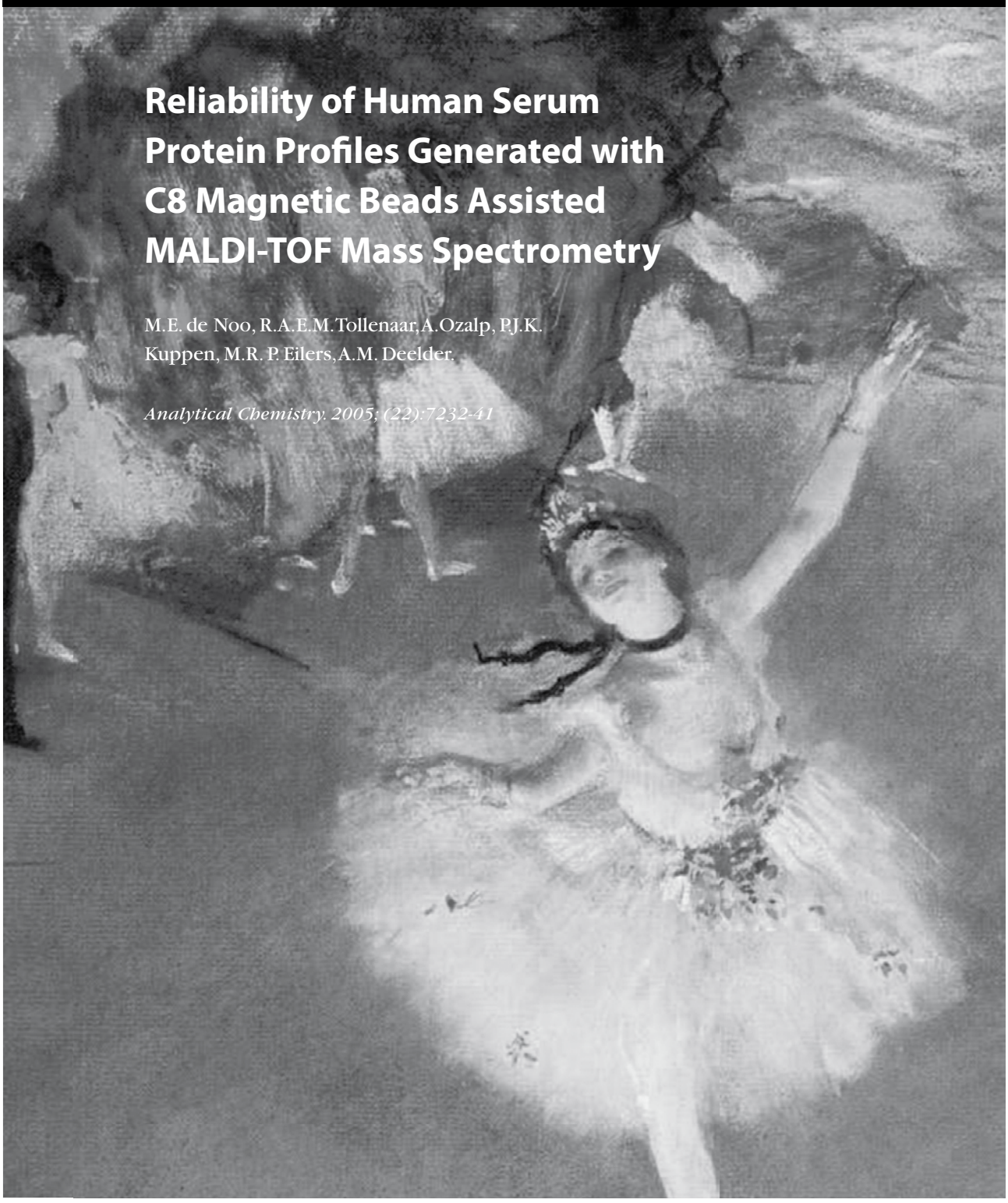
Clinical proteomics in oncology : a passionate dance between science and clinic

Noo, M.E. de

Chapter 3



Reliability of Human Serum Protein Profiles Generated with C8 Magnetic Beads Assisted MALDI-TOF Mass Spectrometry

M.E. de Noo, R.A.E.M. Tollenaar, A. Ozalp, P.J.K. Kuppen, M.R. P. Eilers, A.M. Deelder.

Analytical Chemistry. 2005; (22):7232-41

ABSTRACT

Protein profiling with mass spectrometry is a promising approach for classification and identification of biomarkers. However, there is debate about measurement quality and reliability. Here we present a pipeline for pre-processing, statistical data analysis and presentation. Serum samples of sixteen healthy individuals are used to generate protein profiles with a high-resolution MALDI-TOF after isolation of peptides with C8 magnetic beads. Analysis of variance (ANOVA) was performed after binning, normalization and baseline correction of the mean spectra. Relative variations in the spectra are expressed as coefficient of variation (CV), which depending on the respective preanalytical variation parameter investigated, was found to range between 0.15 and 0.67 in this study. With this novel method the reproducibility of our protein profiling procedure could be quantified. We showed that circadian rhythm and the number of freeze-thaw cycles had relatively limited influence on serum protein profiles, whereas the period between collection and serum centrifugation had a more pronounced effect.

INTRODUCTION

Proteomic pattern diagnostics is a recent and potentially revolutionary technology and approach for early disease detection, surveillance, and monitoring in oncology.[1] In proteomics proteins and functional protein networks as well as their dynamic alteration during physiological and pathological processes are characterised. It is a potential powerful tool in the discovery of disease biomarkers, as the proteome reflects both the intrinsic genetic program of an organism and the impact of its immediate environment.[2] Human serum contains thousands of peptides, most of which are thought to be fragments of larger proteins, but their precise nature remains largely undetermined. High throughput mass spectrometry can generate a proteome/peptidomic fingerprint of a given body fluid, such as serum. Patterns of these peptides can be correlated to biological events occurring in the entire organism and are likely to change in the presence of disease. In oncology new types of bioinformatic pattern recognition algorithms have been used to identify patterns of protein changes in order to discriminate cancer patients from healthy individuals.[3] Furthermore, different profiles may be associated with varying responses to therapeutics and other clinically relevant parameters and may also serve as prediction for treatment outcome. Several studies have shown that biomarkers can be identified on the basis of the presence/absence of multiple low-molecular-weight serum components using time-of-flight (TOF) mass spectrometry technologies such as SELDI-TOF and MALDI-TOF.[4-7] In general, although most studies measure serum components in a range in which primarily peptides and protein degradation products as well as small proteins are detected, the term protein profiling is generally accepted to describe this approach. Although essentially imprecise, this term will also be used in this study. Petricoin et al. showed that patterns of low-molecular-weight serum proteins reflect the pathological state of organs. In addition, these disease-related protein patterns could be useful in the early detection of ovarian cancer.[8] Based on discriminating serological protein profiles that study showed a sensitivity of 100%, specificity of 95% and a positive predictive value of 94% for the detection of ovarian cancer.

Although serum protein patterns have shown high sensitivity and specificity as an early diagnostic tool in several studies, critical notes have been made on biological variation, pre-analytical conditions and analytical reproducibility of serum protein profiles, which would make it difficult to differentiate a normal from a pathological and/or malignant status.[9] In addition, the reproducibility of serum protein profiles has been questioned, however more with respect to the bioinformatical analysis of the measured protein profiles than to the capturing and measuring techniques itself.[10-12] Thus, if proteomics spectra are ultimately to be applied in a routine clinical setting, collection and processing of the data will need to be subject to stringent

quality control procedures.[13] In fact, some critics argue that discriminating protein profiles are so far based more on experimental artefacts than on real biological differences.[14]

There are many factors that are thought to have an influence on serum protein profiles, complicating clear and unambiguous study findings. These factors include environmental and individual factors such as race, age, diet, smoking, stress, general physical condition and use of drugs, which all may influence serum protein profiles. Pre-analytical conditions of human serum also appear to influence protein pattern outcomes. So far, only a few studies have reported on the effects of different serum sample preparations and the use of a magnetic beads based approach to capture and concentrate serum proteins for MALDI-TOF mass spectrometry.[15-17] Since data processing and statistical analysis of protein spectra are essential elements in clinical proteomics, the objective of this study was to quantify the relative contributions of sources of variability on the protein spectra. To this end we developed a novel data processing pipeline, which was performed with an analysis of variance (ANOVA) of the spectra, after the spectra had been made comparable, reduced to common mass channels and the noise had been filtered. Strong baselines were always present in the spectra and had to be removed. This novel analysis method was used to assess the effect of variable pre-analytical conditions on human serum protein profiles, and their effect on reproducibility. In contrast to the above-mentioned study, we have chosen to primarily focus on assaying serum with C8 magnetic beads with hydrophobic functionality, followed by MALDI-TOF analysis. In line with the logistic conditions in a routine clinical setting, the effects of sample handling and storage, and also circadian rhythm factors on the serum protein profiles were analysed.

MATERIAL AND METHODS

Serum samples

Blood was collected from 16 healthy adult volunteers, 8 men and 8 women, by antecubital venipuncture. All blood samples were drawn from the left arm while the volunteers were seated. Approximately 10 ml venous blood was collected in a 10 cc Serum Separator Vacutainer Tube (BD Vacutainer Systems, Preanalytical Solutions, Plymouth, UK) at three different time points throughout the day. The first sample was drawn between 8 and 9 a.m. when all individuals had been fasting since midnight. The second specimen was obtained half an hour after lunch, between 1 and 2 p.m. and the last sample between 5 and 6 p.m. Thirty minutes after collection serum was separated by centrifugation at 3,000rpm for 10 minutes, divided into aliquots (Greiner) and stored at -70°C. The serum procurement, data management and blood

collection protocol were according to the guidelines of the Medical Ethical Committee of the Leiden University Medical Center. Informed consent was obtained from all subjects.

Protein profiling

To enhance signal quality magnetic beads based on hydrophobic interaction chromatography (MB-HIC kit, Bruker Daltonics, Leipzig, Germany) were used for sample preparation prior to MALDI-TOF mass spectrometry analysis. Five µl of serum was diluted with 10 µl binding solution and 5 µl magnetic beads were added. The solution was mixed by carefully pipetting five times. After 30 seconds supernatant was separated from the magnetic beads in a magnetic beads separator (MBS, Bruker) and discarded. This was followed by three washing steps with 100 µl wash solution (MB-HIC kit, Bruker) and supernatant was discarded each time. After 1 minute in 10µl e lution solution (50% Acetonitrile) the magnetic beads were separated in the MBS from the elution solution. An amount of 1 µl of this eluate, containing the captured peptides/proteins, was mixed with 10 µl matrix solution and 1 µl of this mixture was transferred to an Anchor Chip target plate ™ (Bruker Daltonics, Bremen, Germany) and allowed to dry before introduction into the mass spectrometer. Alpha-cyano-4-hydroxycinnamicacid (HCCA) was used as matrix (0.3 mg/ml in Ethanol: Acetone 2:1). Each sample was deposited onto four spots of the target plate. Matrix Assisted Laser Desorption Ionisation Time-Of-Flight (MALDI-TOF) mass spectrometry measurements were performed using an Ultraflex TOF/TOF instrument (Bruker Daltonics, Bremen, Germany) equipped with a SCOUT ion source, operating in linear mode. Ions formed with a N2 pulse laser beam (337 nm) were accelerated to 25 kV. With the employed serum preparation peptide/protein peaks in the m/z range of 1500 to 10,000 were measured. An independent mass spectrometer operator performed all measurements with blinded samples. Hereafter the entire process of capturing and concentrating serum proteins using C8 magnetic beads including the generation of readouts of the MALDI-TOF spectra will be designated as the protein profiling procedure.

Data processing

All spectra were compiled, and qualified mass peaks with mass-to-charge ratios (m/z) between 1500 and 10,000 were auto-detected. Each mass spectrum, as exported in an ASCII file, consisted of approximately 45,000 pairs of mass-to-charge values (Dalton) and ion counts. As we preferred to analyze the data using the intensity of the mass spectra per bin, the first processing step was to collect and average the data in bins of 1 Dalton wide. To reduce noise the Whittaker smoother was applied, using second differences, $\lambda = 100$ and weights proportional to the number of raw data points per

bin.[18] The resulting spectra generally showed strong baselines, which had to be removed before further processing. We used the asymmetric least squares algorithm as described in the appendix of Eilers 2004 [19]; figure 1 shows a typical example. The intensity scale of the baseline-corrected spectra was un-calibrated. To normalize the spectra we divided each mass spectrum by the median of the intensities. We consider this to be more robust than normalization on the average (or equivalently, the area under the curve), as the median is less sensitive to spurious large peaks. While this is an ad-hoc solution, we hope to find relatively stable regions in the spectra, so that we can normalize on medians over these regions in further research.

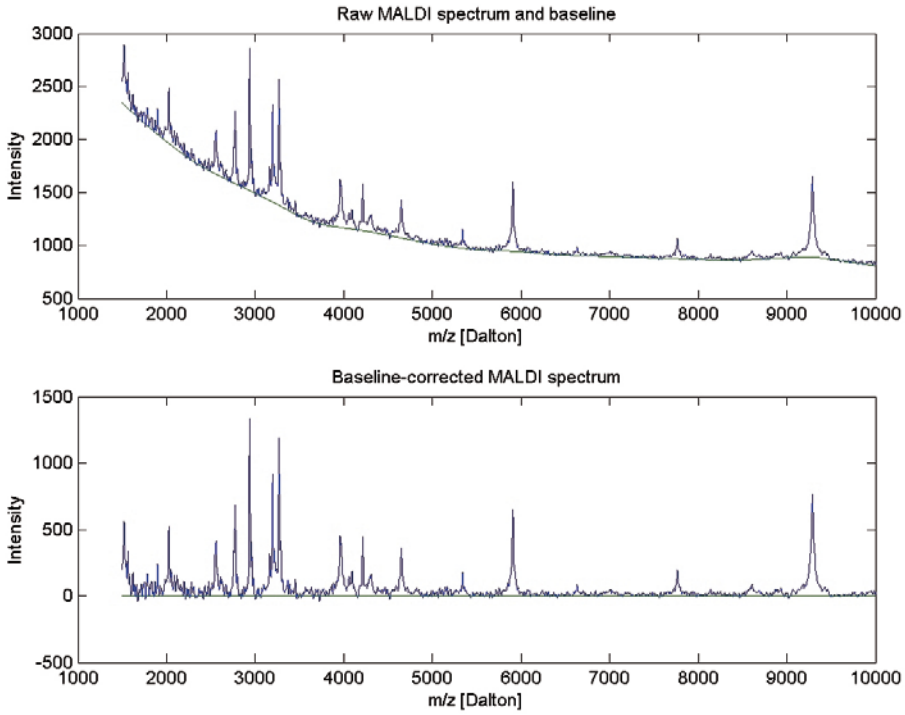


Figure 1. MALDI-TOF spectrum before and after baseline correction.

Statistical data analysis

To quantify the effects of experimental conditions, variability between individual persons and noise, we applied analysis of variance (ANOVA). Consider, as an example, an experiment in which we have P subjects, T storage times and C storage temperatures points. For each combination of subject and time we have measured a spectrum. First, we concentrate on only one, arbitrary, mass channel. We have PTC measurements, which we indicate by Y_{ptc}. The ANOVA model assumes that $Y_{ptc} = \mu+\alpha_p+\beta_t+\gamma_c+e_{ptc}$. Here μ is the overall mean, α_p is the effect of person P, β_t the ef-

fect of storage time t, γ_c is the effect of storage temperature s and e_{ptc} is random variation. The values of μ and the vectors α, β and γ that minimize the sum of the squares of the elements of e are the so-called least squares estimates and the standard result of ANOVA. If all combinations of persons, storage times and temperatures are present, they are the averages per person (storage time, temperature) over all spectra. When some combinations are missing, a somewhat more complicated regression approach has to be used.

The ANOVA was performed for each bin on the mass axis. This results in 1) one spectrum for μ, the average spectrum; 2) P spectra of person effects; 3) T spectra of time effects; 4) C spectra of storage temperature effects and a spectrum of s, the standard deviation of the noise for each sample. The single spectra of μ and s are easy to present and study, but the multiple spectra of the effects can be voluminous. We summarised them by computing standard deviations of α, β and γ per mass bin. The final results are a plot of five spectra for each of the performed experiments, but only shown in figure 3. The plot shows that, generally, the standard deviations increase when the overall mean increases. A simple measure of this relationship would be the coefficient of variation, like s/μ or s_α/μ, where s_α indicates the standard deviation of α. Unfortunately, this can provide wildly fluctuating results when μ is near zero. Therefore, we computed $cv = \Sigma s_i \mu_i / \Sigma_i \mu_i^2$, which is the slope of a regression line through the origin in a scatter plot of s vs. μ. The summation can be over the whole mass range; this result is reported as a number in the title of each graph of standard deviations for all experiments. In addition, we graphically present CV as computed in m/z windows 500 Dalton wide.

To investigate the influence of the effective bin width on computed CVs, we varied the smoothing parameter λ over a large range, artificially increasing peak width up to five times.

EXPERIMENTS

Reproducibility

In a first set of experiments both the reproducibility of repeated measurements of the same eluate and the reproducibility of repeated analysis of the same samples on four different days were determined. Serum samples of 8 randomly chosen individuals, drawn at one time point during the day were used. Each of these serum samples was processed only once, and measured 8 times with MALDI-TOF according to the standard protocol. Additionally, to determine the inter-measurement variation, protein profiling from 4 of the 8 serum samples was performed on 4 consecutive days. In all experiments samples were preparated just before each MALDI-TOF measurement.

Sample handling

To simulate 'realistic' logistical factors, the effects of sample handling and storage temperature prior to serum centrifugation on serum protein profiles were studied. Serum samples of 4 out of the 16 randomly chosen individuals, all drawn at the same time point were used for this experiment. From each individual 7 aliquots were stored at both room temperature and the same number at 4 °C. After a period of 30 minutes, 1 hr, 2, 4, 8, 24 and 48 hours serum samples were processed according to the standard protocol and protein profiles of all samples were compared for each individual.

Freeze-thaw cycles

To determine the utility of (archival) serum banking, effects of multiple freeze-thaw cycles on serum protein profiles were determined. Serum, drawn at one time point, of 8 randomly chosen individuals was used. Serum of each individual was divided into 11 primary aliquots. From each serum sample one aliquot was measured within 30 minutes after blood collection. The remaining ten sets were immediately frozen at -70 °C. Four hours after the initial freezing, all aliquots were removed from the freezer. Two aliquots of each sample were left at room temperature and the rest on ice for approximately 2 hours until completely thawed. Following the first freeze-thaw cycle, two samples, one thawed on ice and one at room temperature, were assayed. The remaining sets of aliquots were refrozen at –70 °C for 4 hours. Again one sample of each individual was allowed to stand at room temperature and the rest on ice for 2 hours until completely thawed. Subsequently, two samples were processed and the rest refrozen. This was repeated after respectively three and four freeze-thaw cycles, but all samples were thawed on ice.

Circadian rhythm

In a last set of experiments, effects of at which moment of the day blood was drawn on serum protein profiles were studied by analyzing serum samples of 16 individuals, drawn at three different times over the day. All samples were frozen and thawed once and assayed on one day according to the standard protocol.

RESULTS

The data processing pipeline described above was applied to all our experiments. In a preliminary step the influence of effective bin width was studied. We found that stronger filtering, which corresponds to increasing the effective bin width, broadens peaks in both mean and standard deviation spectra, but that the CV did not change much (less than 20%). Therefore, the subsequent experiments were analysed with bins of 1 m/z.

Reproducibility

A test concerning intra-measurement reproducibility was performed by determining the coefficient of variation (CV) over 8 MALDI-TOF spectra for each subject, as shown in figure 2. The CV of the reproducibility within one measurement was less than 20% for 6 out of 8 subjects. Subject D2 and D5 showed slightly higher CV's of 22% and 29%, respectively.

The inter-measurement reproducibility of 4 serum samples performed on 4 different days is shown in table 1. The range in CV between the spectra within one individual (14-23%) is similar to the CV between the consecutive days after correction for differences between individuals (17-26%). However, the variation in spectra between the 4 consecutive days was minor, with an increase in CV on day 4 - 26% (Figure 3).

Table 1. Coefficient of variation (CV) for inter-measurement reproducibility. The CV was determined over 4 MALDI-TOF spectra of each individual, all measured at consecutive days.

Subject	CV (in %)
F	23
G	20
M	22
R	14

Sample handling

To establish the effects of serum sample handling, an ANOVA was performed for effects of persons, time and temperature and residual variation (Figure 4). After correction for inter-individual differences and residual standard deviations with ANOVA, CV between storages at room temperature or at 4 °C was calculated to be 45 and 50%, respectively. The CV of the samples stored for different periods of time before centrifugation ranged from 42% to 67% (Table 2). There was no correlation between the storage time and the coefficient of variation.

Freeze-thaw cycles

The effects of multiple freeze-thaw cycles on serum protein profiles were determined for 10 sets with various storage circumstances, as set 5 had to be left out of the analysis due to technical problems. Table 3 shows the coefficient of variation between persons for different freeze-thaw cycles. In fresh serum samples (set 1) the CV was highest with 64%. With the growing number of hours that serum samples were stored in the fridge at 4 °C, the CV decreased to a minimum of 24% after 8

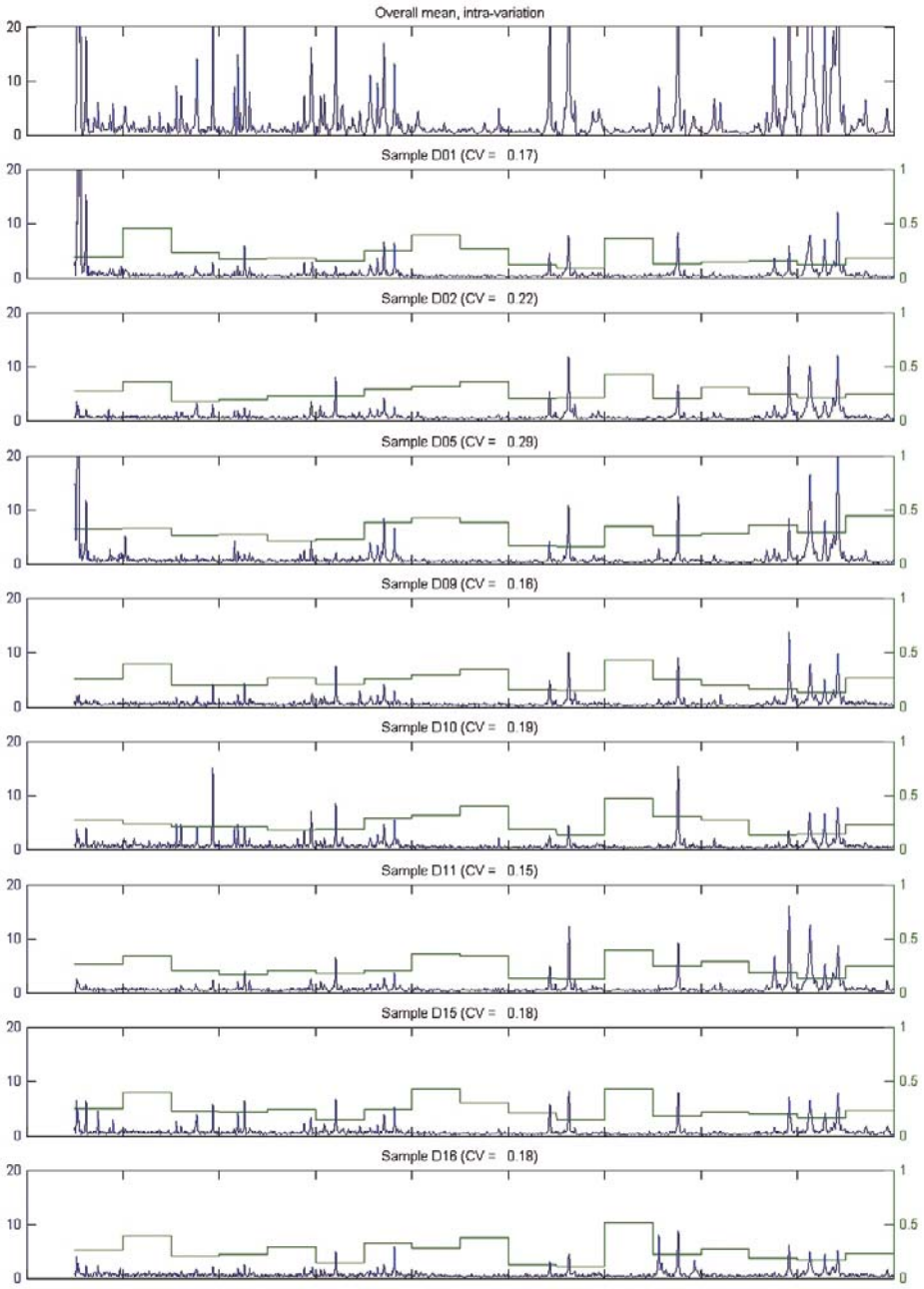


Figure 2. Intra-measurement reproducibility. The CV was determined over 8 MALDI spectra of each individual, all processed in one run.

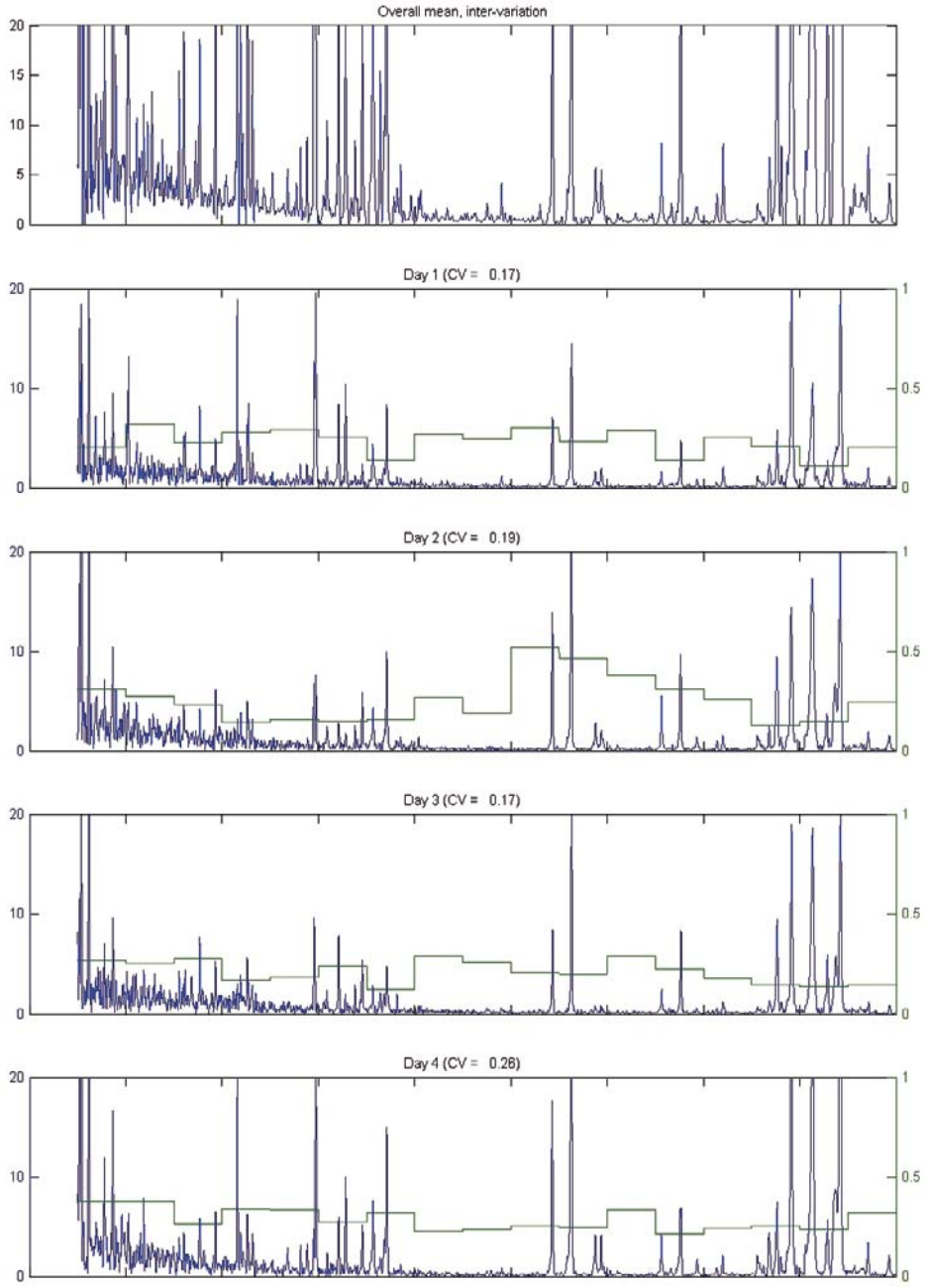


Figure 3. ANOVA for inter-measurement reproducibility. The CV is calculated for spectra that are measured on the same day, after correction for inter individual differences.

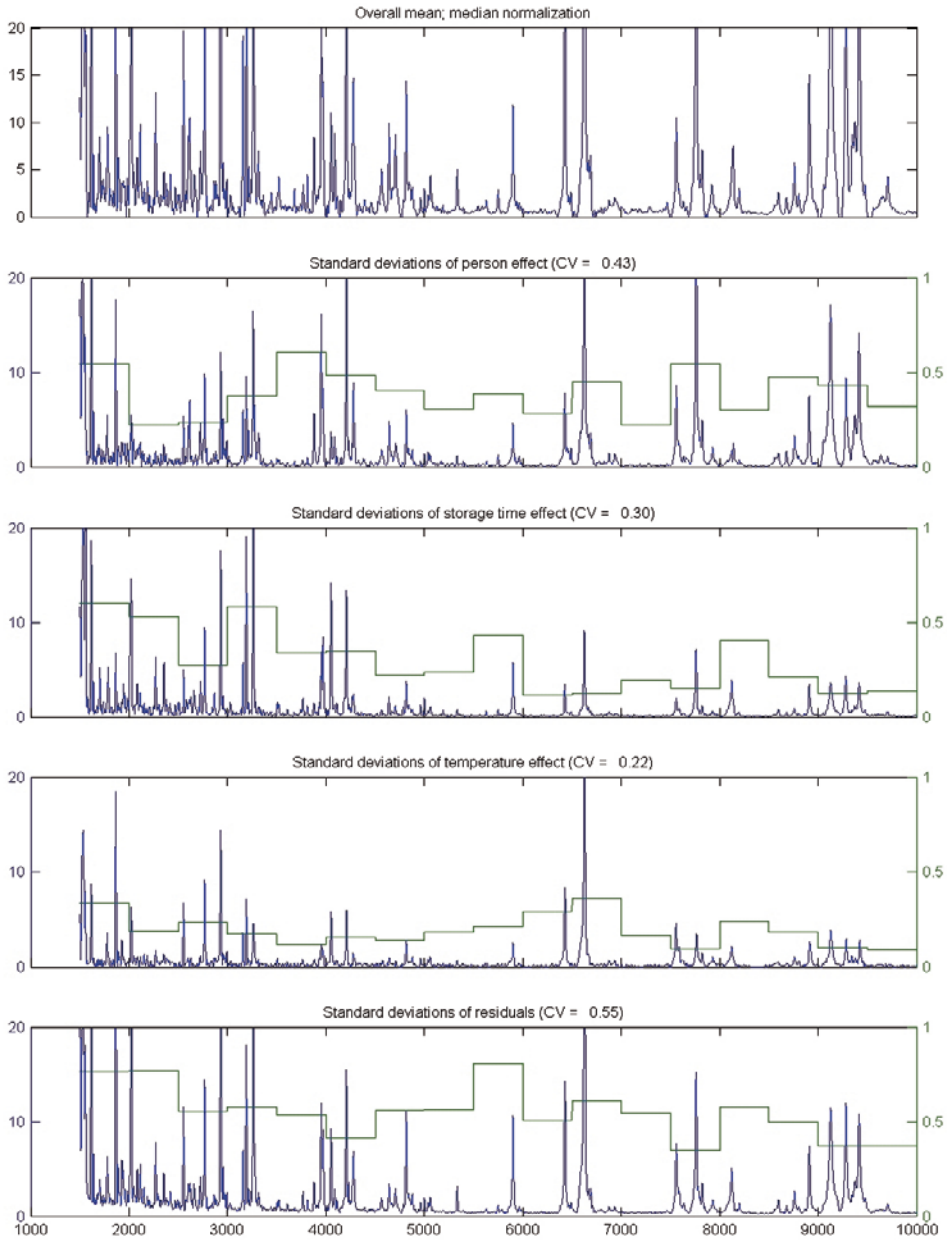


Figure 4. ANOVA of sample handling. From top to bottom: the average spectrum; the variation in spectra due to person's effect; the variation in spectra due to time effects is shown. Finally, the effects of storage temperature variation and the standard deviation of the noise for each sample are presented.

Table 2. Coefficient of variation (CV) between storage times of venous blood before serum centrifugation, regardless the temperature of storage

Storage time	CV (in %)
30 min	52
1 hr	42
2 hrs	63
4 hrs	53
8 hrs	49
24 hrs	52
48 hrs	67

Table 3. Coefficient of variation (CV) between persons per freeze-thaw set. Each set consisted of serum samples of 8 subjects. Each set was stored under different circumstances, namely after none or 1 to 4 freeze-thaw cycles. Sets 1 to 4 were not frozen at all, but stored at 4 ° C during different periods of time.

Set	No freeze-thaw cycles	Temp	CV (in %)
1	0	21 ° C	64
2	0 (2 hrs)	4 ° C	40
3	0 (4 hrs)	4 ° C	39
4	0 (8 hrs)	4 ° C	24
6	1	on ice	26
7	1	21 ° C	39
8	2	on ice	28
9	2	21 ° C	25
10	3	on ice	28
11	4	on ice	28

hours. The number of freeze-thaw cycles had no influence on the CV. All CV of sets 6 to 11 were smaller than 28%, with exception of set 7 (thawed at room temperature after one cycle) with a CV of 39%.

To get an impression of the patterns of change with freeze-thaw cycles, we applied the following procedure to each of the 8 subjects: 1) selected the 8 spectra of the reference set spectra; 2) subtracted these reference spectra from the individual spectra of all sets; 3) regressed (per spectrum) the absolute value of the corresponding reference spectrum, to calculate a CV. The so computed CVs are presented in figure

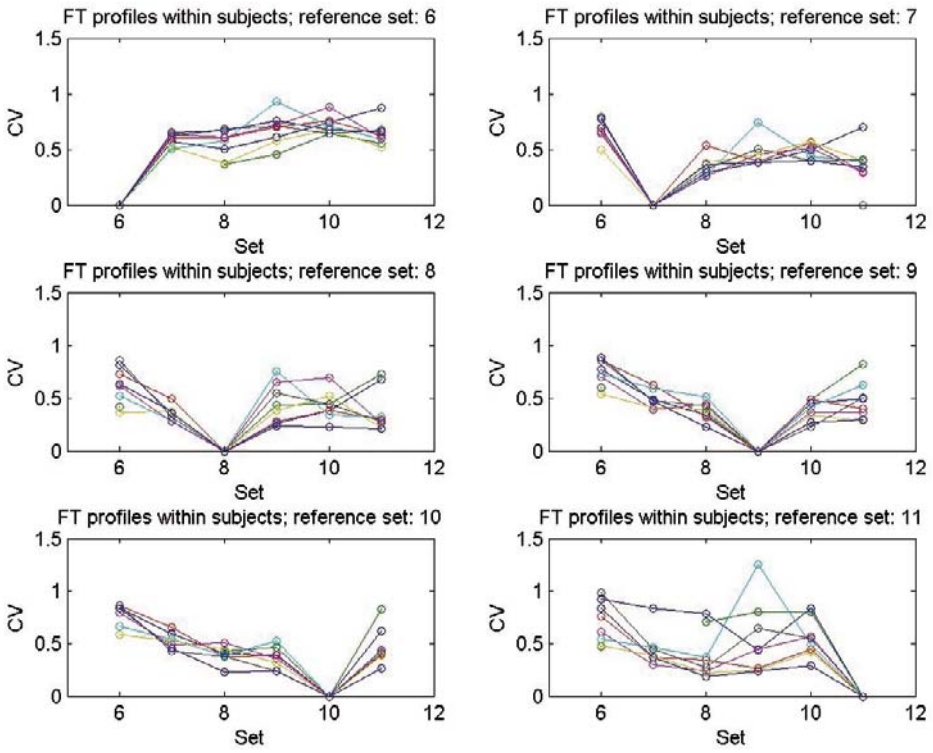


Figure 5. Coefficient of variation between the samples of one reference set and set 6 to 11. On the Y-axis the CV is stated . The sets of the freeze-thaw experiment, as described in table 3, are represented on the X-axis.

5. Generally, all 8 subjects showed the same patterns per reference set, with one or two outliers. In contrast to the other reference sets, 6 showed a continuous increase in CV per extra set. Between the reference sets 7 and 10, the patterns became more identical and the CV was decreased over all sets. In set 11, the variation between the subjects increased, became more variable.

Circadian rhythm

The effect of time variation of blood drawing on serum protein profiles is shown in table 4. Spectra in set 1, collected at 8 a.m. when subjects were fasting, showed a CV between the 8 individuals of 51%. Set 2, collected half an hour after lunch, and set 3, drawn at the end of the afternoon, non-fasting, resulted in 44% and 55%, respectively. No large difference in CV was found between the three sets.

Table 4. Coefficient of variation (CV) between individuals per time point of blood collection. Each set consists of serum samples of 16 subjects, all drawn at time point as indicated in the table.

Set	Time of blood drawing	Fasting	CV
1	8-9 a.m.	Yes	0.51
2	13-14 p.m.	No	0.44
3	18-19 p.m.	No	0.55

DISCUSSION

So far, in a limited number of studies, proteomics-based approaches have shown promising results for the generation of diagnostic profiles in serum. Substantial attention was given to analyze low molecular weight protein patterns from easy-accessible body fluids. To qualify as a future diagnostic test, the entire procedure of protein profiling should be easy to use, robust, reproducible and affordable.[15] High-throughput will also be essential for embedding protein profiling in the clinical setting. The use of fractionation protocols, such as reversed phase magnetic beads, to reduce the complexity of biological samples in MALDI-TOF is needed to avoid signal suppression effects.[20] Therefore, direct analysis of serum is not feasible. In this study we have chosen to use the increasingly accepted C8 magnetic bead capturing technique, taking into consideration that only a small fraction of proteins, from the potential ten thousands of proteins and peptides in human serum can be analysed with this approach. In future studies we will evaluate capturing techniques with different functionalities. In our MALDI-TOF experiments we obtained 'rich' mass spectra, containing many peaks and showing much detail. Our novel data processing pipeline proved to be an effective tool for quality assessment. Baseline correction, binning and filtering provided uniformly structured data in which most typical artefacts had been removed. The ANOVA algorithm separates the sources of variation and provides easily understood numerical summaries of their relative strength.

There is much room for further improvement and refinement. Calibration of the spectra is now based on the median over the domain of interest (1500 to 10,000 Dalton). This is a natural, but rather arbitrary choice. It would be attractive if stable areas in spectra could be located on which to base calibration, or if a reliable spiking procedure was available.

The ANOVA assumed an additive model for the spectral intensities, which is acceptable to compare the relative influence of logistical factors. However, one could argue that a multiplicative model might hold as well, or perhaps even better. It is not possible to simply take logarithms and replicate the ANOVA, as many mass channels

contain negative numbers after baseline correction, caused by noise. A threshold may solve this problem, but overall coefficients of variation depend on the level of this threshold.

We have analysed the data in the form of binned spectra. An alternative approach is to detect individual peaks and analyze peak lists.[21;22] For our purpose, quantification of the reproducibility and of the effects of logistical factors, this would offer no advantages. The experiments with increased smoothing showed only a small influence of the effective bin width on the CV. In a peak list, each peak acts like one 'bin' representing a group of highly correlated intensities around it. Whether we compute a local coefficient of variation by averaging over these individual intensities or over a smaller set of representative peak heights makes little difference. A disadvantage of peak lists is the need for finding complete lists for all spectra, because missing peaks complicate the ANOVA. Furthermore, we used the Whittaker smoother to remove noise and in baseline removal.[18] Compared to wavelets, it has the following advantages: one has continuous control over smoothness and one very short Matlab function does all the work, eliminating any need for toolboxes.[22]

With the employed statistical data analysis the intra-measurement experiments showed a good reproducibility. It is generally accepted that factors like matrix composition and ionisation suppression influence the quality of the MALDI spectra, which in turn will always result in a certain degree of variance in intensity of the generated spectra. This phenomenon can be seen in spectra of subject D5. All spectra of this individual were of inferior quality, possibly due to ionisation suppression or poor matrix solvent composition.[23] Ion suppression results from the presence of less volatile compounds that can change the efficiency of droplet formation or droplet evaporation. This in turn affects the amount of charged ion in the gas phase that ultimately reaches the detector and may result in lower quality spectra.[24;25] To minimize these influences, we used HCCA as a matrix and each sample was spotted four times. However, differences in ionisation rate and thus in peak intensity are intrinsic to the technique and have to be accounted for in the statistical analysis.

The inter-measurement reproducibility within one individual corresponded to the intra-measurement reproducibility for all 4 individuals. However, there seems to be a very small but acceptable day-to-day variation between the different experiments. Therefore, we recommend performing all experiments on one day or to correct for day-to-day variation. To further enhance intra and inter-measurement reproducibility application of robotics for sample processing is recommended. Indeed, implementation of an automated procedure on an 8-channel Hamilton STAR® pipetting robot (Hamilton, Martinsried, Germany) did result in a further reduction of the CV (data not shown).

Ultimately, it might be advisable to include a synthetic peptide mix in the generated spectra for external calibration. In larger profiling studies, batch effects should be taken into account in the design of the study.

Moreover, it is interesting to speculate on the potential discriminating power of MALDI spectra. In the reproducibility experiment we found the overall CV of the error to be 0.18 and that of the person effect 0.33. This can be expressed as a reliability coefficient $r = 0.33^2/(0.33^2 + 0.18^2) = 77\%$. This indicates that nearly 80% of variation in spectra is related to differences between persons. This is not a percentage that indicates that on the basis of whole spectra discrimination between individual spectra will be possible. The graphs of CV as computed for windows of 500 Dalton show strong variations, suggesting that better discrimination could in principle be achieved by using selected parts of the m/z domain. Of course, our data were generated from healthy volunteers, so it remains to be determined how much spectra will differ between healthy and diseased persons.

In this study the largest effect was observed for sample handling conditions. There was no correlation between the increasing number of hours before centrifugation and the variation between the serum protein profiles, but the overall variation was larger. This would already justify acceptance of a certain time range after blood collection and before centrifugation. Furthermore it is unlikely that in a hospital's daily practice this factor could be rigorously standardised. Thus, although a standard time period would be ideal, we accept a delay of 0-4 hours between the moment of blood collection and serum centrifugation. In view of the fact that there was no large difference between the storage temperatures and logistical factors, leaving all blood samples to stand at room temperature before centrifugation seems justified.

The effect of increasing numbers of freeze-thaw cycles was small and consistent, with the exception of set 7, in which serum samples were thawed only once at room temperature. The coefficient of variation in this set was larger than in all other sets, as shown in table 3. This might be explained by the fact that protein degradation occurs sooner at room temperature, as also demonstrated in sets 2-4. This phenomenon might be explained by proteolytic activity and the fact that hydrophobic interactions are strengthened, while with increasing temperature the hydrogen bonding is weakened and the electrostatic interactions are not changed due to its entropic origin.[26] Whereas the range in coefficients of variation between increasing numbers of freezes and thaw cycles is small, fresh serum samples provided the largest variation between persons, almost double in comparison to other sets. Furthermore, in fresh serum samples the number of peaks observed was less than 50, as also reported by other groups.[15;27] We suggest that in this early stage of defining optimal parameters/conditions for serum pattern diagnostics the use of fresh serum samples is better avoided. This seems contradictory, as proteolytic activity after thawing implicates a

loss of proteins and peptides and thus of information. However, on the condition that all samples are treated according to a standard protocol, this would not be critical for a black box approach. Thus it would seem that the use of archival material is safe with respect to the effect of freezing and thawing; nevertheless it remains of paramount importance that the entire sample handling and storage procedure is standardised. Based on the fact that the coefficient of inter-group variation in reference set 8 is lower than in the other sets (Figure 5), we prefer to use serum samples for further studies, which have undergone two freeze-thaw cycles. Moreover, our choice is mainly rooted in practical and logistical reasons, as in many large hospitals; sample collection is centralised in the clinical chemical laboratory.

With only minimal variation observed between protein profiles from samples collected at three different time points over the day, circadian rhythm seems to have limited effect on individual serum protein profiles. This is an encouraging fact, as blood samples can be collected all over the day, which increases the future applicability of serum protein profiling in the clinic. Furthermore, there is no indication that fasting has any influence on serum protein profiles, which also facilitates future clinical use.

All together, we have presented a method to assess the reproducibility of a protein profiling procedure using a high-end MALDI-TOF. Our appliance of ANOVA over the mean spectra allowed analysis of the effects of handling and storage procedures on serum protein profiles. The results from this study stress the importance of a standardised collection of all blood samples, from the moment of sample handling and storage until freezing the samples in order to prevent bias in classification studies. Although the importance of homogeneity and uniformity within sample groups must be stressed, variation of such factors can not totally be excluded in a clinical setting. The most important issues for discriminating studies at this moment are a standardised and well-documented sample collection and a thorough study design. Based on the present data and those of Villanueva et al.[15], we feel that the methodology can be standardised to a level which allows application as a tool in biomarker discovery. Although it remains to be seen whether actual biomarkers can reliably be identified with the current technique, we are now in the process of carrying out a study to determine whether serum protein profiles can differentiate colorectal cancer patients from individuals with benign bowel disorders and healthy subjects. To this end and to facilitate high-throughput studies, we developed an automated platform for our capturing technique with C8 magnetic beads with reverse-phase based functionality and we used the MS instrument's AutoXecute function to further enhance reproducibility (data not shown). In addition to large clinical studies as mentioned above, such a platform would also be valuable for more large-scale studies as e.g. inter group variance (cases versus controls) under different experimental setups.

REFERENCES

1. Poon,T.C. and Johnson,P.J. (2001) Proteome analysis and its impact on the discovery of serological tumor markers. *Clin.Chim.Acta*, 313, 231-239.
2. Srinivas,P.R., Srivastava,S., Hanash,S., and Wright,G.L., Jr. (2001) Proteomics in early detection of cancer. *Clin.Chem.*, 47, 1901-1911.
3. Wulfkuhle,J.D., Liotta,L.A., and Petricoin,E.F. (2003) Proteomic applications for the early detection of cancer. *Nat.Rev.Cancer*, 3, 267-275.
4. Adam,B.L., Qu,Y., Davis,J.W., Ward,M.D., Clements,M.A., Cazares,L.H., Semmes,O.J., Schellhammer,P.F., Yasui,Y., Feng,Z., and Wright,G.L., Jr. (2002) Serum protein fingerprinting coupled with a pattern-matching algorithm distinguishes prostate cancer from benign prostate hyperplasia and healthy men. *Cancer Res.*, 62, 3609-3614.
5. Petricoin,E.F., III, Ornstein,D.K., Paweletz,C.P., Ardekani,A., Hackett,P.S., Hitt,B.A., Velassco,A., Trucco,C., Wiegand,L., Wood,K., Simone,C.B., Levine,P.J., Linehan,W.M., Emmert-Buck,M.R., Steinberg,S.M., Kohn,E.C., and Liotta,L.A. (2002) Serum proteomic patterns for detection of prostate cancer. *J.Natl.Cancer Inst.*, 94, 1576-1578.
6. Rai,A.J., Zhang,Z., Rosenzweig,J., Shih,I., Pham,T., Fung,E.T., Sokoll,L.J., and Chan,D.W. (2002) Proteomic approaches to tumor marker discovery. *Arch.Pathol.Lab Med.*, 126, 1518-1526.
7. Yanagisawa,K., Shyr,Y., Xu,B.J., Massion,P.P., Larsen,P.H., White,B.C., Roberts,J.R., Edgerton,M., Gonzalez,A., Nadaf,S., Moore,J.H., Caprioli,R.M., and Carbone,D.P. (2003) Proteomic patterns of tumour subsets in non-small-cell lung cancer. *Lancet*, 362, 433-439.
8. Petricoin,E.F., Ardekani,A.M., Hitt,B.A., Levine,P.J., Fusaro,V.A., Steinberg,S.M., Mills,G.B., Simone,C., Fishman,D.A., Kohn,E.C., and Liotta,L.A. (2002) Use of proteomic patterns in serum to identify ovarian cancer. *Lancet*, 359, 572-577.
9. Boguski,M.S. and McIntosh,M.W. (2003) Biomedical informatics for proteomics. *Nature*, 422, 233-237.
10. Somorjai,R.L., Dolenko,B., and Baumgartner,R. (2003) Class prediction and discovery using gene microarray and proteomics mass spectroscopy data: curses, caveats, cautions. *Bioinformatics.*, 19, 1484-1491.
11. Yasui,Y., Pepe,M., Thompson,M.L., Adam,B.L., Wright,G.L., Jr., Qu,Y., Potter,J.D., Winget,M., Thornquist,M., and Feng,Z. (2003) A data-analytic strategy for protein biomarker discovery: profiling of high-dimensional proteomic data for cancer detection. *Biostatistics.*, 4, 449-463.
12. Baggerly,K.A., Morris,J.S., and Coombes,K.R. (2004) Reproducibility of SELDI-TOF protein patterns in serum: comparing datasets from different experiments. *Bioinformatics.*, 20, 777-785.
13. Coombes,K.R., Fritsche,H.A., Jr., Clarke,C., Chen,J.N., Baggerly,K.A., Morris,J.S., Xiao,L.C., Hung,M.C., and Kuerer,H.M. (2003) Quality control and peak finding for proteomics data collected from nipple aspirate fluid by surface-enhanced laser desorption and ionization. *Clin. Chem.*, 49, 1615-1623.
14. Sorace,J.M. and Zhan,M. (2003) A data review and re-assessment of ovarian cancer serum proteomic profiling. *BMC.Bioinformatics.*, 4, 24.
15. Villanueva,J., Philip,J., Entenberg,D., Chaparro,C.A., Tanwar,M.K., Holland,E.C., and Tempst,P. (2004) Serum Peptide profiling by magnetic particle-assisted, automated sample processing and maldi-tof mass spectrometry. *Anal.Chem.*, 76, 1560-1570.
16. Baumann,S., Ceglarek,U., Fiedler,G.M., Lembcke,J., Leichtle,A., and Thiery,J. (2005) Standardized Approach to Proteome Profiling of Human Serum Based on Magnetic Bead Separation and Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry. *Clin. Chem.*, 51, 973-980.

17. Yamanishi,H., Kimura,S., Iyama,S., Yamaguchi,Y., and Yanagihara,T. (1997) Fully automated measurement of total iron-binding capacity in serum. *Clin.Chem.*, 43, 2413-2417.
18. Eilers,P.H. (2003) A perfect smoother. *Anal.Chem.*, 75, 3631-3636.
19. Eilers,P.H. (2004) Parametric time warping. *Anal.Chem.*, 76, 404-411.
20. Richter,R., Schulz-Knappe,P., Schrader,M., Standker,L., Jurgens,M., Tammen,H., and Forssmann,W.G. (1999) Composition of the peptide fraction in human blood plasma: database of circulating human peptides. *J.Chromatogr.B Biomed.Sci.Appl.*, 726, 25-35.
21. Semmes,O.J., Feng,Z., Adam,B.L., Banez,L.L., Bigbee,W.L., Campos,D., Cazares,L.H., Chan,D.W., Grizzle,W.E., Izbicka,E., Kagan,J., Malik,G., McLerran,D., Moul,J.W., Partin,A., Prasanna,P., Rosenzweig,J., Sokoll,L.J., Srivastava,S., Srivastava,S., Thompson,I., Welsh,M.J., White,N., Winget,M., Yasui,Y., Zhang,Z., and Zhu,L. (2005) Evaluation of serum protein profiling by surface-enhanced laser desorption/ionization time-of-flight mass spectrometry for the detection of prostate cancer: I. Assessment of platform reproducibility. *Clin.Chem.*, 51, 102-112.
22. Morris,J.S., Coombes,K.R., Koomen,J., Baggerly,K.A., and Kobayashi,R. (2005) Feature extraction and quantification for mass spectrometry in biomedical applications using the mean spectrum. *Bioinformatics.*.
23. Knochenmuss R., Dubois F., Dale M.J., and Zenobi R. The matrix Suppression Effect and Ionization Mechanisms in Matrix-assisted Laser Desoprtion/ Ionization. Rapid Commun.Mass Spectrom. 10, 871-877. 9-5-1996. John Wiley & Sons. Ltd. Ref Type: Generic
24. Cohen,L.H. and Gusev,A.I. (2002) Small molecule analysis by MALDI mass spectrometry. *Anal.Bioanal.Chem.*, 373, 571-586.
25. Annesley,T.M. (2003) Ion suppression in mass spectrometry. *Clin.Chem.*, 49, 1041-1044.
26. Jaenicke,R. and Zavodszky,P. (1990) Proteins under extreme physical conditions. *FEBS Lett.*, 268, 344-349.
27. Wang,M.Z., Howard,B., Campa,M.J., Patz,E.F., Jr., and Fitzgerald,M.C. (2003) Analysis of human serum proteins by liquid phase isoelectric focusing and matrix-assisted laser desorption/ionization-mass spectrometry. *Proteomics.*, 3, 1661-1666.