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

by Rikke Raaen Lund, Marie Grimstrup, Ejvind Mørtz

Email correspondence: info@alphalyse.com

The content of low molecular weight host cell proteins (HCPs) in purified protein drugs is often difficult to evaluate by ELISA and 2D-PAGE, due to their low immunogenicity and poor ability to be visualized in gel-based total protein stains. The proteomes of commonly used expression organisms, such as E. coli and Chinese Hamster cells, contain 30-40% proteins with molecular weights below 20 kDa, and these are easily missed in both 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.

METHODS

Proteins from the in-process protein drug as well as the corresponding null cell lysate were separated by 1D-PAGE and stained for visualization of the high concentration protein drug and for isolation in separate gel fractions. 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. figures 1 and 2).

HCPs are identified by comparing the mass data to the theoretical masses of the host cell proteome by a Mascot database search. Peptides are assigned a score according to how well they match and these are summarized into a protein score.

Figure 1: MS/MS spectrum of a peptide from Ferric Uptake Regulator (in-process sample)

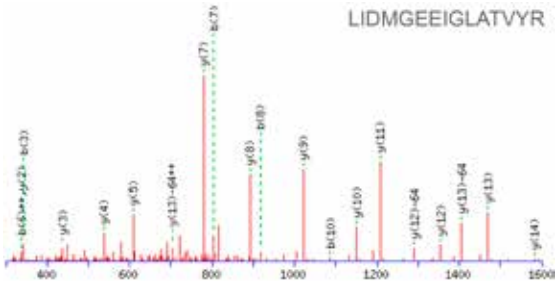


Figure 2: CSH column and qTOF mass spectrometer

Since small proteins have fewer peptides they also have lower protein scores even at the same molar level as larger proteins. We have used conservative inclusion criteria that account for this: 1% false discovery rate, a minimum score of 25 for all peptides, and a two peptide minimum for proteins larger than 20 kDa. These criteria allow highly confident HCP identification with a low false positive rate.

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

ADVANTAGES OF GeLC-MS/MS

- 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

RESULTS

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 light blue bars, figures 3 and 4, respectively). Further, the fractionation lead to a high number of protein identifications in the fractions without the protein drug (figure 4).

Protein separation by 1D-PAGE prior to LC-MS/MS lead 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 20 kDa (figure 5 and 6, tables 1 and 2 - page 3). Further, the HCPs that were identified covered the entire pI range and 99% of the molecular weight range of the total *E. coli* proteome (figure 7). Examples of all HCPs as well as small HCPs are given in tables 3-6 (page 4).

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 (box, figure 8 - page 5).

Figure 3: GeLC-MS/MS of null cell lysate
No. of proteins by gel fraction

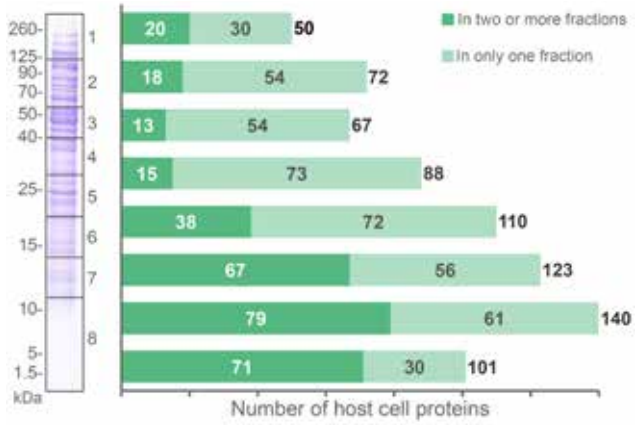


Figure 4: GeLC-MS/MS of the in-process sample
No. of proteins by gel fraction

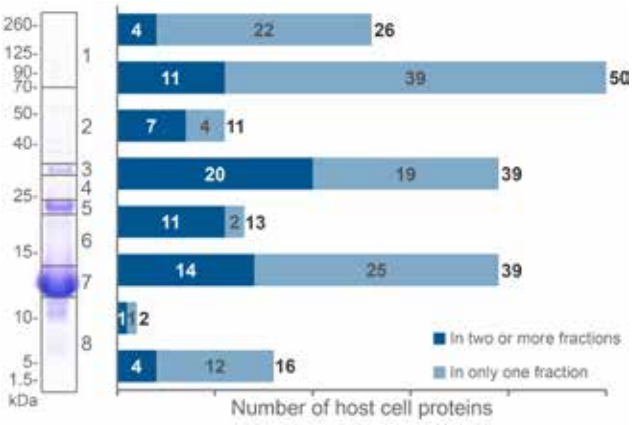


Figure 5: GeLC-MS/MS of the null cell lysate
No. of proteins by theoretical mass

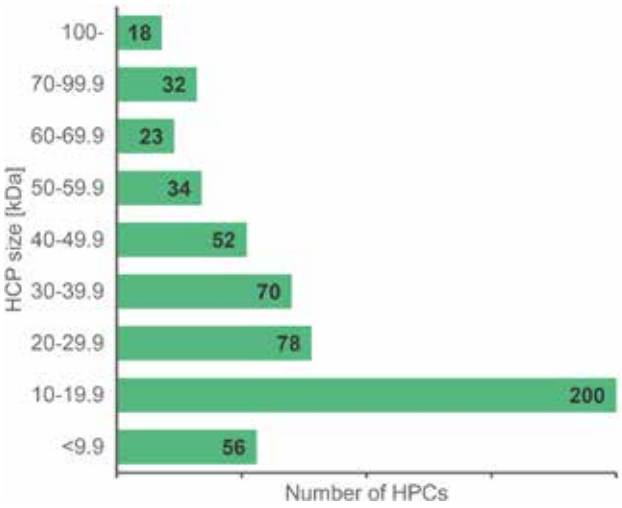


Figure 6: GeLC-MS/MS of the in-process sample
No. of proteins by theoretical mass

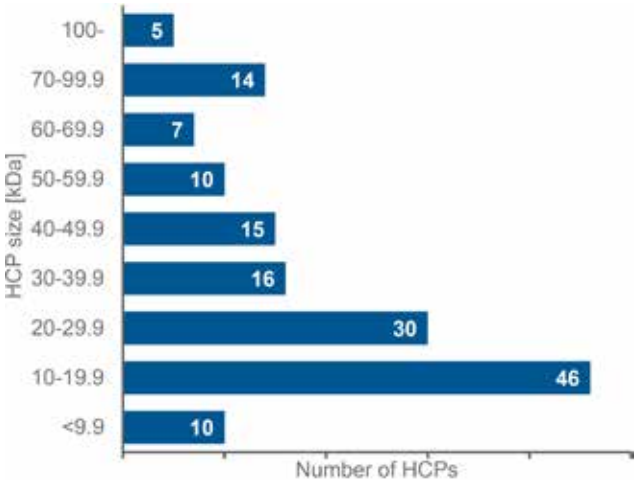


Table 1: Distribution of high and low molecular weight HCPs in the null cell lysate

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

Figure 7: Molecular weight and pI of the HCPs identified by GeLC-MS/MS and the *E. coli* proteome

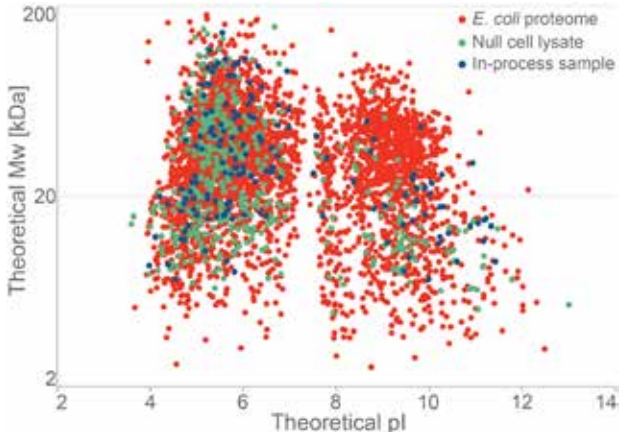


Table 3: Top 30 HCPs identified by GeLC-MS/MS in the null cell lysate

HCP no	Accession no	Protein Name	Mass	pI	Score
1	tr|C6EE31	DNA-directed RNA polymerase, beta subunit	155918	6.67	8620
2	tr|C6EGM8	Aldehyde-alcohol dehydrogenase	96580	6.32	9750
3	tr|C6EAU3	Pyruvate dehydrogenase E1 component	99978	5.46	8381
4	tr|C6EE32	DNA-directed RNA polymerase, beta subunit	150937	5.15	5614
5	tr|C6EFF7	Alpha-1,4 glucan phosphorylase	90865		
6	tr|C6EI53	Porin Gram-negative type	39309		
7	tr|C6EJT4	Alanine--tRNA ligase	96314		
8	tr|C6EAT9	Aconitate hydratase B	94009		
9	tr|C6EJL2	2-oxoglutarate dehydrogenase, E1 subunit	105566		
10	tr|C6ECC93	Valine--tRNA ligase	108536		
11	tr|C6EIK5	Glycine dehydrogenase (decarboxylating)	105078		
12	tr|C6EJI5	DNA-directed RNA polymerase	99477		
13	tr|C6EH57	Translation initiation factor IF-2	97461		
14	tr|C6EK51	Leucyl-tRNA synthetase	97814		
15	tr|C6ELN6	Beta-galactosidase	117331		
16	tr|C6EB27	Isoleucine--tRNA ligase	105042		
17	tr|C6EB19	Carbamoyl-phosphate synthase (glu-hydrolyzing)	118594		
18	tr|C6EE53	Phosphoenolpyruvate carboxylase	99470		
19	tr|C5W720	NADH-quinone oxidoreductase	101078		
20	tr|C6EGF2	Ribosomal protein S3	25967		
21	tr|C6EE39	Elongation factor Tu	43457		
22	tr|C6EDY6	Maltoporin	49995		
23	tr|C6EGC4	Translation elongation factor G	77704		
24	tr|C6EJK5	Cytochrome bd ubiquinol oxidase subunit I	58338		
25	tr|C6EEC0	Glycine--tRNA ligase beta subunit	76936		
26	tr|C6EI79	Formate acetyltransferase	85588		
27	tr|C6EF39	DNA polymerase I	103168		
28	tr|C6EE05	Methionine synthase	136639		
29	tr|C6EJI7	Transcriptional regulator, LacI family	39049		
30	tr|C6ELF6	Phosphate acetyltransferase	77466		

Table 4: Top 30 HCPs identified by GeLC-MS/MS in the in-process sample

HCP no	Accession no	Protein Name	Mass	pI	Score
1	tr|C6EH79	Chaperone protein DnaK	69130	4.83	6210
2	tr|C6EJQ1	Uroporphyrin-III C/tetrapyrrole methyltransferase	31500	5.83	3484
3	tr|C6EHN7	Ferric uptake regulator, Fur family	17012	5.68	3455
4	tr|C6EFE8	Transcriptional regulator, LysR family	33266	6.05	3404
5	tr|C6EGI0	Glycerol-3-phosphate dehydrogenase	56886	6.97	3034
6	tr|C6EHJ3	Peptide deformylase	19430	5.23	2765
7	tr|C6EE39	Bifunctional protein PutA	144393	5.55	2696
8	tr|C6EAG0	Elongation factor Tu	43457	5.3	2522
9	tr|C6EKZ7	GTP cyclohydrolase 1	24929	6.79	2389
10	tr|C6EA14	Primosomal replication priB and priC	20534	10.01	2371
11	sp|P10145	Pseudouridine synthase	25963	5.75	2370
12	tr|C6EGF2	ATP synthase F1, alpha subunit	55416	5.8	2005
13	tr|C6EC51	Ribosomal protein S3	25967	10.27	1918
14	tr|C6EGQ9	Methionine-R-sulfoxide reductase	15783	5.58	1881
15	tr|C6EGE6	Ribose-phosphate pyrophosphokinase	36854	5.48	1842
16	tr|C6EGC4	50S ribosomal protein L3	22230	9.91	1701
17	tr|C6EJI7	Translation elongation factor G	77704	5.24	1555
18	sp|P62577	Transcriptional regulator, LacI family	39049	6.39	1407
19	tr|C6EE34	Chloramphenicol acetyltransferase	25931	5.91	1393
20	tr|C6EB63	50S ribosomal protein L10	17757	9.04	1367
21	tr|C6EGC3	Histidine biosynthesis bifunctional protein HisB	40591	5.76	1144
22	tr|C5W720	Ribosomal protein S7	17593	10.3	1082
23	tr|C6ELS2	NADH-quinone oxidoreductase	101078	5.89	1059
24	tr|C5W9B3	Uncharacterized protein GN=ECBD_3349	25539	4.96	1040
25	tr|C6EL29	Polyribonucleotide nucleotidyltransferase	77110	5.11	1038
26	tr|C6EIL4	Trigger factor	48163	4.83	1006
27	tr|C6EG71	Tyrosine recombinase XerD	34225	8.74	998
28	tr|C6EAA4	ATP synthase F1, beta subunit	50351	4.9	997
29	tr|C6EGB5	Acyl-ACP-UDP-N-acetylglucosamine O-acyltransferase	28348	6.63	982
30	tr|C6ECU9	Peptidyl-prolyl cis-trans isomerase	21182	4.86	967

Table 5: 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|C6ELG9	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	17995		
48	tr|C6EE34	50S ribosomal protein L10	17757		
50	tr|C6ECC2	Inorganic diphosphatase	19833		
55	tr|C6EGN0	DNA-binding protein	15587		
58	tr|C6EE18	Histone family protein DNA-binding protein	9529		
59	tr|C6EGG1	50S ribosomal protein L6	18949		
60	tr|C6EE33	Ribosomal protein L7/L12	12288		
61	tr|C6EGG3	30S ribosomal protein S5	17534		
62	tr|C6EGE5	Ribosomal protein S10	11728		
70	tr|C6EGF1	50S ribosomal protein L22	12219		
72	tr|C6EG32	Thioredoxin	11913		
73	tr|C6EG67	ATP synthase FO, B subunit	17310		
75	tr|C6ECE9	30S ribosomal protein S6	15163		
78	tr|C6ECU9	Glutaredoxin	13042		
79	tr|C6ERW8	Virulence-related outer membrane protein	18648		
82	tr|C6EAL7	Protein-N(Pi)-p-histidine-sugar phosphotransferase	10491		
83	tr|C6EIX0	DNA protection during starvation protein	18684		
86	tr|C6EDI0	Protein-export protein SecB	17494		
88	tr|C6EAL6	Putative PTS IIA-like nitrogen-regulatory protein PtsN	17110		
96	tr|C6EAC8	Iron-sulfur cluster assembly accessory protein	12264		
99	tr|C6EA12	50S ribosomal protein L25	10687		
105	tr|C6ECY3	10 kDa chaperonin	10381		
107	tr|C6EJY6	50S ribosomal protein L19	13135		
108	tr|C6ECD7	Cold-shock DNA-binding domain protein	7398		
110	tr|C6EGB8	ATP synthase F1, delta subunit	19434		
112	tr|C6EGF4	Ribosomal protein L29	7269		

Table 6: 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|C6EGI0	Peptide deformylase	19430	5.23	2765
14	tr|C6ECS1	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|C6EGG8	ATP synthase F1, delta subunit	19434	4.94	254
88	tr|C6EF45	Uncharacterized protein GN=ECBD_4172	10323	5.18	232
90	tr|C6EJQ0	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|C6EK04	Thioredoxin	15887	5	147

HCP 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

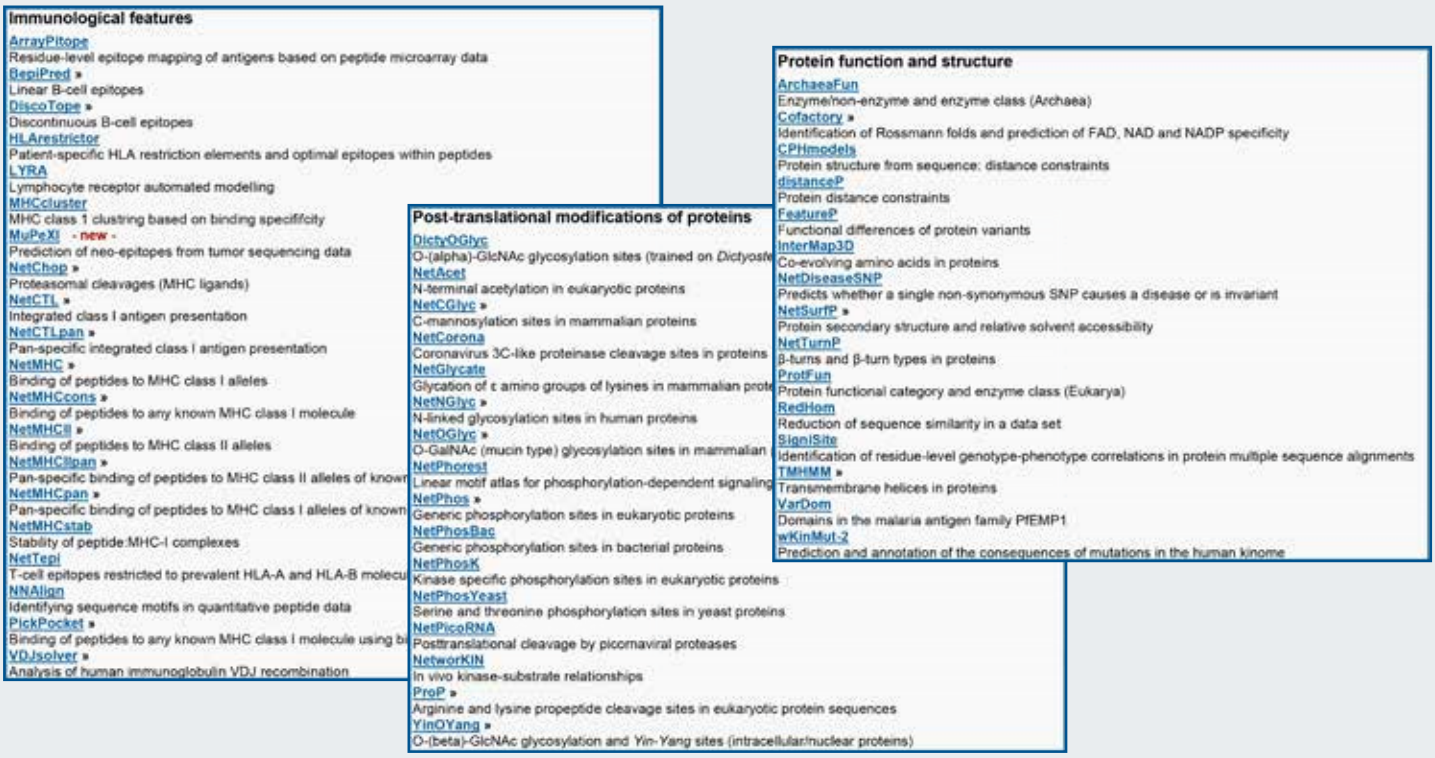
Known information can be found at www.uniprot.org, including:

- 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 www.cbs.dtu.dk/services, including (figure 8):

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

Figure 8: Examples of services available for prediction of protein features



ACKNOWLEDGEMENTS

We thank Maheen Sayeed, Karin Abarca Heidemann, David Chimento, and Carl Ascoli, Rockland Immunochemicals Inc., Limerick, PA, USA for providing the protein samples.