

HIGHER SEQUENCE COVERAGE OF PROTEINS WITH UPLC/MS^E

dr. sc. Mario Cindrić (mcindric@irb.hr, tel: +385 99 218 32 80)

mag. sci. Marija Nišavić (mnisavic@irb.hr)

dr. sc. Amela Hozic (ahozic@irb.hr)

RUĐER BOŠKOVIĆ INSTITUTE, ZAGREB, CROATIA

BIJENICKA CESTA 54

- ✓ *Protein identification method for biopharmaceutical industry designed to become an industrial standard for protein expression / production quality control*

Introduction

Currently used LC/MS (MS) technique for protein identification is based on database matching of non-derivatized peptide signals recorded in the positive ion mode. Although widely used, such an approach does not always provide satisfactory sequence coverage to unambiguously identify a protein sample.

We have introduced a peptide derivatization method that allows for spectra recording in both - positive and negative ion mode, in order to increase peptide sequence coverage. This method for unambiguous protein identification could be utilized in wide variety of industrial and research applications.

What we are presenting is:

- Negative/positive Chemically Activated Fragmentation (**CAF-/CAF+**) reagent
- Make-up solvent with addition of **reagent A^a** to enhance ionization in negative ion mode

The method comprises of four main steps (*Figure 1*):

1. Single protease digestion with trypsin
2. N-terminal peptide derivatization with **5-formylbenzene-1,3-disulfonic acid (EP2710380 A1, US 8647880)**
3. UPLC/MS^E experiments in positive and negative ion mode
4. Data processing (PLGS)

^a - Reagent A is 0.05 % (w/w) of CAF reagent (5-formylbenzene-1,3-disulfonic acid) in isopropanol

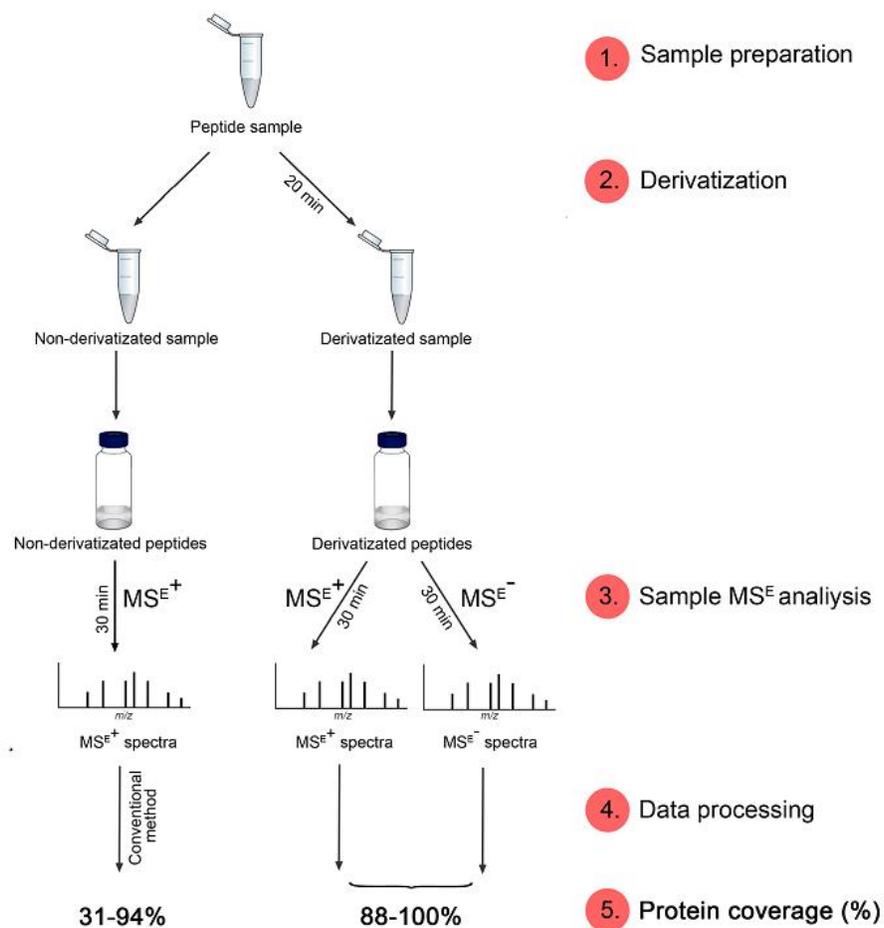


Figure 1. Schematic representation of the methods workflow.

Post-column addition of **reagent A** enables negative ionization of derivatized and non-derivatized peptides. Examples of UPLC chromatograms with and without post-column introduction of reagent A (0.4 μ L flow rate) are shown in Figures 2 and 3.

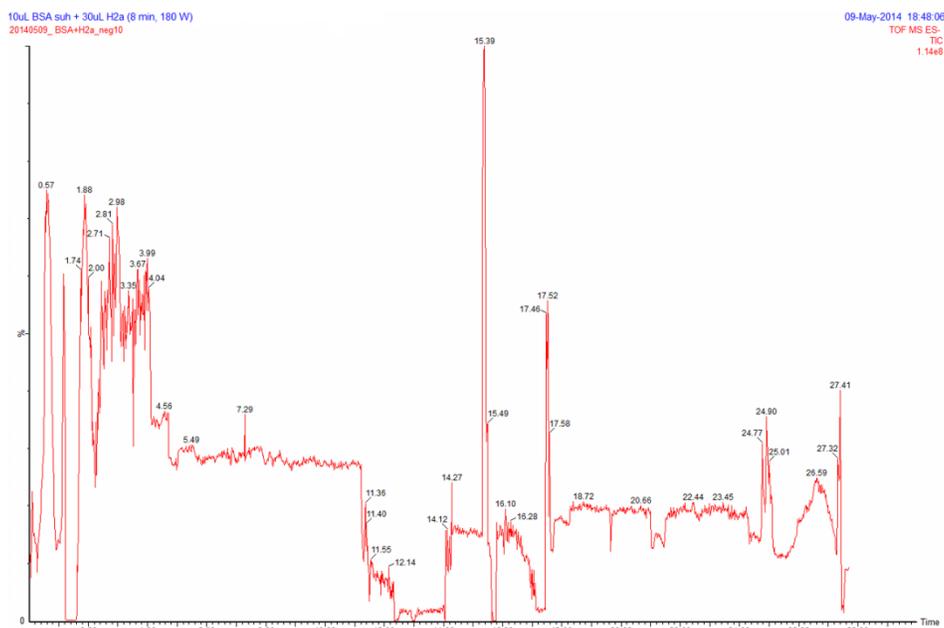


Figure 2. UPLC chromatogram of derivatized BSA digest in negative ion mode **without** post-column addition of reagent A.

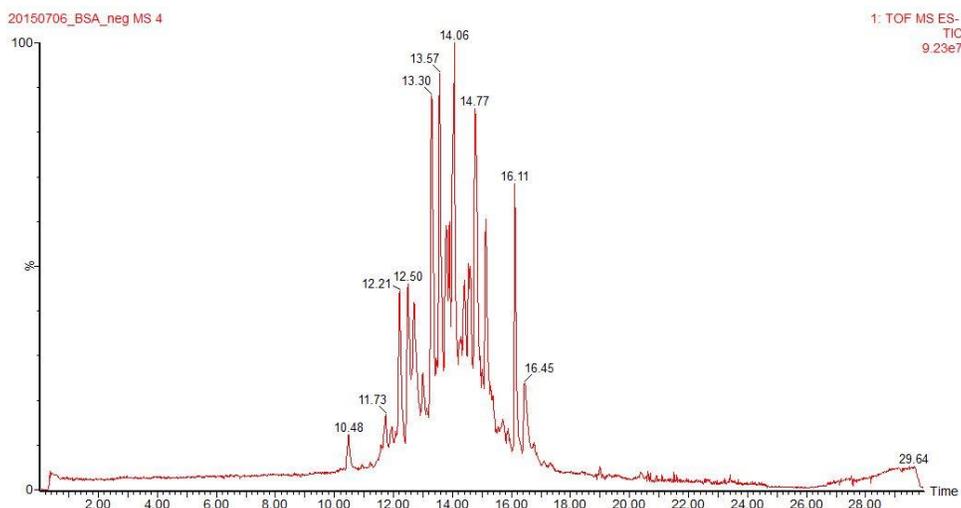


Figure 3. UPLC chromatogram of derivatized BSA digest **with** post-column addition of reagent A.

Experimental

In order to compare sequence coverage obtained using our method with the conventional one, 12 protein samples: bovine serum albumin (BSA), transferrin (Trn), erythropoietin (Epo), lysozyme (Lyz), aldolase (Als), β -casein (Csn), Isoleucyl-tRNA synthetase (IleRS), nucleotidyl transferase (CCA), TEV protease (TEV), Leucyl-tRNA synthetase (LeuRS), Elongation factor Tu (EF-Tu) and Seryl-tRNA synthetase (SerRS) were analyzed. In case of Epo and Trn, glycan moieties were enzymatically removed with PNGase F prior to trypsin digestion using protocol described by Cindrić *et al*¹.

■ Sample Preparation

Is given in APPENDIX under section **Sample preparation**

■ System and Method conditions

1. LC

LC system: nanoACQUITY UPLC® (Waters, Milford, MA, USA)

Trapping column: nanoACQUITY UPLC® 2G-V/M Symmetry® C18 Trap Column, 100Å, 5µm, 180 µm x 20 mm (Waters, Milford, MA, USA, p/n 186006527)

Trapping conditions: Isocratic delivery of aqueous 0.1% formic acid, at 15 µL/min for two minutes

Analytical column: ACQUITY UPLC® BEH130 C₁₈, 130 Å, 1.7 µm, 100 µm × 100 mm Column (Waters, Milford, MA, USA, p/n: 186003546)

Column temperature: 40 °C

Elution flow rate: 1 µL/min

Mobile phase A: aqueous 0.1% formic acid

Mobile phase B: 0.1% formic acid in 95% acetonitrile

Gradient: 0.1% to 99% solvent B in 30 minutes

Run time: 30 minutes

Sample injection volume: 4 µL

2. MS

Instrument: SYNAPT G2-Si mass spectrometer (Waters, Milford, MA, USA)

Instrument parameters:

Analyser mode: Resolution

Capillary: 3.5 kV

Source temperature: 80 °C

Nanoflow gas pressure: 1 bar

Cone voltage: 40 V

MS^E parameters:

Acquisition time: 30 min

Mass range: 50 to 4000 Da

Data format: Continuum

Scan time: 1 s

MS^E data were acquired in positive ion mode for non-derivatized samples and positive and negative ion modes for derivatized samples with collision cell energy alternating between low energy (4 eV) to collect peptide precursor (MS) data, and elevated energy (ramping from 20 to 40 eV) to obtain peptide fragmentation (MS^E) data (standard MS^E procedure). Sampling of the lock spray channel (1 ng/ μ L leucine enkephalin in 50:50 isopropyl alcohol/water containing 0.1% formic acid) was performed every 1 min to ensure high mass accuracy.

■ Data processing

The acquired data were processed by ProteinLynx Global Server software (PLGS; v. 3.0.1, Waters). Peak lists were generated after deisotoping and deconvolution. Separate databases containing sequence of each of the investigated protein were created and the data were searched with trypsin as a digestion reagent and three potential miscleavages. Peptide and fragment tolerance were set to automatic. Oxidation M and dehydration ST were allowed as variable modifications in all protein data sets, while deamidation N was added for Epo and Trn data and phosphorylation for Csn data set.

For the derivatized peptides, N-term reagent modifier was created for **5-formylbenzene-1,3-disulfonic acid** and used as fixed modification in workflow parameters.

Results and Discussion

By comparing the percentage of protein sequence coverage obtained using conventional method (non-derivatized + mode) with the percentage obtained using our method (derivatized +/- mode), we have found that our method provided higher sequence coverage for each of 12 analyzed proteins. The results are shown in *Figure 4* and summarized in *Table 1*. More detailed information about coverage of the analyzed proteins can be found in separate document (Appendix_2015-05-19).

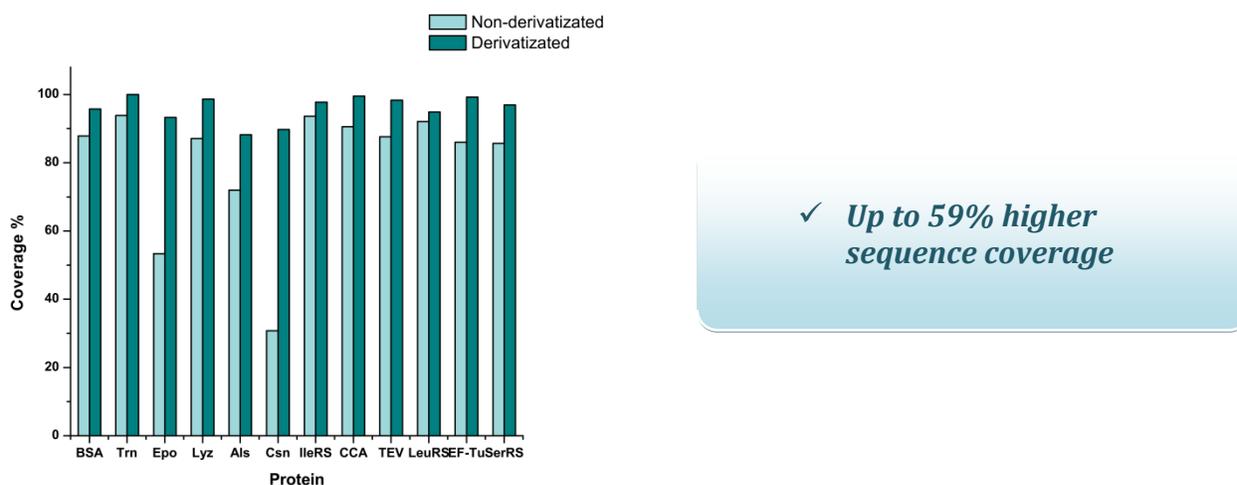


Figure 4. Protein sequence coverage (%) calculated after PLGS data processing and database matching of non-derivatized and derivatized peptides acquired in MS^E analyses in positive and positive/negative ion mode, respectively.

Our method provided sequence coverage that ranged from **88-100% (96% on average)**, in contrast to conventional method that ranged from **31-94% (80% on average)**. For instance at Csn sample the difference in sequence coverage reached 59%.

Table 1. Protein sequence coverage (%) calculated after database matching of non-derivatized and derivatized peptides acquired in MS^E analyses in positive and positive/negative ion modes, respectively.

Protein:	BSA	Trn	Epo	Lyz	Als	Csn	IleRS	CCA	TEV	LeuRS	EF-Tu	SerRS
Non-derivatized (+ mode) %	87.81	93.84	53.37	87.07	71.98	30.80	93.60	90.53	87.60	92.09	86.04	85.63
Derivatized (+/- mode) %	95.72	100	93.26	98.64	88.19	89.73	97.76	99.51	98.35	94.88	99.24	96.93

Conclusions

- Protein sequence coverage calculated after database matching (PLGS) of 12 analyzed proteins reached 31-94% (**80%** on average) when using conventional method. Our method **increased sequence coverage** in only 30 min run to 88-100% (**96%** on average). Protein sequence coverage obtained with this method could be further improved to 100% with minimal gradient and elution time method modification.
- Post-column addition of **reagent A** enables ESI peptide mapping and ionization in negative ion mode at conventional LC conditions (water/acetonitrile/0.1% formic acid).
- **CAF-/CAF+** peptide derivatization was not affected by addition of surfactant or chaotropic agents. Non-specific labeling was not noticed in 3 years of the method developing.
- Validity of the protein sequence coverage method is significantly improved. Positive ion mode results can be checked by negative ion mode results and vice versa. MS^E methodology gains an upgrade to the level of maximum validity. Purpose of the presented method is to **unambiguously determine protein sequence**. Application of the method could be utilized in research and pharmaceutical industry as an industrial standard for protein expression/production quality control.

References

1. M. Cindrić, L. Bindila, T. Čepo, J. Peter-Katalinić. Mass Spectrometry-Based Glycoproteomic Approach Involving Lysine Derivatization for Structural Characterization of Recombinant Human Erythropoietin. *J. Proteome Res.*, 2006, 5 (11), 3066–3076.
2. Mario Cindric, Zdenko Hamersak, Ivana Dodig. Method of detection of amino acid sequence and/or identification of peptides and proteins, by use of a new derivatization reagent and synthesis of 5-formyl-benzene-1,3-disulphonic acid as derivatization reagent, EP2710380 A1.

APPENDIX

Sample preparation

1. For tryptic digestion, 0.05 M Na_2HPO_4 buffer solution of pH 8.0 containing 25 μg of protein was introduced into a polypropylene tube. 25 μL of a 0.02 mg/mL trypsin solution was added to the sample and incubated at 37 °C for 18 h.
2. For denaturing, 125 μL of an 8 M guanidine hydrochloride solution was added to the sample and mixed well.
3. For reduction, 10 μL of a 1 M solution of dithiothreitol was added and mixed well.
4. The capped tube was placed in boiling water for 1 min, then allowed to cool to room temperature.
5. The sample containing tryptic peptide mixture was divided in two equal volumes, purified using 100 μL ZipTips and dried in SpeedVac. After drying, each tube contained 10 μg of the evaporated sample.
6. One sample was resuspended in 100 μL 0.1% formic acid to reach the final concentration of 0.1 mg/mL and placed in a vial for the UPLC/MS analysis.
7. Other sample was reconstituted in 30 μL of **CAF** derivatization solution. The derivatization solution contains 12.5 mM 5-formylbenzene-1,3-disulfonic acid disodium salt hydrate (p.a. synthetic product, Ruđer Bošković Institute, Cindrić *et al*²) and 95.5 mM of NaBH_3CN in 10 mM KH_2PO_4 , pH 5. The sample containing tube was closed, placed in the styrofoam holder and derivatized in a household microwave oven at 180 W for **20 minutes (2x10 min cycles)**. Finally, the sample was diluted with ultrapure water to a final concentration of 0.1 mg/mL.