doi:10.4149/neo_2010_06_537

Optimizing of MALDI-ToF-based low-molecular-weight serum proteome pattern analysis in detection of breast cancer patients; the effect of albumin removal on classification performance.

M. PIETROWSKA[1,5], Ł. MARCZAK[2,5], J. POLANSKA[3,5], E. NOWICKA[1], K. BEHRENT[1], R. TARNAWSKI[1], M. STOBIECKI[2], A. POLANSKI[3,4], P. WIDLAK[1]

[1]Maria Skłodowska-Curie Memorial Cancer Center and Institute of Oncology, 44-100 Gliwice, Poland, e-mail: widlak@io.gliwice.pl [2]Polish Academy of Science, Institute of Bioorganic Chemistry, Poznan, Poland; [3]Silesian University of Technology, Gliwice, Poland; [4]Polish-Japanese Institute of Information Technology, Bytom, Poland; [5]These authors contributed equally

Mass spectrometry-based analysis of the serum proteome allows identifying multi-peptide patterns/signatures specific for blood of cancer patients, thus having high potential value for cancer diagnostics. However, because of problems with optimization and standardization of experimental and computational design, none of identified proteome patterns/signatures was approved for diagnostics in clinical practice as yet. Here we compared two methods of serum sample preparation for mass spectrometry-based proteome pattern analysis aimed to identify biomarkers that could be used in early detection of breast cancer patients. Blood samples were collected in a group of 92 patients diagnosed at early (I and II) stages of the disease before the start of therapy, and in a group of age-matched healthy controls (104 women). Serum specimens were purified and analyzed using MALDI-ToF spectrometry, either directly or after membrane filtration (50 kDa cut-off) to remove albumin and other large serum proteins. Mass spectra of the low-molecular-weight fraction (2-10 kDa) of the serum proteome were resolved using the Gaussian mixture decomposition, and identified spectral components were used to build classifiers that differentiated samples from breast cancer patients and healthy persons. Mass spectra of complete serum and membrane-filtered albumin-depleted samples have apparently different structure and peaks specific for both types of samples could be identified. The optimal classifier built for the complete serum specimens consisted of 8 spectral components, and had 81% specificity and 72% sensitivity, while that built for the membrane-filtered samples consisted of 4 components, and had 80% specificity and 81% sensitivity. We concluded that pre-processing of samples to remove albumin might be recommended before MALDI-ToF mass spectrometric analysis of the low-molecular-weight components of human serum

Key words: albumin removal; breast cancer; clinical proteomics; mass spectrometry; pattern analysis; serum proteome.

The recent progress of molecular cancer diagnostics is related to achievements of genomics and proteomics. Identifying and understanding dynamic changes in the proteome related to disease development and therapy progression is the subject of clinical/disease proteomics [1, 2]. The low molecular weight (<15 kDa) component of the blood proteome is a promising source of previously undiscovered biomarkers. Since this protein fraction is below the limit of effective resolution of conventional gel electrophoresis, mass spectrometric analyses appear to be emerging methods of clinical proteomics and cancer diagnostics [3–6]. The approach that takes into consideration protein fingerprints/profiles defined by mass spectra but does not rely on particularly identified protein(s), could be called proteome pattern analysis or proteome profiling. In this approach multi-component sets of peptides or proteins (which are exemplified by ions registered at defined m/z values in the mass spectrum) define specific proteomic patterns (or profiles) that can be used for sample identification and classification, even though their particular components may lack differentiating power when analyzed separately [7–9]. Mass spectrometric methods particularly suitable for proteome pattern analysis rely on Matrix-Assisted Laser Desorption Ionization spectrometry (MALDI) and its derivative Surface-Enhanced Laser Desorption Ionization spectrometry (SELDI) coupled to a Time-of-Flight (ToF) analyzer [10, 11]. The milestone paper in this field was published in 2002 by the group of Petricoin and Liotta, who showed that components of the serum proteome identified by mass spectrometry differentiate patients with ovarian cancer from healthy individuals [12]. Since that time, in spite of a certain controversy regarding this pioneering work [13], numerous papers have been published that aimed to verify the applicability of mass spectrometric analyses of the serum (or plasma) proteome for cancer diagnostics [14–17]. The relevance of MALDI- and SELDI-based serum (or plasma) proteome pattern analysis has been already successfully tested for several type of human malignancies, yet none of identified peptide patterns/signatures was approved for diagnostics in clinical practice so far [17–22].

Breast cancer is the most common malignancy in women comprising about 18% of all female cancers, which accounts for about a fifth of all deaths among women aged 40-50 [23]. The most important tools in screening and early detection of breast cancer are mammography, ultrasonography and magnetic resonance imaging. Unfortunately however, up to 20% of new breast cancer incidents cannot be detected by these imaging methods [24], indicating a need for novel molecular markers suitable for screening and early detection of this cancer. A few studies have addressed the possible application of mass spectrometric serum analyses in diagnostics of breast cancer [25–30]. These studies proposed several proteome patterns specific for patients with breast cancer, especially at late clinical stages. In addition, mass spectrometry analyses of the blood proteome allowed the identification of proteome patterns specific for breast cancer patients with different outcome and response to therapy [31–33]. Unfortunately, proposed cancer signatures consisted of different peptide sets, which most apparently limited direct applicability of such findings in diagnostic practice. The most obvious reasons for discrepancy among published data is lack of compatibility of methodological approaches, both experimental and computational, that were implemented in such studies. Noteworthy, however, several peptides that differentiated cancer and control samples appeared reproducibly when comparative analysis across different studies was performed [34, 30], demonstrating the high potential of mass spectrometry-based analyses of the blood proteome pattern in diagnostics of breast cancer once problems with optimization and standardization of experimental and computational design are solved.

Here we optimized experimental design of mass spectrometry-based proteome pattern analysis aimed to identify a potential biomarkers that could be applied in early detection of breast cancer patients. Serum peptides in the 2-10 kDa range were analyzed using MALDI-ToF spectrometry, the components of mass spectra were identified using the Gaussian mixture modeling, and then classifiers were built that allowed differentiation between cancer patients and healthy persons. Proteome patterns specific for cancer patients were established for either complete serum samples or samples depleted of albumin, and reliability of corresponding cancer biomarkers was compared.

Materials and methods

*Characteristics of patients and control groups***.** This study was carried out at the Maria Skłodowska-Curie Memorial Cancer Center and Institute of Oncology, Gliwice Branch, between May 2006 and January 2008. Ninety-two patients diagnosed with clinical stage I or II breast cancer were included in the study, of average age 58.5 years (range 31-74 years). Patients were classified according to the TNM scale; the majority were scored as T1 and T2 (47% and 45%, respectively) as well as N0 and N1 (75% and 24%, respectively) and none had diagnosed metastases (all M0). Serum samples were collected before the start of therapy. One hundred and four female volunteers were included as a control group; they were required to be free of any known acute or chronic illness and were not treated with any anticancer therapy in the past. The average age in this group was 54 years (range 32-77 years). The study was approved by the appropriate Ethics Committee and all participants provided informed consent indicating their voluntary participation.

Preparation of serum samples. Samples were collected and processed following a standardized protocol. Blood was collected in a 5 ml Vacutainer Tube (Becton Dickinson), incubated for 30 min. at room temperature to allow clotting, and then centrifuged at 1000g for 10 min. to remove the clot. The serum was aliquoted and stored at -70°C. Directly before analysis, samples were diluted 1:5 with 20% acetonitrile (ACN). One half of each sample was applied onto an Amicon Ultra-4 membrane (50 kDa cut-off) in a spin column and centrifuged at 3000g for 30 min. Different concentrations of ACN and different types of membranes were tested; the best efficiency of removal of high-molecular weight proteins and reproducibility of registered spectra was obtained for the combination described above (experimental data regarding all performed comparisons are not shown).

Registration of mass spectra. Samples were analyzed using an Autoflex MALDI-ToF mass spectrometer (Bruker Daltonics, Bremen, Germany); the analyzer worked in the linear mode and positive ions were recorded in the mass range between 2 and 10 kDa. Mass calibration was performed after every four samples using standards in the range of 2.8 to 16.9 kDa. Prior to analysis each sample was loaded onto a ZipTip C18 tip-microcolumn by passing it through repeatedly 10 times, column was washed with water and then eluted with 1 μl of matrix solution (30 mg/ml sinapinic acid in 50% ACN/H_2O and 0.1% TFA with addition of 1 mM n-octyl glucopyranoside) directly onto the 600 μm AnchorChip (Bruker Daltonics) plate. ZipTip extraction/loading was repeated twice for each sample and for each spot on the plate two spectra were acquired after 120 laser shots (i.e. four spectra were recorded for each sample). Spectra were exported from the Bruker FlexAnalysis 2.2 software in standard 8-bit binary ASCII format; they consisted of approximately 45,400 measurement points describing mass to charge ratios (m/z) for consecutive $[M+H]^+$ ions and the corresponding signal abundances, covering the range of analyzed m/z values. For each sample the average of 4 replicate measurements was taken for further computational analyses.

Data processing and statistical analysis. The pre-processing procedures that included interpolation of missing or non-aligned points, binning, trimming, removal of the baseline and the total ion current normalization were performed according to procedures considering to be standard in the field [35, 36]. In the second step the spectral components, which reflected $[M+H]^+$ ions recorded at defined m/z values, were identified using decomposition of mass spectra into their Gaussian components. The spectra were modeled as a sum of

Gaussian bell-shaped curves, then models were fitted to the experimental data by a variant of the expectation maximization algorithm [37], as described by Pietrowska et al. [30]. Based on the decomposition of the spectra into Gaussian components, the candidate classifier features were obtained by treating the Gaussian curves as kernel functions and using the scalar product operator. The classification relied on the naive Bayesian discrimination rule with the entropy-based feature selection principle [37] and in the constructed classifier a value of 0.5 for the probability threshold was used to discriminate between cancer and healthy states. The size of the training sample was changed from 20% to 90% of the whole dataset, and for each size the two-step procedure, training/validation, was repeated 1000 times to estimate the average error rate and its 95% confidence interval, which characterized the accuracy of classification. In order to further characterize the quality of classification, receiver operating curves (ROC) [38] were computed by changing the value of the probability threshold in the Bayesian rule from 0.0 to 1.0, and averaging the obtained specificity/sensitivity proportions over 1000 random validation experiments. We tested the performance of classification with classifiers built of different numbers of spectral components by estimating the level of total errors, as well the number of false positive and false negative classifications. Construction and validation of a classifier is a statistical process, i.e. many different classifiers built of a given number of spectral components were tested (1000 random splits of the dataset), and those which pass the quality threshold could be built of different spectral components. Thus, to identify the components that are the best determinants of a specific proteome pattern we looked for the most frequent components in classifiers that correctly classified samples.

Results and discussion

Mass spectra of complex protein mixtures such as serum can be obtained after direct application of a sample to the spectrometer, or samples can be pre-processed to deplete or enrich certain protein fractions as exemplified by the SELDI type of analyses or by approaches where selected protein fractions are first separated or purified by chromatography (e.g. a LC-MS approach). Albumin is the most abundant serum protein, and together with 9 other major proteins comprises more than 90% of serum proteins by mass [39]. Albumin is a typical carrier protein that binds numerous low-molecular-weight peptides, and for this reason a strategy based on purification of albumin followed by recovery of cargo peptides and their mass spectrometric analysis has been proposed [40]. On the other hand, although albumin is not directly targeted in the majority of spectral analyses it apparently affects the sensitivity of the low-molecular-weight serum peptide analysis because of its high abundance. For this reason removal of albumin, as well as other high-molecular-weight serum proteins, could be a useful approach to enhance the sensitivity and facilitate the reliability of MS analyses of the low-molecular-weight serum compartment. Here we implemented dilution of serum samples with a denaturing organic solvent (acetonitrile) that destroyed the majority of protein interactions and allowed analysis of individual peptides dissociated from (not interacting with) other proteins (e.g., albumin). Albumin and other high-molecular-weight serum proteins were removed by membrane filtration through Amicon Ultra-4 spin columns with a nominal cut-off of 50 kDa. Both complete input and filtered samples were analyzed by SDS-PAGE and subjected to MALDI-ToF analysis in the range of 2-10 kDa. Figure 1 shows results of comparative analyses of complete non-filtered (CS) and filtered (FS) serum from the same healthy donor, performed by either SDS-PAGE (panel A) or mass spectrometry (panel B). Filtering allowed removing of 80-90% of the albumin (Fig. 1A) and, as expected, affected the structure of the mass spectra; more peptides in the low-molecular-weight serum component could be detected in the filtered sample. This difference was clearly visible when the differential spectrum was obtained by subtracting the spectrum of the non-filtered from that of the filtered sample (Fig. 1B, bottom graph). The average of the differential spectra for all the 196 individuals is shown in Figure 1C; statistical analysis by the Lilliefors normality test showed a high significance ($p<0.00001$) of the structural differences between the spectra for filtered and non-filtered serum samples. For this reason, both type of serum samples were used in parallel for identification of breast cancer markers, and the quality of cancer classifiers obtained for such samples was compared.

Computational processing of protein profiles registered as mass spectra is always a multi-step process where spectral features are extracted after the pre-processing operations, and then they are used for further analyses, e.g., constructing spectral classifiers. Characteristic feature of MALDI ionization is that majority of registered peaks correspond to mono-protonated peptide/protein molecular ions $[M+H]^+$ described by m/z values that reflect actual molecular weights increased by the mass of the proton. However, when MALDI mass spectra are recorded in a linear mode over a wide range of m/z values, like the 2-10 kDa range in this study, the expected mass accuracy is relatively low and corresponds up to a few Daltons. In consequence, the relative broadening of spectral peaks recorded for the $[M+H]^+$ ions could reflect the low resolution of the analyzer and might result in overlapping of ions originating from protein/peptides of very similar molecular masses. In addition, because of technological imperfections there might be some shift in the positions of peptide ions between measurements, which adds more complexity to analyses of large datasets. For this reason, some approaches used for extraction of spectral features from large datasets relay on alignment of identified spectral peaks [35], which requires numerical "stretching" of spectra before further analyses. Here we decided to implement an original mathematical procedure based on modeling average spectra as the sum of Gaussian components then fitting actual experimental spectra into such a model [30]. For the mass spectra analyzed in the present work we tested models with different numbers of components

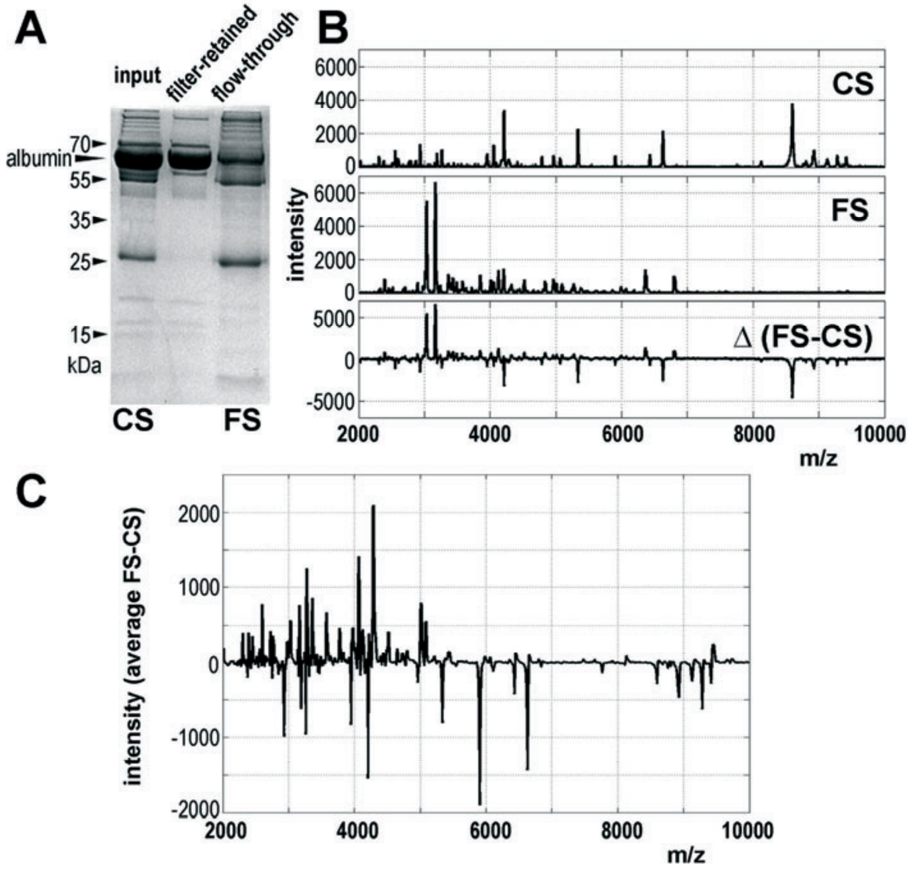

Figure 1. Characteristics of serum samples. Serum specimen from the same individual healthy person, both complete input (CS) and membrane-filtered (FS), was analyzed by SDS-PAGE (panel A) or MALDI-ToF mass spectrometry in the low molecular-weight range (panel B). To obtain the differential spectrum (bottom graph) the spectrum of the complete sample was subtracted from that of the membrane-filtered sample. The average of differences between the spectra of complete and membrane-filtered samples for all 196 individuals is also shown (panel C).

(up to 800), and found that 300 components ensured both sufficient fidelity of the model and its efficient computation (not shown). As a result of computation an "average" spectrum was decomposed into spectral components characterized by the molecular weight (m/z values of recorded $[M+H]^+$ ions) and the interval where fit corresponding peaks in at least 95% of actual spectra expected in the dataset (+/-95% CI). The resulting spectral components reflect peaks recorded in multiple samples during mass spectrometric analysis, which contained either single peptide/protein ions or a combination of a few ions of very similar m/z values. This approach allowed us to avoid artifacts resulting from the peak alignment and facilitated quantitative analysis of data by simple assessment of signal volumes that fitted to a given component within its 95% CI. Having "extracted" and quantified spectral components, one could find certain whose abundances were significantly different between groups of samples (e.g. between cancer patient and healthy controls), which could be defined as "differentiating". However, to obtain more reliable classification of samples we used spectral components to build multi-component classifiers that determined proteome patterns characteristic for defined groups, and looked for the most frequent components in classifiers that annotated samples correctly.

The performance of classification with classifiers built of different number of components (features) was tested by estimating the level of total errors, as well the number of false positive and false negative classifications. Figure 2 shows estimations of the total error rate as a function of the component/feature number for classification that used either complete (panel A) or membrane-filtered albumin-depleted samples (panel B). The best performance was observed with classifiers built of 8-12 components for the non-filtered samples (Fig. 2A) and of 3-6 components for the membrane-filtered samples (Fig. 2B). Such classifiers also represented minima in the numbers of false positive and false negative classifications (not shown). For further analyses we selected classifiers built of either 8 components or 4 components, for spectra of complete and of membrane-filtered albumin-de-

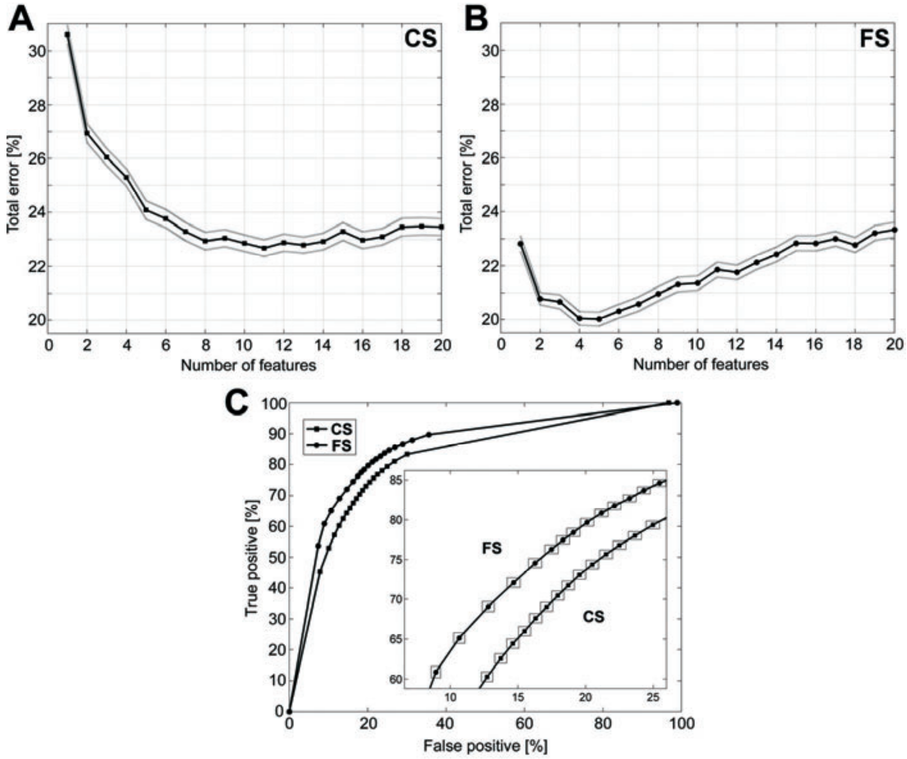

Figure 2. Estimation of the performance of cancer classification. The error rate was plotted against the number of features in the classifier for complete (panel A) and membrane-filtered samples (panel B). Shown are average error rates and 95% confidence intervals calculated based on 1000 random validation experiments with 50:50 training/validation data splits. Panel C shows estimation of the sensitivity and specificity of the classification based on complete (squares) and membrane-filtered (circles) samples. ROC curves were computed by changing the value of the probability threshold in the naive Bayesian classifier from 0.0 to 1.0, and averaging the specificity obtained versus sensitivity rate over 1000 random repeats of training and validation. In the enlarged area 95% confidence regions are shown. The percentage of true positives corresponds to the sensitivity, while (100 minus the percentage of false positives) values correspond to the specificity.

pleted samples, respectively. In order to further characterize the quality of classification, receiver operating curves (ROC), which are routine in medical applications (e.g. for estimation of reliability of bio-markers [38]), were computed to allow estimation of the sensitivity and specificity of the classifier. We computed ROC curves for direct comparison of the accuracy of classification of breast cancer patients based on analyses of either complete or membrane-filtered serum samples (Fig. 2C). According to our estimations, MS analysis that based on complete serum allowed to classify cancer patients with 81% specificity and 72% sensitivity, while that based on membrane-filtered serum had 80% specificity and 81% sensitivity. Importantly, these estimations were based on the highly rigid Monte-Carlo resampling method; when the less rigid, yet more frequently used, leave-one-out method was used the specificity/sensitivity of classification increased to 82%/89% and to 78%/91% for complete serum and membrane-filtered serum, respectively.

Assuming the different structures of mass spectra which apparently reflected the presence of unique peptide peaks/components in complete and membrane-filtered albumin-depleted serum (Fig. 1), we aimed to identify differentiating components that built classifiers and were present in these two types of sample. Essentially, we looked for the most frequent components in classifiers that correctly classified breast cancer samples. Eight most important components of the classifier for complete serum and four most important components of the classifier for filtered serum are characterized in Table 1. All of these were present in at least 40% of classifiers built of a given number of features and were marked along the whole mass spectra of both types of samples (Fig. 3A and 3B). Interestingly, such differentiating components were unique for either complete or membrane-filtered albumin-depleted serum (with possible exception of components 2876.05 Da and 2865.54 Da present in complete and filtered samples, respectively). Importantly, these most frequent components of cancer classifiers had very high potency to differentiate control and cancer samples by themselves; the statistical significance of differences obtained in univariant analyses for these three peaks were at the level of p-values from 10^{-20} to 10^{-5} (they remained highly

Table 1. Characteristics of spectral components that differentiated samples from breast cancer patients and healthy controls. Shown are the most frequent components (m/z values) and their 95% confidence intervals, and relative frequencies in cancer classifiers built of 8 or 4 features. The p-values are for differences between patients and healthy controls measured by the Mann-Whitney test for each individual feature (also shown after the Bonferroni correction against multiple testing).

m/z	- 95% CI	+ 95% CI	p-value	corrected p-value	frequency
Complete serum sample (CS)					
3314.50	3313.69	3315.31	5.37e -13	1.79e-11	96.8%
2518.60	2516.55	2520.65	3.48e -07	1.04e-04	78.6%
7649.70	7648.28	7651.12	3.47e -11	1.04e-08	76.6%
2509.27	2508.63	2509.91	3.38e-06	1.01e-03	62.4%
4363.02	4362.02	4364.02	1.28e-08	3.84e-06	61.7%
2876.05	2874.13	2877.97	6.85e-09	2.06e-06	60.6%
4967.33	4965.29	4969.38	3.71e-08	1.11e-05	47.6%
9936.24	9932.63	9939.85	1.45e-10	4.35e-08	41.9%
Membrane-filtered albumin-depleted sample (FS)					
2865.54	2864.46	2866.62	4.19e-20	1.26e-17	99.1%
3578.73	3577.42	3580.04	5.84e-18	1.75e-15	91.4%
3895.05	3894.12	3895.98	1.58e-11	4.74e-09	41.7%
7473.82	7473.01	7474.63	1.05e-05	3.15e-03	41.0%

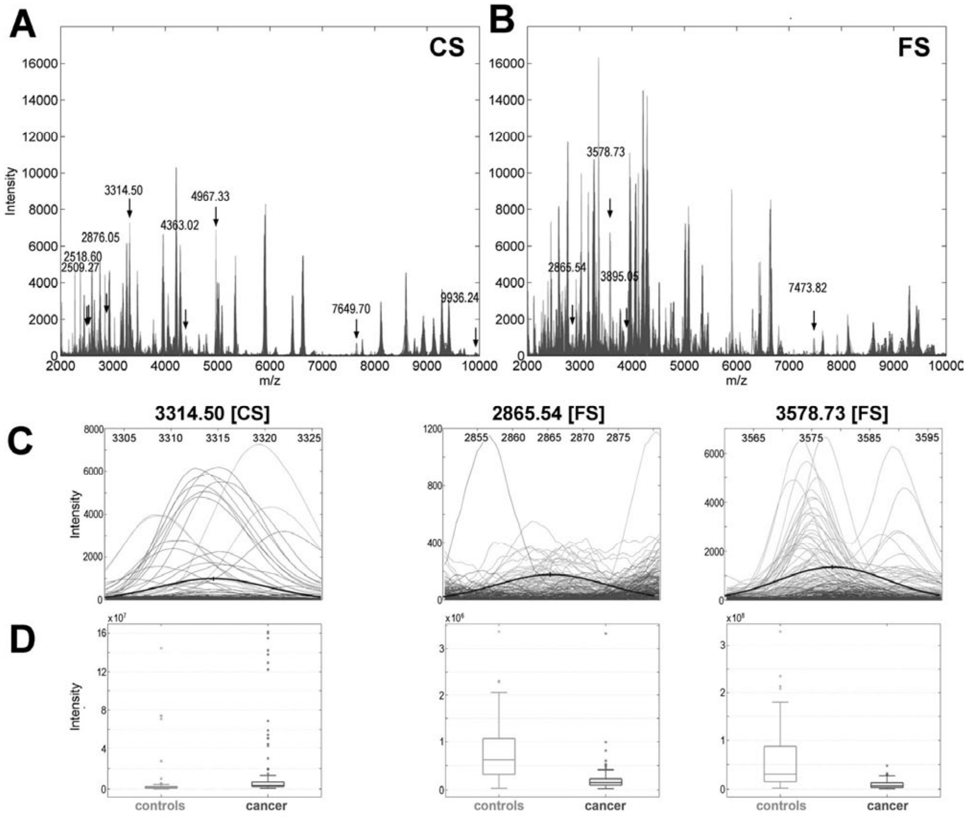

Figure 3. Characterization of essential differentiating components of the spectra. The most frequent differentiating components are marked with arrows along average mass spectra of complete (panel A) and membrane-filtered (panel B) serum samples. Panel C presents selected spectral components; shown are actual spectral plots for cancer patients (red/black lines) and healthy controls (green/grey lines), as well as modeled Gaussian kernels (blue/solid curves). X-axes represent the m/z values, Y-axes represent intensities in arbitrary units. Box-plots (panel D) represent quantification of the level of Gaussian kernel-based spectral features in samples from cancer patients (red/black) and healthy controls (green/grey) (shown are minimum, lower quartile, median, upper quartile and maximum values; outliers are marked by asterisks).

significant after application of the Bonferroni correction for multiple testing; Table 1). Almost all classifiers built for complete serum (97%) contained 3314.50 Da component, while 2865.54 and 3578.73 components were present in almost all classifiers built for filtered albumin-depleted serum (99% and 91%, respectively). The frequency of particular spectral components in cancer classifiers could be different when different algorithm was selected for classification in the same dataset. Noteworthy, however, irrespective of used algorithm, i.e. the naïve Bayesian used in this work versus a modification of Support Vector Machine used in Pietrowska *et al.* [30], frequencies of the most essential components (namely m/z=2865.54 and m/z=3578.73) remained essentially the same. Figure 3C shows fragments of mass spectra in the near vicinity of the model components that were the most frequent features of the breast cancer classifiers; actual spectral lines for samples from all 196 individuals are shown together with the Gaussian kernel. The levels of Gaussian kernel-based spectral features in samples from individual breast cancer patients and healthy controls were quantified and are shown as box-plots (Fig. 3D).

In summary, when either complete or albumin-depleted (i.e. membrane-filtered) serum samples were used for classification distinct spectral components built classifiers that differentiated cancer and control samples. Such differences could result from a different structure and/or composition of complete and albumin-depleted serum due to the depletion procedure, and might be an intrinsic feature of the spectrum generation process as well. One should expect that essential components of the low-molecular-weight part of the serum mass spectrum are “cargo” peptides normally carried by albumin, whose removal would result in their depletion and contribute to the lack or lowered levels of certain components in membrane-filtered samples. On the other hand, the presence of albumin in a sample apparently reduces the efficiency of ionization and detection of less abundant peptides [39], which would contribute to the appearance or increased levels of some components in membrane-filtered samples. Noteworthy, MALDI-ToF mass spectra of peptides present in both type of preparation could be used for identification of serum from patients with early stages of breast cancer, and that both types of sample could be used for classification with rather similar sensitivity and specificity. However, classification based on membrane-filtered albumin-depleted serum performs slightly better and relies on a less complex classifier, and thus pre-processing of human serum to remove albumin before MALD-ToF mass spectrometric analysis of the low-molecular-weight components could be recommended.

Acknowledgements**.** This work was supported by the Polish Ministry of Science, Grant 2P05E06730.

References

[1] HANASH S Disease proteomics. Nature 2003; 422: 226–232 doi:10.1038/nature01514

[2] WULFKUHLE JD, LIOTTA LA, PETRICOIN EF Proteomic applications for the early detection of cancer. Nature Rev Cancer 2003; 3: 267–275 doi:10.1038/nrc1043

[3] AEBERSOLD R, MANN M Mass spectrometry-based proteomics. Nature 2003; 422: 198–207 doi:10.1038/nature01511

[4] LIOTTA LA, FERRARI M, PETRICOIN EF Clinical proteomics: written in blood. Nature 2003; 425: 905 doi:10.1038/425905a

[5] ROSENBLATT KP, BRYANT-GREENWOOD P, KILLIAN JK, MEHTA A, GEHO D, et al. Serum proteomics in cancer diagnosis and management. Annu Rev Med 2004; 55: 97–112 doi:10.1146/annurev.med.55.091902.105237

[6] LIOTTA LA, PETRICOIN EF Serum peptidome for cancer detection: spinning biological trash into diagnostic gold. J Clin Invest 2006; 116: 26–30 doi:10.1172/JCI27467

[7] DWORZANSKI JP, SNYDER AP Classification and identification of bacteria using mass spectrometry-based proteomics. Expert Rev Proteomics 2005; 2: 863–878 doi:10.1586/14789450.2.6.863

[8] SOMORJAI RL Pattern recognition approaches for classifying proteomic mass spectra of biofluids. Methods Mol Biol 2008; 428: 383–396 doi:10.1007/978-1-59745-117-8_20

[9] LI L, TANG H, WU Z, GONG J, GRUIDL M, et al. Data mining techniques for cancer detection using serum proteomic profiling. Artif Intell Med 2004; 32: 71–83 doi:10.1016/j.artmed.2004.03.006

[10] HUTCHENS TW, YIP TT New desorption strategies for the mass spectrometric analysis of macromolecules. Rapid Commun Mass Spectrom 1993; 7: 576–580 doi:10.1002/rcm.1290070703

[11] PETRICOIN EF, LIOTTA LA SELDI-TOF-based serum proteomic pattern diagnostics for early detection of cancer. Curr Opin Biotech 2004; 15: 24–30 doi:10.1016/j.copbio.2004.01.005

[12] PETRICOIN EF, ARDEKANI AM, HITT BA, LEVINE PJ, FUSARO VA, et al. Use of proteomic patterns in serum to identify ovarian cancer. Lancet 2002; 359: 572–577 doi:10.1016/S0140-6736(02)07746-2

[13] RANSOHOFF DF Lessons from controversy: ovarian cancer screening and serum proteomics. J Natl Cancer Inst 2005; 97: 315–319 doi:10.1093/jnci/dji054

[14] POSADAS EM, SIMPKINS F, LIOTTA LA, MACDONALD C, KOHN EC Proteomic analysis for the early detection and rational treatment of cancer-realistic hope? Ann Oncol 2005; 16: 16–22 doi:10.1093/annonc/mdi004

[15] AZAD NS, RASOOL N, ANNUNZIATA CM, MINASIAN L, WHITELEY G, KOHN EC Proteomics in clinical trials and practice. Mol Cell Proteomics 2006; 5: 1819–1829 doi:10.1074/mcp.R600008-MCP200

[16] CHO WCS Contribution of oncoproteomics to cancer biomarker discovery. Mol Cancer 2007; 6: e25 doi:10.1186/1476-4598-6-25

[17] PALMBLAD M, TISS A, CRAMER R Mass spectrometry in clinical proteomics – from the present to the future. Proteomics Clin Appl 2009; 3: 6–17 doi:10.1002/prca.200800090

[18] CONRADS TP, HOOD BL, ISSAQ HJ, VEENSTRA TD Proteomic patterns as a diagnostic tool for early-stage cancer: a

review of its progress to a clinically relevant tool. Mol Diagn 2004; 8: 77–85 doi:10.2165/00066982-200408020-00001

[19] YANG SY, XIAO XY, ZHANG WG, ZHANG LJ, ZHANG W, et al. Application of serum SELDI proteomic patterns in diagnosis of lung cancer. BMC Cancer 2005; 5: e83 doi:10.1186/1471-2407-5-83

[20] LIU XP, SHEN J, LI ZF, YAN L, GU J A serum proteomic pattern for the detection of colorectal adenocarcinoma using surface enhanced laser desorption and ionization mass spectrometry. Cancer Invest 2006; 24: 747–753 doi:10.1080/07357900601063873

[21] LIN YW, LIN CY, LAI HC, CHIOU JY, CHANG CC, et al. Plasma proteomic pattern as biomarkers for ovarian cancer. Int J Gynecol Cancer 2006; 16 Suppl 1: 139–146 doi:10.1111/j.1525-1438.2006.00475.x

[22] LIM JY, CHO JY, PAIK YH, CHANG YS, KIM HG Diagnostic application of serum proteomic patterns in gastric cancer patients by ProteinChip surface-enhanced laser desorption/ionization time-of-flight mass spectrometry. Int J Biol Markers 2007; 22: 281–286

[23] MCPHERSON K, STEEL CM, DIXON JM Breast cancer - epidemiology, risk factors, and genetics. BrMed J 2000; 321: 624–628 doi:10.1136/bmj.321.7261.624

[24] ASTLEY SM Computer-based detection and prompting of mammographic abnormalities. Br J Radiol 2004; 77: S194-S200 doi:10.1259/bjr/30116822

[25] LI J, ZHANG Z, ROSENZWEIG J, WANG YY, CHAN DW Proteomics and bioinformatics approaches for identification of serum biomarkers to detect breast cancer. Clin Chem 2002; 48: 1296–1304

[26] VLAHOU A, LARONGA C, WILSON L, GREGORY B, FOURNIER K, et al. A novel approach toward development of a rapid blood test for breast cancer. Clin Breast Cancer 2003; 4: 203–209 doi:10.3816/CBC.2003.n.026

[27] LI J, ORLANDI R, WHITE CN, ROSENZWEIG J, ZHAO J, et al. Independent validation of candidate breast cancer serum biomarkers identified by mass spectrometry. Clin Chem 2005; 51: 2229–2235 doi:10.1373/clinchem.2005.052878

[28] VILLANUEVA J, SHAFFER DR, PHILIP J, CHAPARRO CA, ERDJUMENT-BROMAGE H, et al. Differential exoprotease activities confer tumor-specific serum peptidome patterns. J Clin Invest 2006; 116: 271–284 doi:10.1172/JCI26022

[29] BELLUCO C, PETRICOIN EF, MAMMANO E, FACCHIANO F, ROSS-RUCKER S, et al. Serum proteomic analysis identifies a highly sensitive and specific discriminatory pattern in stage 1 breast cancer. Ann Surg Oncol 2007; 4: 2470–2476 doi:10.1245/s10434-007-9354-3

[30] PIETROWSKA M, MARCZAK L, POLANSKA J, BEHRENDT K, NOWICKA E, et al. Mass spectrometry-based serum proteome pattern analysis in molecular diagnostics of early stage breast cancer. J Translat Med 2009; 7: e60 doi:10.1186/1479-5876-7-60

[31] PUSZTAI L, GREGORY BW, BAGGERLY KA, PENG B, KOOMEN J, et al. Pharmacoproteomic analysis of prechemotherapy and postchemotherapy plasma samples from patients receiving neoadjuvant or adjuvant chemotherapy for breast carcinoma. Cancer 2004; 100: 1814–1822 doi:10.1002/cncr.20203

[32] HEIKE Y, HOSOKAWA M, OSUMI S, FUJII D, AOGI K, et al. Identification of serum proteins related to adverse effects induced by docetaxel infusion from protein expression profiles of serum using SELDI ProteinChip system. Anticancer Res 2005; 25: 1197–1203

[33] GONCALVES A, ESTERNI B, BERTUCCI F, SAUVAN R, CHABANNON C, et al. Postoperative serum proteomic profiles may predict metastatic relapse in high-risk primary breast cancer patients receiving adjuvant chemotherapy. Oncogene 2006; 25: 981–989 doi:10.1038/sj.onc.1209131

[34] CALLESEN AK, VACH W, JØRGENSEN PE, COLD S, MOGENSEN O, et al. Reproducibility of mass spectrometry based protein profiles for diagnosis of breast cancer across clinical studies: a systematic review. J Proteome Res 2008; 7: 1395–1402 doi:10.1021/pr800115f

[35] KARPIEVITCH YV, HILL EG, SMOLKA AJ, MORRIS JS, COOMBES KR, et al. PrepMS: TOF MS Data Graphical Preprocessing Tool. Bioinformatics 2007; 23: 264–265

[36] HILARIO M, KALOUSIS A, PELLEGRINI C, MÜLLER M Processing and classification of protein mass spectra. Mass Spectrom Rev 2006; 25: 409–449 doi:10.1002/mas.20072

[37] HASTIE T, TIBSHIRANI R, FRIEDMAN JH The Elements of Statistical Learning. Springer Verlag, 2001

[38] ZWEIG MH, CAMPBELL G ROC plots: a fundamental evaluation tool in clinical medicine. Clin Chem 1993; 39: 561–577

[39] TIRUMALAI RS, CHAN KC, PRIETO DRA, ISSAQ HJ, CONRADS TP, VEENSTRA TD Characterization of the low molecular weight human serum proteome. Mol Cell Proteomics 2003; 2: 1096–1103 doi:10.1074/mcp.M300031-MCP200

[40] MEHTA AI, ROSS S, LOWENTHAL MS, FUSARO V, FISHMAN DA, et al. Biomarker amplification by serum carrier protein binding. Dis Markers 2003; 19: 1–10