Accepted Manuscript
Version of Record at: https://doi.org/10.1002/btpr.3343

Analytical characterization of host-cell-protein-rich aggregates in monoclonal antibody solutions

Chase E. Herman[a], Lie Min[a], Leila H. Choe[a], Ronald W. Maurer[b], Xuankuo Xu[b], Sanchayita Ghose[b], Kelvin H. Lee[a] and Abraham M. Lenhoff[a]*

[a] Department of Chemical and Biomolecular Engineering, University of Delaware, Newark, DE 19716, USA

[b] Biologics Process Development, Bristol-Myers Squibb, Devens, MA 01434, USA

* Correspondence to: Abraham M. Lenhoff

Department of Chemical and Biomolecular Engineering

University of Delaware

150 Academy St.

Newark, DE 19716

USA

lenhoff@udel.edu

Accepted Manuscript
Version of Record at: https://doi.org/10.1002/btpr.3343

Abstract

Host-cell proteins (HCPs) and high molecular weight (HMW) species have historically been treated as independent classes of impurities in the downstream processing of monoclonal antibodies (mAbs), but recent indications suggest that they may be partially linked. We have explored this connection with a shotgun proteomic analysis of HMW impurities that were isolated from harvest cell culture fluid (HCCF) and protein A eluate using size-exclusion chromatography (SEC). As part of the proteomic analysis, a cross-digest study was performed in which samples were analyzed using both the standard and native digest techniques to enable a fair comparison between bioprocess pools. This comparison reveals that the HCP profiles of HCCF and protein A eluate overlap substantially more than previous work has suggested, because hundreds of HCPs are conserved in aggregates that may be up to ~50 nm in hydrodynamic radius and that persist through the protein A capture step. Quantitative SWATH proteomics suggests that the majority of the protein A eluate's HCP mass is found in such aggregates, and this is corroborated by ELISA measurements on SEC fractions. The SWATH data also show that intra-aggregate concentrations of individual HCPs are positively correlated between aggregates that were isolated from HCCF and protein A eluate, and species that have generally been considered difficult to remove tend to be more concentrated than their counterparts. These observations support prior hypotheses regarding aggregate-mediated HCP persistence through protein A chromatography and highlight the importance of this persistence mechanism.

Keywords

Aggregates, host-cell proteins, proteomics, digest method, protein A chromatography

Accepted Manuscript
Version of Record at: https://doi.org/10.1002/btpr.3343

1. Introduction

The purification of monoclonal antibody (mAb) therapeutics produced in Chinese hamster ovary (CHO) cells typically relies on platform processes that are designed to deliver a consistent product composition, which is characterized by critical quality attributes (CQAs) under the quality by design (QbD) guidance.[1–3] The QbD rationale has led bioprocessing research to focus on developing a more fundamental understanding of the relationship between process parameters and CQAs, one of which is the host-cell protein (HCP) content of drug substance. CHO cells secrete native proteins along with the recombinant product, and intracellular HCPs may also be liberated into harvest cell culture fluid (HCCF) via necrosis and apoptosis.[4]

There are thousands of species in the CHO proteome and those in HCCF often possess diverse biophysical properties.[5,6] Protein A chromatography usually provides substantial HCP clearance but some species manage to persist. This is an incompletely understood phenomenon that may be due in part to product association or co-elution, and the former has been suggested as the primary contributor in the capture step.[7] While product association may be mAb-specific, there is a set of HCPs that have been commonly encountered in mAb downstream processing studies and are generally considered difficult to remove.[6,8–14] Additionally, a subset of the CHO proteome is known to constitute a high therapeutic risk due to deleterious impacts on product safety or stability.[8] Some HCPs, such as cathepsins and PLBL2, fall into both the high-risk and difficult-to-remove categories and they may require targeted removal strategies.[8,15,16]

The HCP concentration of mAb solutions is usually measured using an enzyme-linked immunosorbent assay (ELISA); values are typically 10^5–10^6 ppm in HCCF and < 100 ppm in drug substance.[17,18] Although ELISA has served as a *de facto* standard analytical technique, it suffers from inaccuracies that arise from coverage mismatches between the sample's HCPs and the assay's

polyclonal antibody probes.[19–21] The ELISA measurements used for overall HCP assays also do not identify individual HCP species, which is clearly relevant to monitoring the clearance of high-risk impurities. Shotgun proteomic techniques using liquid chromatography-tandem mass spectrometry (LC-MS/MS) provide an alternative for HCP profiling that is finding increasing use, with both instrumentation and methodological advances contributing to improvements in HCP coverage. These methods are implemented by enzymatically digesting the proteins in a sample and identifying peptides that are detected by MS/MS, which are then queried against sequence databases to identify the original proteins.

The standard digest technique of shotgun proteomic analysis uses trypsin for the simultaneous proteolysis of HCP and mAb molecules, but the LC co-elution of mAb and HCP peptides can obscure HCP signals in MS/MS data. The native digest method was developed to circumvent this problem by leaving the mAb largely intact, thus decreasing the required dynamic range for the MS/MS detection of HCPs.[22] This is accomplished by omitting the standard digest's reduction and denaturation steps prior to digestion and then precipitating the mAb after digestion. Several mAb depletion (HCP enrichment) strategies have been derived from the native digest method to enhance coverage of HCPs in purified mAb solutions.[23–27] However, a concern with all these techniques is that HCP peptides may be lost during the mAb removal step, leading to uncertainty as to which digest method is better suited for the analysis of incompletely purified mAb solutions, such as HCCF.

Another CQA of mAb products is the high molecular weight (HMW) content in the drug substance.[28] HMW species have historically been conceptualized primarily as mAb oligomers and treated independently from HCPs[28] but there are indications that the two impurity classes may be related. Gagnon et al.[29–34] have shown populations of mAb aggregates to be relatively rich in DNA

and histone HCPs, which they attributed to nucleosomal arrays acting as nucleation centers for aggregation, where the proximity of positively-charged histones to negatively-charged DNA may create an environment that attracts mAb molecules as well as non-histone HCPs. They noted that cell viabilities at harvest were relatively low in their studies, which is known to worsen the HCCF's HCP profile due to the release of intracellular proteins.[35–40] The observation of HCPs in mAb aggregates nonetheless warrants deeper investigation, as improvements in aggregate clearance have appeared to reduce HCP levels generally.[29,41–43]

HCP persistence in protein A chromatography was also recently correlated with the self-association propensity of mAbs and was attributed to the transient formation of positively-charged mAb clusters that attract HCPs.[44] It seems reasonable that HCPs may likewise associate with more permanent mAb aggregates and their survival through protein A chromatography could be construed as an extension of the product association mechanism. Separate work recently confirmed this idea, showing that hundreds of HCPs may be present in aggregates in mAb solutions.[45,46] The biophysical basis of this phenomenon was attributed at least in part to the association between mAbs and HCPs involved in the unfolded protein response. Oh et al.[46] suggested that the resulting aggregates may directly mediate the persistence of many HCPs through protein A chromatography, but analysis of the extent of this phenomenon was limited by the different digest methods that were used to analyze HCCF and protein A eluate samples.

To better understand the persistence of HCP and HMW impurities and their possible relationship, we analyzed the HCP composition of HCCF and protein A eluate from a commercial mAb manufacturing campaign as well as subsequent processing of these streams. An enabling step in all these studies was the use of size-exclusion chromatography (SEC) to isolate HMW impurities, allowing characterization of different components of the process streams involved and

Accepted Manuscript
Version of Record at: https://doi.org/10.1002/btpr.3343

deeper profiling of HCP compositions. A key element is a cross-digest proteomics study that compares the relative merits of the standard and native digest methods for detecting HCPs in these samples, which may inform the construction of spectral libraries prior to quantitative proteomic analysis. Complementary ELISA and dynamic light scattering (DLS) analyses were also performed to provide additional context within which to interpret the findings. Although this observational study is limited in scope to one mAb product, we expect the trends to be transferrable to other mAb solutions that contain HMW impurities given similar observations that were made in a recent study of several mAb products.[46]

2. Materials and Methods

2.1. Materials

Arginine and ethylenediaminetetraacetic acid (EDTA) were purchased from Acros Organics, and triethylamine hydrochloride was purchased from Alfa Aesar. Tris-HCl, iodoacetamide, bovine gamma globulin, bovine ribonuclease A, and hen ovalbumin were purchased from Sigma-Aldrich. Sodium bicarbonate, 50% w/w NaOH, 12.1 N HCl, and acetonitrile (ACN) were purchased from Fisher Scientific. 2-(N-morpholino)ethanesulfonic acid hydrate (MES), tris(2-carboxyethyl)phosphine (TCEP), and formic acid (FA) were purchased from Thermo Scientific. Dithiothreitol (DTT) was purchased from Bio-Rad and sodium azide was purchased from Fluka. Sequencing-grade trypsin was purchased from Promega and pre-digested yeast alcohol dehydrogenase (ADH) was purchased from Waters. Buffer solutions were prepared using deionized water from an EMD Millipore Milli-Q system (> 18.2 MΩ cm), and pH and conductivity measurements were made using a Cole-Parmer PC200 meter**.**

Accepted Manuscript
Version of Record at: https://doi.org/10.1002/btpr.3343

Bristol Myers Squibb (BMS, Devens, MA) supplied HCCF (5.4 mg/ml) from an IgG1 mAb (145 kDa, pI 8.9) manufacturing process that used CHO cells as the host. As part of its preparation, the HCCF had been clarified using an Emphaze AEX Hybrid Purifier (3M) that terminates in 0.2 μm sterile filtration. BMS also supplied protein A eluate that was purified from the HCCF using MabSelect SuRe LX (Cytiva). Viral inactivation was performed by titrating the eluate to pH 3.5 with 0.1 N HCl and incubating quiescently for 60 min. The pool was neutralized to pH 5.4 with 2 M tris and filtered through an Emphaze AEX Hybrid Purifier before custody transfer (PAVIN, 26.9 mg/ml).

2.2. SEC fractionation

Prior to SEC fractionation the HCCF feedstock was concentrated approximately four-fold using Amicon 100 kDa MWCO centrifugal ultrafiltration units (EMD Millipore) at 7500 g. The concentrated HCCF and the unconcentrated PAVIN were fractionated using a TSKgel G3000SWxl SEC column (Tosoh Bioscience) on an Äkta Explorer (Amersham Biosciences) equipped with a 10 mm UV flow cell and a fraction collector. Nominal injections of 150 μl were made cyclically using a 50 ml Superloop assembly (GE Healthcare) at a flow rate of 0.6 ml/min in a pH 6.5 running buffer of 50 mM MES, 200 mM arginine, 5 mM EDTA, and 0.05% sodium azide. A mAb fraction and two HMW impurity fractions, which are referred to as large and small aggregates, were collected from both the HCCF and PAVIN; two small-protein fractions (SPFs) were also collected from the HCCF. The column was washed according to the manufacturer's recommendations after every ~100 HCCF injections and ~150 PAVIN injections. The approximate molecular masses of the SEC fractions were estimated using a standard curve provided by the

Accepted Manuscript
Version of Record at: https://doi.org/10.1002/btpr.3343

column manufacturer. Gamma globulin, ribonuclease A, and ovalbumin calibrants were used to validate the standard curve and adjust for the Äkta Explorer's extra-column volume.

All SEC fractions and unprocessed feedstocks were subjected to centrifugal ultrafiltration for concentration and buffer exchange. Filters with a 10 kDa MWCO were used for the aggregate and mAb fractions whereas 3 kDa MWCO filters were used for the SPFs and unprocessed feedstocks. Samples for the standard digest were exchanged into 100 mM triethylammonium bicarbonate at pH 8.0 and concentrated to > 0.5 mg/ml and those for the native digest were exchanged into 25 mM tris-HCl at pH 8.0 and concentrated to > 5.0 mg/ml. Total protein concentrations were estimated in triplicate using the Bradford assay (Thermo Scientific) with bovine serum albumin as the calibrant, and the concentrations of HCPs in the SEC fractions were estimated using an F550-1 3G CHO HCP ELISA kit (Cygnus), where duplicate measurements were taken from different concentrations in a single 10-fold serial dilution for each sample.

2.3. Dynamic light scattering

DLS was performed on a 3D LS Spectrometer (LS Instruments) equipped with a 660 nm laser, a motor-driven goniometer, two avalanche photodiode detectors, and a two-channel digital correlator. All measurements were made in aqueous solution (refractive index $n = 1.330$, viscosity $\eta = 0.891$ mPa · s) at an angle of 104º, corresponding to a reciprocal scattering vector magnitude of 50 nm. The buffer-exchanged SEC fractions were diluted to < 1 mg/ml, filtered through 0.8 μm low protein binding Acrodisc syringe filters (Pall), and analyzed in auto- and pseudo-cross correlation modes; unprocessed HCCF and PAVIN feedstocks were analyzed in the instrument's modulated 3D cross-correlation mode to suppress the effects of multiple scattering. Data collection proceeded until the product of count rate and sampling time exceeded 1.0×10^8 for SEC fractions

and 4.5×10^7 for unprocessed feedstocks. The data were interpreted using the CONTIN algorithm with a curvature minimization constraint to impose smoothness; hydrodynamic radius bounds of [0.5 nm, 1 μm] were specified and the L-curve method was used to select the optimal value of the regularization parameter independently for each data set.[47,48]

2.4. Proteomic analysis

2.4.1. Native digest

Native digestion was performed under non-reducing conditions as described previously.[22] Samples of volume 100 μl containing 500 μg of total protein were digested with trypsin at an enzyme to substrate mass ratio of 1:400 for 16 hr at 37 °C. Undigested mAb was reduced and precipitated by adding 1 μl of 50 mg/ml DTT and incubating for 10 min at 90 °C. Samples were centrifuged at 13,000 g for 2 min and the supernatants were filtered through 10 kDa MWCO centrifugal ultrafiltration units (VWR) at 5000 g for 3.5 min. The filtration units were rinsed with 100 μl of 25 mM tris-HCl and centrifuged for an additional 4 min. The filtrates were acidified with 3 μl of 20% FA and samples were desalted using OMIX C18 pipette tips (Agilent). The C18 tips were conditioned with 50% ACN, equilibrated with 1% FA, loaded with sample, washed with 0.1% FA, and eluted with 0.1% FA in 50% ACN. Three bind-and-elute cycles were completed on the same tip for each sample. Desalted samples were dried using a SpeedVac vacuum concentrator (Thermo Scientific) and redissolved in 49.5 μl of 2% ACN with 0.1% FA. Samples were spiked with pre-digested ADH to a concentration of 6.675 fmol/μl and retention time calibrants (iRT, Biognosys, Schlieren, Switzerland), and the volume equivalents of 90.91 μg of digested proteins were analyzed in technical triplicate in LC-MS/MS.

2.4.2. Standard digest

Trypsin digestion was performed under standard conditions as described previously.[49,50] Samples of volume 100 µl containing 50 µg of protein were reduced by adding 2.5 µl of 100 mM TCEP and incubated at 60 °C for 1 hr. Samples were alkylated by adding 5 µl of 150 mM iodoacetamide and incubating in the dark at room temperature for 30 min. Trypsin was added at an enzyme to substrate mass ratio of 1:50 and digestion proceeded at 37 °C for 16 hr. Samples were acidified with 4 µl of 20% FA, desalted and dried as described above, and redissolved in 45 µl of 2% ACN with 0.1% FA. Samples were spiked with pre-digested ADH to a concentration of 5 fmol/µl and with retention time calibrants, and the volume equivalents of 5 µg of digested proteins were analyzed in technical triplicate in LC-MS/MS.

2.4.3. LC-MS/MS

LC-MS/MS analysis was performed as described previously.[51] A TripleTOF 6600 Sciex instrument equipped with an Eksigent Nano 425 LC was operated in microflow mode with a ChromXP C18CL Sciex column (3 µm, 120 Å, 150 mm x 0.3 mm). Mobile phases A and B consisted of 0.1% FA in water and 0.1% FA in ACN, respectively, and a flow rate of 5 µl/min was used. Elution was performed via gradients from 3 to 25% B over 68 min, 25 to 35% B over 5 min, and then 35 to 80% B over 2 min. Data-dependent acquisition (DDA) was performed in positive ion mode with a 250 ms MS1 scan over a mass range of 400-1250 *m/z*, and the top 30 precursor ions were selected for fragmentation and MS/MS detection with a 50 ms scan over a mass range of 100-1500 *m/z*. Data-independent acquisition (DIA) was performed using sequential windowed acquisition of all theoretical fragment ion spectra (SWATH) with a full MS1 scan and 64 variable

Accepted Manuscript
Version of Record at: https://doi.org/10.1002/btpr.3343

MS/MS windows for standard digestions and 32 variable MS/MS windows for native digestions.[52–54]

Database searches were performed using the Paragon algorithm in the Sciex ProteinPilot software (v5.1) as described previously,[51,55] using a local copy of the CH RefSeq 2019 database supplemented with ADH, retention time calibrants, and common contaminants. Search parameters were set to specify trypsin as the enzyme, iodoacetamide for cysteine alkylation of standard digest samples, and a 10% detected protein threshold. HCP identifications with at least two peptides at 95% confidence in all three technical triplicates were accepted at a local false discovery rate of 1%.

DDA data from all native digest samples and the standard digest of SPF 2 were used to construct a consolidated spectral library in Skyline as described previously (v20.2.0.343; MacCoss Lab, University of Washington).[51,56,57] Triplicate SWATH data for each sample were processed together using the Skyline command-line interface, and peaks were selected and integrated automatically using the mProphet algorithm based on a target decoy approach.[58] The consolidated spectral library was used in the analysis of all samples except the PAVIN mAb SEC fraction, which contained too few HCP peptides for mProphet automatic peak picking. A spectral library based only on the PAVIN mAb DDA data was therefore used in the analysis of that sample. A detection q value was assigned to each peak and peptides with $q > 0.01$ in any of the triplicate measurements were removed from the analysis. Following summarization by Tukey's median polish, protein peak areas were calculated from peptide peak areas using MSstats (v3.18.5; Olga Vitek Lab, Northeastern University) and normalized against the ADH peak area.[59] The spiked mass of ADH was used to estimate HCP masses from normalized peak areas, based on the

Accepted Manuscript
Version of Record at: https://doi.org/10.1002/btpr.3343

assumption that all proteins of a given concentration generate the same response, and these were divided by the total injected sample mass to estimate fractional HCP mass concentrations.

2.4.4. Data analysis

It is often possible for multiple HCP candidates to explain the spectral data that identify a single protein, so a candidate selection algorithm was developed with the objective of minimizing the total number of HCPs in the set union of all DDA results. Only candidates with the most peptide identifications were considered and those found in other samples as the only possible option for any protein identification were given selection preference. Selection preference was then given to candidates occurring with the highest frequency in the sample set union and previously selected HCPs were always given preference in new selections.

The Statsmodels Python package was used to perform one-way and two-way ANOVA tests as well as post-hoc pairwise comparisons using Tukey's HSD test, with a threshold of $p < 0.01$ considered significant. The Seaborn Python module was used to interface with SciPy for hierarchical clustering analysis, which was performed using Ward's minimum variance method based on the Euclidean distances between binary identification vectors (DDA data) or species-normalized concentration vectors (DIA SWATH data), where unidentified HCPs were assigned a concentration of zero.

3. Results

3.1. SEC analysis

To enable investigation of how HMW impurities may contribute to HCP persistence, mAb HCCF and protein A eluate that had been viral-inactivated and neutralized (PAVIN) were fractionated

Accepted Manuscript
Version of Record at: https://doi.org/10.1002/btpr.3343

using SEC via serial injections (Figure 1). In the course of the injection sequence there was a lower apparent injection mass during the first few cycles, which was attributed to pressure increases in the Superloop assembly in later cycles, leading to higher injection masses. The column performance appeared to be stable otherwise and peaks containing large and small aggregates were collected from both feedstocks along with the mAb fraction. Two small-protein fractions (SPFs) were also collected from the HCCF. A subtle but nonetheless observable shoulder was present in the PAVIN chromatograms at the location of the first SPF peak (Supplementary Figure S1) but material requirements precluded its characterization. The presence of this shoulder suggests that mAb fragments and some HCPs in free solution likely survived the protein A capture step but this peak is appreciably smaller than its aggregate counterparts. The logarithmic mass scale in Figure 1 should be interpreted with caution, as the excluded hydrodynamic radii of non-spherical structures may not correlate well with molecular mass. Given this uncertainty, as well as column dispersion, the small aggregate peak is expected to contain primarily mAb dimers and small oligomers. The large aggregate peak is expected to contain molecular structures that are on the order of hundreds of kDa as well as materials that exceeded the column's exclusion limit of ~1.2 MDa.

Accepted Manuscript
Version of Record at: https://doi.org/10.1002/btpr.3343

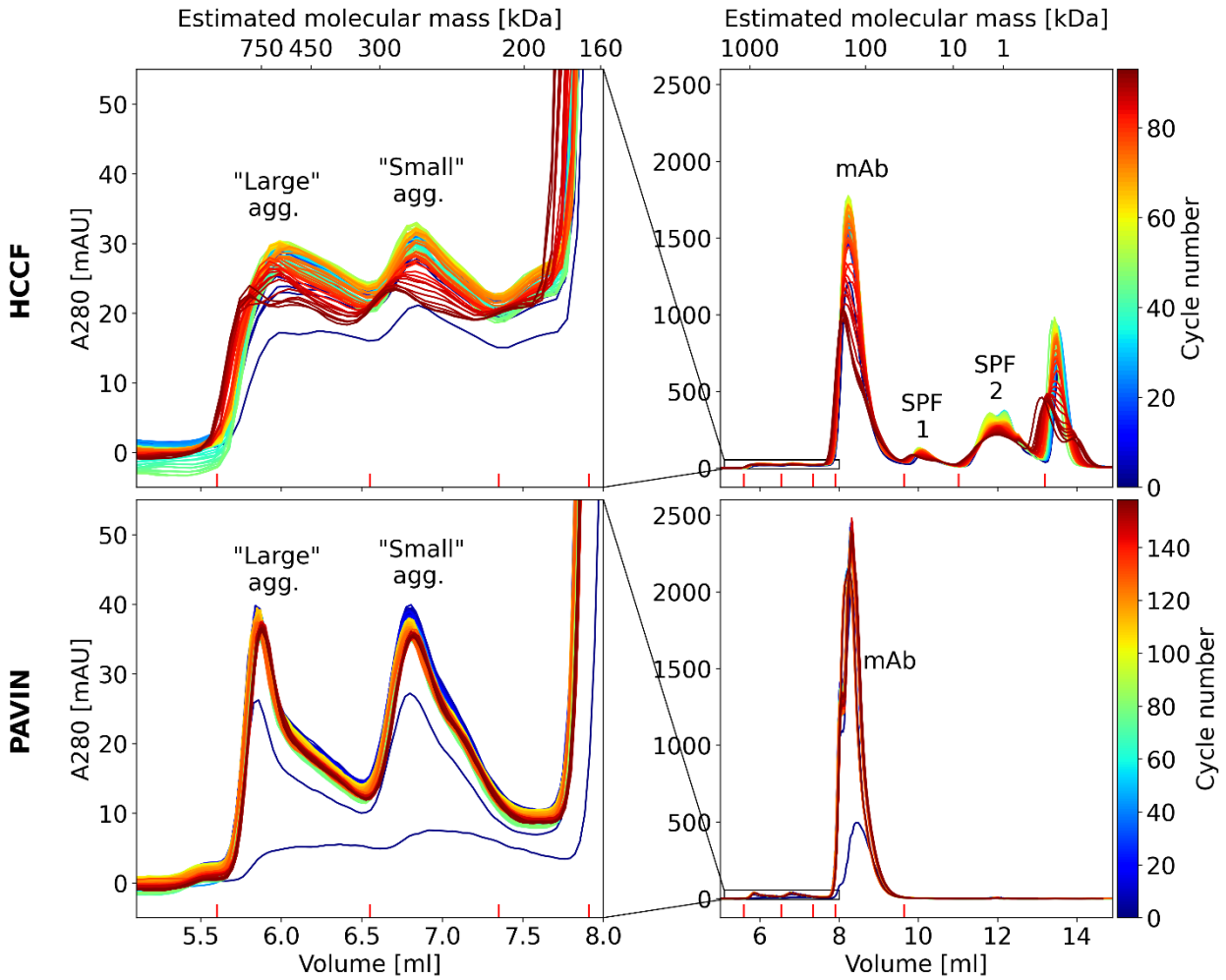

Figure 1. Overlay of SEC chromatograms from the fractionation of HCCF (top row) and PAVIN (bottom row). Full chromatograms are shown on the right and the aggregate portions are magnified on the left. Vertical red bars indicate fraction bounds and the series color corresponds to the cycle number. The molecular mass abscissa was obtained by transforming the volume abscissa by the column's standard calibration curve. A280, absorbance at 280 nm; agg., aggregate; mAb, monoclonal antibody; SPF, small-protein fraction; HCCF, harvest cell culture fluid; PAVIN, protein A eluate that had been viral-inactivated, neutralized, and filtered.

Accepted Manuscript
Version of Record at: https://doi.org/10.1002/btpr.3343

3.2. Size analysis by DLS

Previous estimates have suggested that the hydrodynamic diameters of HMW impurities may be as much as several hundred nm,[29,41,60] which would not be expected to survive 0.2 μm sterile filtration. However, the HCCF and PAVIN aggregates observed here had already been subjected to depth filtration through an assembly that terminated in a 0.2 μm membrane. To provide a broader context for understanding their process behavior, therefore, the hydrodynamic radius (R_h) distributions of aggregate fractions were estimated using DLS and are shown in Figure 2. As with all DLS measurements, an inverse Laplace transform is required to interpret the measured time correlation function, which can be solved analytically for strictly monodisperse systems but is ill-posed otherwise.[47] This necessitates a regularization constraint, which is usually chosen to impose smoothness on the result, as well as the *a priori* selection of a regularization parameter α. Unfortunately the spread of the final R_h distribution can be sensitive to α, with values that are too low leading to disproportionate sensitivity and excessively narrow R_h peaks and values that are too high leading to inaccurately broad R_h peaks. The L-curve method was used to select the optimal value of α for the CONTIN analysis of each data set independently; these values are given in Supplementary Table S1 and are generally on the order of 0.1.[47,61]

The maximum R_h in all of the measured distributions appears to be appropriately < 100 nm and the mAb fractions appear to have monomodal distributions located at the correct R_h of ~5 nm. The main peaks in unfractionated feedstock distributions are also located at this size, although they are broader due to their inclusion of aggregates. The scattering intensity scales with R_h^6, so it is generally expected that even trace amounts of aggregates should be perceptible in the DLS analysis of bioprocess samples. This was the case for the HCCF feedstock, which has a subtle R_h peak around 50 nm, but no such peak was observed in the PAVIN, perhaps due to its relatively

Accepted Manuscript
Version of Record at: https://doi.org/10.1002/btpr.3343

lower aggregate concentration. Analysis of the small-aggregate SEC fractions reveals bimodal distributions with both a free mAb peak and an aggregate peak, the latter of which was estimated to be ~15 nm in the HCCF and ~25 nm in the PAVIN. The presence of the free mAb peak may be due to the incomplete SEC chromatographic resolution between the mAb and small-aggregate fractions. The aggregate subpopulation that was observed in the DLS analysis of the small-aggregate SEC fraction would be large enough to encompass several mAb molecules that are packed tightly or a few mAb molecules that are packed loosely, which seems generally consistent with size indications from the SEC chromatograms (Figure 1). The size distributions of the large-aggregate fractions appear to be broader than those of the small-aggregate fractions and shifted to slightly higher R_h. These were estimated for the HCCF and PAVIN to be unimodal and bimodal, respectively, but the apparent difference may be an artifact of the measurement noise and the smoothing, as they both span comparable R_h domains. Although the precise value depends on α, the largest aggregates have R_h values on the order of 50 nm.

Accepted Manuscript
Version of Record at: https://doi.org/10.1002/btpr.3343

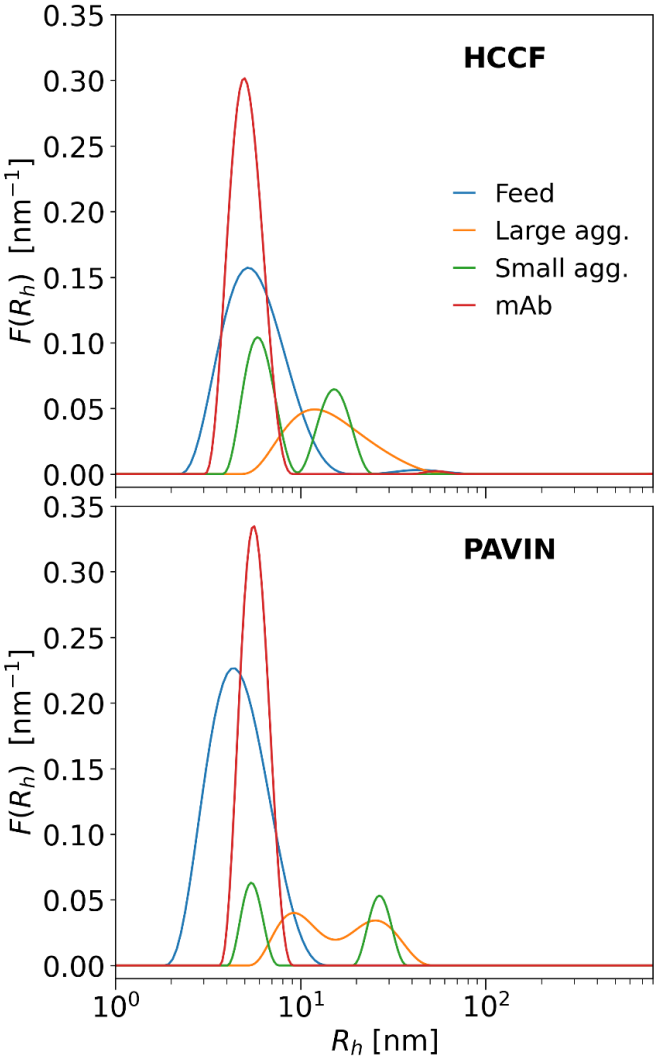

Figure 2. Normalized distributions of hydrodynamic radii (R_h) of unprocessed feedstocks and derivative SEC fractions from harvest cell culture fluid (HCCF, top) and protein A eluate that had been viral-inactivated, neutralized, and filtered (PAVIN, bottom).

Accepted Manuscript
Version of Record at: https://doi.org/10.1002/btpr.3343

3.3. Cross-digest proteomic analysis of SEC fractions

It was unclear which digest method would be better suited for the proteomic analysis of each of the different SEC fractions. Compared with the standard digest, the native digest can offer superior coverage by theoretically removing spectral interference between co-eluting mAb and HCP peptides. However, the lack of reduction and denaturation steps prior to digestion could hinder the identification of HCPs that are part of mAb solution aggregates. To resolve this uncertainty a cross-digest study was performed in which all SEC fractions and the unfractionated feedstocks were analyzed using both techniques in data-dependent acquisition (DDA) mode, which accomplishes HCP identification but not quantification. The HCCF SPF 2 fraction was the only exception, for which there was insufficient material to perform the native digest. Interference from mAb peptides is expected to influence the analysis of this fraction the least, as it should contain primarily HCPs that are < 15 kDa, protein fragments, and other low molecular weight absorbing impurities from the cell culture media.

After acquisition of the DDA data, alternative HCP candidates that could explain single-protein identifications were selected so as to minimize the total number of HCPs in the sample set union irrespective of biological function, biophysical properties, etc. This objective was chosen to avoid artificial inflation of the total number of HCPs identified, which may be trivial in the union of only a few samples that have little HCP overlap but more appreciable when comparing many samples. Relative to arbitrarily selecting the first HCP candidate in each protein identification, the selection algorithm reduced the total number of species in the sample set union by 15% (3118 vs. 2662).

An overview of all data from the cross-digest study is provided in Figure 3, which permits comparisons between the: digest methods, HCCF and PAVIN, feedstocks and derivative SEC

Accepted Manuscript
Version of Record at: https://doi.org/10.1002/btpr.3343

fractions, and the individual fractions. In the analysis of unfractionated HCCF only ~40% of the HCPs in the aggregates could be identified, whereas ~70% of HCPs in the mAb and SPF fractions were found. Although this would appear to suggest an identification bias toward free-solution HCPs, it may be simply a consequence of the larger number of species present in the aggregates and the apparent relative concentrations of the fractions (cf. the aggregate and SPF 1 peak sizes in Figure 1). The ~70% of HCPs that could be identified from the mAb and SPF fractions were mostly present in aggregate fractions as well.

Some of the multiple comparisons that Figure 3 enables are unpacked in Figure 4, including a comparison of the total coverage of the two digest methods that reveals the extent to which the standard digest is limited by the co-elution of mAb and HCP peptides. The native digest identified roughly twice as many HCPs as the standard digest in all HCCF samples (Figure 4A), and the difference was unexpectedly more pronounced in the PAVIN, for which the standard digest co-elution problem is greater due to the higher sample purity. However, there were 81 species in the sample set union that were able to be identified only using the standard digest (Figure 3 and 4B). The mean cysteine content of these species is significantly higher than that of the HCPs identified by only the native digest or those that were identified by both digest methods (based on one-way ANOVA and post-hoc pairwise comparisons), and the cysteine content is expected to correlate with disulfide bond content. The higher mean in the group that was uniquely identified by the standard digest is consistent with the difference in reduction between the two methods but the effect size is small.

Accepted Manuscript
Version of Record at: https://doi.org/10.1002/btpr.3343

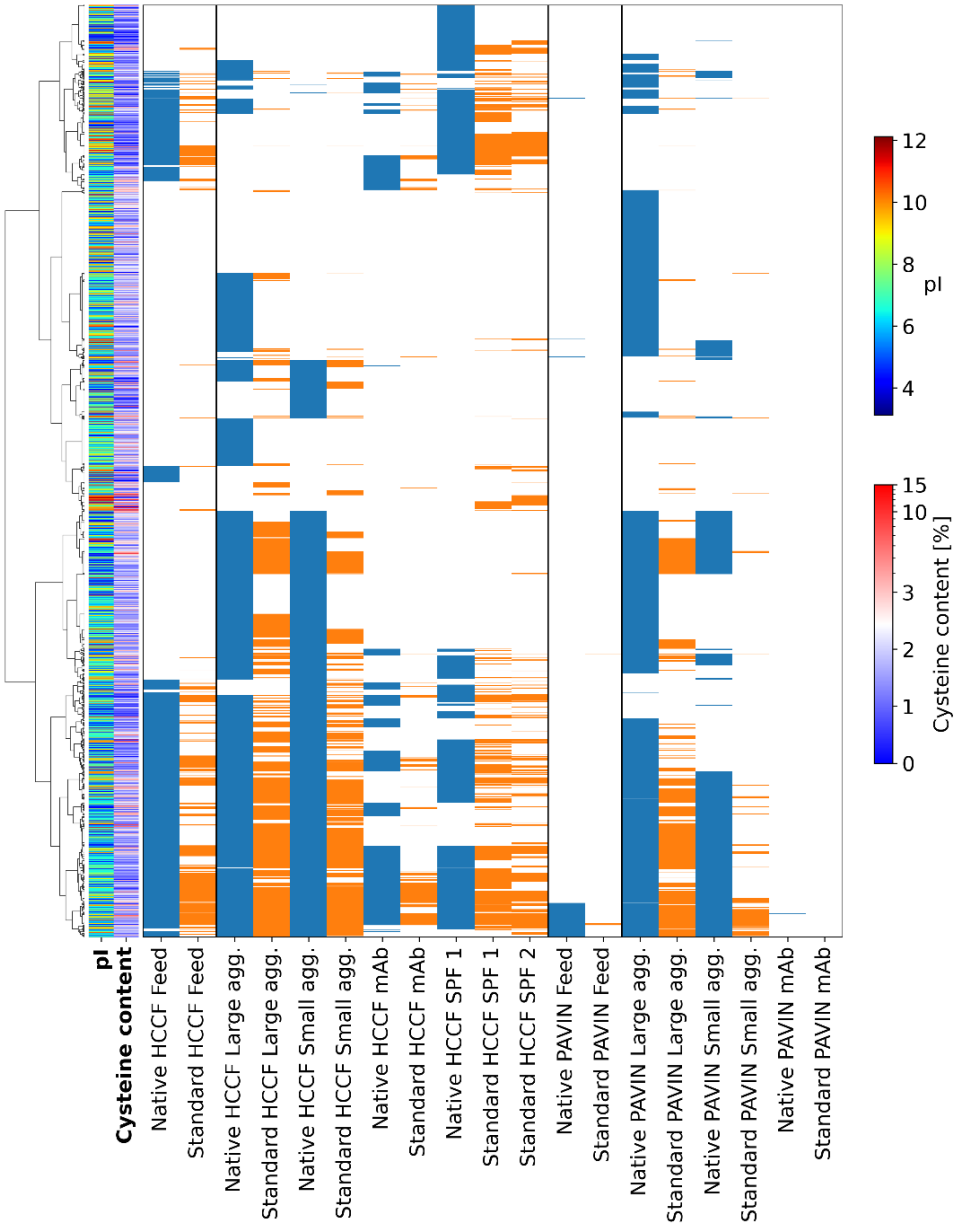

Figure 3. Clustered heat map of all binary HCP identifications from the data-dependent acquisition (DDA) cross-digest study. Blue and orange indicate positive identifications using the native and standard digest methods, respectively, whereas white represents no identification. HCPs (rows) are clustered, whereas samples (columns) are not, and boxes around the feedstock columns are intended only to facilitate comparison with the derivative SEC fractions. Each row is annotated with the calculated isoelectric point (pI) and cysteine content of the HCP, and the color bar for cysteine content is on a linear scale from 0% to 3% but logarithmic thereafter. The HCCF SPF 2 fraction was analyzed using only the standard digest due to material constraints.

Accepted Manuscript
Version of Record at: https://doi.org/10.1002/btpr.3343

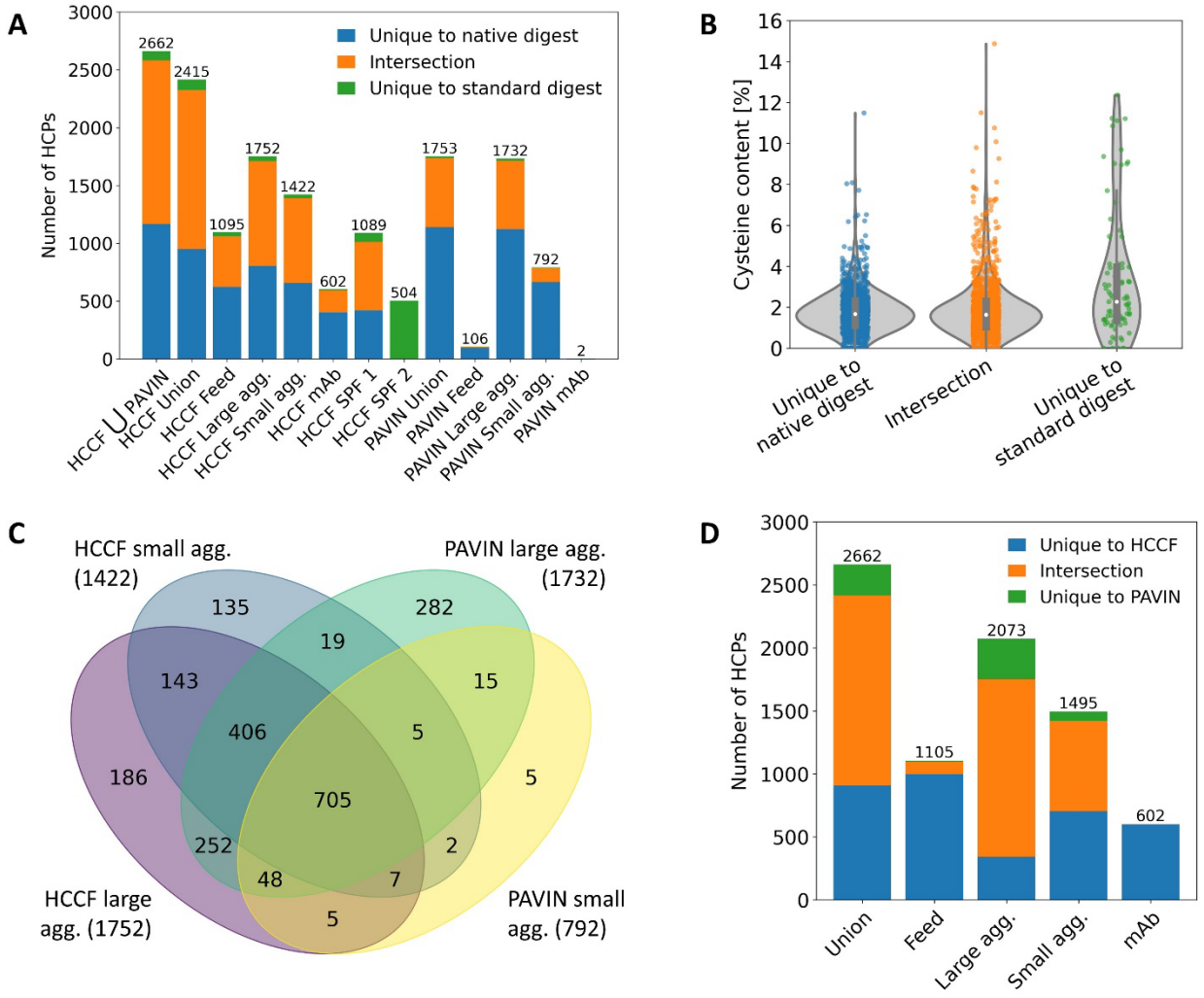

Figure 4. (A) Stacked bar chart of the number of HCPs that were uniquely identified by each digest method and those that were identified by both digest methods (intersection), where values above bars represent the total. HCCF ∪ PAVIN refers to the combination of all samples, HCCF union refers to the combination of all HCCF samples, and PAVIN union is analogous. The HCCF SPF 2 fraction was analyzed only using the standard digest due to material constraints. (B) Distribution of the cysteine content of HCPs in the digest comparison groups within HCCF ∪ PAVIN. Data points are superimposed on the density distributions and box plots are superimposed to show distribution quartiles, where the white point represents the median. (C) Overlap of HCPs in HCCF and PAVIN aggregate fractions after HCP identifications by the native and standard digest methods are united for each sample. Values in parentheses denote the total number of species that were identified in each fraction. (D) Stacked bar chart of the number of HCPs that were uniquely identified in each process pool and those that were identified in both pools (intersection) after HCP identifications by the native and standard digest methods are united for each of the common samples analyzed. Union represents the combination of all data.

Accepted Manuscript
Version of Record at: https://doi.org/10.1002/btpr.3343

The cross-digest study shows that roughly 80% of the identifiable HCPs were found in aggregate fractions and that there were hundreds of common HCPs among those fractions (Figure 3 and 4C). The majority of aggregate HCPs were shared by three or more fractions, as shown by the extensive overlap and comparably large numbers of HCPs in both HCCF large and small aggregates as well as the PAVIN large aggregates. The large aggregates appear to have been implicated in the persistence of more than a thousand HCPs through protein A chromatography, as hypothesized in prior work.[46] This is more clearly shown in Figure 4D, which compares the HCP overlap between the common samples derived from HCCF and PAVIN. Roughly 80% of HCPs in PAVIN aggregates were also identified in the corresponding HCCF aggregates, which is about double what was observed previously.[46]

Comparisons of HCP biophysical properties across SEC fractions reveal significant but relatively small effect sizes. About 10% of the HCPs identified in the SPFs have unphysically large molecular masses of > 100 kDa, which is presumably due to protein fragmentation (Supplementary Figure S2). This may be the result of CHO lysosomal protein degradation as well as protease activity in free solution.[46,62] The mean HCP mass in the SPF fractions is significantly lower than that in the HCCF aggregates, but the effect size is small. Similarly, a comparison of isoelectric point distributions does not reveal significant differences aside from a slightly higher mean pI in the HCCF SPF 2 fraction than in the other HCCF samples (Supplementary Figure S3).

Prior work has identified numerous HCPs that are difficult to remove in downstream bioprocessing.[6,8–14] About 200 of those species were identified in this study, the majority of which were present in both the HCCF and PAVIN (Supplementary Figures S4 and S5). Consistent with the trends observed in the full data set, the native digest offered superior coverage for the difficult-to-remove species, but the difference between the two digest methods was less pronounced than

Accepted Manuscript
Version of Record at: https://doi.org/10.1002/btpr.3343

for the full set of HCPs. Additionally, 56 HCPs that are considered to represent a high therapeutic risk were observed (Figure 5 and Supplementary Figure S6),[8] most of which appear to have persisted through protein A chromatography via aggregate association. Some of the high-risk species were not identifiable in the PAVIN without SEC fraction enrichment, including protein S100, which is suspected to elicit an immunogenic response.[63] Matrix metalloproteinases, peptidyl-prolyl cis-trans isomerase A, and protein disulfide-isomerase were also undetected in the PAVIN feed without SEC fractionation, and these can affect the drug quality deleteriously by promoting degradation and aggregation.[8]

As previous knowledge of HCP persistence is preconditioned by detectability, it may be useful to consider broader classes of HCPs that are generally of interest to downstream bioprocess development, such as those included in Supplementary Figure S7. The aggregates contain a heterogeneous mixture of HCPs with diverse biological functions, including heat shock proteins and other chaperones that may enable the CHO host to cope with the stress of high mAb production.[46,64] The association of these species with incorrectly-folded mAb molecules has been suggested as one possible nucleation mechanism for HCP-rich aggregates.[46] Nucleosomal arrays represent another putative nucleation center,[29] and this is supported by the presence of histones H2A and H2B in the large aggregates.

Accepted Manuscript
Version of Record at: https://doi.org/10.1002/btpr.3343

Figure 5. Sorted binary heat map of high-risk HCPs.[8] Blue and orange indicate positive identifications with the native and standard digest methods, respectively, whereas white represents no identification. Each HCP (row) is annotated with the type of its known or expected deleterious impact on the therapeutic quality.[8]

Accepted Manuscript
Version of Record at: https://doi.org/10.1002/btpr.3343

3.4. Quantitative SWATH proteomics

Given its general superiority, the native digest was used for the quantitative SWATH proteomic analysis of all SEC fractions except HCCF SPF 2, which was analyzed using the standard digest. SWATH identified slightly more species than the DDA analysis of each sample using the same digest method (Supplementary Figure S8), despite identifying slightly fewer species overall (2514 vs. 2581), which suggests that the SWATH spectral library constructed from the DDA results was appropriately comprehensive. The measured HCP masses are expressed as a fraction of the total analyzed mass of each sample, rather than the mAb mass, because some SEC fractions are expected to contain relatively little mAb. All SWATH data are depicted in Figure 6, which shows sample-specific HCP concentrations that have been normalized across the SEC fractions for each HCP individually. The normalized concentrations are generally highest in the large aggregates, although there are appreciable clusters where the highest concentration is observed in the HCCF small aggregates and SPF fractions.

Figure 6 suggests that HCP concentrations may be correlated between certain fractions, and this is investigated for all sample pairs in Figure 7, where the fraction-specific HCP concentrations have been logarithmically transformed because they span several orders of magnitude. The histograms show that HCP concentrations are approximately log-normally distributed within each SEC fraction, and the scatter plots reveal that HCP concentrations are positively correlated between all sample pairs. The strongest correlations are those among the aggregates as well as those between the PAVIN feed and its derivative aggregate fractions. HCP concentrations in the HCCF feed are also correlated relatively strongly with those in the HCCF mAb and HCCF SPF 1 samples.

Accepted Manuscript
Version of Record at: https://doi.org/10.1002/btpr.3343

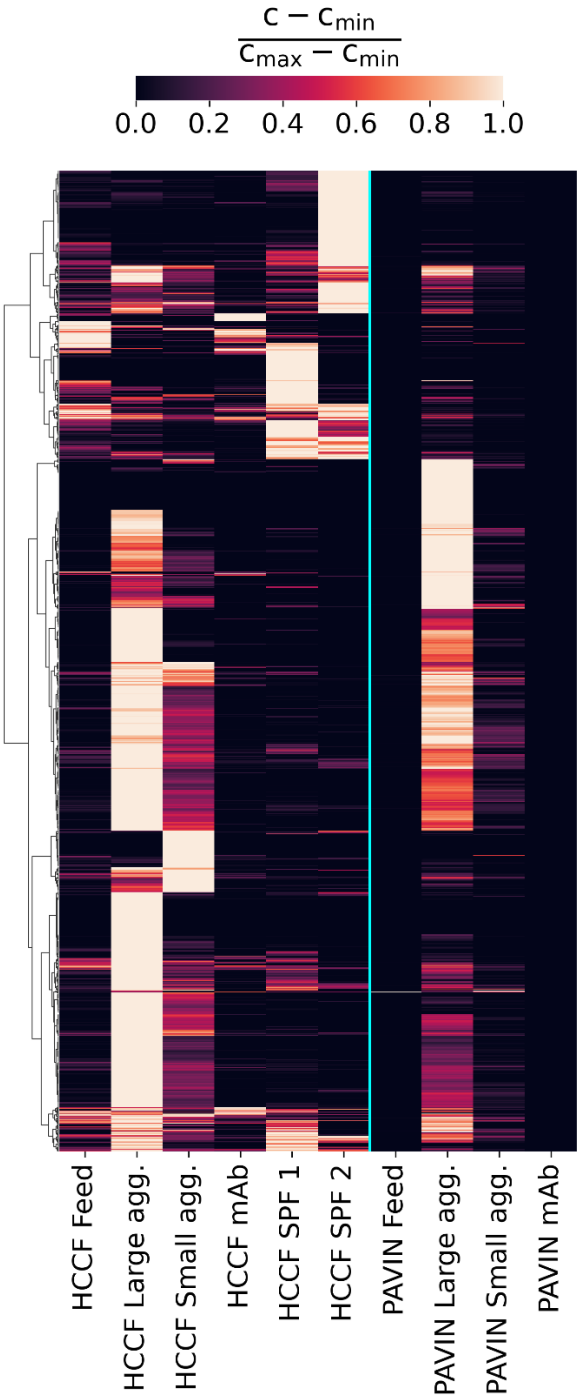

Figure 6. Clustered heat map of row-normalized HCP concentrations in the SEC fractions. HCPs (rows) are clustered, whereas samples (columns) are not. Negative identifications were assigned a concentration of zero and HCP concentrations were computed in units of ppm on a total mass basis (i.e. ng HCP / mg total for each SEC fraction) and row-normalized using the minimum and maximum concentrations observed across all samples.

Accepted Manuscript
Version of Record at: https://doi.org/10.1002/btpr.3343

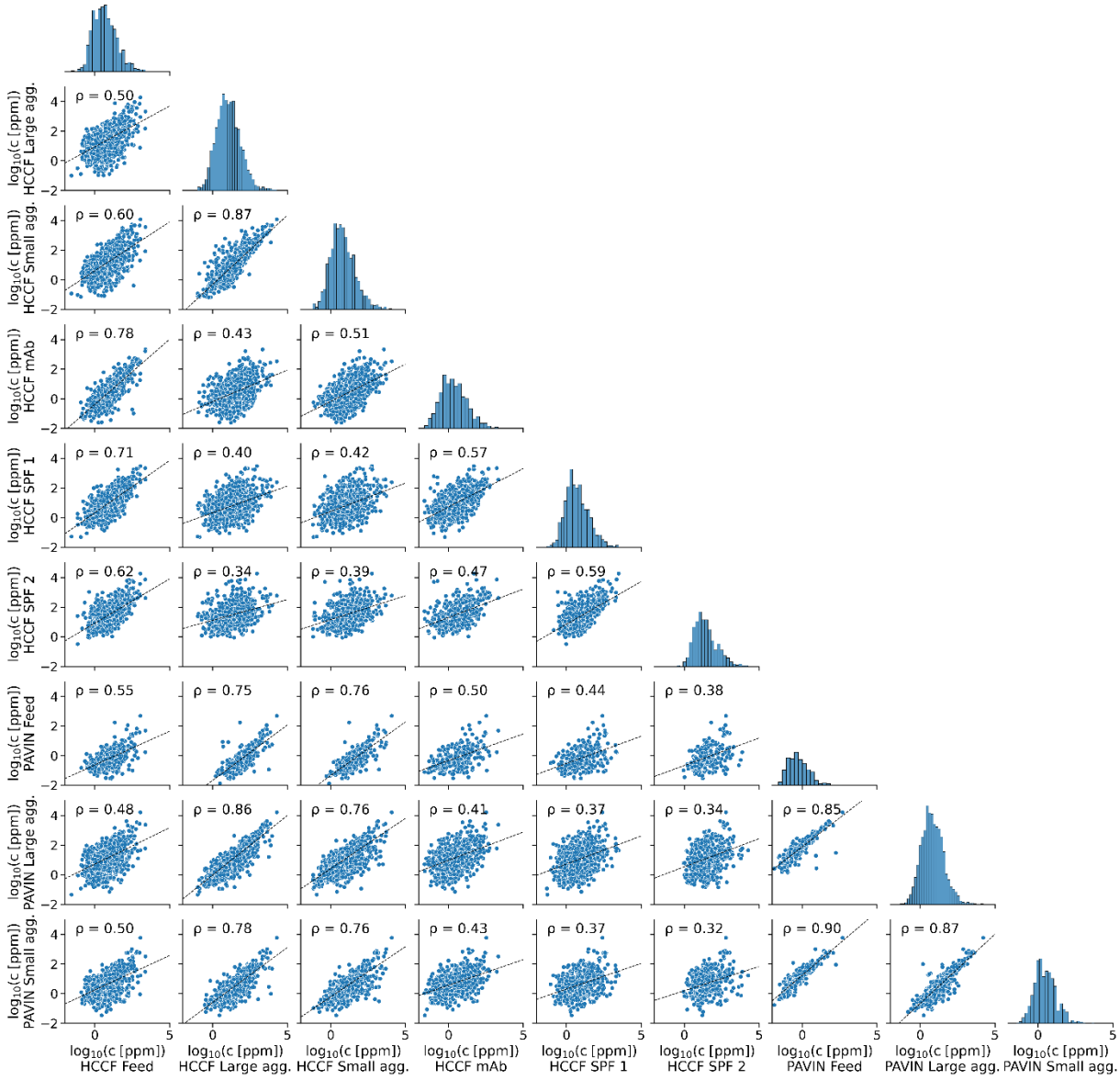

Figure 7. Pair-correlation plots of logarithmically transformed HCP concentrations in the SEC fractions. The abscissas of all plots and the ordinates of off-diagonal plots represent $\log_{10} c$, where the concentration c of each HCP has units of ppm on a total mass basis (i.e., ng HCP / mg total for each SEC fraction). Histograms of logarithmically transformed HCP concentrations in each sample are shown on the diagonal. Dashed lines represent linear regressions and the Pearson correlation coefficient ρ is shown for each sample pair. The PAVIN mAb sample is not included because it contained only four identifiable HCPs.

Accepted Manuscript
Version of Record at: https://doi.org/10.1002/btpr.3343

It is also informative to compare the abundances of difficult-to-remove HCPs with those of their counterparts (Figure 8). Within each fraction, the mean of the logarithmically transformed concentration distribution of difficult-to-remove species is higher than that of the other HCPs by about one log, a difference that was found to be statistically significant (based on two-way ANOVA). In general, the highest fraction-specific concentrations of difficult-to-remove species were observed in the large aggregates and the HCCF SPF fractions (Supplementary Figure S9), as was also the case for the subset of high-risk species (Figure 9). The high-risk species in HCCF SPF 1, which are primarily glutathione S-transferase isoforms, were mostly cleared by the protein A capture step. Many of the high-risk species in HCCF SPF 2 were detected in the HCCF feed but also cleared by the protein A step. Aggregate association appears to be responsible for the persistence of most high-risk HCPs into the PAVIN (Figures 5 and 9) and this is also the case for the broader classes of HCPs that may be of general interest to downstream bioprocessing (Supplementary Figure S10).

Accepted Manuscript
Version of Record at: https://doi.org/10.1002/btpr.3343

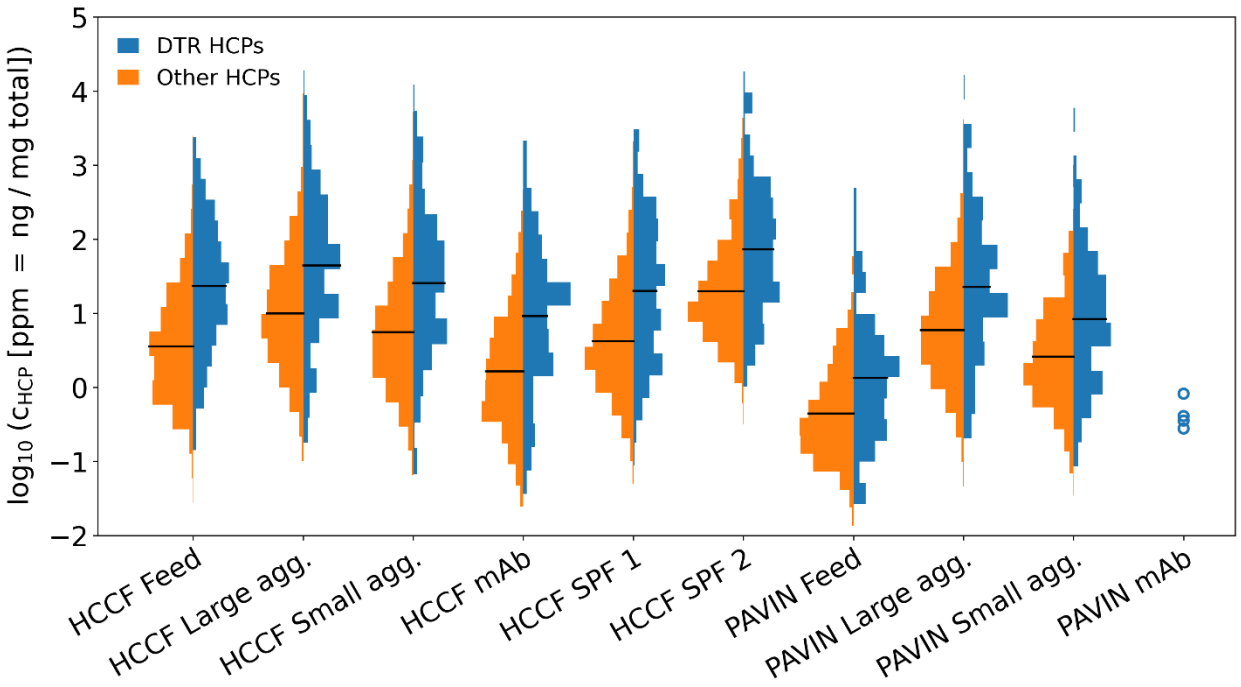

Figure 8. Split histogram plots comparing the logarithmically transformed HCP concentrations of species that have been generally considered difficult to remove (DTR) with their non-DTR counterparts in each SEC fraction.[6,8–12,14,65] Black lines represent the distribution means. The PAVIN mAb sample contained only four identifiable HCPs, all of which are considered DTR and are shown as individual points.

Accepted Manuscript
Version of Record at: https://doi.org/10.1002/btpr.3343

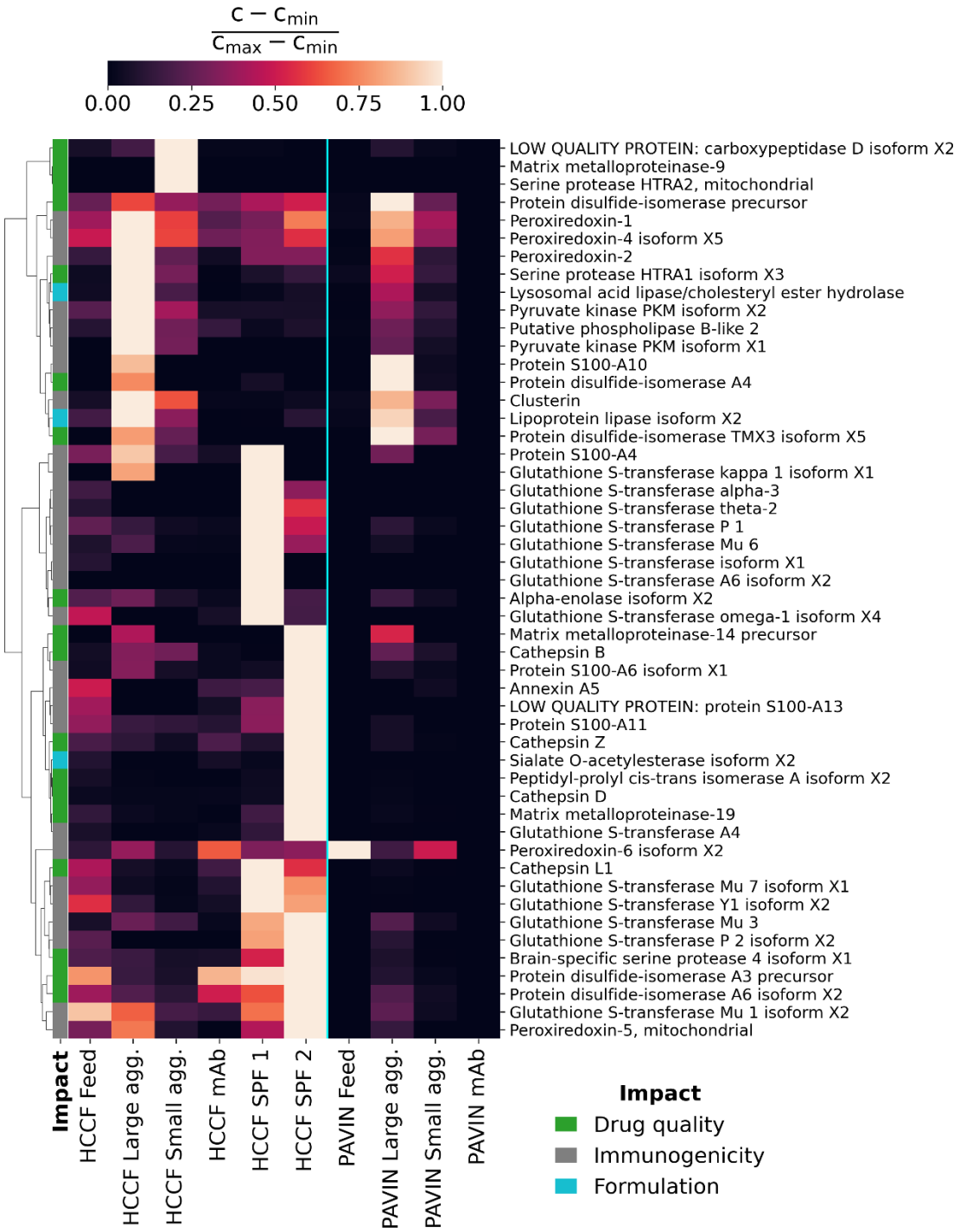

Figure 9. Clustered heat map of row-normalized HCP concentrations for species that are considered to represent a high therapeutic risk.[8] HCPs (rows) are clustered, whereas samples (columns) are not, and each HCP is annotated with the type of its known or expected deleterious impact on the therapeutic quality. Row-normalized HCP concentrations were computed as described in Figure 6.

Accepted Manuscript
Version of Record at: https://doi.org/10.1002/btpr.3343

3.5. SWATH vs. ELISA

To assess the distribution of HCP mass among the SEC fractions, the HCP concentrations that were measured by SWATH were added for each sample (Figure 10A). These were compared to ELISA measurements of the total HCP content of each fraction and were found to agree generally for the aggregates and the PAVIN feed. ELISA appeared to overestimate the HCP content of the HCCF feed, mAb and SPF fractions, which SWATH measured to have HCP concentrations comparable to or lower than those of the HCCF aggregates. The HCP content of the PAVIN large aggregates was comparable to that of the HCCF aggregates but the HCP concentration in the PAVIN small aggregates was noticeably lower. SDS-PAGE gels (not shown) suggest that mAbs or mAb fragments are the principal constituent of the aggregates and the HCCF SPF 1 fraction, consistent with the SWATH HCP concentrations that sum to $< 5 \times 10^5$ ppm (on a total mass basis).

An approximate mass balance was performed using the fractional SEC peak areas (Figure 1) and assuming equivalent extinction coefficients; the relative standard deviation was 11.6%, which includes propagated error estimates from Bradford, SWATH and SEC analyses. Comparison with SWATH-determined HCP concentrations (Figure 10A) would suggest that the PAVIN aggregates and mAb fraction contain 5.1% more HCP mass than was measured in the PAVIN feed, which is unphysically high but within experimental error. When normalized by the sum of HCP masses in the PAVIN SEC fractions rather than the HCP mass in the PAVIN feed, the aggregates are shown to contain 99.9% of the HCP mass. However, this value may be overestimated because HCP mass contributions from the shoulder in the SPF 1 peak location (Figure S1) were not included; material requirements precluded the proteomic analysis of that sample. A lower-bound value of 87.9% may be obtained by assuming the shoulder in the PAVIN

Accepted Manuscript
Version of Record at: https://doi.org/10.1002/btpr.3343

chromatogram has the same fraction-specific HCP concentration as the SPF 1 peak in the HCCF, although the substantial convolution of the SEC peak areas of the PAVIN mAb fraction and SPF 1 (Figure S1) likely causes an appreciable underestimation of the aggregate contributions to the total PAVIN HCP mass, which inflates the relative importance of SPF 1. Alternatively, if the ratio of HCP mass in SPF 1 to that in the mAb fraction is assumed to be the same in the HCCF and PAVIN, the aggregates are again estimated to contain 99.9% of the PAVIN HCP mass. The contributions of SPF 1 are negligible under this assumption largely as a result of the much lower HCP concentration in the PAVIN than in the HCCF mAb fraction (Figure 10A).

The assumption of equivalent extinction coefficients is not tenable for the HCCF because the SPF 2 fraction contains low molecular weight absorbing impurities, which precludes a comparable assessment of the HCP mass distribution within the HCCF. However, it is possible to compare the relative HCP content of the two aggregate fractions within each process stream, which shows that large aggregates harbor 1.9 and 3.9-fold more HCP mass than small aggregates in the HCCF and PAVIN, respectively. A comparison between the number of protein species detected in each sample and the total HCP concentration measured by the SWATH analysis (Figure 10B) shows that these metrics appear to correlate positively for all but one of the SEC fractions. The outlier is HCCF SPF 2, which has a higher SWATH HCP concentration than the other samples with similar numbers of HCPs. However, this point may be overestimated because full protein molecular masses were used in the computation of the HCP mass content, but the SPF 2 fraction is expected to contain primarily protein fragments.

Accepted Manuscript
Version of Record at: https://doi.org/10.1002/btpr.3343

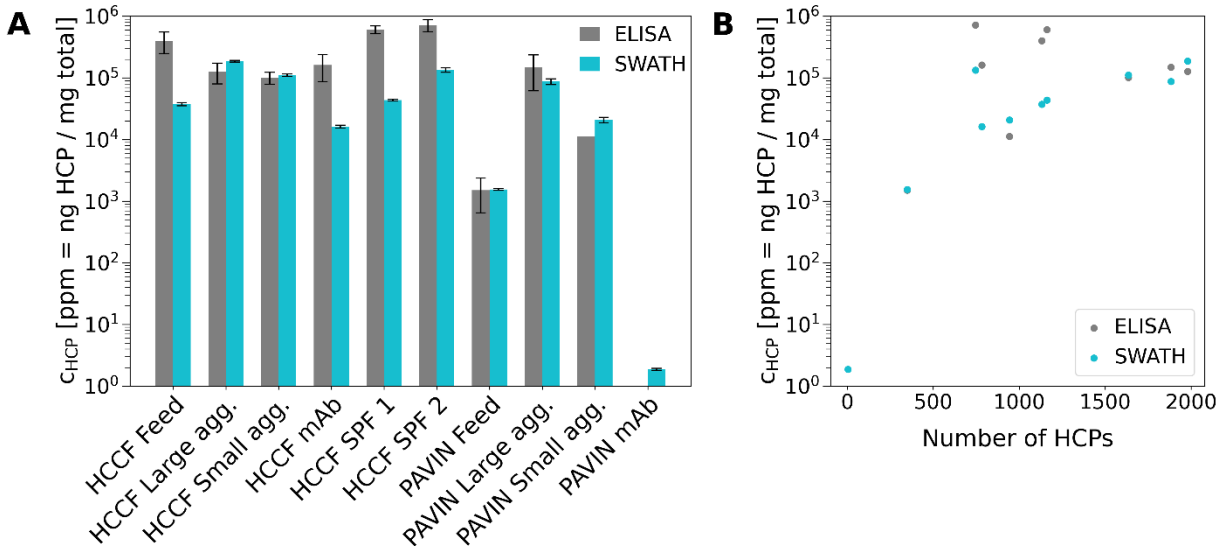

Figure 10. (A) Comparison of ELISA and SWATH quantifications of the total HCP content in feedstocks and derivative SEC fractions. Total protein masses (i.e., the denominators) were estimated using the Bradford assay and error bars represent 1× sample standard deviations, which were computed using standard error propagation rules to combine the uncertainties from Bradford and ELISA / SWATH measurements. No error is shown for the ELISA PAVIN small-aggregate sample because only one serial dilution point fell on the ELISA calibration curve. The HCP content of the PAVIN mAb fraction also did not fall on the ELISA calibration curve but contained < 125 ppm HCP based on the assay's limit of detection and the sample dilution. (B) Scatter plot of the total HCP concentration that was determined using ELISA or SWATH vs. the total number of HCPs that were identified in the SWATH analysis of each SEC fraction.

Accepted Manuscript
Version of Record at: https://doi.org/10.1002/btpr.3343

4. Discussion

The mechanisms of HCP persistence through protein A chromatography remain incompletely understood but aggregate association appears to mediate the survival of thousands of HCPs. Although it is unknown to what extent mAb properties influence this phenomenon, similar observations in a study of several mAb products suggest that this is not product-specific.[46] Aggregate-associated HCPs comprise the majority of the PAVIN HCP mass because the protein A step appears to clear SPF impurities much more effectively than aggregates (Figure 1). Such aggregates must bind to the protein A resin and partially co-elute with the product to persist into the PAVIN. The HCCF large- and small-aggregate fractions contained comparably large numbers of HCPs but the PAVIN small aggregates contained only about half as many (Figure 4A). It is possible that the protein A step may have cleared an HCP-rich subpopulation of the small aggregates while enriching mAb oligomers that are unassociated with HCPs, thus leading to the observed reduction in the number of identified species and the total HCP concentration (Figure 10A). This may also explain the increase in the ratio of total HCP mass in the large aggregates relative to that in the small aggregates (1.9-fold in HCCF vs. 3.9-fold in PAVIN).

Although the SPF 1 fraction from the PAVIN could not be analyzed, it seems reasonable to expect that its HCP content may decrease similarly because the protein A step should increase the mAb Fc fragment content of SPF 1 while clearing some HCPs. A subpopulation of the SPF 1 HCPs may nonetheless survive via nonspecific adsorption and co-elution. This subpopulation may contain species that are more basic than the quadrivalent Z-domain ligand in the MAbSelect SuRe LX resin, which has a pI ~5.1 and therefore presents a cation-exchange-like surface for nonspecific adsorption at neutral pH.[66–68] Bracketing the mass contribution of these HCPs to the total PAVIN HCP content requires alternative assumptions, which introduces ambiguity in the assessment of

Accepted Manuscript
Version of Record at: https://doi.org/10.1002/btpr.3343

the HCP mass distribution among the aggregates and other SEC fractions. The value of 87.9% is considered to represent an underestimate of the PAVIN aggregate contributions to the total PAVIN HCP mass, not only because the SPF 1 peak area is substantially convoluted with the tail of the mAb peak (Figure S1) but also because the protein A step is expected to increase the relative mAb Fc fragment content of SPF 1. The estimates that are closer to 99% are expected to be more accurate, although the accumulated error in the mass balance (relative standard deviation of 11.6%) makes all of these values comparably large.

Many of the HCPs present in aggregates were not necessarily identified in the parent unfractionated feedstocks. This illustrates one of the salient challenges of shotgun proteomic analysis: the well-known and unavoidable problem of false negatives. Positive HCP identifications may be asserted with statistical confidence but the absence of an HCP cannot be confidently concluded from a negative result. In principle, a negative result would suggest that an HCP of interest is not present above its limit of detection but this interpretation is obfuscated by co-elution of mAb peptides in the standard digest and sample loss during undigested mAb precipitation in the native digest. This complication could hypothetically be avoided using positive controls that are spiked with HCPs of interest to determine species-specific limits of detection in the experimental sample but doing so is generally infeasible for the thousands of HCPs that may be present and also requires prior knowledge of which HCPs are of interest. This constitutes a nontrivial problem for mAb downstream processing studies that seek to demonstrate the clearance of dilute but high-risk HCPs based on negative identification results.

SEC fractionation may be construed as an HCP enrichment technique, albeit a laborious one, that reduces the extent of the false negatives problem. Removal of the mAb peak enables a deeper analysis of aggregate fractions and the SPFs that contain many HCP species. Comparison

Accepted Manuscript
Version of Record at: https://doi.org/10.1002/btpr.3343

of feedstocks with the respective sample unions (Figure 4A) shows that SEC fractionation led to more than 2-fold and 16-fold increases in the HCP coverage for the analysis of HCCF and PAVIN, respectively. The more substantial benefit for the PAVIN is due to its higher relative mAb content, because far fewer HCPs could be detected in the PAVIN feed than in the HCCF feed, but comparable numbers were present in the PAVIN and HCCF large aggregates. In general, SEC fractionation coupled with the native digest substantially improves HCP coverage and may be useful for constructing relatively comprehensive spectral libraries to support quantitative SWATH proteomics. Unfortunately, these techniques do not eliminate the false negatives problem entirely, as evidenced by the identification of some species in the PAVIN that were not identified in the HCCF (Figure 4D). The false negatives problem may also partially explain the apparent positive correlation between the total HCP concentration and the number of species that were identified in SWATH proteomics (Figure 10B), as has also been reported previously for protein A eluate pools and in spiking studies.[10,22]

The SWATH analysis reveals that HCP concentrations in downstream bioprocess pools are log-normally distributed, and the abundance of individual HCPs is positively correlated across HCCF and PAVIN SEC fractions (Figure 7). Species that have previously been recognized as difficult to remove are about one order of magnitude more concentrated than their counterparts on average (Figure 8), which may make them easier to detect and may partially explain why they have been found to persist in prior studies. HCPs that have higher starting concentrations in the HCCF are also expected to be more challenging for downstream separations generally, which seems consistent with their difficult-to-remove designation. The aggregates contain both difficult-to-remove and high-risk HCPs as well as several other proteases and lipases. It is unclear whether the enzymatic activity of these species is mitigated by their association with the HMW structures, as

Accepted Manuscript
Version of Record at: https://doi.org/10.1002/btpr.3343

would be suggested by the apparent stability of the aggregates. It is also unknown whether HCP dissociation from the aggregates proceeds to any appreciable extent, which could make the aggregates function like HCP reservoirs in purified mAb solutions. Regardless, the adsorption of HMW impurities on protein A chromatography resins could contribute to column fouling and capacity loss by hindering the diffusional transport of mAbs into resin particles.[29] This possibility is supported by the observation that the largest aggregates in this study have hydrodynamic radii that are comparable to the apparent pore radii of several commercial protein A chromatography resins.[69]

5. Conclusions

The complexity of bioprocess samples and the inherent ambiguities of shotgun proteomic techniques lead to inevitable false negatives in the observed HCP composition of mAb solution impurities. This problem becomes more pronounced as the mAb purity increases and it constitutes a nontrivial challenge for confidently demonstrating HCP clearance through downstream operations. Given these limitations, it is essential to incorporate coverage enhancement techniques like the native digest, which appears to offer greater HCP coverage than the standard digest, not only in purified mAb solutions but also in HCCF. Coupled with SEC fractionation, this technique reveals broader overlap among the HCP profiles of HCCF and PAVIN than has been observed previously, because the majority of identifiable HCPs are conserved in HMW impurities that are concentrated through the capture step (Figure 1), whereas the concentration of SPF impurities is substantially reduced. This is reflected in SWATH and ELISA measurements that indicate that most of the PAVIN HCP mass is present in aggregate structures. Although the SEC fractionation

represents an additional analytical step, it has the added benefit of explaining differences in persistence behavior of the constituent HCCF fractions through protein A chromatography.

Disclosure of interest

The authors report no conflict of interest.

Data availability

Proteomics data are available in 1_proteomics_data_v_05_01.xlsx.

Supplementary material

Supplementary tables and figures may be found in 1_proteomics_supp_v_05_01.docx.

ORCID

Chase E. Herman: https://orcid.org/0000-0002-9989-9604

Xuankuo Xu: https://orcid.org/0000-0002-5557-0284

Kelvin H. Lee: http://orcid.org/0000-0003-0908-2981

Abraham M. Lenhoff: https://orcid.org/0000-0002-7831-219X

Author contributions

Chase E. Herman: conceptualization, data curation, formal analysis, investigation, methodology, project administration, validation, visualization, writing – original draft, writing – review & editing. **Lie Min:** data curation, investigation, methodology, validation. **Leila H. Choe:** data curation, investigation, methodology, validation. **Ronald W. Maurer:** resources. **Xuankuo Xu:**

funding acquisition, project administration, resources, supervision, writing – review & editing. **Sanchayita Ghose:** funding acquisition. **Kelvin H. Lee:** supervision. **Abraham M. Lenhoff:** conceptualization, formal analysis, funding acquisition, project administration, supervision, writing – review & editing.

Acknowledgments

We thank Bristol Myers Squibb for providing materials and financial support.

References

1. U.S. Food and Drug Administration. *Q8(R2) Pharmaceutical Development*.; 2009. https://www.fda.gov/media/71535/download
2. Szkodny AC, Lee KH. Biopharmaceutical manufacturing: Historical perspectives and future directions. *Annu Rev Chem Biomol Eng*. 2022;13:141-165. doi:10.1146/annurev-chembioeng-092220-125832
3. Walsh G. Biopharmaceutical benchmarks 2018. *Nat Biotechnol*. 2018;36(12):1136-1145. doi:10.1038/nbt.4305
4. Albrecht S, Kaisermayer C, Gallagher C, Farrell A, Lindeberg A, Bones J. Proteomics in biomanufacturing control: Protein dynamics of CHO-K1 cells and conditioned media during apoptosis and necrosis. *Biotechnol Bioeng*. 2018;115(6):1509-1520. doi:10.1002/bit.26563
5. Baycin-Hizal D, Tabb DL, Chaerkady R, et al. Proteomic analysis of Chinese hamster ovary cells. *J Proteome Res*. 2012;11(11):5265-5276. doi:10.1021/pr300476w

6. Aboulaich N, Chung WK, Thompson JH, 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. doi:10.1002/btpr.1948
7. Shukla AA, Hinckley P. Host cell protein clearance during protein A chromatography: Development of an improved column wash step. *Biotechnol Prog*. 2008;24(5):1115-1121. doi:10.1002/btpr.50
8. Jones M, Palackal N, Wang F, et al. "High-risk" host cell proteins (HCPs): A multi-company collaborative view. *Biotechnol Bioeng*. 2021;118(8):2870-2885. doi:10.1002/bit.27808
9. Levy NE, Valente KN, Choe LH, Lee KH, Lenhoff AM. Identification and characterization of host cell protein product-associated impurities in monoclonal antibody bioprocessing. *Biotechnol Bioeng*. 2014;111(5):904-912. doi:10.1002/bit.25158
10. Zhang Q, Goetze AM, Cui H, et al. Characterization of the co-elution of host cell proteins with monoclonal antibodies during protein A purification. *Biotechnol Prog*. 2016;32(3):708-717. doi:10.1002/btpr.2272
11. Levy NE, Valente KN, Lee KH, Lenhoff AM. Host cell protein impurities in chromatographic polishing steps for monoclonal antibody purification. *Biotechnol Bioeng*. 2016;113(6):1260-1272. doi:10.1002/bit.25882
12. Molden R, Hu M, Yen E S, et al. Host cell protein profiling of commercial therapeutic protein drugs as a benchmark for monoclonal antibody-based therapeutic protein development. *MAbs*. 2021;13(1):e1955811. doi:10.1080/19420862.2021.1955811

13. Falkenberg H, Waldera-Lupa DM, Vanderlaan M, et al. Mass spectrometric evaluation of upstream and downstream process influences on host cell protein patterns in biopharmaceutical products. *Biotechnol Prog*. 2019;35(3):1-12. doi:10.1002/btpr.2788

14. Kreimer S, Gao Y, Ray S, 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*. 2017;89(10):5294-5302. doi:10.1021/acs.analchem.6b04892

15. Bee JS, Tie L, Johnson D, Dimitrova MN, Jusino KC, Afdahl CD. Trace levels of the CHO host cell protease cathepsin D caused particle formation in a monoclonal antibody product. *Biotechnol Prog*. 2015;31(5):1360-1369. doi:10.1002/btpr.2150

16. Fischer SK, Cheu M, Peng K, et al. Specific immune response to phospholipase B-like 2 protein, a host cell impurity in Lebrikizumab clinical material. *AAPS J*. 2017;19(1):254-263. doi:10.1208/s12248-016-9998-7

17. Gilgunn S, El-Sabbahy H, Albrecht 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. doi:10.1016/j.chroma.2019.02.056

18. Chon JH, Zarbis-Papastoitsis G. Advances in the production and downstream processing of antibodies. *N Biotechnol*. 2011;28(5):458-463. doi:10.1016/j.nbt.2011.03.015

19. Zhu-Shimoni J, Yu C, Nishihara J, et al. Host cell protein testing by ELISAs and the use of orthogonal methods. *Biotechnol Bioeng*. 2014;111(12):2367-2379. doi:10.1002/bit.25327

20. Wang X, Hunter AK, Mozier NM. Host cell proteins in biologics development: Identification, quantitation and risk assessment. *Biotechnol Bioeng*. 2009;103(3):446-458. doi:10.1002/bit.22304

21. Eaton LC. Host cell contaminant protein assay development for recombinant biopharmaceuticals. *J Chromatogr A*. 1995;705(1):105-114. doi:10.1016/0021-9673(94)01249-E

22. Huang L, Wang N, Mitchell CE, Brownlee T, Maple SR, DeFelippis MR. A novel sample preparation for shotgun proteomics characterization of HCPs in antibodies. *Anal Chem*. 2017;89(10):5436-5444. doi:10.1021/acs.analchem.7b00304

23. Johnson RO, Greer T, Cejkov M, Zheng X, Li N. A combination of FAIMS, protein A depletion, and native digest conditions enables deep proteomic profiling of host cell proteins in monoclonal antibodies. *Anal Chem*. 2020;92(15):10475-10484. doi:10.1021/acs.analchem.0c01175

24. Nie S, Greer T, Johnson RO, Zheng X, Torri A, Li N. Simple and sensitive method for deep profiling of host cell proteins in therapeutic antibodies by combining ultra-low trypsin concentration digestion, long chromatographic gradients, and BoxCar mass spectrometry acquisition. *Anal Chem*. 2021;93(10):4383-4390. doi:10.1021/acs.analchem.0c03931

25. Li D, Farchone A, Zhu Q, et al. Fast, robust, and sensitive identification of residual host cell proteins in recombinant monoclonal antibodies using sodium deoxycholate

assisted digestion. *Anal Chem*. 2020;92(17):11888-11894. doi:10.1021/acs.analchem.0c02258

26. Chen IH, 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. doi:10.1021/acs.analchem.9b05081
27. Yang F, Li D, Kufer R, et al. 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. doi:10.1021/acs.analchem.1c03095
28. Vázquez-Rey M, Lang DA. Aggregates in monoclonal antibody manufacturing processes. *Biotechnol Bioeng*. 2011;108(7):1494-1508. doi:10.1002/bit.23155
29. Gagnon P, Nian R, Lee J, et al. Nonspecific interactions of chromatin with immunoglobulin G and protein A, and their impact on purification performance. *J Chromatogr A*. 2014;1340:68-78. doi:10.1016/j.chroma.2014.03.010
30. Gagnon P, Nian R, Tan L, et al. Chromatin-mediated depression of fractionation performance on electronegative multimodal chromatography media, its prevention, and ramifications for purification of immunoglobulin G. *J Chromatogr A*. 2014;1374:145-155. doi:10.1016/j.chroma.2014.11.052
31. Gagnon P, Nian R, Yang Y, Yang Q, Lim CL. Non-immunospecific association of immunoglobulin G with chromatin during elution from protein A inflates host contamination, aggregate content, and antibody loss. *J Chromatogr A*. 2015;1408:151-160. doi:10.1016/j.chroma.2015.07.017
32. Nian R, Gagnon P. Advance chromatin extraction enhances performance and productivity of cation exchange chromatography-based capture of immunoglobulin

G monoclonal antibodies. *J Chromatogr A*. 2016;1453:54-61. doi:10.1016/j.chroma.2016.05.029

33. Chen Q, Latiff SMA, Toh P, et al. A simple and efficient purification platform for monoclonal antibody production based on chromatin-directed cell culture clarification integrated with precipitation and void-exclusion anion exchange chromatography. *J Biotechnol*. 2016;236:128-140. doi:10.1016/j.jbiotec.2016.08.014

34. Nian R, Zhang W, Tan L, et al. Advance chromatin extraction improves capture performance of protein A affinity chromatography. *J Chromatogr A*. 2016;1431:1-7. doi:10.1016/j.chroma.2015.12.044

35. Goey CH, Alhuthali S, Kontoravdi C. Host cell protein removal from biopharmaceutical preparations: Towards the implementation of quality by design. *Biotechnol Adv*. 2018;36:1223-1237. doi:10.1016/j.biotechadv.2018.03.021

36. Farrell A, Mittermayr S, Morrissey B, et al. Quantitative host cell protein analysis using two dimensional data independent LC-MS^{E}. *Anal Chem*. 2015;87(18):9186-9193. doi:10.1021/acs.analchem.5b01377

37. Hogwood CEM, Ahmad SS, Tarrant RD, Bracewell DG, Smales CM. An ultra scale-down approach identifies host cell protein differences across a panel of mAb producing CHO cell line variants. *Biotechnol J*. 2016;11(3):415-424. doi:10.1002/biot.201500010

38. Park JH, Jin JH, Lim MS, An HJ, Kim JW, Lee GM. Proteomic analysis of host cell protein dynamics in the culture supernatants of antibody-producing CHO cells. *Sci Rep*. 2017;7:44246. doi:10.1038/srep44246

39. Tait AS, Hogwood CEM, Smales CM, Bracewell DG. Host cell protein dynamics in the supernatant of a mAb producing CHO cell line. *Biotechnol Bioeng*. 2012;109(4):971-982. doi:10.1002/bit.24383

40. Jin M, Szapiel N, Zhang J, Hickey J, Ghose S. Profiling of host cell proteins by two-dimensional difference gel electrophoresis (2D-DIGE): Implications for downstream process development. *Biotechnol Bioeng*. 2010;105(2):306-316. doi:10.1002/bit.22532

41. van de Velde J, Saller MJ, Eyer K, Voloshin A. Chromatographic clarification overcomes chromatin-mediated hitch-hiking interactions on Protein A capture column. *Biotechnol Bioeng*. 2020;117(11):3413-3421. doi:10.1002/bit.27513

42. Ichihara T, Ito T, Kurisu Y, Galipeau K, Gillespie C. Integrated flow-through purification for therapeutic monoclonal antibodies processing. *MAbs*. 2018;10(2):325-334. doi:10.1080/19420862.2017.1417717

43. Ichihara T, Ito T, Gillespie C. Polishing approach with fully connected flow-through purification for therapeutic monoclonal antibody. *Eng Life Sci*. 2019;19:31-36. doi:10.1002/elsc.201800123

44. Luo H, Du Q, Qian C, et al. Formation of transient highly-charged mAb clusters strengthens interactions with host cell proteins and results in poor clearance of host cell proteins by protein A chromatography. *J Chromatogr A*. 2022;1679:463385. doi:10.1016/j.chroma.2022.463385

45. Oh YH, Becker ML, Mendola KM, et al. Factors affecting product association as a mechanism of host-cell protein persistence in bioprocessing. *In preparation*. 2023.

46. Oh YH, Becker ML, Mendola KM, et al. Characterization and implications of host-cell protein aggregates in biopharmaceutical processing. *Biotechnol Bioeng*. 2023;120(4):1068-1080. doi:10.1002/bit.28325

47. Scotti A, Liu W, Hyatt JS, et al. The CONTIN algorithm and its application to determine the size distribution of microgel suspensions. *J Chem Phys*. 2015;142(23):234905. doi:10.1063/1.4921686

48. Stetefeld J, McKenna SA, Patel TR. Dynamic light scattering: a practical guide and applications in biomedical sciences. *Biophys Rev*. 2016;8(4):409-427. doi:10.1007/s12551-016-0218-6

49. Hou S, Jones SW, Choe LH, Papoutsakis ET, Lee KH. Workflow for quantitative proteomic analysis of *Clostridium acetobutylicum* ATCC 824 using iTRAQ tags. *Methods*. 2013;61(3):269-276. doi:10.1016/j.ymeth.2013.03.013

50. Valente KN, Lenhoff AM, Lee KH. Expression of difficult-to-remove host cell protein impurities during extended Chinese hamster ovary cell culture and their impact on continuous bioprocessing. *Biotechnol Bioeng*. 2015;112(6):1232-1242. doi:10.1002/bit.25515

51. Hamaker NK, Min L, Lee KH. Comprehensive assessment of host cell protein expression after extended culture and bioreactor production of CHO cell lines. *Biotechnol Bioeng*. 2022;119(8):2221-2238. doi:10.1002/bit.28128

52. Gillet LC, Navarro P, Tate S, et al. Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: A new concept for consistent and accurate proteome analysis. *Molecular and Cellular Proteomics*. 2012;11(6). doi:10.1074/mcp.O111.016717

53. Ludwig C, Gillet L, Rosenberger G, Amon S, Collins BC, Aebersold R. Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial. *Mol Syst Biol*. 2018;14(8):e8126. doi:10.15252/msb.20178126

54. Husson G, Delangle A, O'Hara J, et al. Dual data-independent acquisition approach combining global HCP profiling and absolute quantification of key impurities during bioprocess development. *Anal Chem*. 2018;90(2):1241-1247. doi:10.1021/acs.analchem.7b03965

55. Shilov I v, Seymour SL, Patel AA, et al. The paragon algorithm, a next generation search engine that uses sequence temperature values and feature probabilities to identify peptides from tandem mass spectra. *Molecular and Cellular Proteomics*. 2007;6(9):1638-1655. doi:10.1074/mcp.T600050-MCP200

56. MacLean B, Tomazela DM, Shulman N, et al. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics*. 2010;26(7):966-968. doi:10.1093/bioinformatics/btq054

57. Younghoon Oh. *Mechanisms of Host-Cell Protein Persistence in Monoclonal Antibody Processing*. Ph.D. Thesis. University of Delaware; 2022.

58. Reiter L, Rinner O, Picotti P, et al. mProphet: automated data processing and statistical validation for large-scale SRM experiments. *Nat Methods*. 2011;8(5):430-435. doi:10.1038/nmeth.1584

59. Choi M, Chang CY, Clough T, et al. MSstats: An R package for statistical analysis of quantitative mass spectrometry-based proteomic experiments. *Bioinformatics*. 2014;30(17):2524-2526. doi:10.1093/bioinformatics/btu305

60. Bermudez O, Forciniti D. Aggregation and denaturation of antibodies: A capillary electrophoresis, dynamic light scattering, and aqueous two-phase partitioning study. *Journal of Chromatography B*. 2004;807(1):17-24. doi:10.1016/j.jchromb.2004.01.029

61. Hansen PC. Analysis of discrete ill-posed problems by means of the L-curve. *SIAM Review*. 1992;34(4):561-580. doi:10.1137/1034115

62. Dikic I. Proteasomal and autophagic degradation systems. *Annu Rev Biochem*. 2017;86:193-224. doi:10.1146/annurev-biochem-061516-044908

63. Gilgunn S, Bones J. Challenges to industrial mAb bioprocessing—removal of host cell proteins in CHO cell bioprocesses. *Curr Opin Chem Eng*. 2018;22:98–106. doi:10.1016/j.coche.2018.08.001

64. Oh YH, Mendola KM, Choe LH, et al. Identification and characterization of CHO host-cell proteins in monoclonal antibody bioprocessing. *In preparation*. 2023.

65. Falkenberg H, Waldera-Lupa DM, Vanderlaan M, et al. Mass spectrometric evaluation of upstream and downstream process influences on host cell protein patterns in biopharmaceutical products. *Biotechnol Prog*. 2019;35(3):1-12. doi:10.1002/btpr.2788

66. Deisenhofer J. Crystallographic refinement and atomic models of a human Fc fragment and its complex with fragment B of protein A from *Staphylococcus aureus* at 2.9- and 2.8-Å resolution. *Biochemistry*. 1981;20(9):2361-2370. doi:10.1021/bi00512a001

67. Ultsch M, Braisted A, Maun HR, Eigenbrot C. 3-2-1: Structural insights from stepwise shrinkage of a three-helix Fc-binding domain to a single helix. *Protein*

Engineering, Design and Selection. 2017;30(9):619-625. doi:10.1093/protein/gzx029

68. Gasteiger E, Hoogland C, Gattiker A, et al. Protein identification and analysis tools on the ExPASy server. In: Walker JM, ed. *The Proteomics Protocols Handbook*. Humana Press; 2005:571-608. doi:10.1385/1592598900

69. Pabst TM, Thai J, Hunter AK. Evaluation of recent Protein A stationary phase innovations for capture of biotherapeutics. *J Chromatogr A*. 2018;1554:45-60. doi:10.1016/j.chroma.2018.03.060