Improving the Dynamic Range of Host Cell Proteins Analysis Using an HRAM Orbitrap Mass Spectrometer

Stephane Houel, Romain Huguet, Susan Abbatiello, David Sarracino, Jonathan Josephs, Thermo Fisher Scientific, Cambridge, MA, USA, 02139

INTRODUCTION

Host cell proteins (HCPs) are process-related impurities that can affect the safety and efficacy of biotherapeutics. Therefore, the clearance of HCPs to a safe level is required by regulatory agencies. HCP analysis is challenging because up to six orders of magnitude of dynamic range are required to be able to detect low ppm concentrations of residual HCPs in biotherapeutics. One strategy is to load a large amount of digested biotherapeutic protein on column to push HCPs above the limit of detection, but low level HCPs are still close to the noise level. In this study, we tested different parameters such as amount of digested sample on-column, length of the LC gradient, peak capacity and the use of a FAIMS device to evaluate their effect on the dynamic range of host cell protein identification.

MATERIALS AND METHODS

Sample Preparation

NIST mAb was digested overnight with trypsin in non denaturing condition. Undigested antibody was then precipitated by heat (90 °C for 10 min). This novel sample preparation was published by Huang et al., 2017 and allows the digestion of host cell proteins while the antibody is minimally digested (see Huang et al for details).

Liquid Chromatography

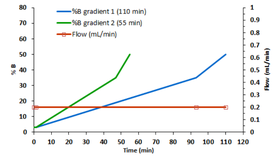
LC: Thermo Scientific™ Vanquish™ UHPLC system

Columns: Except for one experiment, three Thermo Scientific™ Acclaim™ columns (C18, 250 mm * 2.1 mm; 2.2 µm) were connected together. For one experiment, only one column was used.

Mobile phase:

Buffer A: 0.1% Formic acid, in H_2O

Buffer B: 0.1% Formic acid, in CH_3CN



Mass Spectrometry

Mass spectrometer: Thermo Scientific™ Orbitrap Fusion™ Lumos™ Tribrid™ mass spectrometer

Data were acquired in data-dependent acquisition (DDA) mode with HCD fragmentation and an isolation window of 1.6 Da, A FAIMS prototype interface was also tested.

TopS (2s) DDA method	
MS1 OT res: 120K at 200 m/z	MS2 OT res: 30K at 200 m/z
AGC MS1: 8E5	AGC MS2: 5E4
MS1 Max Inject time: 500 ms	MS2 Max Inject time: 500 ms
m/z: 300-1500	Min Intensity threshold 1E4

Data Analysis

Spectral .raw files from data-dependent acquisition were analyzed using Thermo Scientific™ Proteome Discoverer™ 2.2 software with SEQUEST® HT search. Data was searched against CHO database at a 1% spectrum level FDR. MS1 and MS2 mass tolerances were respectively 10 ppm and 0.02 Da. Deamidation of asparagine (+0.984) and/or oxidation of methionine (+15.995) were considered as variable modifications.

Figure 1. Data were collected using a Vanquish UHPLC system (a) coupled to a Orbitrap Fusion Lumos mass spectrometer (b) and processed with Proteome Discoverer software (c). The 3 columns (d) fit in the Vanquish UHPLC column compartment.

RESULTS

1) Different loading of digested NIST mAb on-column

Table 1. Number of identified host cell proteins and peptides at different calculated loading on-column.

	calculated loading on-column (ug)			
	125	60	30	5
Host Cell Proteins (HCPs)	54	54	40	13
Unique Peptides	218	203	140	50

Figure 2. Venn diagram for the identified host cell proteins between 125 µg, 60 µg, and 30 µg calculated loading on-column

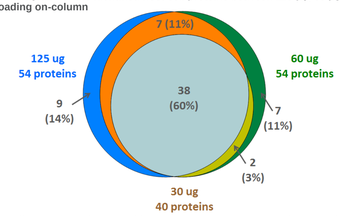


Figure 3. Venn diagram for unique peptides between 125 µg, 60 µg, and 30 µg calculated loading on-column

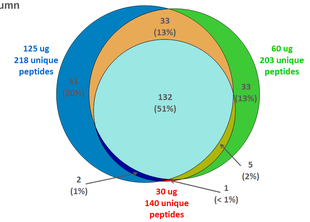
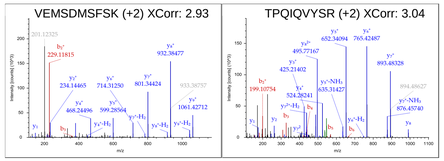


Figure 4. Unique peptides of beta-2 microglobulin (P01887)



The sample is digested in non-denaturing conditions so the NIST mAb will only be partially digested. The calculated loading on-column corresponds to the amount that will be injected if one hundred percent of the NIST mAb was digested. Two injections for each condition were acquired using the 110 min LC gradient with the three columns set-up. Each condition was searched individually and only proteins with at least two unique peptides were counted. Fifty four proteins were identified for the 125 µg and 60 µg calculated loading on column (Table 1).

The number of unique peptides were respectively 218 and 203 for the 125 µg and 60 µg calculated loading on-column. No unique HCP to the 30 µg calculated loading on column was observed while 7 and 9 HCPs were respectively unique to 60 µg and 125 µg calculated loading on column. The same trend is observed at the peptide level. Even if significantly fewer proteins and peptides were identified at 30 µg than at 125 µg or 60 µg calculated loading on column, low level HCPs were still identified. For example, HCPs such as serine/arginine-rich splicing factor 7 (Q8BL7), ARF GTPase activating protein ((Q68FF6), or bifunctional glutamate/proline (Q8CGC7) were identified with at least two unique peptides at 30 µg and they were reported previously (Huang et al., 2017) at less than 3 ppm.

In this study, beta2-microglobulin was identified with two unique peptides (Figure 4) from 30 µg to 125 µg calculated loading on column. The fact that beta2-microglobulin was not identified previously with two unique peptides is probably not related to the novel sample preparation.

2) Two different LC gradient lengths

Table 2. Number of MS1 and MS2 for the 55 min and 110 min LC gradient

	55 min	110 min
MS1	2350	5834
MS2	8270	19640

Table 3. Number of identified host cell proteins and peptides for the two different LC gradients

	55 min	110 min
Host Cell Protein (HCP)	39	54
Unique Peptides	144	218
peptide/HCP	3.7	4.0

Figure 5. Venn diagram for the identified host cell proteins between the 55 min and 110 min LC gradient

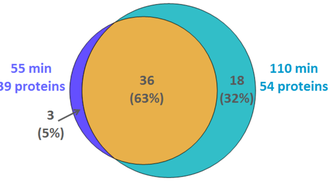
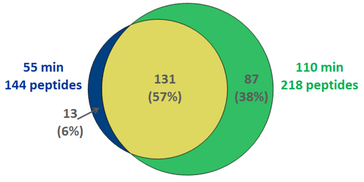


Figure 6. Venn diagram for unique peptides between the 55 min and 110 min LC gradient



To test the impact of a shorter LC gradient on HCP identification, 125 µg of calculated loading on-column was used with the three columns set-up. Two raw files were acquired for each gradient. One of the first impact of a shorter LC gradient is less MS1 and MS2 scans (Table2). Even with the novel sample preparation when most of the mAb is not digested, more HCPs and related peptides are identified with a longer LC gradient. With the 55 min gradient, around 30% fewer HCPs and related peptides were not identified (Table 3). For both gradients, the maximum fill time of 500 ms was reached for 80% of the identified peptide spectrum matches (PSM) linked to HCPs. Peptides from HCPs are often just above the noise and the complexity at low levels is still significant. Therefore a longer gradient will generate a higher number of identified HCPs.

Even for the 55 min LC gradient, three HCPs were uniquely identified (Figure 5). These three HCPs were also identified in the 60 µg calculated loading on-column experiment. Even if fewer HCP were identified with the 55 min LC gradient, nevertheless low level HCPs were identified. For example, HCPs such as flavin reductase (Q923D2) that was previously identified at 3 ppm or peroxiredoxin-5 (P99029) and dystroglycan (Q62165) that were identified at 1 ppm.

3) Using 1 column versus 3 columns for a 110 min LC gradient.

Table 4. Number of MS1 and MS2 for one or 3 columns connected together

	Number of columns	
	1	3
MS1	5562	5834
MS2	16067	19640

Table 5. Number of identified host cell proteins and peptides for the two different column settings

	1 column	3 columns
Host cell Protein (HCP)	30	54
Unique Peptides	115	218
peptide/HCP	3.8	4.0

Figure 7. XIC for the six most intense peptides of the NIST mAb for the three columns (a) and one column (b) settings

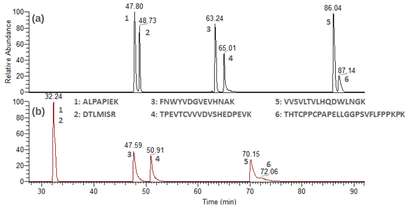


Figure 8. Venn diagram for the identified host cell proteins between the one or the three column settings

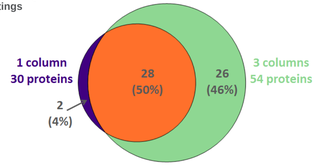
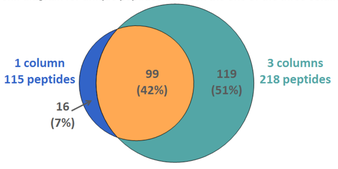


Figure 9. Venn diagram for unique peptides between the one or the three column settings



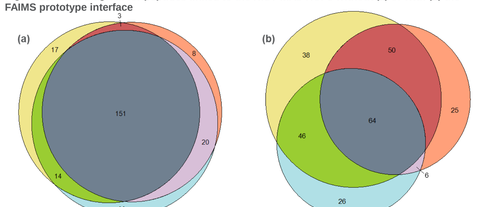
To test the effect of column length, 125 µg of calculated loading on-column was used with the 110 min LC gradient. For each condition two raw files were acquired. The number of MS1 and MS2 scans is slightly higher for the three column setting (Table 4), but the number of identified HCPs is drastically lower for the one column setting. For the three column setting the peak width at the base for the most abundant peptides is around 40 s, whereas for the same peptides with the one column setting the peak width at base is around 2 min (Figure 7). The peak tailing or the overloading of the column has a strong negative effect on the number of identified HCPs.

3) Combining data acquisition without or with a FAIMS prototype interface

Table 6. Number of identified host cell proteins and peptides for without or with a FAIMS prototype interface

	no FAIMS	FAIMS
Host cell Protein (HCP)	56	62
Unique Peptides	224	255

Figure 10. Venn diagram for peptides linked to the NIST mAb HCPs without (a) or with (b) the FAIMS prototype interface



To test the erect of FAIMS on HCP identification, the three column setting,110 min LC gradient and 125 µg calculated loading on-column were used. In one case, three raw files were acquired without FAIMS. In the other case, two raw files were acquired with the FAIMS prototype interface at two different compensation voltages and combined with one raw file acquired without FAIMS. Sixty-two HCPs were identified when FAIMS was used. Without using FAIMS, approximately the same peptides are identified in the three runs (Figure 10b). With FAIMS, different populations of peptides were triggered and the number of peptides and HCPs increased.

CONCLUSIONS

- With the novel sample preparation, the optimal sample loading for the NIST mAb with the three column setting is between 60 µg and 125 µg on-column.
- Using only one column has a bigger effect on HCP identification than reducing the LC gradient length.
- A FAIMS interface can be a powerful tool for HCP identification.

For Research Use Only. Not for use in diagnostic procedures.

REFERENCE

1. A Novel Sample Preparation for Shotgun Proteomics Characterization of HCPs in Antibodies. Lihua Huang, Ning Wang, Charles E. Mitchell, Tammy Brownlee, Steven R. Maple, and Michael R. De Felippis, Analytical Chemistry 2017 89 (10), 5436-5444.

TRADEMARKS/LICENSING

PO65278-EN 0518S