Biomarker discovery for cervical cancer
Govorukhina, Natalia I.

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2007

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
Chapter I.III

Label-Free Proteomics of Serum

N.I. Govorukhina, P. Horvatovich and R. Bischoff

1. Introduction

The comparative analysis of serum samples from patients and healthy controls requires highly standardized operating procedures that produce reproducible data [1]. The generated data need to be processed in a manner to bring the significant, disease-related differences in protein or peptide profiles forward and to reduce non-related noise [2]. Processed data have to be analyzed in a statistically rigorous fashion and subjected to both statistical and biological validation.

In this chapter we present a protocol to perform proteomics of serum samples obtained from cancer patients but the protocol is generic enough to be also applicable to sera from patients with other diseases. This is obviously just one way of proceeding and there are quite a number of other approaches, some of which can be found in this series. In order to enhance the concentration sensitivity of this method, we remove high-abundance proteins by immunoaffinity chromatography. We have recently shown that this can be done efficiently and with high repeatability [3,4]. The subsequent trypsin digestion step and reversed-phase HPLC-MS analysis are performed in a reproducible manner and controlled with standard samples at regular intervals. Concentration sensitivity of this method is app. 0.5 μM for the added Cytochrome C. In order to enhance concentration sensitivity further, it is optional to include an additional protein separation step. We describe the use of a recently developed reversed-phase material that can be run at 80°C [5].

Although we touch upon data processing and statistical analysis, we cannot go into the methodological details due to limited space. We refer the reader to the cited references as well as to a dedicated book in this series focusing on bioinformatics.

2. Materials

2.1. Depletion of the 6 most abundant proteins on a Multiple Affinity Removal column
1. Store serum samples at -80°C in aliquots.
2. Buffer A (# 5185-5987, Agilent, Palo Alto, California, USA).
4. 0.22 μm spin filters (Part # 5185-5990, Agilent).
5. Multiple Affinity Removal column (Agilent, 4.6 x 50 mm, Part # 5185-5984 Palo Alto, California, USA).

2.2. SDS-PAGE
1. All chemicals for Polyacrylamide gel were from BIO-RAD (Bio-Rad, www.biorad.com)
2. PageRuler™ Prestained Protein Ladder (Fermentas, #SM0671)

2.3. HPLC-MS
1. Atlantis™ dC 18 (1.0 x 150 mm, 3 μm) column for cap-LC-MS, (Waters, Milford, Massachusetts, USA, www.waters.com).
3. Chip for chip-LC-MS with a 40 nl trap column (75 μm × 11 mm) and a 75 μm × 43 mm analytical column both containing C-18SB-ZX 5 μm chromatographic material (Cat. # G4240-62001, Agilent, Palo Alto, California, USA). The chip is equipped with a nanoelectrospray tip of 2 mm length with conical shape: 100 μm OD × 8 μm ID.
5. Sequencing-grade modified trypsin (Promega, Cat# V5111, U.S.A.).
6. Acetonitrile HPLC-S (ACN) gradient-grade was from (Biosolve, Valkenswaard, The Netherlands).
7. Formic acid, FA, 98-100% pro analysis (Cat# 1.00264.1000., Merck, Darmstadt, Germany).

2.4. Pre-fractionation of depleted serum on an mRP-C18 Macroporous Reversed-Phase column
1. Macroporous Reversed-Phase mRP-C18 column (Agilent, 4.6 x 50mm, Part # 5188-5231 Palo Alto, California, USA).
2. Trifluoroacetic Acid, TFA, sequencing grade (# 28902, Pierce).
5. Glacial acetic acid (Cat#1.00063.1000, Merck, Darmstadt, Germany)
6. Solvent A for mRP column (97% water/0.1% TFA)
7. Solvent B for mRP column (97% AcN/0.1% TFA)
3. Methods

3.1. Preparation of Samples

1. Mix 20 μL of crude serum with 80 μL of buffer A (Agilent). Filter through 0.22 μm spin filters at 13000g and 4°C for 10 min to remove particulates.

2. Inject 80 μL (80% of the total amount of diluted crude serum) on a Multiple Affinity Removal column for depletion according to the manufacturer’s instructions (with detection at 280 nm using the following timetable: 0-9 min, 100% buffer A (0.25 mL/min); 9.0-9.1 min, linear gradient 0-100 B % (1 mL/min), 9.1-12.5 min, 100% buffer B (1 mL/min); 12.5-12.6 min, linear gradient 100-0% buffer B (1 mL/min); 12.6-20 min, 100% buffer A (1 mL/min). Removal of abundant proteins, as described above, was performed on a LaChrom HPLC System (Merck Hitachi, www.merck.com) or on an AKTA FPLC system (GE Healthcare).

3. Collect the flow-through fraction (depleted serum collected between 2-6 min) of a total volume of appr. 1 mL.

4. Determine protein concentrations with the Micro BCA™ protein assay reagent kit (www.piercenet.com) and calculate for an average protein molecular weight of 50 kDa. Use BSA as the calibration standard.

5. Digest 100 μL (~10% of the total amount, which corresponds to ~8 μg or 160 pmol of total protein) of depleted serum with trypsin (1:20 wt/wt enzyme to substrate) at 37°C overnight with shaking at 400 rpm.

3.2. SDS-PAGE

1. SDS PAGE was performed in a Mini-Protein III cell (Bio-Rad, www.biorad.com) using 12% gels with 0.1% SDS according to the manufacturer’s instructions.

2. Boil samples with sample buffer containing 0.02 M DTT for 1 min, cool down and apply directly to the gel.

3. Perform staining with Coomassie Brilliant Blue R concentrate (Sigma, www.sigmaaldrich.com) diluted and used as instructed by the manufacturer.

3.3. HPLC-MS

3.3.1. Capillary- and chip-LC-MS

1. All LC-MS analyses were performed on an Agilent 1100 capillary (cap) HPLC system coupled on-line to an SL ion-trap mass spectrometer (www.home.agilent.com). In the case of cap-LC-MS, the instrument was equipped with an Atlantis™ dC 18 (1.0 x 150 mm, 3 μm) column that was protected by an Atlantis™ dC 18 in-line trap column. 40 μL of the pretreated (depleted and digested) serum corresponding to ~8 μg or 160 pmol of total protein digest (calculation based on a 50kDa protein) were injected. An autosampler (cat. # G1367A) equipped with a 100 μL injection loop was used for cap-LC-MS. For chip-LC-MS the same mass spectrometer was used but
equipped with a microfluidics (chip-cube) interface (cat. # G4240A) including a chip microfluidic device. The injected sample amount was 0.25 µg (3.4 - 5.1 µL; 5 pmol) of depleted, trypsin-digested serum, 10-times diluted with 0.1% aq. FA. Injections were performed with an autosampler (Agilent, cat. n° G1389A) equipped with an injection loop of 8 µL (this includes also the dead volume up to the trapping column). In both case, the autosampler was temperature-controlled using a cooler (cat. # G1330A) maintaining the samples at 4°C.

The HPLC system for cap-LC-MS had the following additional modules: capillary pump (cat. #, G1376A), solvent degasser (cat. #, G1379A), UV detector (cat. # G1314A) and column holder (cat. #, G1316A). The sample was injected and washed in the back-flush mode for 30 min (0.1% aq. FA and 3% acetonitrile at a flow-rate of 50 µL/min). Peptides were eluted in a linear gradient from 0 to 70% (0.5%/min) acetonitrile with 0.1% formic acid at a flow-rate of 20 µL/min. After each injection the in-line trap and the analytical column were equilibrated with eluent A for 20 min prior to the next injection.

The chip-LC-MS system contained the following additional modules: nanopump (cat. n° G2226A), capillary loading pump and solvent degasser. The sample was injected and washed in the back-flush mode for 4 min (0.1% aq. FA, 2 µL/min) and then the on-chip trapping column was switched in-line with the analytical column on the microfluidics device. For these separations, the same eluents A and B as for the cap-LC-MS system were used at a flow-rate of 0.3 µL/min. After elution for 6 minutes with eluent A, a linear gradient from 0 to 50% eluent B at 0.5 %/min followed by a step gradient from 50 to 70% at 1 %/min of eluent B was run. 70% eluent B was maintained for 10 min. After each injection, the in-line trap and the analytical column were equilibrated with eluent A for 20 min at 2 and 0.3 µL/min, respectively.

2. In the MS acquisition parameters only the ionization voltage and the use of nebulizer gas were different between the two systems (1800-2000 V of ionization voltage and no use of nebulizer gas for chip-LC-MS; 16.0 psi N₂ nebulizer gas and 3500 V of ionization voltage for cap-LC-MS). The following general settings were used for mass spectrometry during LC-MS: drying gas: 6.0 L/min N₂, skimmer: 40.0 V, cap. exit: 158.5 V, Oct. 1: 12.0V, Oct. 2: 2.48 V, Oct. RF: 150 Vpp (Voltage, Peak Power Point), Lens 1: -5.0 V, Lens 2: -60.0 V, Trap drive: 53.3, T°: 325°, Scan resolution: enhanced (5500 m/z per second scan speed). Target mass: 600. Scan range: 100-1500 m/z. Spectra were saved in centroid mode. LC-MS chromatographic data were analyzed with Bruker Data Analysis software, version 2.1 (Build 37).

3.3. Data Processing

The original Bruker Daltonics LC-MS data files were converted into ASCII-format with the Bruker data analysis software. For further data analysis Matlab (version 7.2.0.232 (R2006a), Mathworks, Natick, Massachusetts, USA) and the PLS toolbox (version 3.5.2, Eigenvector Research Inc., Wenatchee, Washington,
USA) were used. Centroid data were smoothed and reduced using a normalized two-dimensional Gaussian filter with rounding of the nominal m/z ratios to 1 m/z (the original data had a resolution of 0.1 m/z). After meshing the data files of all chromatograms, they were time-aligned to a reference data file using Correlation Optimized Warping (COW) based on Total Ion Currents (TICs) constructed from signals in the range 100-1500 m/z.

A modified M-N rule was applied for peak detection by first calculating a median local baseline using a sliding window technique separately for each m/z trace. A median window size of 1200 data points, corresponding to 20.84 min for chip-LC-MS and 20.17 min for cap-LC-MS, was used with a moving rate of 10 points and a minimum median value of 200 counts. According to the M-N rule, a threshold of M-times the local baseline was used and a peak was assigned if, within one m/z trace, the signal exceeded this threshold for at least N consecutive points. For each detected peak the m/z value, the mean retention times of the three highest measured intensities (within the same peak reduced by the local baseline) were stored in a peak list created for every chromatogram.

In order to combine the peak lists from different samples with each other, one-dimensional peak matching was achieved by using the sliding window technique, in which the same m/z traces were evaluated for peaks that are proximate in time (step size 0.1 min; search window 1.0 min; maximal accepted standard deviation for all retention times within a group of matched peaks was 0.75 min). Missing peak locations were filled with extracted local signals reduced with the local baseline at a given m/z retention time location. The generated peak matrix, created from the peak lists of the individual samples, consisted of a peak(row)-sample(column)-intensity(value) matrix. This peak matrix was used for multivariate statistical analysis.

A Nearest Shrunken Centroid (NSC) supervised classification algorithm in conjunction with leave-one-out cross-validation (LOOCV) was applied to select the most discriminating compounds. The selected compounds were then subjected to autoscaled Principal Component Analysis (PCA) and visualized using biplots of the first two principal components. All data processing and statistical analyses were done on a personal computer equipped with a dual core +3800 MHz AMD 64 X2 processor equipped with 4 GB of RAM. Figure 1 shows an example of data obtained by chip- and capillary-LC-MS.

3.4. Pre-fractionation of depleted serum by reversed-phase HPLC on an mRP column at 80°C
1. Add to ~300μg (about 300μL) of depleted serum 0.48 g urea and 13μL of glacial acetic acid, according to the manufacturer’s instructions (www.agilent.com/chem/bioreagents).
2. Add solvent A to a final volume of 1mL and inject the total volume with a 1mL loop onto the column.
3. Fractionate at 80 °C (pH <5.0) at a flow rate of 0.75mL/min with UV detection at 280nm.
4. Run gradient from 3 to 30% B in 6 min, to 55% solvent B in 40 min and up to 100% in 53 min.
5. Collect fractions of 0.75mL (see Figure 2a).
6. Compare fractions after pre-fractionation by SDS-PAGE (in our case pairwise before and after medical treatment for each patient) (see Figure 2b).

Figure 1. Raw LC-MS data of depleted and trypsin-digested serum analyzed by chip- (a) or capillary LC-MS (b) represented as “heat map”. The horizontal axis represents the m/z values in amu and the vertical axis shows the retention time in min. Peak intensity is coded as indicated (white: high; black: low). Panels (c) and (d) show the same data in the conventional representation as Total Ion Chromatograms (TICs). The dashed lines depict the calculated baseline. Data were collected in centroid mode and meshed using a data reduction of 1:10.
Figure 2. (a) 300 μg of depleted serum were pre-fractionated on an mRP column at 80°C (example of a sample from a cervical cancer patient before treatment). (b) 12% SDS-PAGE of serum from a cervical cancer patient before “A” and after “B” medical treatment. 30A and 31A: fractions 30 and 31 from the mRP column of patient serum before medical treatment; 30B and 31B: fractions 30 and 31 of the serum from the same patient after treatment. Note the clear difference at about 35 kDa in fraction 30.

4. Notes

Prefractionating proteins in depleted serum/plasma on a newly developed “Macroporous Reversed-Phase C18 High-Recovery Protein Fractionation HPLC column (mRP)” prior to digestion is reproducible and enables high resolution to be achieved. High protein recoveries allow the complexity of the sample to be significantly reduced. Robustness and high recovery mRP fractionation makes higher-quality protein identification by coupled LC-MS methods [5]. Based on our experiences and the work of Martosella et al. [5], elevated mRP column temperature (80 degrees C) is a very critical operational parameter, while poor control of the temperature could result in poor reproducibility and bad chromatographic resolution.

The performance of the described methodology was evaluated by comparing the ability of cap- and chip-LC-MS to find discriminating features. For this purpose, 5 serum samples, spiked with 21 pmol of horse heart Cytochrome C in 2 μL serum, were analyzed next to 5 non-spiked serum samples. Due to losses during immunoaffinity depletion of high-abundance proteins, the actual amount of Cytochrome C that was analyzed was 4.2 pmol [3], corresponding to about 3% (molar) of the total protein content. The raw data obtained were subjected to data processing as described [6], followed by supervised classification and selection of discriminating features using the Nearest Shrunken Centroid (NSC) algorithm [7]. The shrinkage parameter was optimized using a “leave one out” cross validation (LOOCV), strategy with the aim of reaching the lowest cross validation error. Although we applied a rather low threshold (M = 2, N = 5) for peak picking, which introduced more noise in the peak list, a large domain of shrinkage showed no cross-validation error (0.90-29.51 for chip- and 0.61-16.80 for cap-LC-MS, Figures 3a and b,
respectively), indicating a robust classification model. Evaluating the 16 most discriminating features selected at shrinkages of 10 and 8.5 for chip- and cap-LC-MS, respectively, resulted in 6 different peptides. Six peptides selected from the chip-LC-MS and 5 of the 6 peptides selected from the cap-LC-MS data corresponded to \textit{in-silico} predicted tryptic peptides of horse heart Cytochrome C. Figure 3 shows that correct discrimination between spiked versus non-spiked serum samples was easily possible based on the selected peaks (Figures 3c and d). PCA analysis of the selected features (Figures 3e and f) revealed that almost all variability in the data can be explained by Principal Component 1 (99% for chip- and 98% for cap-LC-MS). Visualization of the Extracted Ion Chromatograms (EICs) of some of the selected peaks (Figure 4) confirmed that highly discriminating peaks had been correctly found within the complex mixture of digested serum proteins. Figure 4 shows also the generally good time alignment using COW.

The results show that the integration of nanoLC into microfluidic devices enables quantitative profiling studies to find protein expression differences using ~30 times less sample with higher separation efficiency compared to capillary LC.
Figure 3. Representation of the “leave-one out” cross-validation (LOOCV) error and the number of selected variables as a function of the shrinkage for chip-LC-MS (a) and cap-LC-MS (b). The selected variables, where the shrinkage domain has no cross-validation error, are indicated with arrows. For these domains, the selected variables enabled a perfect separation of the two classes (see panels (c) and (f)). PCA plots using all peaks obtained with M = 2, N = 5 for chip-LC-MS (c) and cap-LC-MS (d), (14091 for chip-LC-MS and 11256 for cap-LC-MS) did, however, not allow discrimination between the classes. In the figures PC 1 and PC 2 refer to the Principal Component axis 1 and 2.
Figure 4. Examples of Extracted Ion Chromatograms (EICs) of NSC-selected peaks corresponding to tryptic fragments of horse heart Cytochrom C from datasets obtained with chip-LC-MS (left) and cap-LC-MS (right). The green upper traces were obtained from spiked, the blue lower traces were obtained from non-spiked samples.
5. References


