ASMS 2013

MP-512

Determination of Contaminant Bacterial Host Cell Proteins in Recombinant Proteins Expressed in *E. coli* by LC-QQQ

Oscar Potter; Gregory Staples; Yanan Yang; Hongfeng Yin; Kevin Killeen, Agilent Technologies, 5301 Stevens Creek Boulevard, Santa Clara, CA 95051

Introduction

The biopharmaceutical industry requires methods for the quantification of host cell proteins (HCPs) that contaminate purified recombinant therapeutic proteins in the parts per million (ppm) range. The industry currently depends on ELISA kits which quantify total HCP concentration with dubious accuracy and precision because their selectivity cannot readily be matched to the population of HCP contaminants in the purified product.

Protein identification and quantitation based on LC-MS analysis of protease digested samples have become well established techniques over the past two decades. Together, they hold great potential for HCP analysis because they can identify and specifically quantify numerous particular proteins without the need to generate antibodies.

This study demonstrates quantitative and qualitative analysis of HCPs in recombinant human growth hormone (rhGH) expressed in *E. coli*.

The key challenge in applying LC-MS to HCP analysis is in overcoming the interferences that arise from the very high concentration of therapeutic protein compared to HCPs. This situation presents unique challenges and opportunities compared to quantitative and qualitative proteomics of more heterogeneous samples.

This study also demonstrates a new approach to increasing the sensitivity of LC-MS HCP assays by the application of high resolution selective sample enrichment.

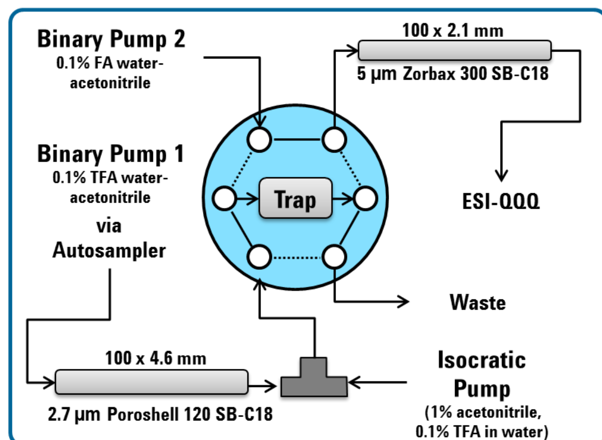
Experimental

Sample preparation

A sample of rhGH expressed in *E. coli* was obtained from MBL Internatiional (Woburn, MA). 600 µg rhGH was digested in 500 µl of solution prepared according the Rapigest protocol (Waters, Milford, MA) using a 1:30 ratio of Trypsin Gold (Promega, Fitchburg, WI) to sample.

Analytical LC-QQQ

An Agilent 1290 infinity pump system was operated with a 2.1 x 100 mm RRHD Zorbax 300 SB-C18 column. This system was plumbed to the source of an Agilent 6490A QQQ which was used in dynamic MRM mode for the 40 transition program. MRM mode was used for the high resolution selective enrichment method. This method used the configuration illustrated below:



Analytical LC-QTOF

An Agilent 6520 accurate mass QTOF was connected to an Agilent 1290 infinity HPLC pump system. 24 µg of rhGH were injected onto the 2.1 x 100 mm RRHD Zorbax 300 SB-C18 1.8µm column. The sample was eluted using a 0.1% formic acid in water-acetonitrile gradient over 82 minutes.

High resolution selective sample enrichment:

120 µg rhGH were injected onto a 4.6 x 100 mm Poroshell 120 2.7 µm SB-C18 column. The sample was eluted using a 0.1% trifluoroacetic acid in water-acetonitrile gradient. The column outlet was connected to a 1:3 flow splitter such that the smaller part of the effluent was directed to the ESI source of a QTOF instrument for monitoring whilst the larger part was directed via a switching valve into either a fraction collection vial or to waste. The effluent in the fraction collection vial was dried down to 20 µL.

HPLC-chip-MS/MS

10 µL of the 20 µL remaining in the fraction collection vial after drying down was injected using an HPLC-Chip-MS/MS system in conjunction with a 1200 series nano/cap HPLC system.

Data analysis

QTOF results analyzed using the feely available *Morpheus* software:

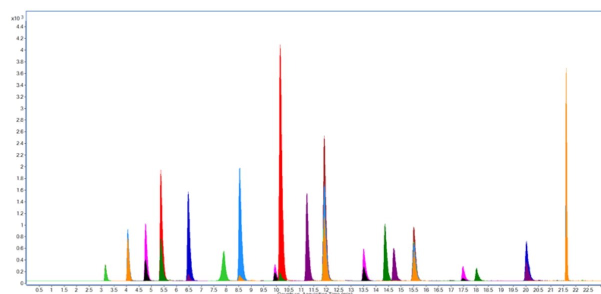
http://www.ncbi.nlm.nih.gov/pubmed/23323968

Database: ecoli-K12 ref uniprot-taxonomy_83333.fasta

Quantitation of contaminant *E. coli* proteins on LC-QQQ

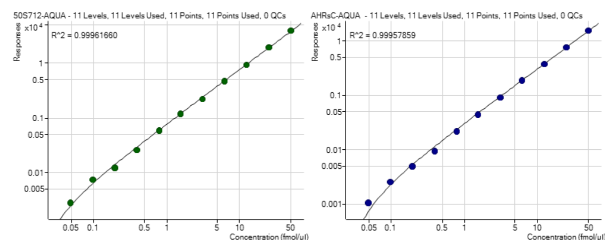
Sensitive determination of 10 *E. coli* proteins by LC-QQQ

A multiple reaction monitoring (MRM) program was developed for 10 of the most abundant proteins found in *E. coli* cytosol. The *Skyline* software package was used to select two tryptic digest peptides for each protein and two transitions (a qualifier and a quantifier) for each peptide, giving a total of 40 reactions. Transitions that gave weak signals were replaced with alternative transitions or by alternative peptide digest products of the respective proteins. Hydrophilic peptides were preferred over more hydrophobic alternatives because of their greater stability in solution during sample handling.

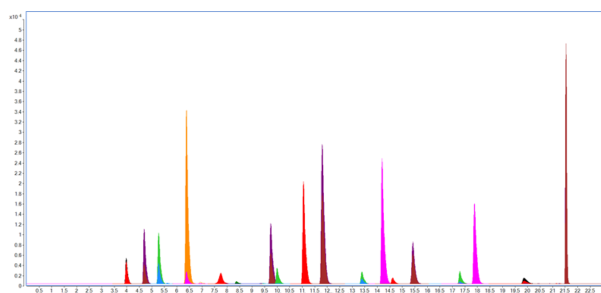


10 *E. coli* cytosol proteins were determined in 22 minutes

Stable isotope labeled standards were acquired for two of the *E. coli* peptides. These were used to show the linear response of the system across three orders of magnitude (50 fmol/µl down to 50 amol/µl).

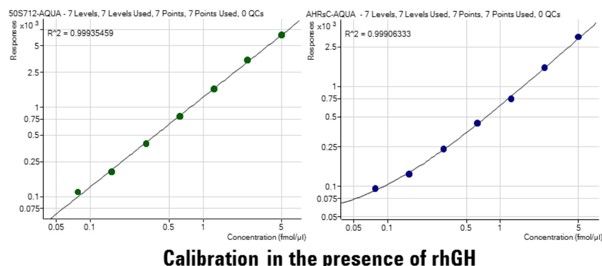


All 10 of the monitored *E. coli* proteins were detected in the commercial rhGH sample at levels far greater than the LOQ.



Determination of *E. coli* proteins in commercial rhGH

Two stable isotope labeled peptides were spiked into the rhGH in order to prepare calibration curves.



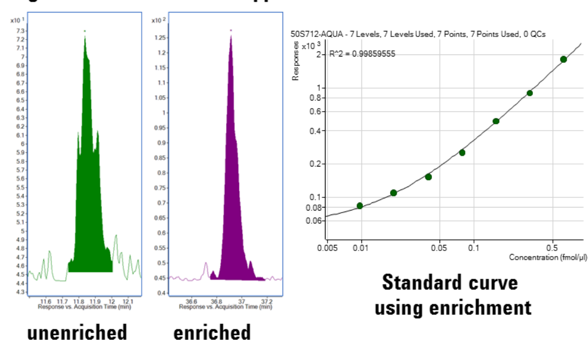
Calibration in the presence of rhGH

These data were used to quantify two *E coli.* proteins in ppm weight versus rhGH. The sample contained 1547 ppm of 50S ribosomal protein L7/L12 and 1735 ppm of alkyl hydroperoxide reductase subunit C. The method can quantify proteins in the single digit ppm range, limited mostly by interference from unidentified compounds, likely including minor degradation products of rhGH itself. Whilst this sensitivity is probably sufficient, improvements were sought so that highly undesirable contaminants might be quantified at even lower concentrations.

High resolution selective sample enrichment

The selective sample enrichment scheme shown in the experimental section was demonstrated by using it to capture calibration standards of labeled peptides in the presence of rhGH. This set-up allowed larger volumes of sample to be injected whilst reducing interference by allowing co-enriched interfering compounds to separate from the enriched peptide during analysis due to their differing selectivity in the presence of TFA vs. FA.

Signal/interference at 1.6 ppm



unenriched **enriched**

Standard curve using enrichment

This approach improved the ratio of signal to interference and thereby extended the sensitivity down to the sub-ppm level. Such sensitivity is greater than anticipated requirements and would most likely satisfy even the potentially more stringent purity requirements of the future.

Agilent Technologies

Identification of additional *E. coli* protein contaminants by LC-QTOF

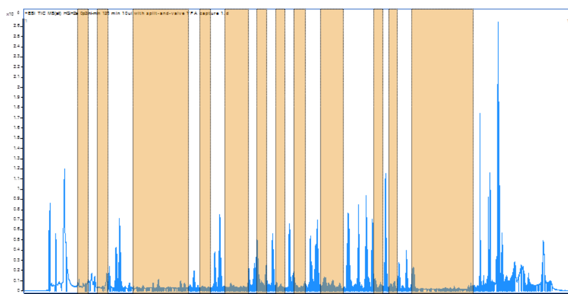
It is not practical to monitor the proteins from all 4290 genes possessed by *E. coli* by MRM. Rather, proteomic LC-MS/MS can be used to identify the major contaminant proteins in the sample. These results can be used to determine which proteins to monitor by MRM.

A standard proteomics method, employing a 2.1 x 100 mm analytical column, was used to identify HCPs in the sample. 17 unique target peptides corresponding to 9 *E coli.* proteins were identified in the commercial rhGH sample.

Boosting IDs with high resolution selective enrichment

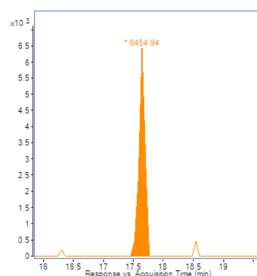
High abundance peptides from the digested therapeutic protein dictate the amount of sample that can be loaded onto the column, severely limiting the quantity of HCPs which enter the mass spectrometer. Peptides from some HCPs may either not be detected or may have insufficient intensities to yield processable MS/MS spectra.

A new approach was devised to overcome this problem. A large quantity of digested rhGH sample was injected onto a 4.6 mm ID column. Fractions of effluent that contained high abundance rhGH peptides were shunted off to waste whilst fractions containing only low abundance peptides were collected together and dried down to a volume of 20 µL.

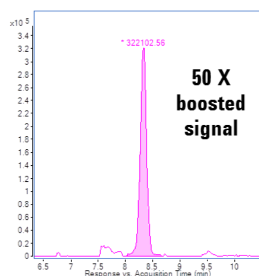


TIC of digested rhGH; shaded fractions were collected

This enriched sample was then analyzed using a sensitive HPLC-Chip-MS system. Signals from many HCP peptides were boosted by factors as high as 50.

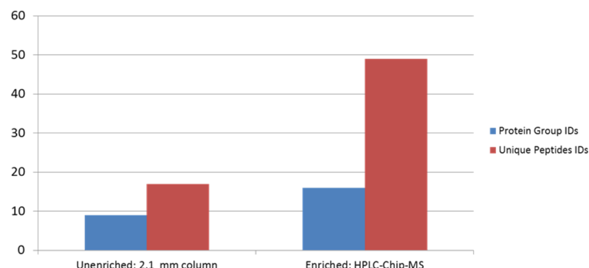


Unenriched sample on 2.1 x 100 mm column



Enriched sample on HPLC-Chip-MS

The increased HCP peptide signals afforded the selective enrichment approach resulted in a significant increase in unique peptide IDs and protein group IDs.



List of *E. coli* proteins identified in commercial rhGH using high resolution selective enrichment approach

Aconitate hydratase 2
ATP synthase subunit beta
Alkyl hydroperoxide reductase subunit C
6-phosphogluconate dehydrogenase, decarboxylating
Phospho-2-dehydro-3-deoxyheptonate aldolase, Phe-sensitive
Acetate kinase
Aerobic respiration control protein ArcA
Phospho-2-dehydro-3-deoxyheptonate aldolase
Aspartate ammonia-lyase
Aminopeptidase N
Xaa-Pro aminopeptidase
ATP synthase subunit delta
Transcriptional regulatory protein BaeR
Fructose-bisphosphate aldolase class 2
Acetylornithine deacetylase
N-succinylarginine dihydrolase

Conclusions

LC-MS/MS of digested recombinant protein samples is a promising approach to contaminant HCP analysis. A triple quadrupole mass spectrometer can quantify contaminant proteins present at concentrations at least six orders of magnitude lower than the recombinant product. Given that 10 proteins (40 transitions) were very easily accommodated within in 22 minute analysis, it will be possible to quantify many more HCPs in a single method. Meanwhile, a QTOF instrument can be used to identify the most abundant HCPs. In both cases, high resolution selective sample enrichment of the digested samples can be used to overcome some of the limitations and interferences that arise from the high concentration of therapeutic protein compared to HCPs.

Acknowledgements:
Pat Perkins kindly donated digested *E. coli* cytosol proteins that were used to develop the MRM program.
Craig Wenger provided a tutorial on the *Morpheus* software.