Comprehensive and automated profiling of host cell proteins using the Proteograph™ XT workflow

Shao-Yung (Eric) Chen, Max Mahoney, Aaron S. Gajadhar

Introduction

The presence of host cell proteins (HCPs) in biopharmaceuticals is a critical concern due to their potential immunogenicity and potential to diminish therapeutic efficacy, which pose detrimental risks to patients. To address this, liquid chromatography–mass spectrometry analysis (LC–MS) has emerged as a novel and versatile analytical technique for comprehensive HCP analysis.

While this approach offers exceptional sensitivity, specificity, and comprehensive accuracy in identification and quantification of HCPs, conventional LC–MS is still challenged by the inherently large dynamic range of HCPs compared to biologics. Regardless, LC–MS has the potential to detect a wide range of HCPs, delivers precise and quantitative results, and resolve the identity of individual proteins within a sample. Notably, LC–MS obviates the reliance on specific antibodies, a limitation inherent in conventional ELISA methods.

Monoclonal antibody–based detection generally requires tedious development time and resources, and while polyclonal antibody–based detection offers a broader coverage, it often lacks specificity. Nonetheless, traditional LC–MS–based HCP analysis encounters challenges stemming from the intricacy and heterogeneity of the sample matrix, as well as the prerequisite for advanced instrumentation and expertise in data analysis.

In this study we investigated the performance of the Proteograph™ Product Suite, an automated and standardized nanoparticle–based sample preparation and data analysis workflow, integrated with state-of-the-art microflow liquid chromatography and mass spectrometry technology. In this application note we subject the Proteograph XT workflow to rigorous evaluation of HCP identification depth, qualitative and quantitative reproducibility of the platform, and measurement reliability toward a list of common problematic HCPs. We performed the evaluation using a widely accepted community benchmark standard, the NIST monoclonal antibody (NIST Reference Material® 8671) (NIST 8671).

Study design

Sample preparation

Sample preparation using the Proteograph XT is shown in Figure 1 with the following steps. (1) Upon addition of biologic samples to Seer's NP suspensions, a stable and reproducible protein corona is formed based on the particle physicochemical properties. Protein corona–containing NPs are pulled down and washed, taking advantage of the paramagnetic core. (2) Proteins are then denatured, reduced, alkylated, and digested directly on the particles using an automated one-pot sample preparation workflow, resulting in tryptic peptides released into the supernatant. The resulting peptide mixture is then desalted using solid phase extraction

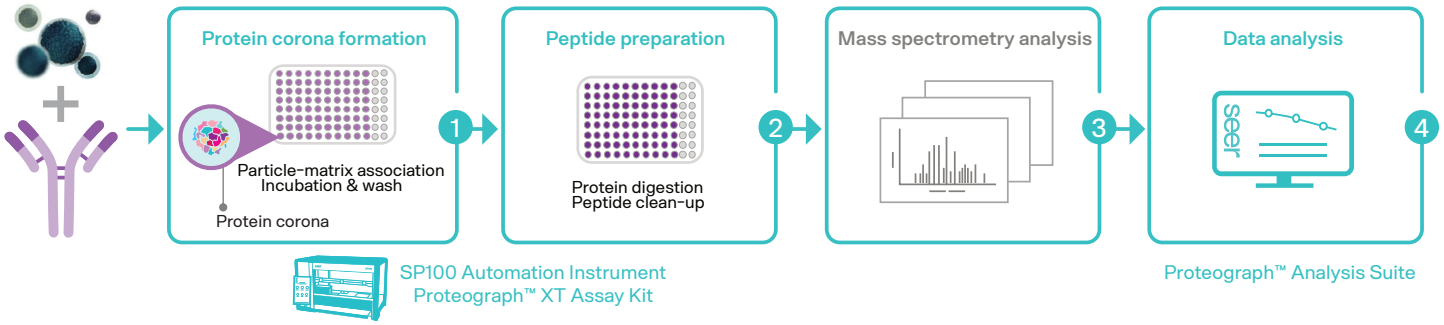

Figure 1. The Proteograph XT Workflow. **(1)** Samples are added to Seer's NP suspensions and a protein corona is formed. **(2)** Proteins are processed then the resulting peptide mixture is desalted and quantified. **(3)** Peptides are injected onto an LC–MS system. **(4)** LC–MS raw data are transferred to the Proteograph™ Analysis Suite ("PAS") for peptide and protein identification.

on the SP100 Automation Instrument. Peptides are then quantified using a fluorescence spectrometer, dried, and resuspended on the SP100 Automation Instrument before injection onto a (3) LC-MS system. (4) LC-MS raw data were transferred directly to the Proteograph™ Analysis Suite (PAS) for peptide and protein identification, quantification, and other biological insights.

LC-MS data acquisition

The performance evaluation of the Proteograph XT workflow was conducted using NIST 8671 samples obtained from Sigma Aldrich. We investigated the performance of the Proteograph XT workflow with LC-MS/MS. Two distinct acquisition strategies were employed: a 60-minute sample-to-sample Data-Independent Acquisition (DIA) and a 120-minute sample-to-sample Data-Dependent Acquisition (DDA) approach. Samples were directly processed using the Proteograph Product Suite (Seer, Inc.), while a conventional denaturing digest (neat digest) sample approach was employed as a control in parallel. Tryptic peptides derived from the samples were subjected to DIA/DDA LC-MS analysis using a Thermo Fisher Scientific Orbitrap™ Exploris™ 480 mass spectrometer, and subsequent data processing was performed using PAS.

LC-MS acquisition and data analysis

For both DDA and DIA approaches, 240 ng of peptides were reconstituted in 4 µL with a solution of 0.1% formic acid (FA) and 3% acetonitrile (ACN) spiked with 5 fmol/µL PepCalMix from SCIEX for constant mass MS injection between replicates and denaturing digests. Raw DIA MS data was processed using the DIA-NN search engine (version 1.8.1) in library-free mode searching MS/MS spectra against an in silico generated spectral library of mouse protein entries (UP000000589_10090). Raw DDA MS data was processed using MSFragger 3.5 in PAS. Both DDA and DIA searches included a Match-Between-Run identification approach for Proteograph XT and denaturing digests. All identifications were filtered to peptide and protein group FDR of 0.01

Results

Proteograph XT enhances HCP detection in NIST 8671 by 4-6X

Proteograph XT results were obtained using both DDA and DIA approaches, each performed in triplicate. Comparative analysis revealed that the Proteograph XT DDA results exhibited a 5.5-fold increase in protein group identifications compared to traditional denaturing digest (211 versus 38, Figure 2A). Similarly, the Proteograph XT DIA results demonstrated a 4-fold enhancement in protein group identifications compared to traditional denaturing digest (862 versus 215, Figure 2B). This expanded list of proteins suggests an additional depth of impurities is being missed without the Proteograph XT workflow.

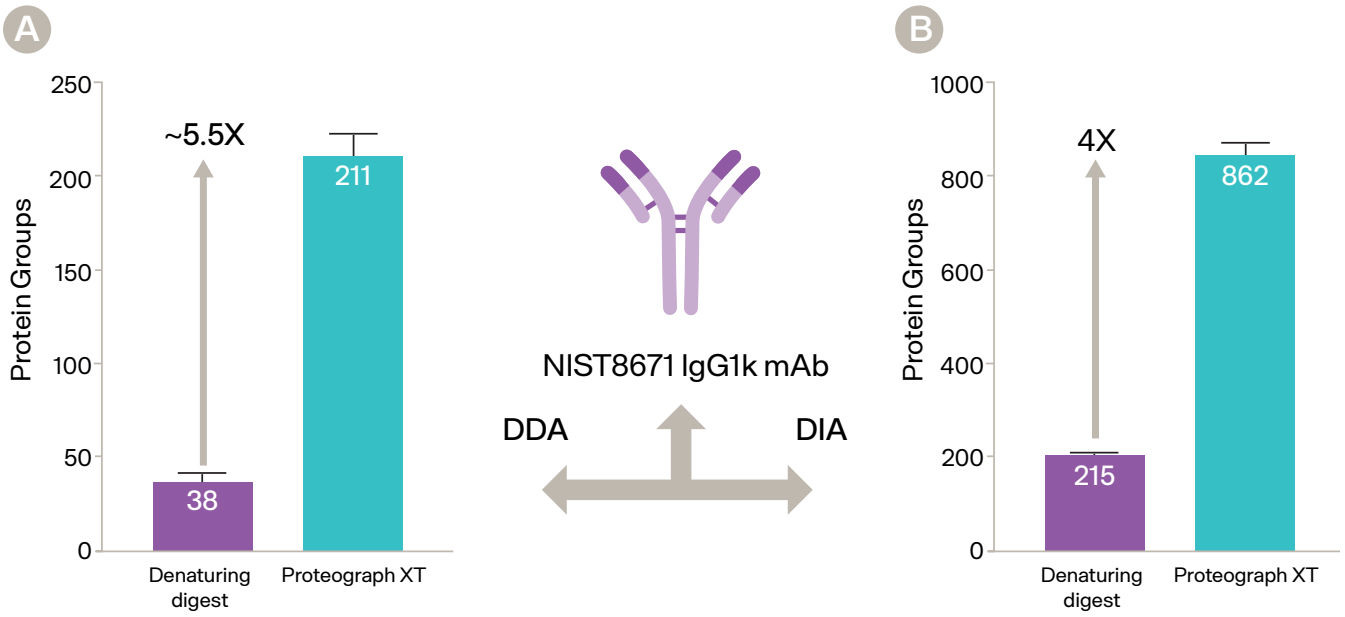

Figure 2. Protein group identifications for NIST 8671 samples for **(A)** DDA and **(B)** DIA using the Proteograph XT standard protocol compared to a conventional denaturing digestion workflow, with error bars denoting standard deviation (n= 3).

Proteograph XT label-free quantitation is highly reproducible across replicates

Pearson correlation analysis was conducted using label-free quantitative data obtained from Proteograph XT and denaturing digest approaches, employing both DDA and DIA methods. The analysis aimed to assess the reproducibility between any two replicates within each approach. The results revealed high levels of reproducibility across all combinations, with Pearson correlation coefficients for Proteograph exceeding 0.94 ($r > 0.94$). Notably, despite Proteograph XT's capability to detect a greater number of low-abundance HCPs in both DDA (as shown in Figure 3A compared to Figure 3B) and DIA (as shown in Figure 3C compared to Figure 3D), the reproducibility remained comparable to that of the denaturing digest method.

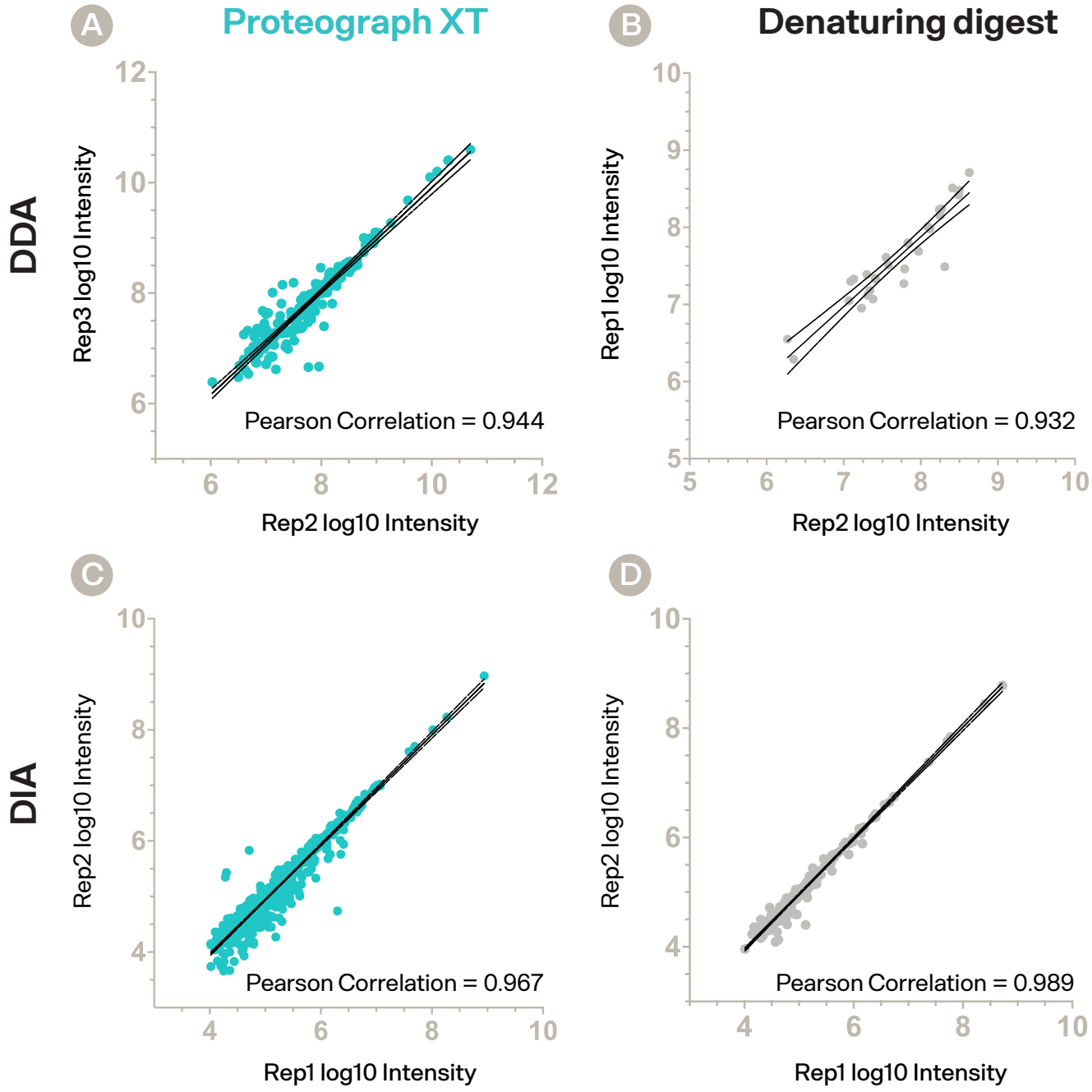

Figure 3. Pearson correlation analysis for protein group intensity between replicates of NIST 8671 for **(A)** Proteograph XT DDA workflow, $r = 0.944$, **(B)** Denaturing digest DDA workflow, $r = 0.932$, **(C)** Proteograph XT DIA workflow, $r = 0.967$, **(D)** Denaturing digest DIA workflow, $r = 0.989$.

Highly reproducible Proteograph XT workflow allows detection of hundreds of additional HCPs compared to conventional sample preparation

Protein group overlap analysis (Jaccard Index) was conducted to assess the overlap of identifications by Proteograph across all three replicates and provide a measure of the run-to-run consistency. The analysis revealed a substantial overlap of over 74% among the protein groups identified using the DDA approach across the three replicates (Figure 4A). Similarly, DIA approach demonstrated an overlap of over 93% across the three replicates (as shown in Figure 4B). Furthermore, a comparison was made between workflows to understand the overlap between Proteograph XT and denaturing digest identifications. The results highlighted that Proteograph XT not only detected nearly all the identifications made by denaturing digest, but also provided a significant number of additional identifications beyond denaturing digest. This finding emphasizes the superior HCP detection performance of Proteograph XT and the ability to characterize HCP samples more comprehensively.

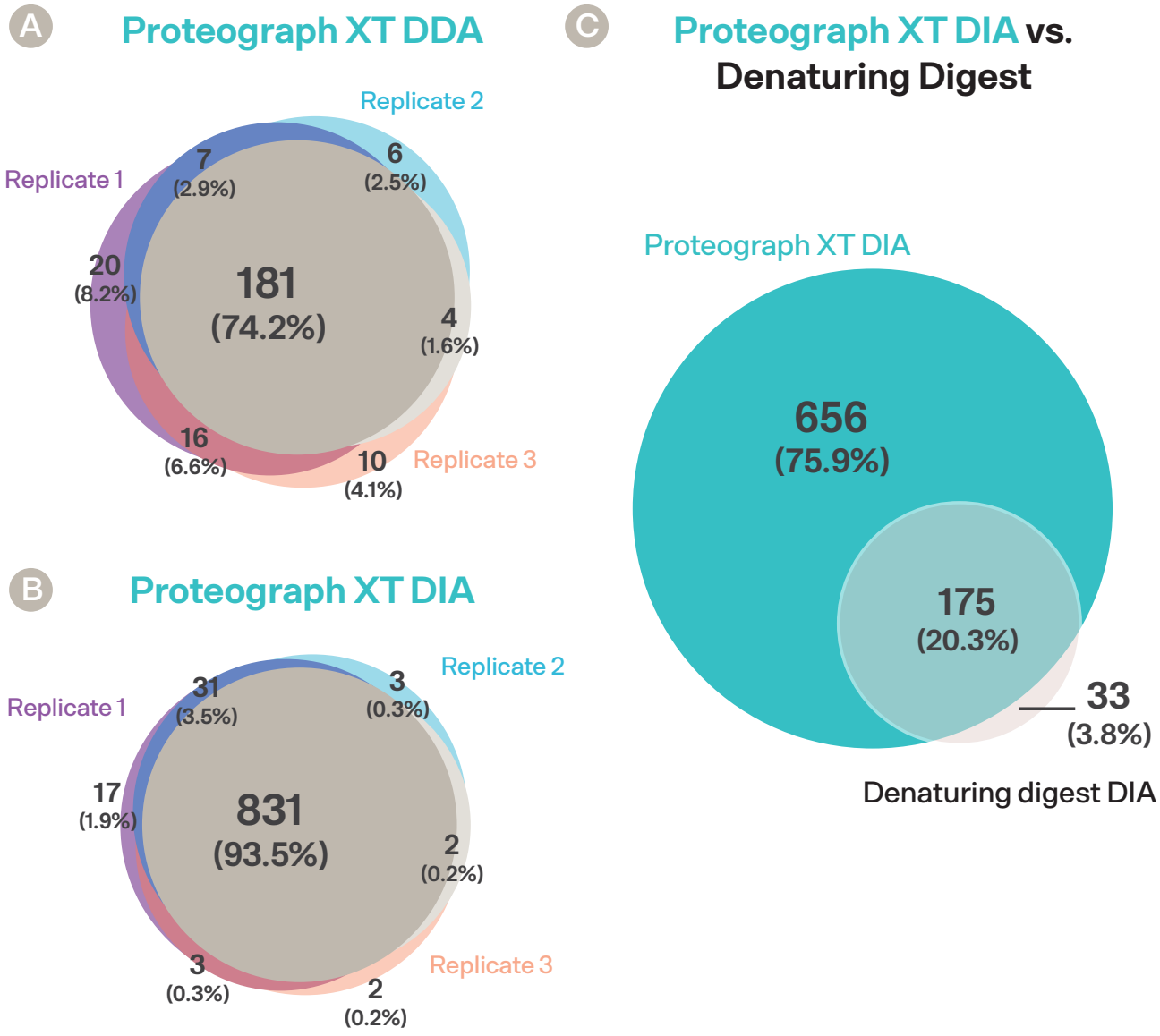

Figure 4. Venn diagram showing process replicate reproducibility for NIST 8671 samples run with **(A)** Proteograph XT DDA workflow or **(B)** Proteograph XT DIA workflow. **(C)** Venn diagram demonstrating overall identification overlap between Proteograph XT/denaturing digestion DIA workflow.

Proteograph XT workflow detects more HCPs than conventional workflows while maintaining comparable precision

To assess the capability of the Proteograph platform's measurement precision in detecting low abundant or potentially problematic HCPs, we referred to a peer-reviewed publication documenting a comprehensive list of 101 HCPs detected and reported through various LC-MS/MS analyses and publications.[1] The Proteograph XT demonstrated robustness in detecting 95 out of the 101 reported HCPs consistently (i.e., 95 HCPs detected across the triplicate injections), without missing values (Figure 6). This surpassed the denaturing digest approach which only detected 33 of 101 HCPs. To highlight the classes of common HCPs detected, we focused on the top 18 proteins from the aforementioned list and analyzed the corresponding coefficient of variation (% CV) for the quantitation values in both Proteograph XT DIA and denaturing digest DIA (Table 1). Importantly, the results show Proteograph XT precision that is equivalent, if not superior to, conventional denaturing digest in terms of quantitative precision for some of these high-value HCPs.

Protein Name	Accession	% CV		Referenced PPM[1]
		Proteograph	Denaturing Digest	
Fructose-bisphosphate aldolase A	P05064	8.4	2.3	932
Protein disulfide-isomerase A6	Q922R8	6.2	34.5	652
Fructose-bisphosphate aldolase C	P05063	6.2	2.7	249
Beta-2-microglobulin	P01887	9.4	2.7	226
Glucose-6-phosphate isomerase	P06745	5.4	3.5	200
Heterogeneous nuclear ribonucleoproteins A2/B1	O88569	15.3	6.5	66
Protein ABHD11	Q8K4F5	6.5	3.5	66
NSFL1 cofactor p47	Q9CZ44	4.6	4.6	58
Low affinity immunoglobulin gamma Fc region receptor II	P08101	5.8	6	54
Granulins	P28798	15.7	Not Detected	46
Prostaglandin reductase 1	Q91YR9	8.9	6.2	42
Peptidyl-prolyl cis-trans isomerase FKBP2	P45878	10.4	12.6	41
Serine/arginine-rich splicing factor 7	Q8BL97	18.3	8.4	37
Syntaxin-12	Q9ER00	4.6	9.4	33
THO complex subunit 4	O08583	14.2	Not Detected	29
Semaphorin-4B	Q62179	8.5	17	28
Heterogeneous nuclear ribonucleoprotein A1	P49312	12.7	48.5	25
Ubiquitin-conjugating enzyme E2 variant 2	Q9D2M8	7.3	Not Detected	24

Table 1. List of common HCPs in NIST 8671 material, quantitative coefficient of variation (%CV), and reference mole ppm (micromole HCP/mole drug substance) for Proteograph XT and denaturing digestion workflows (n=3). Reference mole ppm was taken from experimental data published by Molden et al.[1] The reported ppm are estimations of expected impurity levels of HCPs present in NIST 8671 based on computed values in NIST 8671 data from other studies.[1]

Moreover, by referencing recent publications from other experts in the field[1] we were able to demonstrate the estimated quantitation range of HCPs identified with Proteograph XT workflow. Although Table 1 only demonstrates the top 18 proteins from the commonly detected HCP in NIST 8671, ranging from sub-1000 ppm to ~20-ppm, the overall identified HCPs from the Proteograph XT platform managed to detect lower abundant HCPs, reported to be present in the sub-ppm range (Figure 6, on page 7).

Furthermore, a comprehensive evaluation of quantitation %CV for all proteins detected was conducted for both Proteograph XT and denaturing digest across both DDA and DIA methods. The analysis revealed that the median %CV for Proteograph XT was comparable to that of traditional denaturing digest (Figure 5). Notably, Proteograph XT provided a substantially higher number of HCP identifications (3.8- & 7.7-fold) across all replicates for DDA and DIA, respectively, with no missing values.

Conclusions

The Proteograph XT Assay, used with the off-the-shelf protocol, proves to be compatible with HCP analysis of biopharmaceuticals, offering several advantages over traditional denaturing digestion methods. Notably, it yields 4-6 times more HCP identifications for NIST 8671 without the need for additional sample pre-processing steps.

This workflow demonstrates consistent reproducibility, effectively capturing the vast majority of HCPs detected through denaturing digest approaches while also detecting hundreds of additional HCPs that would have been overlooked. Importantly, the Proteograph XT Assay enhances the sensitivity of HCP detection without compromising assay precision, highlighting its efficacy as an advanced analytical tool. With its ability to achieve deep HCP coverage, high assay reproducibility, and the integration of automated sample preparation and data analysis, the Proteograph Product Suite presents a comprehensive and reliable solution for HCP analysis in the field of biopharmaceutical development.

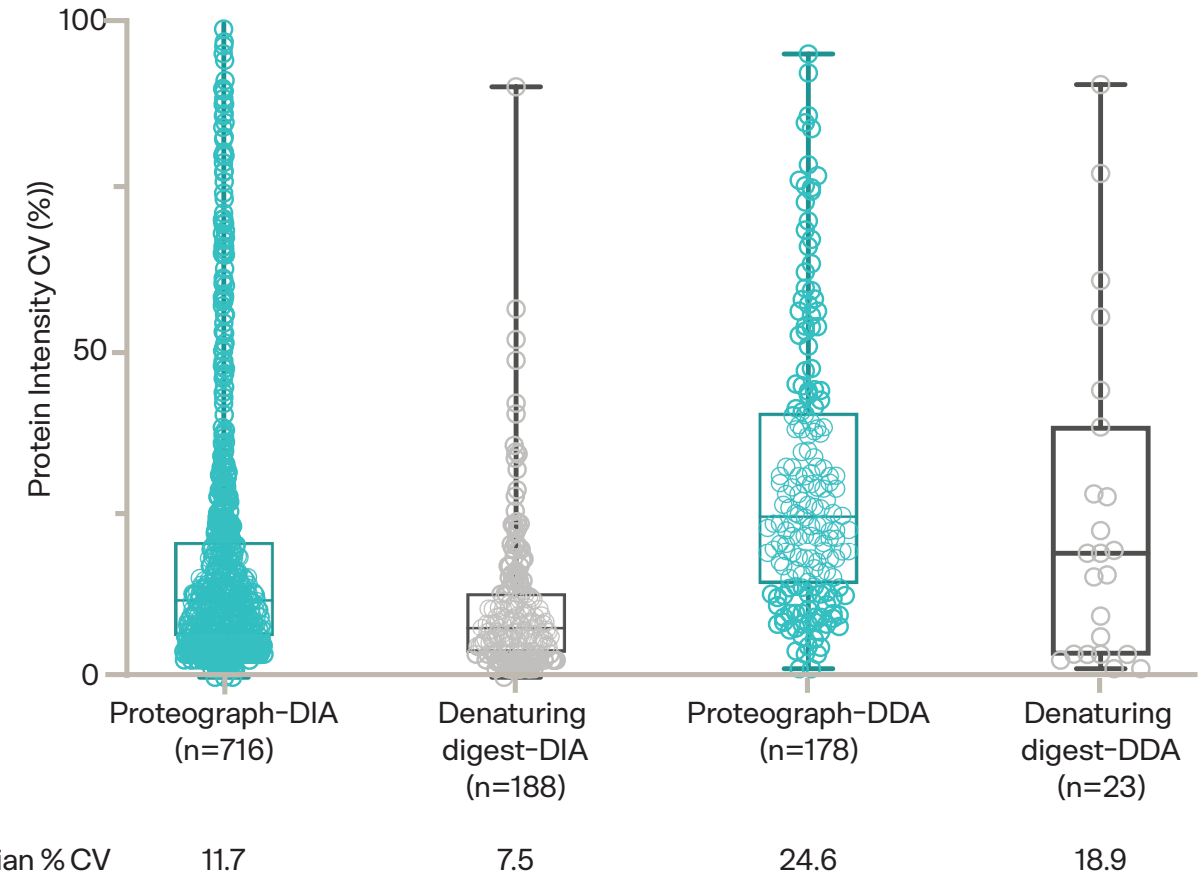

Figure 5. Protein intensity %CV for Proteograph XT and denaturing digestion workflows with DDA or DIA for proteins measured in across all experiment replicates (n=3).

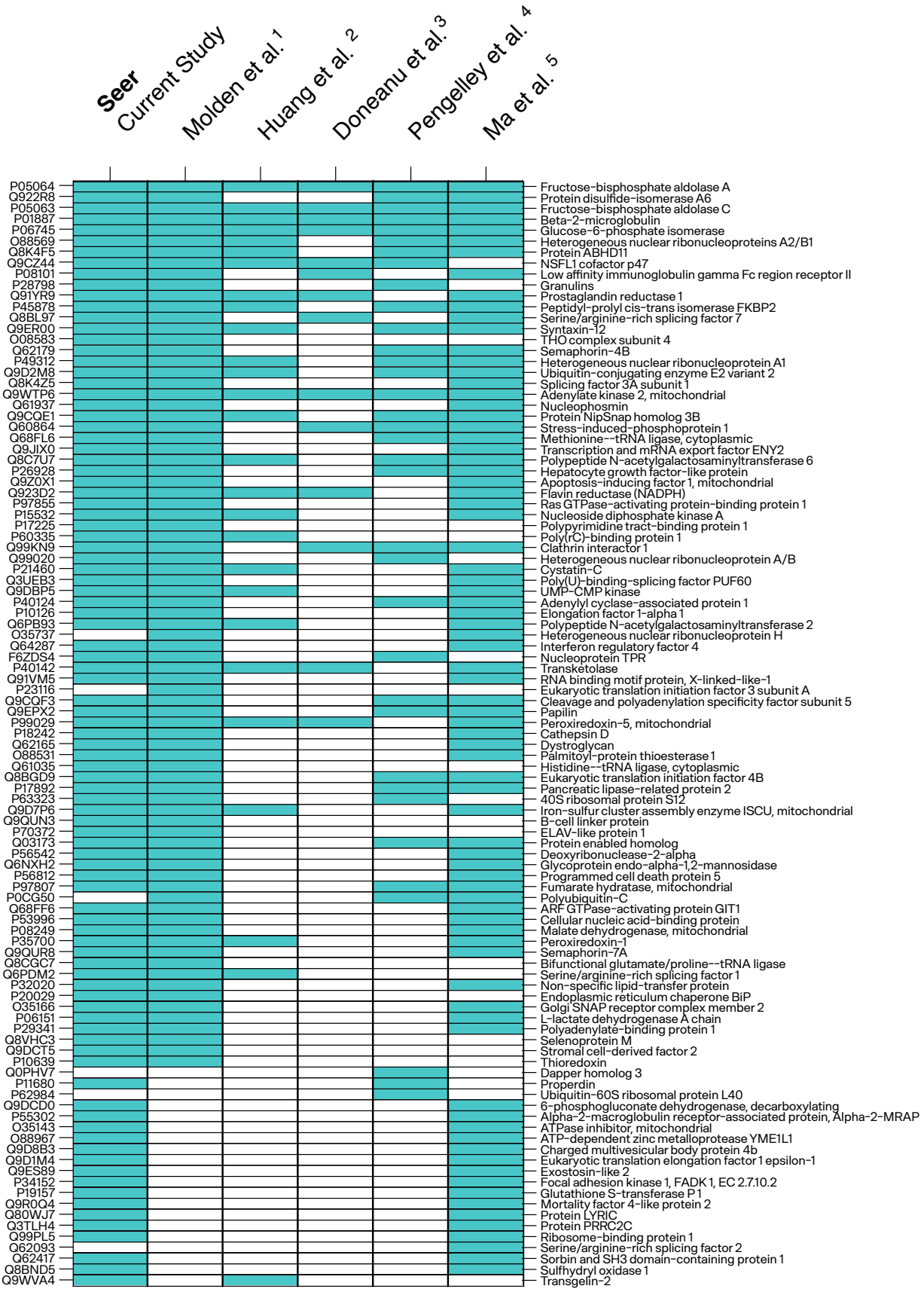

Figure 6. List of 101 commonly detected HCP and reported across current study and reference literature [1-5] through various LC-MS/MS analysis. Teal bars denote detected proteins in specific studies, white bars denote not detected in specific literatures/study.

References

1. Molden, R, et al. Host cell protein profiling of commercial therapeutic protein drugs as a benchmark for monoclonal antibody-based therapeutic protein development. *mAbs* 13:1 (2021).
2. Huang et al. A Novel Sample Preparation for Shotgun Proteomics Characterization of HCPs in Antibodies. *Analytical Chemistry*; 89(10) (2017)
3. Doneanu et al. Enhanced Detection of Low-Abundance Host Cell Protein Impurities in High-Purity Monoclonal Antibodies Down to 1 ppm Using Ion Mobility Mass Spectrometry Coupled with Multidimensional Liquid Chromatography. *Analytical Chemistry*; 87(20) (2015)
4. Pengelley et al. In Depth Analysis of Host Cell Proteins from Antibody Preparations using PASEF. *Bruker Application Note*; LCMS–143 (2018)
5. Ma et al. Sensitive, Rapid, Robust, and Reproducible Workflow for Host Cell Protein Profiling in Biopharmaceutical Process Development. *J Proteome Res*; 19(8) (2020)

Learn more at
seer.bio/product/proteograph-product-suite