Identification and Absolute Quantification of Individual Host Cell Proteins by SWATH® LC-MS

Rikke Raaen Lund, Solveig Beck Nielsen, Janne Crawford, Jakob Bunkenborg, Anette Rasmussen, Ejvind Mørtz, and Thomas Kofoed
Alphalyse, www.alphalyse.com

Introduction

Host cell proteins (HCPs) are the most critical process-related impurity in recombinant biopharmaceuticals. ELISA is widely used for measuring HCP levels in batch-consistency and release testing of drug substance batches. ELISA gives total amount of HCP in "immuno-equivalent ng/ml" depending on animal immune-response, and no information about the identity of individual HCPs.

Therefore, FDA and EMA are increasingly aware of mass spectrometry (MS) as an orthogonal method to ELISA. We here present a sensitive mass spectrometry method based on SWATH® LC-MS for identification and absolute quantification of HCPs down to low ppm concentrations to be used for process-development samples and batch-consistency of drug substance batches.

What is SWATH® LC-MS?

SWATH® is data independent acquisition (DIA) mass spectrometry where the Sciex® mass spectrometer divides the mass range into small mass windows and performs MS/MS analysis on all peptides in each window. With dynamic size mass windows, small windows are used in areas with many peptides and larger windows in areas with few peptides.

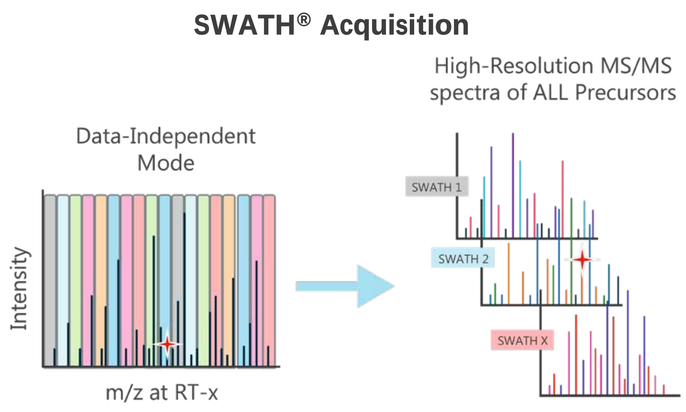

Figure 1: The principle behind the SWATH® LC-MS technology.

Key advantages of SWATH® for HCP analysis:

- Low interference from the high amount of the drug substance protein, because the data on low level HCP peptides are acquired in data-independent mode.
- Highly reproducible HCP identification and quantification.
- All recorded SWATH® data can be re-analyzed – when the ion library is updated with more known compounds and their fragmentation pattern.

The Ion Library

Interpretation of a complex SWATH® mass spectrum of multiple peptides requires comparison to an ion library with MS/MS data of known peptides. The ion library contains information about each peptide; HPLC retention time, MS/MS data and amino acid sequence. The ion library can be build using data/information dependent acquisition (DDA or IDA) analysis of the same samples. It is of key importance that the database used for protein identification analysis contains all the potential proteins in the samples to be analyzed. This includes:

- The complete proteome of the host cell.
- The sequence of the drug substance, including variants.
- Internal standard proteins.
- Possible contaminating protein like keratin, trypsin, protein A etc.

The Method

The analysis consists of a three step procedure:

1. Sample preparation – in Eppendorf® epMotion pipetting robot

- Add internal standard proteins (n=7).
- Precipitate all proteins.
- Digest all proteins into peptides.

2. Mass spec analysis – in Sciex® TripleTOF 6600

- Microflow LC – 80 min gradient, 5 µl/min, CSH column.
- Triplicate sample analysis in data dependent mode .
- Triplicate sample analysis in SWATH® mode.

3. Data analysis and reporting

- Identify and quantify HCP using MS/MS data – see infobox.

HCPs in Process Samples Quantified by SWATH® LC-MS

Host Cell Protein	Step 1	Step 2	Step 3	Step 4	Step 5	Step 6	Mass	pI	Protein name
sp|P0C058|IBPB_ECOLI	4,274	2,905	2,154	186	229	111	16,093	5.2	Small heat shock protein IbpB
sp|P0A9A9|FUR_ECOLI	158	284	296	142	147	94	16,795	5.7	Ferric uptake regulation protein
sp|P0ABK5|CYSK_ECOLI	597	913	711	618	200	68	34,490	5.8	Cysteine synthase A
sp|P69783|PTGA_ECOLI	33	250	378	256	185	62	18,251	4.7	PTS system glucose-specific EIIA component
sp|P0A8J4|YBED_ECOLI	432	215	253	222	112	21	9,827	5.5	UPF0250 protein YbeD
sp|P02930|TOLC_ECOLI	41	283	187	417	57	11	53,741	5.2	Outer membrane protein TolC
sp|P62623|ISPH_ECOLI	312	1,146	855	231	62	15	34,775	5.2	4-hydroxy-3-methylbut-2-enyl diphosphate reductase
sp|P0ADP9|YIHD_ECOLI	33	28	32	11	10	18	10,275	5.1	Protein YihD
sp|P0A763|NDK_ECOLI	106	240	100	349	113		15,463	5.6	Nucleoside diphosphate kinase
sp|P35340|AHPF_ECOLI	67	291	171	174	48		56,177	5.5	Alkyl hydroperoxide reductase subunit F
sp|P08200|IDH_ECOLI	390	271	166	355	42		45,757	5.2	Isocitrate dehydrogenase
sp|P69797|PTNAB_ECOLI	284	339	240	25	26		35,048	5.7	PTS system mannose-specific EIIAB component
sp|P0A717|RIBA_ECOLI	345	297	106	48	45		21,836	5.6	GTP cyclohydrolase-2
sp|P0AEN1|FRE_ECOLI	741	870	849	1,404	33		26,242	5.3	NAD(P)H-flavin reductase
sp|P36683|ACNB_ECOLI	129	87	82	9	8		93,498	5.2	Aconitate hydratase B
sp|P0ADE8|YGFZ_ECOLI	18	113	45	150	26		36,094	5.2	tRNA-modifying protein YgfZ
sp|P0AB91|AROG_ECOLI	56	231	188	42	22		38,010	6.1	Phospho-2-dehydro-3-deoxyheptonate aldolase
sp|P0A825|GLYA_ECOLI	128	332	204	263	13		45,317	6.0	Serine hydroxymethyltransferase
sp|P0A6K3|DEF_ECOLI	129	136	55	61	26		19,328	5.2	Peptide deformylase
sp|P0ABP8|DEOD_ECOLI	33	113	78	19	20		25,950	5.4	Purine nucleoside phosphorylase DeoD-type
Number of HCPs	562	245	206	67	25	8			
Total HCP ppm (w/w)	193,169	48,548	33,391	9,599	1,493	401			
HCP % (w/w)	**19.32%**	**4.85%**	**3.34%**	**0.96%**	**0.15%**	**0.04%**			

Table 1: List of HCPs identified in six process samples from different purification steps. Each individual HCP and total HCP are quantified in ppm.

Method Characteristics

Seven internal standard proteins are added at either 100 ppm or 2,000 ppm. The intact proteins are added before precipitation and digestion, and are therefore "true" internal standards that undergo exactly the same treatment as the HCPs. The signal of the standards are used for a calibration curve to quantify the HCPs. Dilution series of all standard proteins show that the quantification is linear within the measurement range from about 10 ppm - 100,000 ppm.

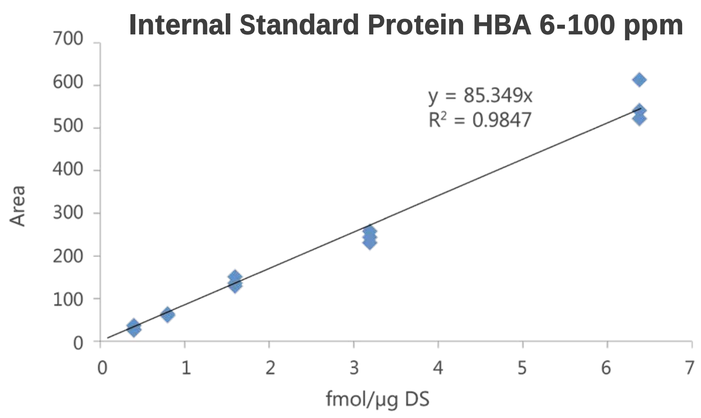

Figure 2: The linearity of one internal standard protein from 6 ppm to 100 ppm.

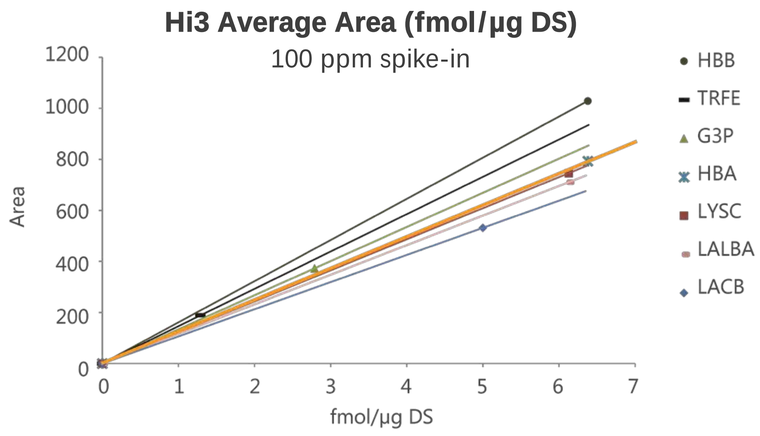

Figure 3: The slopes obtained for the seven standard proteins using a spike-in level of 100 ppm following the Hi3 quantification method.

Looking at the linearity curves of the standard proteins it is clear that the Hi3 measurement method does not give exactly the same slopes. The intensities of the three highest scoring peptides depend on the specific protein sequence. By using the median curve (orange line in figure 3) the HCPs with an unknown response curve can be quantified.

The label-free quantification method provides absolute quantification within the spread of the response curve. An *E. coli* cell lysate was analyzed by the method and 1064 proteins were identified and quantified. By summing the calculated ppm amount for each protein the total was very close to 1,000,000 ppm (1,004,620).

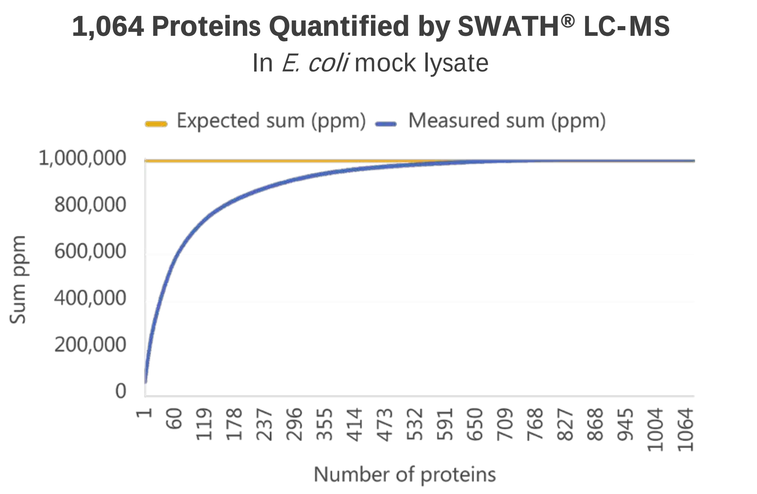

Figure 4: Quantification of E. coli cell lysate using the Hi3 method in combination with the median response curve of seven internal standard proteins.

Conclusion

The presented HCP analysis method based on SWATH® mass spectrometry is:

- Sensitive – down to low ppm.
- Robust – no carry over problem using microflow HPLC.
- Reproducible – using SWATH® data independent acquisition method.
- Absolute quantification – using internal standard proteins and Hi3.

The method is applicable for:

- Preclinical & clinical batches.
- Bioprocess optimization.
- Evaluation of ELISA assays.
- Biosimilar versus originator.
- Characterization of specific HCPs of concern.
- All types of expression systems – *E. coli*, *CHO*, yeast etc.

The method is highly flexible and can be used to optimize processes for clearance of individual HCPs. The analysis can be established for a new biologic within weeks and samples analyzed on an ongoing basis.

HCP Data for Process Samples

The SWATH® method was applied to process samples in a six-step purification process of a small protein drug substance (16 kDa). The number of identified HCPs decreased from 562 to eight HCPs in the final drug substance, and the total HCP amount from 19% down to 0.04% (see table 1).

The MW and pI for each individual HCPs are plotted as a virtual 2D gel for each process step. Comparison with the full *E. coli* proteome shows that the method covers the full MW and pI range.

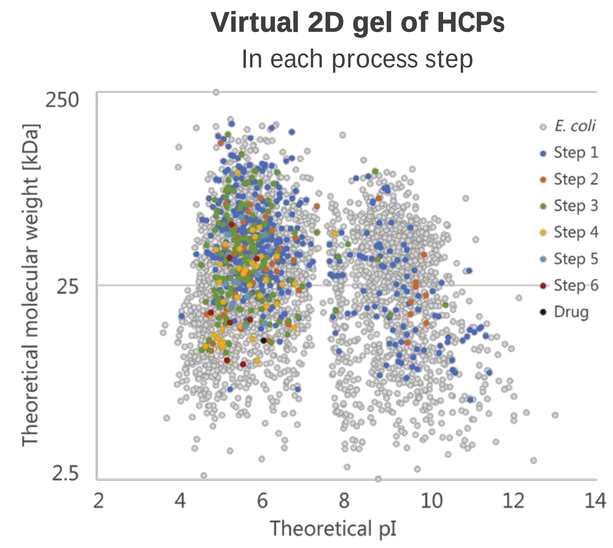

Figure 5: Virtual 2D gel of MW and pI of each identified HCP in six different process samples compared to the proteome of E. coli.

Infobox

What is label free protein quantification using Hi3?

Hi3 (High-3) quantitation takes advantage of the observation that MS/MS signals compared to their corresponding MS precursors are less likely to lead to signal saturation. Peptides are assigned to proteins via the process of generating the ion library.

Hi3 is performed for each identified protein by summing the product ion (MS/MS) intensity of all associated peptides. The top three most abundant peptides are selected and the MS/MS intensities of these three peptides (MS/MS pep) are summed to give each protein a product ion intensity value used as a quantity measurement.

Silva *et al.*, 2006: "Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition." Mol Cell Proteomics 5 (1):144-56.