L. et al. Host cell proteins in biotechnology-derived products: A risk assessment framework. *Biotechnol. Bioeng.* **2015**, *112*, 2284–2291.
2. IHC, *Development and Manufacture of Drug Substances (Chemical Entities and Biotechnological/biological Entities) Q11.*, 2012
3. ICH, *Guidance for Industry Q6B Specifications: Test Procedures and Acceptance Criteria for Biotechnological/Biotechnology Products*, 1999.
4. Liu, H.F. et al. Recovery and purification process development for monoclonal antibody production. *MAbs* **2010**, *2*(5), 480–499.
5. Albrecht, S. et al. Proteomics in biomanufacturing control: protein dynamics of CHO-K1 cells and conditioned media during apoptosis and necrosis. *Biotechnol. Bioeng.* **2018**, *115*(6), 1509–1520.
6. [Online] <https://www.uniprot.org/proteomes/UP000001075>
7. Comet open source tandem mass spectrometry (MS/MS) sequence database search engine. <http://comet-ms.sourceforge.net/>
8. Van de Peer, Y. Calculate and draw custom Venn diagrams (Bioinformatics and Evolutionary Genomics group, UGent/VIB, 2009). <http://bioinformatics.psb.ugent.be/webtools/Venn/>
9. Gilgunn, S. et al. Identification and tracking of problematic host cell proteins removed by a synthetic, highly functionalized nonwoven media in downstream bioprocessing of monoclonal antibodies. *J. Chromatogr. A*, **2019**, *1595*, 28–38.
10. Grossmann, J. et al. Implementation and evaluation of relative and absolute quantification in shotgun proteomics with label-free methods. *J. Proteomics*, **2010**, *73*, 1740–1746.

Find out more at thermofisher.com