Small HCPs in a 12 kDa Protein Drug Analyzed by GeLC-MS/MS

Rikke Raaen Lund[1], Marie Grimstrup[1], Maheen Sayeed[2], Karin Abarca Heidemann[2], Ejvind Mørtz[1]
[1]Alphalyse A/S, Odense, Denmark, [2]Rockland Immunochemicals Inc., Limerick, PA, USA

ALPHALYSE™ protein analysis services

ROCKLAND™ antibodies & assays

Introduction

The content of low molecular weight host cell proteins (HCPs) in purified protein drugs is often difficult to evaluate, due to their low immunogenicity and poor ability to be visualized in gel-based total protein stains. The proteome of commonly used expression organisms, such as *E. coli* and Chinese Hamster cells, contains 30-40% proteins with a molecular weight below 20kDa, and these are easily missed in both in gel separations, Western blots and ELISA quantitation of the total HCP-content. To provide unbiased analysis of small as well as larger HCPs, we introduce the use of a mass spectrometry-based orthogonal method, well known from proteomics, called GeLC-MS/MS.

Here, we analyze an in process protein drug, a 12 kDa protein produced in *E. coli*, as well as the corresponding null cell lysate, using 1D-PAGE and nano-flow LC-MS/MS (GeLC-MS/MS) to achieve high coverage of small HCPs.

Figure 1: GeLC-MS/MS of null cell lysate

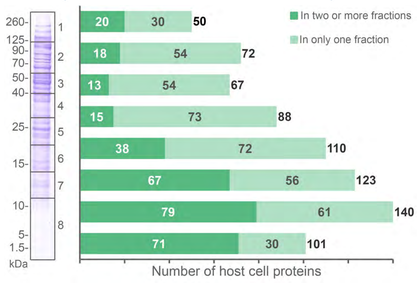
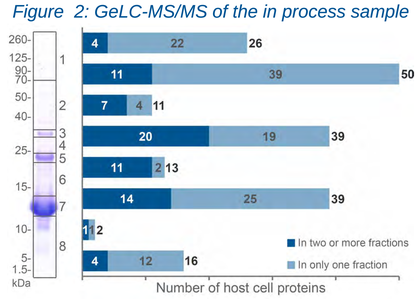


Figure 2: GeLC-MS/MS of the in process sample



1D-PAGE

Proteins were separated by 1D-PAGE and stained for isolation and visualization of the high concentration protein drug in separate gel fractions. Most proteins, 78% in the null cell lysate and 82% in the in process sample were only identified in a single gel fraction (light green and blue bars, fig. 1 and 2, respectively). Further, the fractionation leads to a high number of protein identifications in the fractions without the protein drug (fig 2).

Protein separation by 1D-PAGE prior to LC-MS/MS leads to identification of a high percentage of small HCPs: 46% of the HCPs identified in the null cell lysate and 37% in the in process sample were smaller than 20kDa (fig. 3, table 1 and 2). Further, the HCPs that were identified covered the entire pI range and 99% of the molecular weight range of the total *E. coli* proteome (fig. 3). Examples of small HCPs is given in table 3 and 4.

The sensitivity of the method was estimated by parallel analysis of the purified protein drug with two spiked-in standards, these standards were both identified at 50ppm.

Figure 3: Molecular weight and pI of the HCPs identified by GeLC-MS/MS and the E.coli proteome

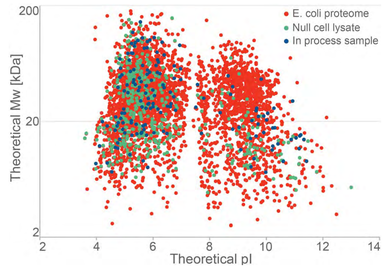


Table 1: Distribution of high and low molecular weight HCPs in the null cell lysate

Protein size	HCPs identified
HMW (≥20kDa)	297
LMW (<20kDa)	256
Total	553

Table 2: Distribution of high and low molecular weight HCPs in the in process sample

Protein size	HCPs identified
HMW (≥20kDa)	96
LMW (<20kDa)	56
Total	152

Nano Flow LC-MS/MS

The proteins contained in the gel fractions were digested enzymatically into peptides using trypsin. Separation of the complex peptide mixture was achieved by nanolitre flow HPLC using a CSH (charged surface hybrid, Waters) column. This material has a high loading capacity and excellent peak shape in formic acid mobile phases. This enables highly sensitivity and accurate MS detection of low level HCPs using a qTOF mass spectrometer (Bruker Corp. fig. 4 and 5).

GeLC-MS/MS advantages

- 1D-PAGE requires very little sample preparation and has a wide mass and pI range, leading to unbiased sample analysis
- Protein digest combined with nano flow LC-MS/MS provides sensitive and accurate measurement of multiple peptides from each protein as well as sensitive and accurate measurement of peptide fragment masses

Figure 4: MS/MS spectrum of the a peptide from Ferric Uptake Regulator (in process sample)

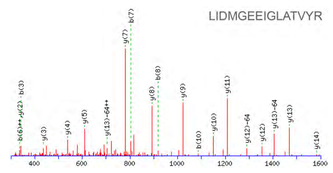
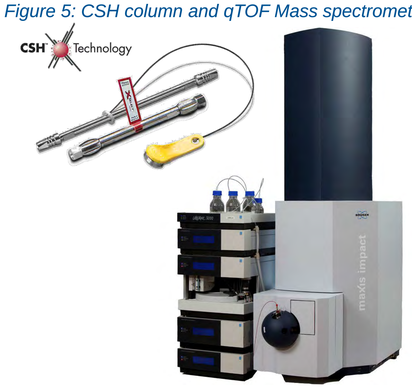


Figure 5: CSH column and qTOF Mass spectrometer



HCP Identification Criteria

HCPs are identified by comparing the mass data to the theoretical masses of the host cell proteome. Peptides are assigned a score according to how well they match and these are summarized to a protein score. Since small proteins have fewer peptides they also have lower proteins scores even at the same molar level as larger proteins.

We have used conservative inclusion criteria that accounts for this: 1% false discovery rate, a minimum score of 25 for all peptides, and a two peptide minimum for proteins larger than 20kDa. These criteria allows highly confident HCP identification with a low false positive rate.

Table 3: Top 30 small proteins in the null cell lysate

HCP no	Accession no	Protein Name	Mass	pI	Score
22	tr|C6ECE6	Ribosomal protein L9	15759	6.17	3350
29	tr|C6EGC3	Ribosomal protein S7	17593	10.3	2701
31	tr|C6EL69	6,7-dimethyl-8-ribityllumazine synthase	16147	5.15	2656
42	tr|C6EL60	PTS system, glucose subfamily, IIA subunit	18240	4.73	2336
46	tr|C6EFQ8	Redoxin domain protein			
48	tr|C6EE34	50S ribosomal protein L10			
50	tr|C6ECC2	Inorganic diphosphatase			
55	tr|C6EGN0	DNA-binding protein			
58	tr|C6EE18	Histone family protein			
59	tr|C6EGG1	50S ribosomal protein L6			
60	tr|C6EE33	Ferric uptake regulator, Fur family			
61	tr|C6EGG3	Peptide deformylase			
62	tr|C6EGE5	Methionine-R-sulfoxide reductase			
70	tr|C6EGF1	50S ribosomal protein L10			
72	tr|C6EG32	Ribosomal protein S7			
73	tr|C6EG67	Glutaredoxin			
75	tr|C6ECE9	Ribosomal protein L29			
78	tr|C6ECU9	30S ribosomal protein S5			
79	tr|C6EIW8	50S ribosomal protein L11			
82	tr|C6EAL7	DNA protection during starvation protein			
83	tr|C6EIX0	Regulator of sigma D			
86	tr|C6EDI0	Molybdopterin synthase sulfur carrier subunit			
88	tr|C6EAL6	Ferritin			
96	tr|C6EAC8	Iron-sulfur cluster assembly scaffold protein IscU			
99	tr|C6EA12	50S ribosomal protein L31			
105	tr|C6ECY3	30S ribosomal protein S6			
107	tr|C6EJY6	ATP synthase F1, delta subunit			
108	tr|C6ECD7	Uncharacterized protein GN=ECBD_4172			
110	tr|C6EG68	Flavodoxin			
112	tr|C6EGF4	50S ribosomal protein L18			

Table 4: Top 30 small HCPs in the in process sample

HCP no	Accession no	Protein Name	Mass	pI	Score
3	tr|C6EJQ1	Ferric uptake regulator, Fur family	17012	5.68	3455
6	tr|C6EG80	Peptide deformylase	19430	5.23	2765
14	tr|C6EC51	Methionine-R-sulfoxide reductase	15783	5.58	1881
20	tr|C6EE34	50S ribosomal protein L10	17757	9.04	1367
22	tr|C6EGC3	Ribosomal protein S7	17593	10.3	1082
31	tr|C6ECU9	Glutaredoxin	13042	4.75	906
40	tr|C6EGF4	Ribosomal protein L29	7269	9.98	625
46	tr|C6EGG3	30S ribosomal protein S5	17534	10.23	591
63	tr|C6EE36	50S ribosomal protein L11	14923	9.64	399
65	tr|C6EIX0	DNA protection during starvation protein	18684	5.72	390
70	tr|C6EE23	Regulator of sigma D	18288	5.65	371
71	tr|C6EIZ8	Molybdopterin synthase sulfur carrier subunit	8734	4.38	370
73	tr|C6EBF4	Ferritin	19468	4.77	332
75	tr|C6EKH8	Iron-sulfur cluster assembly scaffold protein IscU	14011	4.82	301
80	tr|C6EE73	50S ribosomal protein L31	8094	9.46	271
81	tr|C6ECE9	30S ribosomal protein S6	15163	5.25	262
83	tr|C6EG68	ATP synthase F1, delta subunit	19434	4.94	254
88	tr|C6EF45	Uncharacterized protein GN=ECBD_4172	10323	5.18	232
90	tr|C6EFQ0	Flavodoxin	19896	4.21	224
91	tr|C6EGG2	Ribosomal protein L18	12762	10.41	222
95	tr|C6EGN0	DNA-binding protein	15587	5.43	202
96	tr|C6EGG1	50S ribosomal protein L6	18949	9.71	201
97	tr|C6EK87	UspA domain protein	15925	6.03	200
102	tr|C6EHC0	Acyl carrier protein	8634	3.98	190
103	tr|C6EAC8	Iron-sulfur cluster assembly accessory protein	12264	4.11	189
105	tr|C6EGF1	50S ribosomal protein L22	12219	10.23	185
111	tr|C6EGG9	Ribosomal protein S11	13950	11.33	158
112	tr|C6EL60	PTS system, glucose subfamily, IIA subunit	18240	4.73	153
113	tr|C6EGK5	Acetyl-CoA carboxylase, biotin carboxyl carrier protein	16733	4.66	152
115	tr|C6EKO4	Thioredoxin	15887	5	147

Conclusion

The HCPs that were identified covered 99% of the entire *E. coli* proteome in terms of molecular weight and pI. This shows that the developed GeLC-MS/MS method has no inherent limitations with respect to pI or molecular weight for HCP identification. The obtained protein identity enables an in-depth analysis of each individual HCP and a more detailed risk assessment.

HCPs Characteristics and Individual Risk Assessment

Identification of the protein names and database accession numbers enables an in-depth analysis of each individual HCP present in the drug sample. Important features for a drug risk assessment include:

- Immunological properties and presence of human B- and T-cell epitopes
- Homology to human proteins with important biological function
- Homology to the drug protein
- Enzymatic activity to modify or cleave the drug product constituents
- Hormone or hormone-like activity

Known information can be found at http://www.uniprot.org/ including:

- Molecular function, biological process, ligand binding and cellular component
- Post translational modifications and processing
- Expression, interaction and structure
- Protein family and domains
- Sequence variations
- Publications

Prediction of protein features can be investigated at http://www.cbs.dtu.dk/services, including (fig. 6):

- Immunological features
- Post translational modifications
- Protein structure and function

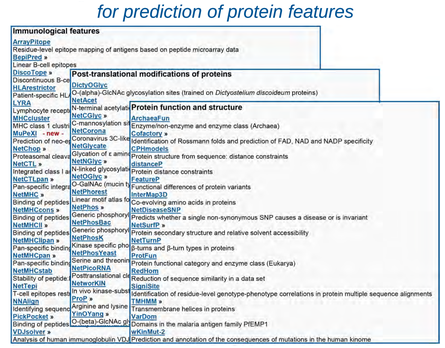



Figure 6: Examples of services available for prediction of protein features