Search for Breast Cancer Biomarkers in Fractionated Serum Samples by Protein Profiling With SELDI-TOF MS

Annemieke W.J. Opstal-van Winden,^{1,2} Jos H. Beijnen,^{2,3} Arnoud Loof,⁴
Waander L. van Heerde,⁴ Roel Vermeulen,^{1,5} Petra H.M. Peeters,^{1,6}
and Carla H. van Gils^{1*}

¹Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht,
Utrecht, The Netherlands

²Department of Pharmacy & Pharmacology, The Netherlands Cancer Institute/Slotervaart Hospital,
Amsterdam, The Netherlands

³Beta Faculty, Department of Pharmaceutical Sciences, Division of Biomedical Analysis, Section of Drug
Toxicology, Utrecht University, Utrecht, The Netherlands

⁴Central Laboratory for Haematology, Radboud University Nijmegen Medical Centre, Nijmegen,
The Netherlands

⁵Institute for Risk Assessment Sciences, Utrecht University, The Netherlands

⁶Department of Epidemiology, Public Health and Primary Care, Faculty of Medicine, Imperial College,
London, United Kingdom

Background: Many high-abundant acute phase reactants have been previously detected as potential breast cancer biomarkers. However, they are unlikely to be specific for breast cancer. Cancer-specific biomarkers are thought to be among the lower abundant proteins. **Methods:** We aimed to detect lower abundant discriminating proteins by performing serum fractionation by strong anion exchange chromatography preceding protein profiling with SELDI-TOF MS. In a pilot study, we tested the different fractions resulting from fractionation, on several array types. Fraction 3 on IMAC30 and Fraction 6 on Q10 yielded the most discriminative proteins and were used for serum protein profiling of 73 incident breast cancer cases and 73 matched controls. **Results:** Eight peaks showed statistically significantly different intensities between cases and controls ($P < 0.05$), and had less than

10% chance to be a false-positive finding. Seven of these were tentatively identified as apolipoprotein C-II (m/z 8,909), oxidized apolipoprotein C-II (m/z 8,925), apolipoprotein C-III (m/z 8,746), fragment of coagulation factor XIIIa (m/z 3,959), heterodimer of apolipoprotein A-I and apolipoprotein A-II (m/z 45,435), hemoglobin B-chain (m/z 15,915), and post-translational modified hemoglobin (m/z 15,346). **Conclusion:** By extensive serum fractionation, we detected many more proteins than in previous studies without fractionation. However, discriminating proteins were still high abundant. Results indicate that either lower abundant proteins are less distinctive, or more rigorous fractionation and selective protein depletion, or a more sensitive assay, are needed to detect lower abundant discriminative proteins. *J. Clin. Lab. Anal.* 26:1–9, 2012. © 2012 Wiley Periodicals, Inc.

Key words: breast cancer; proteomics; SELDI-TOF MS; fractionation; biomarker

INTRODUCTION

In the last few years, many proteomics studies using surface-enhanced laser desorption/ionization time of flight mass spectrometry (SELDI-TOF MS) have been

Additional Supporting Information may be found in the online version of this article.

*Correspondence to: Carla H. van Gils, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Mail-drop Str 6.131, P.O. Box 85500, 3508 GA Utrecht, The Netherlands. E-mail: C.vanGils@umcutrecht.nl

Received 20 July 2011; Accepted 31 August 2011

DOI 10.1002/jcla.20492

Published online in Wiley Online Library (wileyonlinelibrary.com).

Grant sponsors: University Medical Center Utrecht; Julius Center for Health Sciences and Primary Care ('Strategische Impuls'); Netherlands Laboratory for Anticancer Drug Formulation (NLADF), Amsterdam, The Netherlands.

carried out in search of diagnostic blood markers for several types of cancer (1–12). The discovery of such markers could enable the detection of tumors in early stages of the disease, with easy to perform and less-invasive blood test. Diagnosis in earlier stages gives patients higher chances of survival with possibly less invasive therapies.

However, the potential biomarkers for breast cancer, that have been detected with SELDI-TOF MS so far, such as apolipoprotein C-I, complement component 3a, fibrinogen, haptoglobin and inter- α -trypsin inhibitor heavy chain 4, are mainly high-abundant blood proteins involved in coagulation and acute phase responses (13). These responses are not likely to be specific for cancer, let alone for one type of cancer. This limits the usefulness of these markers as breast-cancer-specific markers.

The fact that primarily acute phase reactants are found is most likely due to the complexity of the serum proteome. The serum proteome contains a large number of proteins spanning a wide dynamic range of concentrations ($>10^{10}$) (14). Only a few, high-abundant proteins comprise about 99% of the total amount of proteins (15). Acute phase reactants are examples of such high-abundant proteins, and are thus easily detected. Cancer-specific proteins that are exclusively expressed by one type of malignant cells are expected to be much less abundant (13, 14). Using SELDI-TOF MS, these proteins are largely “masked” by high-abundant proteins, since it only has the capacity to detect proteins in a concentration range of 10^2 (14).

Serum fractionation has been proposed as a promising method to measure low-abundant proteins (16–19). In this study, we performed fractionation by anion-exchange chromatography to detect low-abundant proteins related with breast cancer. Based on differences in isoelectric point (pI) of the proteins, we divided a subjects’ serum proteome into six protein fractions. High-abundant proteins are segregated into a limited number of fractions, which reduces the signal suppression effects on proteins of lower abundance in the other fractions. This facilitates the detection of more and/or low-abundant proteins (14). We did not perform depletion of the most-abundant proteins since this may cause elimination of lower abundant proteins that are bound to the depleted proteins (20). In this study, we compared the protein profile in fractionated serum samples of incident breast cancer patients with that of healthy controls with the aim to detect proteins, other than high-abundant acute phase reactants, that differentiate between these two groups.

MATERIAL AND METHODS

Study Population

To investigate serum protein profiles, we performed a case-control study. Serum samples of both the cases and

the controls were obtained from a serum bank at The Netherlands Cancer Institute (NKI), Amsterdam, The Netherlands. Samples were collected from March 2003 until July 2005, from women who were just diagnosed with primary breast cancer, and from healthy female relatives or friends of the patients. Serum samples of the cases were collected before surgery or the start of any other kind of treatment. All samples were collected after receiving the individuals’ informed consent, under approval of the Institutional Review Board control.

Blood collection, processing, and storage of the serum samples were performed under strictly defined conditions, which were the same for cases and controls (see Supplementary material for details).

Menopausal status at diagnosis was obtained through examination of the cases’ medical records. Tumor stage, tumor size, estrogen receptor status, progesterone receptor status, HER2/neu expression, and p53 expression were determined by pathological examination of the removed tumor. Lymph node involvement and the presence of metastasis were also examined.

From a group of 157 women with primary diagnosed breast cancer and 131 healthy controls, we matched cases and controls for age and serum sample storage duration. Finally, 73 cases and 73 controls could be matched with a maximum age difference of 3 years and a maximum difference in sample storage duration of three months.

Pilot Study for Selection of Experimental Conditions

In this study, we carried out serum fractionation before protein profiling with SELDI-TOF MS. To obtain the most informative testing conditions, different combinations of fraction and array type were tested in a pilot study with nine randomly selected case-control pairs. Serum fractionation was performed with strong anion exchange Q ceramic resin (Bio-Rad Labs, Hercules, CA) according to the manufacturers’ protocol (see Supplementary material), which resulted in six fractions.

These six fractions were applied to three different array types with an appropriate binding buffer to test these combinations. First, we tested the Immobilized Metal Affinity Capture (IMAC30) array (Bio-Rad Labs) charged with copper sulfate (Merck, Darmstadt, Germany) with a binding buffer containing 0.01 M phosphate-buffered saline pH 7.4 (Sigma, St Louis, MO) with 0.5M sodium chloride (Merck). Furthermore, we tested the weak cation exchange (CM10) array (Bio-Rad Labs) with a binding buffer containing 100 mM sodium acetate pH 4 (Sigma). Finally, we tested the strong anion exchange (Q10) array (Bio-Rad Labs) with a binding buffer containing 20 mM Tris-HCl pH 9 (Sigma).

Ultimately, the IMAC30 array with Fraction 3 and the Q10 array with Fraction 6 were the two conditions, which yielded the most discriminative proteins relative to the total number of detected peaks, mainly in the mass range of 2–10 kDa.

Profiling of Fractionated Serum Samples of Cases and Controls

Selected conditions were subsequently used for the analysis of the total sample set. The samples were fractionated and applied to the ProteinChip arrays in three batches, on three consecutive days, with a Biomek pipetting robot (Beckman Coulter, Brea, CA). For application of the serum samples to the arrays, we used the same protocol as in the pilot study (see Supplementary material). Samples of matched cases and controls were analyzed in the same batch. In every batch also two aliquots of two quality control (QC) samples were analyzed. All samples were randomly applied to the well plates. On the fourth day, we performed SELDI-TOF MS on all arrays with the PCS 4000 ProteinChip Reader (Bio-Rad Labs). See Supplementary material for the settings of the ProteinChip Reader for protein profiling, for method of processing the spectra, and for settings of peak detection. Spectra in which normalization revealed too low or too high total ion current (TIC) were excluded from further analysis.

To estimate the reproducibility of the analysis, we calculated the within-batch variation and the between-batch variation in the QC samples. To determine the within-batch variation, we first averaged the peak intensities measured in the duplicates of each QC sample, per batch. Then, we calculated the median coefficient of variance (CV) (the SD as percentage of average) of all peaks detected on one type of array, within every batch. To determine the between-batch variation, we averaged the averaged peak intensities of each QC sample in the three batches for every peak. Subsequently, we calculated the median CV of all peaks detected on one type of array, between batches.

Data Analysis

Peak information was subsequently exported as CSV-files and imported into SPSS 15.0 for statistical analysis. Data analysis was performed separately for the IMAC30 peaks and the Q10 peaks. The sera were fractionated and applied to the arrays on three consecutive days, a parameter likely to influence spectral data (21–23). Therefore, before merging peak intensity data of the three batches, peak intensities were transformed into *Z*-values within each batch (histograms showed normally distributed peak intensities within the batches). *Z*-values expressed the intensities as the number of standard deviations above or

below the mean intensity of that peak across all samples in a batch. Differences in mean *Z*-transformed peak intensities between breast cancer cases and matched healthy controls were tested with a paired samples *T* test. *P*-values <0.05 were considered statistically significant.

Correction for multiple testing was performed on all detected IMAC30 and Q10 peaks together, using the False Discovery Rate (FDR) method suggested by Benjamini and Hochberg (24). The FDR controls the expected proportion of falsely rejected hypotheses. We choose 10% as an acceptable proportion of false-positive results (*q*-value = 0.10) (24).

We also performed a conditional multivariate logistic regression analysis in which we simultaneously included the peaks that had less than 10% chance to be a false-positive finding. We performed backward selection (*P*-value <0.20) to determine which peaks statistically significantly contributed to the discrimination of cases and controls. We subsequently determined the area under the curve (AUC) of the Receiver Operating Characteristic (ROC) curve based on the predicted probabilities resulting from the model, with 95% confidence interval (CI). We executed this analysis to determine which peaks were independently related to breast cancer, and to find a combination of peaks that could optimally distinguish breast cancer cases from healthy controls.

Identification of the Most Discriminative Peaks

Based on their mass-to-charge ratio (*m/z*), the sample type (serum) and fraction in which they were found, the ProteinChip surface used, as well as data from previously performed serum profiling studies, we tentatively identified the most discriminative proteins.

RESULTS

Study Population

Table 1 gives an overview of the characteristics of the breast cancer cases, the matched controls and their serum samples. The median age of both the cases and the controls was 55 years at the time of blood collection. The menopausal status at diagnosis of two women was not reported in their medical record. Information on menopausal status was not available for the controls.

Tumor characteristics of the breast cancer cases are listed in Table 2. Most of the patients were diagnosed with Stage I (23%) or Stage IIA (44%) breast cancer. Six percent of the patients were diagnosed with a carcinoma in situ. More than half of the patients with an invasive tumor had lymph node involvement, but none of the patients was affected with metastases.

TABLE 1. Characteristics of the Breast Cancer Cases, the Healthy Controls, and Their Serum Samples

	Breast cancer cases (<i>n</i> = 73)	Healthy controls (<i>n</i> = 73)
Age at diagnosis (years)		
Median (IQR)	55 (46–60.5)	55 (44.5–60)
Menopausal status, <i>n</i> (%)		
Premenopausal	28 (39.4)	
Postmenopausal	43 (60.6)	
Missing	2	
Sample storage duration (months)		
Median (IQR)	52 (42–59)	52 (42–59)
Time from diagnosis to blood sampling (days)		
Median (IQR)	15 (6–22)	

IQR, interquartile range.

TABLE 2. Tumor Characteristics of the Breast Cancer Cases

All breast cancer cases	<i>n</i> = 73
TNM stage, <i>n</i> (%)	
0	4 (5.5)
I	17 (23.3)
IIA	32 (43.8)
IIB	9 (12.3)
IIIA and IIIC	11 (15.1)
Breast cancer cases with an invasive tumor	<i>n</i> = 69
Tumor size, <i>n</i> (%)	
>0.1–0.5 cm	4 (5.8)
>0.5–1 cm	6 (8.7)
>1–2 cm	29 (42.0)
>2cm	30 (43.5)
Lymph node involvement, <i>n</i> (%)	
No	30 (43.5)
Yes	39 (56.5)
ER status, <i>n</i> (%)	
Negative	16 (23.2)
Positive	53 (76.8)
PR status, <i>n</i> (%)	
Negative	30 (43.5)
Positive	39 (56.5)
HER2/neu expression, <i>n</i> (%)	
Negative	55 (79.7)
Positive	14 (20.3)
P53 expression, <i>n</i> (%)	
Negative	23 (33.3)
Positive	46 (66.7)

ER, estrogen receptor; PR, progesterone receptor.

Peak Detection

After normalization, 22 of the 146 spectra (73 cases and 73 controls) resulting from the analysis using the IMAC arrays showed divergent total ion current (9 control-spectra and 13 case-spectra, belonging to 17 case-control pairs). Twenty of the 146 spectra resulting from the analysis using the Q10 arrays showed divergent total ion current (8 control-spectra and 12 case-spectra, belonging to

16 case-control pairs). All these case-control pairs were excluded from the paired analyses.

Hundred and twenty-nine peaks were detected in the spectra resulting from the IMAC-analysis (84 in the 2- to 12-kDa mass range and 45 in the 12- to 300-kDa mass range). Hundred and fifty peaks were detected in the spectra resulting from the Q10-analysis (83 in the 2- to 12-kDa mass range and 67 in the 12- to 300-kDa mass range). The within-batch reproducibility for all IMAC-peaks was 15, 24, and 15% for Batches 1, 2, and 3, respectively. For all Q10-peaks, the within-batch reproducibility was 11, 14, and 15%, respectively. The between-batch reproducibility was 32% for the IMAC-peaks and 18% for the Q10-peaks.

Relations Between Peak Intensities and Breast Cancer

Based on the paired samples *T* test, the intensities of 16 of the 129 IMAC-peaks were found to be significantly different between cases and controls (*P*-value <0.05). Five of these peaks were detected in the 2- to 12-kDa mass range. Fourteen of the 16 peaks were lower in cases than controls. After correction for multiple testing, only two peaks (*m/z* 15,915 and *m/z* 15,346) had less than 10% chance to be a false-positive finding. The *m/z* of the 16 peaks, their mean *z*-transformed intensities, the results of the *T* test, and the FDR thresholds are listed in Table 3.

Table 4 shows that the intensities of 29 of the 150 Q10-peaks were significantly different between cases and controls. Fifteen of these peaks were detected in the 2- to 12-kDa mass range. Sixteen of the 29 peaks were lower in cases than controls. After FDR correction, six peaks appeared to have less than 10% chance to be a false-positive finding (*m/z* 8,926, *m/z* 8,909, *m/z* 4,162, *m/z* 8,746, *m/z* 45,435, and *m/z* 3,959).

When we performed a paired samples *T* test in which the sets that were used for the pilot study were excluded, the same peaks were in the top of the ranking.

Conditional Multivariate Logistic Regression Analysis

The multivariate analysis, including all peaks that had less than 10% chance to be false-positive result, revealed that *m/z* 3,959, *m/z* 4,162, *m/z* 8,909, and *m/z* 15,915 significantly contributed to the distinction between breast cancer cases and healthy controls. A ROC curve of the predicted probabilities for breast cancer, based on the intensities of these peaks, resulted in an AUC of 0.77 (95%CI: 0.69–0.86).

Proposed Identities of the Peaks

Based on their *m/z*, the sample type and fraction in which they were found, the ProteinChip surface used,

TABLE 3. Mean Z-Transformed Intensities of the Discriminative Peaks Detected on IMAC Arrays in Cases and Controls in Order of Significance

IMAC	Breast cancer cases (<i>n</i> = 56)		Healthy controls (<i>n</i> = 56)		Intensity in cases vs. controls	<i>P</i> -value ^b	FDR threshold	Correlated ^d	Tentative identity		Molecular weight (Da)
	Mean intensity ^a (SD)		Mean intensity ^a (SD)						Protein		
<i>m/z</i>											
15,915	− 0.31 (0.69)		0.33 (1.21)		Lower	0.0013 ^c	0.0022	B	Haemoglobin β-chain		15,867
15,346	− 0.28 (0.70)		0.30 (1.22)		Lower	0.0035 ^c	0.0036	B	Post-translational modified haemoglobin		
15,143	− 0.26 (0.60)		0.30 (1.27)		Lower	0.0043	0.0039	B	Haemoglobin α-chain		15,126
6,923	− 0.24 (0.98)		0.26 (0.79)		Lower	0.0044	0.0043				
94,790	− 0.27 (0.84)		0.27 (1.07)		Lower	0.0051	0.0047	C	Albumin/Apo A-I heterodimer		94,500
28,269	− 0.27 (0.86)		0.25 (1.07)		Lower	0.0081	0.0057	C	Apo A-I +..		
29,114	− 0.25 (0.68)		0.28 (1.22)		Lower	0.0082	0.0061				
55,991	− 0.25 (0.89)		0.24 (1.06)		Lower	0.0173	0.0082				
28,100	− 0.24 (0.85)		0.23 (1.10)		Lower	0.0178	0.0086	C	Apo A-I		28,080
14,061	− 0.24 (0.82)		0.21 (1.10)		Lower	0.0212	0.0104	C	Apo A-I 2+		14,040
3,095	0.20 (1.05)		− 0.21 (0.89)		Higher	0.0259	0.0115	A	Albumine fragment + oxide atom		3,099
5,856	− 0.21 (1.05)		0.21 (0.86)		Lower	0.0290	0.0129				
47,482	− 0.24 (0.81)		0.15 (1.13)		Lower	0.0292	0.0133	C	Albumin/Apo A-I heterodimer 2+		47,250
3,079	0.17 (1.09)		− 0.20 (0.88)		Higher	0.0383	0.0140	A	Albumine fragment		3,083
14,163	− 0.21 (0.85)		0.20 (1.07)		Lower	0.0390	0.0143	C	<i>m/z</i> 28,269 2+		
6,414	− 0.19 (0.87)		0.14 (1.04)		Lower	0.0416	0.0151				

m/z, mass-to-charge ratio; SD, standard deviation; 2+, Double charged ion.

^aZ-transformed peak intensities.

^bPaired samples *T* test.

^cLess than 10% chance to be a false positive finding after FDR correction (*P*-value is lower than FDR threshold, which is the rank of the *P*-value (from smallest to largest) divided by the total number of tested peaks, multiplied by the *q*-value (0.10)).

^dPeaks of which transformed intensities were correlated (A: *R*² = 0.886; B: *R*² > 0.887; C: *R*² > 0.729 (except for *m/z* 47,482: *R*² > 0.652)).

TABLE 4. Mean Z-Transformed Intensities of the Discriminative Peaks Detected on Q10 Arrays in Cases and Controls in Order of Significance

Q10	Breast cancer cases ($n = 57$)		Healthy controls ($n = 57$)		Intensity in cases vs. controls	P -value ^b	FDR threshold	Correlated ^d	Tentative identity		Molecular weight (Da)
	Mean intensity ^a (SD)		Mean intensity ^a (SD)						Protein		
8,926	-0.29 (0.83)		0.28 (1.10)	Lower	0.0003 ^c	0.0004	A		Apo C-II + oxide atom		8,930
8,909	-0.30 (0.62)		0.30 (1.22)	Lower	0.0005 ^c	0.0007	A		Apo C-II		8,914
4,162	0.31 (0.96)		-0.26 (0.88)	Higher	0.0007 ^c	0.0011					
8,746	0.28 (1.07)		-0.29 (0.90)	Higher	0.0008 ^c	0.0014			Apo C-III		8,765
45,435	-0.29 (0.95)		0.31 (0.98)	Lower	0.0010 ^c	0.0018			Apo A-I/Apo A-II heterodimer		45,470
3,959	0.30 (1.18)		-0.24 (0.72)	Higher	0.0015 ^c	0.0025			Factor XHIIa fragment		3,951
28,094	-0.27 (1.06)		0.24 (0.91)	Lower	0.0029	0.0029	C		Apo A-I		28,080
14,058	-0.25 (1.05)		0.23 (0.87)	Lower	0.0033	0.0032	C		Apo A-I 2+		
8,819	0.17 (1.04)		-0.25 (0.76)	Higher	0.0055	0.0050			Apo A-II		8,810
14,158	-0.23 (1.08)		0.21 (0.87)	Lower	0.0078	0.0054	C		m/z 28303 2+		
7,608	0.22 (1.05)		-0.22 (0.77)	Higher	0.0110	0.0065			Apo L-I		7,616
8,191	-0.22 (0.64)		0.22 (1.24)	Lower	0.0112	0.0068	A		Apo C-II truncated		8,204
56,202	-0.23 (1.02)		0.20 (0.95)	Lower	0.0116	0.0072	C		Apo A-I dimer		56,160
28,303	-0.24 (1.06)		0.21 (0.93)	Lower	0.0119	0.0075	C		Apo A-I + ...		
14,261	-0.22 (1.06)		0.22 (0.92)	Lower	0.0138	0.0079	C				
79,213	0.21 (1.06)		-0.18 (0.96)	Higher	0.0184	0.0090			Serotransferrin		79,000
84,751	-0.20 (0.97)		0.18 (0.99)	Lower	0.0186	0.0093					
9,430	-0.22 (0.94)		0.18 (1.05)	Lower	0.0186	0.0097			Apo C-III glycosylated		9,420
124,174	0.24 (0.91)		-0.17 (0.98)	Higher	0.0194	0.0100					
2,008	0.21 (1.03)		-0.24 (0.97)	Higher	0.0237	0.0108					
13,092	-0.22 (0.94)		0.17 (0.96)	Lower	0.0250	0.0111					
33,494	0.20 (1.12)		-0.15 (0.87)	Higher	0.0268	0.0118					
116,138	0.21 (0.94)		-0.21 (1.01)	Higher	0.0278	0.0122					
5,371	0.21 (1.14)		-0.16 (0.85)	Higher	0.0286	0.0125					
7,812	0.16 (0.95)		-0.18 (0.86)	Higher	0.0301	0.0136					
8,209	-0.18 (0.87)		0.17 (1.12)	Lower	0.0394	0.0147	A		Apo C-II truncated + oxide atom		8,220
6,428	-0.18 (1.05)		0.15 (0.95)	Lower	0.0475	0.0154	B		Apo C-I truncated		6,432
23,741	0.13 (1.04)		-0.22 (0.94)	Higher	0.0488	0.0158					
6,622	-0.19 (1.05)		0.13 (0.96)	Lower	0.0494	0.0161	B		Apo C-I		6,630

 m/z , mass-to-charge ratio; SD, standard deviation; 2+, double-charged ion.^aZ-transformed peak intensities.^bPaired samples T test.^cLess than 10% chance to be a false-positive finding after FDR correction (P -value is lower than FDR threshold, which is the rank of the P -value (from smallest to largest) divided by the total number of tested peaks, multiplied by the q -value (0.10)).^dPeaks of which transformed intensities were correlated (A: $R^2 > 0.706$; B: $R^2 > 0.884$; C: $R^2 > 0.702$).

as well as data from previously performed serum profiling studies (N. Harris, unpublished data, and (25–30)), we tentatively identified the most discriminative peaks. The peak with m/z 15,915 that was detected in Fraction 3 on the IMAC30 array is very likely hemoglobin β -chain. Peaks with similar mass were previously structurally identified as hemoglobin β -chain by SDS-PAGE, followed by in-gel trypsin digestion and analysis using tandem MS (e.g., Q-TOF), and/or by an immunoassay (25–29). The molecular weight (MW) of this protein is 15,867 Da and it has a pI of 6.81. This pI fits with the assumption that this peak, detected in Fraction 3, is hemoglobin β -chain, because due to the nature of the method of fractionation, Fraction 3 can only contain proteins with a pI of >5 and <7 . The Z-transformed intensities of m/z 15,915 were highly correlated with the Z-transformed intensities of the peak with m/z 15,346 ($R^2 = 0.887$), which was also found in Fraction 3 on IMAC. This peak is therefore likely a post-translational modified form of hemoglobin. The tentative identities of these two peaks and the other identified IMAC-peaks are listed in Table 3.

The peaks with m/z 8,909 and m/z 8,925, which were both detected in Fraction 6 on the Q10 array, are likely apolipoprotein C-II and its oxidized form. The MW of apolipoprotein C-II is 8,914 Da and its pI is 4.66. The difference in m/z of these two peaks is the exact mass of an oxide-atom (16 Da). The peak intensities of these peaks were also highly correlated ($R^2 = 0.875$), which is to be expected if one is an oxidized form of the other. Another peak that was detected in the same fraction, m/z 8,746, is likely apolipoprotein C-III (MW: 8,765 Da), which has a pI of 4.72. The peak with m/z 3,959, also found in this fraction, is likely a fragment of coagulation factor XIIIa (30). This fragment has a MW of 3,951 Da and a pI of 4.03. A peak with m/z 3,950 was previously structurally identified as a fragment of coagulation factor XIIIa using nanoelectrospray ionization quadrupole time-of-flight mass spectrometry (nESI-qTOF-MS) (30). The peak with m/z 45,435 is likely a heterodimer of apolipoprotein A-I and apolipoprotein A-II. This dimer has a MW of 45,470 Da. The tentative identities of these peaks and the other identified Q10-peaks are listed in Table 4. The fraction in which the Q10-peaks were found is not informative for the identification. This is because Fraction 6 results from the elution with an organic buffer, which elutes all remaining proteins. By this, also strongly bound or high-abundant proteins with pI above 3, which should have been eluted in previous fractions, can end up in Fraction 6. We have no indications for the identity of the last discriminating peak that was detected on Q10 (m/z of 4,162).

DISCUSSION

In this study, we detected eight peaks that were statistically significantly related to the presence of breast cancer with less than 10% chance of being a false-positive finding. The tentative identities of seven of these peaks are: apolipoprotein C-II, an oxidized form of apolipoprotein C-II, apolipoprotein C-III, fragment of coagulation factor XIIIa, a heterodimer of apolipoprotein A-I and apolipoprotein A-II, hemoglobin β -chain, and post-translational modified hemoglobin. A peak with m/z 4,162 could not be identified. Apolipoprotein C-II, fragment of coagulation factor XIIIa, hemoglobin β -chain, and m/z 4,162 appeared to contribute significantly to the discrimination of cases and controls.

Apolipoprotein C-II and apolipoprotein C-III were not previously described in relation to the presence of a primary breast cancer tumor. In this study, a fragment of coagulation factor XIIIa was higher in breast cancer patients. Jiang et al. (31) found coagulation factor XIII itself to be lower in breast tumor tissue compared with normal breast tissue (31). Contrary, another fragment of coagulation factor XIIIa (m/z 2,602) was again higher in serum of breast cancer patients in another study (32). Differences in regulation may be related to the form of the protein; elevated fragment concentrations are not necessarily the consequence of higher precursors as argued in a study by Villanueva et al. (32). Additionally, differences may be due to differences in type of sample that was investigated; tissue vs. serum. Such a difference was also found by Engwegen et al. (11); a specific protein was lower in serum samples of colorectal cancer patients than in controls, while the same protein was higher in colorectal cancer tissue compared with healthy colon tissue of the same subjects (hyperplastic polyps) (11).

The heterodimer of apolipoprotein A-I and apolipoprotein A-II was not previously reported in relation to the presence of a primary breast tumor. In this study, hemoglobin β -chain was lower in breast cancer patients. A peak with a mass likely representing hemoglobin β -chain (15,940 Da), was previously found to be higher in nipple aspirate fluid (NAF) of breast cancer patients compared with NAF of controls (33). This study comprised only 20 breast cancer patients and 13 controls, but the discriminative value of this protein was very high (expressed in 16 cases and only 1 control) (33). Again, difference in type of body fluid as well as differences in the subject characteristics between the studies could have caused this difference in expression.

The discriminative proteins found in this study are involved in lipid metabolism (apolipoprotein C-II, A-I, and A-II), blood coagulation (coagulation factor XIIIa) and

oxygen transport (hemoglobin (β -chain), all processes that do not seem to be cancer specific. Moreover, these proteins are high abundant. Even though it is found by Villanueva et al. (32) that fragments of high abundant, acute phase reactants are cancer specific, the substrates itself are unlikely to be good cancer biomarkers. The hypothesized underlying process, namely the fragmentation by exoproteases released by the tumor, is probably best reflected by the concentration of the end products, instead of by the concentration of the substrates. Unfortunately, the SELDI-TOF MS technique is not sensitive enough to be able to detect these low-mass fragments. Therefore, the usefulness of the detected discriminative proteins for the diagnosis of breast cancer is doubtful as they should be able to discriminate between breast cancer and other types of cancer.

Despite the extensive fractionation of serum samples in this study, we did not detect the expected low-abundant discriminative proteins. Serum fractionation by strong anion exchange chromatography has previously been performed in only a few studies, mainly to increase the number of detectable peaks (25, 34–36), which was achieved in three of the studies (25, 34, 36). Solassol et al. (36) found that prostate-specific antigen, which is a low-abundant protein, only could be detected in fractionated serum. However, they tested this only with two-dimensional gel electrophoresis and not with SELDI-TOF MS (36).

One explanation for our findings could be that the low-abundant proteins are not as distinctive as thought, and that they were therefore not found. Another explanation could be that despite the extensive fractionation, we were not able to detect the low-abundant proteins after all. Possibly, the relatively high storage temperature of our samples (-30°C) has led to degradation of the serum proteins during the 4- to 6-year storage period (37). Concentrations of the already low-abundant proteins would then have been decreased to undetectable levels for SELDI-TOF MS. A third reason could be that despite extensive fractionation, too many abundant proteins were left, which suppressed the signal of low-abundant proteins. The fact that we detected a threefold increased number of peaks in our spectra compared with those in a previous “unfractionated” SELDI-TOF MS study by our group (4), suggests that we were able to eliminate the most-abundant proteins from the investigated fractions.

However, it may not have been sufficient to detect the least abundant, possibly highly discriminative proteins. More thorough removal of the highest abundant proteins may be needed. Immunodepletion of the top six most-abundant proteins (albumin, IgG, IgA, transferrin, haptoglobin, and α -1-antitrypsin) could already remove 83% of the total protein content (16), but this may at the same time cause elimination of discriminative low-abundant

proteins that are bound to the highly abundant ones (20). Other methods of serum fractionation, which separate the most-abundant proteins from the low-abundant proteins, are therefore needed to reduce the high dynamic range of serum protein concentrations and to enable the detection of low abundant, possibly discriminative proteins.

ACKNOWLEDGMENTS

We acknowledge Leo Kruijt from the Animal Science Group Lelystad for his hospitality at his laboratory and for the possibility to make use of the PCS 4000 ProteinChip Reader.

REFERENCES

1. Belluco C, Petricoin EF, Mammano E, et al. Serum proteomic analysis identifies a highly sensitive and specific discriminatory pattern in stage I breast cancer. *Ann Surg Oncol* 2007;14:2470–2476.
2. Fung ET, Yip TT, Lomas L, et al. Classification of cancer types by measuring variants of host response proteins using SELDI serum assays. *Int J Cancer* 2005;115:783–789.
3. 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.
4. van Winden AW, Gast MC, Beijnen JH, et al. Validation of previously identified serum biomarkers for breast cancer with SELDI-TOF MS: A case control study. *BMC Med Genomics* 2009;2:4.
5. Li J, Orlandi R, White CN, et al. Independent validation of candidate breast cancer serum biomarkers identified by mass spectrometry. *Clin Chem* 2005;51:2229–2235.
6. Mathelin C, Cromer A, Wendling C, Tomasetto C, Rio MC. Serum biomarkers for detection of breast cancers: A prospective study. *Breast Cancer Res Treat* 2006;96:83–90.
7. Song J, Patel M, Rosenzweig CN, et al. Quantification of fragments of human serum inter-alpha-trypsin inhibitor heavy chain 4 by a surface-enhanced laser desorption/ionization-based immunoassay. *Clin Chem* 2006;52:1045–1053.
8. Hu Y, Zhang S, Yu J, Liu J, Zheng S. SELDI-TOF-MS: The proteomics and bioinformatics approaches in the diagnosis of breast cancer. *Breast* 2005;14:250–255.
9. Vlahou A, Laronga C, Wilson L, et al. A novel approach toward development of a rapid blood test for breast cancer. *Clin Breast Cancer* 2003;4:203–209.
10. Engwegen JY, Helgason HH, Cats A, et al. Identification of serum proteins discriminating colorectal cancer patients and healthy controls using surface-enhanced laser desorption ionisation-time of flight mass spectrometry. *World J Gastroenterol* 2006;12:1536–1544.
11. Engwegen JY, Depla AC, Smits ME, et al. Detection of colorectal cancer by serum and tissue protein profiling: A prospective study in a population at risk. *Biomark Insights* 2008;3:375–385.
12. Engwegen JY, Mehra N, Haanen JB, Schellens JH, Voest EE, Beijnen JH. Identification of two new serum protein profiles for renal cell carcinoma. *Oncol Rep* 2009;22:401–408.
13. Gast MC, Schellens JH, Beijnen JH. Clinical proteomics in breast cancer: A review. *Breast Cancer Res Treat* 2009;116:17–29.
14. Hoffman SA, Joo WA, Echan LA, Speicher DW. Higher dimensional (Hi-D) separation strategies dramatically improve the potential for cancer biomarker detection in serum and plasma. *J Chromatogr B Anal Technol Biomed Life Sci* 2007;849:43–52.

15. Anderson NL, Anderson NG. The human plasma proteome: History, character, and diagnostic prospects. *Mol Cell Proteomics* 2002;1:845–867.
16. Faca V, Pitteri SJ, Newcomb L, et al. Contribution of protein fractionation to depth of analysis of the serum and plasma proteomes. *J Proteome Res* 2007;6:3558–3565.
17. Vermeulen R, Lan Q, Zhang L, et al. Decreased levels of CXCL chemokines in serum of benzene-exposed workers identified by array-based proteomics. *Proc Natl Acad Sci USA* 2005;102:17041–17046.
18. Restuccia U, Boschetti E, Fasoli E, et al. pI-based fractionation of serum proteomes versus anion exchange after enhancement of low-abundance proteins by means of peptide libraries. *J Proteomics* 2009;72:1061–1070.
19. Sahab ZJ, Iczkowski KA, Sang QX. Anion exchange fractionation of serum proteins versus albumin elimination. *Anal Biochem* 2007;368:24–32.
20. Yocum AK, Yu K, Oe T, Blair IA. Effect of immunoaffinity depletion of human serum during proteomic investigations. *J Proteome Res* 2005;4:1722–1731.
21. Karsan A, Eigel BJ, Flibotte S, et al. Analytical and preanalytical biases in serum proteomic pattern analysis for breast cancer diagnosis. *Clin Chem* 2005;51:1525–1528.
22. Pelikan R, Bigbee WL, Malehorn D, Lyons-Weiler J, Hauskrecht M. Inter-session reproducibility of mass spectrometry profiles and its effect on accuracy of multivariate classification models. *Bioinformatics* 2007;23:3065–3072.
23. Hu J, Coombes KR, Morris JS, Baggerly KA. The importance of experimental design in proteomic mass spectrometry experiments: Some cautionary tales. *Brief Funct Genomic Proteomic* 2005;3:322–331.
24. Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J R Stat Soc B* 1995;57:289–300.
25. Timms JF, Arslan-Low E, Gentry-Maharaj A, et al. Preanalytic influence of sample handling on SELDI-TOF serum protein profiles. *Clin Chem* 2007;53:645–656.
26. Trak-Smayra V, Dargere D, Noun R, et al. Serum proteomic profiling of obese patients: Correlation with liver pathology and evolution after bariatric surgery. *Gut* 2009;58:825–832.
27. Kozak KR, Su F, Whitelegge JP, Faull K, Reddy S, Farias-Eisner R. Characterization of serum biomarkers for detection of early stage ovarian cancer. *Proteomics* 2005;5:4589–4596.
28. Sundsten T, Zethelius B, Berne C, Bergsten P. Plasma proteome changes in subjects with Type 2 diabetes mellitus with a low or high early insulin response. *Clin Sci (Lond)* 2008;114:499–507.
29. Woong-Shick A, Sung-Pil P, Su-Mi B, et al. Identification of hemoglobin-alpha and -beta subunits as potential serum biomarkers for the diagnosis and prognosis of ovarian cancer. *Cancer Sci* 2005;96:197–201.
30. Tammen H, Mohring T, Kellmann M, Pich A, Kreipe HH, Hess R. Mass spectrometric phenotyping of Val34Leu polymorphism of blood coagulation factor XIII by differential peptide display. *Clin Chem* 2004;50:545–551.
31. Jiang WG, Ablin R, Douglas-Jones A, Mansel RE. Expression of transglutaminases in human breast cancer and their possible clinical significance. *Oncol Rep* 2003;10:2039–2044.
32. Villanueva J, Shaffer DR, Philip J, et al. Differential exoprotease activities confer tumor-specific serum peptidome patterns. *J Clin Invest* 2006;116:271–284.
33. Sauter ER, Zhu W, Fan XJ, Wassell RP, Chervoneva I, Du Bois GC. Proteomic analysis of nipple aspirate fluid to detect biologic markers of breast cancer. *Br J Cancer* 2002;86:1440–1443.
34. Rollin D, Whistler T, Vernon SD. Laboratory methods to improve SELDI peak detection and quantitation. *Proteome Sci* 2007;5:9.
35. Harezlak J, Wang M, Christiani D, Lin X. Quantitative quality-assessment techniques to compare fractionation and depletion methods in SELDI-TOF mass spectrometry experiments. *Bioinformatics* 2007;23:2441–2448.
36. Solassol J, Marin P, Demetree E, et al. Proteomic detection of prostate-specific antigen using a serum fractionation procedure: Potential implication for new low-abundance cancer biomarkers detection. *Anal Biochem* 2005;338:26–31.
37. Gast MC, van Gils CH, Wessels LF, et al. Influence of sample storage duration on serum protein profiles assessed by surface-enhanced laser desorption/ionisation time-of-flight mass spectrometry (SELDI-TOF MS). *Clin Chem Lab Med* 2009;47:694–705.